Deep-Sea Research II 134 (2016) 379–389

Contents lists available at ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

Multi-model inference for incorporating trophic and climate uncertainty into stock assessments

James Ianelli[a,*], Kirstin K. Holsman[b], André E. Punt[c], Kerim Aydin[a]

[a] Alaska Fisheries Science Center NOAA Fisheries, 7600 Sand Point Way N.E., Building 4, Seattle, WA 98115, USA
[b] University of Washington JISAO/Alaska Fisheries Science Center NOAA Fisheries, 7600 Sand Point Way N.E., Building 4, Seattle, WA 98115, USA
[c] University of Washington School of Aquatic and Fisheries Sciences, 1122 NE Boat St., Seattle, WA 98105, USA

ARTICLE INFO

Available online 14 April 2015

Keywords:
Model averaging
Model ensemble
Multi-species model
Stock assessment
Fisheries
USA
Alaska Bering Sea
Groundfish
Predator prey
Climate change
Biological reference points
Walleye Pollock
Pacific cod
Arrowtooth flounder

ABSTRACT

Ecosystem-based fisheries management (EBFM) approaches allow a broader and more extensive consideration of objectives than is typically possible with conventional single-species approaches. Ecosystem linkages may include trophic interactions and climate change effects on productivity for the relevant species within the system. Presently, models are evolving to include a comprehensive set of fishery and ecosystem information to address these broader management considerations. The increased scope of EBFM approaches is accompanied with a greater number of plausible models to describe the systems. This can lead to harvest recommendations and biological reference points that differ considerably among models. Model selection for projections (and specific catch recommendations) often occurs through a process that tends to adopt familiar, often simpler, models without considering those that incorporate more complex ecosystem information. Multi-model inference provides a framework that resolves this dilemma by providing a means of including information from alternative, often divergent models to inform biological reference points and possible catch consequences. We apply an example of this approach to data for three species of groundfish in the Bering Sea: walleye pollock, Pacific cod, and arrowtooth flounder using three models: 1) an age-structured "conventional" single-species model, 2) an age-structured single-species model with temperature-specific weight at age, and 3) a temperature-specific multi-species stock assessment model. The latter two approaches also include consideration of alternative future climate scenarios, adding another dimension to evaluate model projection uncertainty. We show how Bayesian model-averaging methods can be used to incorporate such trophic and climate information to broaden single-species stock assessments by using an EBFM approach that may better characterize uncertainty.

Published by Elsevier Ltd.

1. Introduction

The Scientific and Statistical Committees, SSCs, of the Regional Fishery Management Councils