ICES Journal of
Marine Science

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

ICES Journal of Marine Science (2015), 72(1), 93–98. doi:10.1093/icesjms/fsu043

Original Article

Model averaging to streamline the stock assessment process

Colin P. Millar[1,2]*, Ernesto Jardim[1], Finlay Scott[1], Giacomo Chato Osio[1], Iago Mosqueira[1], and Nekane Alzorriz[1]

[1]*European Commission Joint Research Centre, Institute for the Protection and Security of the Citizen (IPSC), Via Enrico Fermi 2749, 21027 Ispra (VA), Italy*
[2]*Marine Scotland Freshwater Laboratory, Faskally, Pitlochry, Perthshire PH16 5LB, UK*

**Corresponding author: tel: +44 1224 294419; fax: +44 1796 473523; e-mail: c.millar@marlab.ac.uk*

Millar, C. P., Jardim, E., Scott, F., Osio, G. C., Mosqueira, I., and Alzorriz, N. Model averaging to streamline the stock assessment process. – ICES Journal of Marine Science, 72: 93–98.

Received 2 September 2013; revised 19 February 2014; accepted 20 February 2014; advance access publication 1 April 2014.

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, fishery policy, and computing. There is a particular focus on producing a single "best" stock assessment model, but as fishery 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 frameworks to translate hypotheses into models. We propose a change to the current stock assessment procedure, driven by the use of model averaging to address a range of plausible hypotheses, where increased collaboration between the varied disciplines within fishery science will result in more robust advice.

Keywords: a4a, fishery management, model averaging, model selection uncertainty, scientific advice, stock assessment.

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 abundance, mortality, and productivity (Hilborn and Walters, 1992). The typical stock assessment process involves a team of scientists generating a suite of 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, e.g. the model likelihood or the examination of residuals. The outcomes of this assessment process are often point estimates of the stock status, with minimal presentation of the form and size of the uncertainty of those estimates. In the United States and Europe, the tendency has been to deal with uncertainty through the definition of stock status limits (Shertzer *et al.*, 2008; Ralston *et al.*, 2011).

Stock assessment 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 (Methot, 2009; Jardim *et al.*, 2014).

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". Although only one model may be the most likely, the others may still represent plausible "states of nature" and contribute to the estimation of the uncertainty about it. Many other authors have raised and discussed issues surrounding the basing of conclusions on a single model (Butterworth *et al.*, 1996; Hilborn, 1997; Patterson, 1999; Hill *et al.*, 2007; Simmonds *et al.*, 2011). The major danger is that ignoring model-selection error leads to underreporting of variability and too narrow confidence intervals (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) specifically within the age-structured a4a framework (section 3; Jardim *et al.*, 2014) to address the two stock assessment issues identified above: (i) the need to choose only a single model, and (ii) 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 wherein 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, e.g. all demersal species in the North Sea. We believe that by the combined effort of experts in fishery and statistical modelling, model averaging, in the context of statistical age-structured stock assessment models, could become an automated process.

In this paper, we propose a change in the stock assessment process from one where model checking and model selection are the focus to one where designing appropriate models is the most important scientific task. We believe that this change is not a revolution, merely an evolution of what is already done. We envisage a process where the full gamut of fishery disciplines—oceanography, genetics, biology, population biology, gear technology, among others—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. We recognize that there are unsolved problems even for model selection when modelling overdispersed-count data; therefore, we restrict our attention to lognormally distributed data. First, we describe in more detail the idea of setting up plausible states of nature and techniques for model averaging, and in the final section, we discuss practical issues.

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 fishery 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 submodels: one for fishing mortality (F), one for survey catchability (Q), one for recruitment (R), and models for observation variances (V). Each submodel is specified as one would specify a linear equation in simple linear regression; in practice, we work on the log scale to ensure that model quantities (F, Q, etc.) are positive. Consider a dataset with age and year as separate covariates, so that a model is specified where log F is constant with age a, but varies linearly with year y:

$$\log F_{a,y} = \alpha + \mathrm{year}_y \times \beta,$$

or in simpler notation

$$\log F \sim 1 + \mathrm{year}, \tag{1}$$

where "1" denotes intercept and "∼" means "is modelled as". Simple extensions to this model are:

$$\log F \sim \mathrm{age} + \mathrm{year}, \tag{2}$$

$$\log F \sim \mathrm{age} + \ \mathrm{factor}\ (\mathrm{year}), \tag{3}$$

$$\log F \sim s(\mathrm{age}) + \ \mathrm{factor}(\mathrm{year}). \tag{4}$$

In Equation (2), log F is now linear in age [i.e. F at age is of the form c exp($b \times$ age)] rather than constant. In Equation (3), log F changes level each year, while log F at age is linear on the log scale. Finally, approaching a more sensible model for F, Equation (4) has a spline function for log F at age, while the level of F changes from year to year. Beyond the very simple models given by Equations (1)–(4), there is the possibility to specify two-dimensional splines to allow a changing F-at-age pattern over time using s(age, year) or more structured interactions such as s(age):year-blocks, where year-blocks define periods of different F-at-age patterns. Further complexity may be added by including time-varying covariates, e.g. it is conceivable that including mean length-at-age in the equation for survey catchability (Q) may capture some temporal variation in survey catches. Models for observation variance are specified in the same way; however, recruitment models are more complex in that random effects are permitted.

The a4a framework gives easy access to a wide variety of fishing mortality, survey catchability, recruitment, and observation variance models, allowing the user to set up a wide range of population dynamics and fishery models. Obviously, the user must be familiar with linear modelling and working with splines to get the most out of this interface.

Most stock assessment models rely on a few basic assumptions: the data correspond to a fully spatially mixed stock, full spatial coverage in terms of catch estimates and surveys, known natural mortality (M), a constant stock–recruitment relationship, and constant survey catchability. These assumptions, depending on the model, can be relaxed or adapted, but often, it is not straightforward to incorporate flexibility with respect to these assumptions, mainly because most stock assessment models have not been designed as exploratory tools, but rather to provide a single assessment. Modelling frameworks, like a4a (Jardim *et al.*, 2014) or Stock Synthesis 3 (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 hypothesis can be valid, e.g. a regime shift in the North Sea in 1988 changed the temperature that may have affected the mortality rate of salmon (*Salmo salar*) via reduced growth (Beaugrand and Reid, 2012). Was there a regime shift effect? Does the regime shift affect mortality or reproductive success? Other examples of situations where fishery 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 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 North Sea cod (*Gadus morhua*),

which has experienced depressed recruitment since the mid-1980s (Beaugrand *et al.*, 2003; Olsen *et al.*, 2010).

- Cases characterized 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 a need for flexible selectivity patterns. Crone *et al.* (2013) suggest including 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.
- About the previous point, 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 affect availability to surveys.
- Increasing efficiency of survey vessels is a known issue. For example, one of the two multinational surveys for North Sea cod was removed from the annual stock assessment due to 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 (Wilberg *et al.*, 2010).
- Developing operating models for southern bluefin tuna (*Thunnus maccoyii*) (R. 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.

In our opinion, these cases fall into two categories:

(i) Anything that could be considered to be on a continuum, which would be dealt with by specifying a range of plausible values. For example, a range of breakpoints in a stock–recruitment relationship, a range of ages above which fishing mortality is constant, a range of M values, or a range of growth scenarios.

(ii) Different hypotheses form categories rather than a range of options. For example, in Wilberg *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.

Each of the situations listed above highlights a recognized need by the fishery/scientific community to address multiple hypotheses about the "states of nature". Some effects may act across a number of stocks, such as temperature effects or variations in primary productivity, while others may be more stock-specific, such as gear mesh size increases affecting smaller or thinner 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 the model-averaging setting, all of these models will contribute to the final estimates and uncertainty of the state of the stock and to other standard outputs from stock assessment. The weight that each model receives in the final output ultimately depends on the fit of the model to the data. A selection of procedures for weighting models is given in the next section.

Model averaging

Model averaging is a technique for incorporating model-selection uncertainty into inference (Buckland *et al.*, 1997). 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 is 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 fixed weights, while the third and fourth methods acknowledge that model weights, like parameter estimates, are random and use stochastic simulation to build the distribution of the weights. Finally, we mention how expert weights can be incorporated into model-averaging schemes. For further reading on model averaging and model discrimination, we suggest Buckland *et al.* (1997), Burnham and Anderson (2002), Claeskens and Hjort (2008), and King *et al.* (2010).

Single-model inference

To describe the model-averaging schemes, we introduce some notation and use it to describe single-model inference. Let S be the collection of plausible models, and we are interested in estimates of a parameter μ. For each model M, there is an estimator which we denote $\hat{\mu}_M$. Fixed-weight model averaging simply takes a weighted average of these estimators to produce the final estimate:

$$\hat{\mu} = \sum_{M \in S} w_M \hat{\mu}_M, \tag{5}$$

where w_M are weights which sum to 1. In single-model inference, the model weights w_M are all 0, apart from the selected model which receives a weight of 1.

Frequentist weights

There are a variety of frequentist model-weighting strategies ranging from giving all models in S a weight related to the AIC (Akaike information criterion) or BIC (Bayesian information criterion) of each model, to an interpolation between two extreme cases. We present an AIC-based weighting strategy called smooth AIC weights that was first presented by Buckland *et al.* (1997). In smooth AIC weighting, the weights are given by:

$$w_M = \frac{\exp(-(1/2)\mathrm{AIC}_M)}{\sum_{M' \in S} \exp(-(1/2)\mathrm{AIC}_{M'})}. \tag{6}$$

It is suggested to subtract the minimum AIC from each model AIC to avoid numerical issues when taking exponents.

To construct confidence intervals for the estimate, Buckland *et al.* (1997) suggest a method for calculating the standard error of $\hat{\mu}$, which can then be used in the usual way: $\hat{\mu} \pm 2se_{\hat{\mu}}$. However, the estimators of interest from stock assessment models are typically non-linear functions of model parameters $(\theta_{1,M}, \theta_{2,M}, \ldots)$, i.e.

$$\hat{\mu}_M = f(\hat{\theta}_{1,M}, \hat{\theta}_{2,M}, \ldots), \tag{7}$$

and are likely to have skewed distributions. Issues of skewness can be addressed if one is willing to assume that model parameters follow a multivariate (MV) normal distribution. The distribution of the model-average estimator would be derived from the fitted models as follows:

(i) Simulate for each model M $(\theta^*_{1,M}, \theta^*_{2,M}, \ldots)$ from an MV normal distribution with mean $(\hat{\theta}_{1,M}, \hat{\theta}_{2,M}, \ldots)$ and variance $\hat{\Sigma}_M$.

(ii) Calculate $\mu^* = \sum_{M \in S} \mu^*_M = \sum_{M \in S} f(\theta^*_{1,M}, \theta^*_{2,M}, \ldots)$.

Confidence intervals can then be derived given sufficient simulations of μ^*. Note that implicit in the derivation of these confidence intervals is the assumption that the estimators $\hat{\mu}_M$ are perfectly correlated (see Buckland *et al.*, 1997).

Bayesian weights

The Bayesian weights we consider are estimates of the posterior-model probability called "harmonic mean estimators" (HME). This estimator, introduced by Newton and Raftery (1994), requires samples from the posterior distribution of the parameters of each model from a Markov Chain Monte Carlo (MCMC) procedure $(\theta^*_{1,M}, \theta^*_{2,M}, \ldots)$, say, for each sample, the log likelihood $l(\theta^*_M)$ is calculated and the weights are given by:

$$w_M \propto \left[\frac{1}{n} \sum_{j=1}^{n} \frac{1}{l(\theta^{j*}_M)} \right]^{-1}, \qquad (8)$$

i.e. the weights are proportional to the harmonic mean of the simulated log-likelihoods.

Confidence intervals can be derived in a similar fashion to the simulation approach described for frequentist weights. There are a variety of options for the distribution of the model parameters. One option would be to assume MV normality using the posterior mean and posterior-covariance matrix derived from the MCMC samples. Another option would be to build a copula approximation to the posterior distribution using an estimate of the posterior-variance matrix and kernel-density estimates of the marginal-posterior distributions of the parameters.

Frequentist simulation-based weighting

We now consider two simulation-based approaches to model averaging. Two reasons to do simulation-based model averaging is to (i) account for the fact that the fixed weights in the previous two examples are, themselves, estimates; and (ii) avoid assuming that the estimators are perfectly correlated.

The first is a frequentist approach developed in Buckland *et al.* (1997), which can be thought of as bootstrapping the full model-selection process, from model fitting through model selection to prediction. It requires that the empirical distribution of the data can be simulated using non-parametric bootstrapping. Single-model inference can then be applied independently to each resample, the estimator $\hat{\mu}^*$ calculated for each resample in turn, and the mean and confidence interval of the estimator derived using standard techniques (Davison and Hinkley, 2003). It should be noted that in fishery contexts, it is often not possible to apply non-parametric bootstrapping. Parametric bootstrapping may be considered, but care should be taken to ensure that the parametric model underlying the parametric bootstrap does not overweight one or more of the models under consideration.

Bayesian simulation-based weighting

The fourth and final method we consider is known as reversible jump Markov Cain Monte Carlo (RJMCMC). We mention this approach for completeness, but the details are beyond the scope of this paper. The improvements that RJMCMC brings over the use of fixed weights, e.g. the HME approach, are that the assumption that the estimators are perfectly correlated is not required, and there is no need to use independent samplers for each model. We refer the reader to King *et al.* (2010) for further details on RJMCMC.

Incorporating expert weights

Expert knowledge of the appropriateness of models can be included in model-averaging procedures by considering them as prior weights on the models $P(M)$. Buckland *et al.* (1997) use a Bayesian analogy to derive the expert weighted version of smooth AIC weights:

$$w_M = \frac{P(M) \exp(-(1/2)\mathrm{AIC}_M)}{\sum_{M' \in S} P(M') \exp(-(1/2)\mathrm{AIC}_{M'})}.$$

For the Bayesian approaches, prior weights on models can be applied directly to the weights w_M, as in the Bayesian setting, model weights are analogous to model probabilities.

Approaches to model averaging can be developed in relative isolation relative to the development of plausible models. An efficient division of labour would be to engage statisticians and programmers to develop tailored methods for model averaging of stock assessment models. This would free up stock assessment scientists, population biologists, fishery biologists, oceanographers, and others to collaborate on proposing a set of plausible models; this task would be periodically reviewed.

Discussion and future challenges

The use of model averaging proposed here should allow fishery scientists to concentrate on fishery science. Within this framework, fishery scientists who are not stock assessment experts in the traditional sense can, along with stakeholders, directly contribute to the stock assessment process. By defining a stock assessment as a (potentially large) set of models, specialized and stakeholder knowledge can be more readily incorporated into the definition of a stock assessment through the development of a list of plausible models and hypotheses. It may be that a stock assessment defined by such a group of models provides more consistent year-to-year advice than a stock assessment defined by a single model, which is periodically refined. It is also possible that this approach would receive more support from the stakeholder community.

As stated previously, the key to the success of the ideas in this paper is a flexible and intuitive interface. In a4a (Jardim *et al.*, 2014), this is achieved through the use of equations for fishing mortality, survey catchability, recruitment, and observation variance, specified as linear models and splines. We think this interface 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, a tier of models could be constructed to give some common structure across stocks. This type of coherence is often sought by having an assessment group be responsible for all flatfish or all demersal fish in an ecoregion. One could set up ecoregion-level scenarios applicable to all stocks, followed by further levels such as demersal and pelagic. This would provide a starting point for any new stock that is assessed, and should improve the coherence and consistency, and perhaps even the transparency, of the assessment process. It may be that the ecoregion level is just too broad to be practical; more appropriate groupings would take into account knowledge of fisheries and the biology of the species, and could be informed by classification analyses.

Model averaging avoids the pitfalls of using a single model: too narrow confidence intervals, overly optimistic tests of significance, and potentially biased results (Claeskens and Hjort, 2008). Dealing with uncertainty in general (model and parameter) in fishery advice is developing, but there needs to be more discussion about the translation of (model and parameter) uncertainty in advice through to the implementation of policy. We do not address these issues here, but instead refer the reader to Hill *et 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 requiring 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 make more assumptions, while a full Bayesian treatment through reversible jump MCMC is very difficult. The smoothed 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). Bootstrap-generated weights are appealing, since they are based on few assumptions, and the algorithm is simple. However, it is typically very difficult to generate bootstrap resamples with fishery data due to the complexity of the underlying sampling schemes.

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 addressing similar issues. In an advisory body, guidelines for practical issues like this could be developed in conjunction with 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 in 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. If we can develop ways to accumulate ideas from a range of sources into a set of stock assessment models, then we could have the basis for a cross-discipline, scientifically defensible, and powerful stock assessment.

Acknowledgements

We thank Ruth King for general discussion on model-averaging methods, and two anonymous reviewers whose helpful comments made for an improved paper.

References

Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S., and Reid, P. C. 2003. Plankton effect on cod recruitment in the North Sea. Nature, 426: 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. 688 pp.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. 1997. Model selection: an integral part of inference. Biometrics, 53: 603–618.

Burnham, K. P., and Anderson, D. R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer-Verlag, New York. 488 pp.

Butterworth, D. S., Punt, A. E., and Smith, A. D. M. 1996. On plausible hypotheses and their weighting, with implications for selection between variants of the Revised Management Procedure. Reports of the International Whaling Commission, 46: 637–642.

Claeskens, G., and Hjort, N. L. 2008. Model Selection and Model Averaging. Cambridge University Press, New York. 320 pp.

Crone, P. R., Maunder, M. N., Valero, J. L., McDaniel, J. D., and Semmens, B. X. (Eds) 2013. 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 pp.

Davison, A. C., and Hinkley, D. V. 2003. Bootstrap Methods and Their Application. Cambridge University Press, New York. 594 pp.

Hilborn, R. 1997. Uncertainty, risk and the precautionary principle. American Fisheries Society Symposium, 20: 100–106.

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

Hill, S. L., Watters, G. M., Punt, A. E., McAllister, M. K., Quéré, C. L., and Turner, J. 2007. Model uncertainty in the ecosystem approach to fisheries. Fish and Fisheries, 8: 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., and Pitois, S. 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: e38758. doi:10.1371/journal.pone.0038758.

Jardim, E., Millar, C. P., Mosqueira, I., Scott, F., Osio, G. C., Ferretti, M., Alzorriz, N., *et al.* 2015. What if stock assessment is as simple as a linear model? The a4a Initiative. ICES Journal of Marine Science, 72: 232–236.

King, R., Morgan, B., Gimenez, O., and Brooks, S. 2010. Bayesian Analysis for Population Ecology. Chapman & Hall/CRC Press, Boco Raton, FL, USA. 456 pp.

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

Methot, R. D. 2009. Stock assessment: operational models in support of fisheries management. *In* The Future of Fisheries Science in North America, pp. 137–165. Ed. by R. J. Beamish, and B. J. Rothschild. Springer Netherlands, Dordrecht. 736 pp.

Methot, R. D., and Wetzel, C. R. 2012. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

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

Newton, M. A., and Raftery, A. E. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society Series B (Methodological), 56: 3–48.

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.

Olsen, E. M., Ottersen, G., Llope, M., Chan, K-S., Beaugrand, G., and Stenseth, N. C. 2010. Spawning stock and recruitment in North Sea cod shaped by food and climate. Proceedings of the Royal Society B: Biological Sciences, 278: 504–510. doi:10.1098/rspb.2010.1465.

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: 208–221.

Poos, J. J., Aarts, G., Vandemaele, S., Willems, W., Bolle, L. J., and 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.

Ralston, S., Punt, A. E., Hamel, O. S., DeVore, J. D., and Conser, R. J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fishery Bulletin US, 109: 217–231.

Reid, P. C., Borges, M. F., and Svendsen, E. 2001. A regime shift in the North Sea circa 1988 linked to changes in the North Sea horse mackerel fishery. Fisheries Research, 50: 163–171.

Shertzer, K. W., Prager, M. H., and Williams, E. H. 2008. A probability-based approach to setting annual catch levels. Fishery Bulletin US, 106: 225–232.

Simmonds, E. J., Campbell, A., Skagen, D., Roel, B. A., and 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, 68: 848–859.

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2010. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science, 18: 7–24.

Handling editor: Emory Anderson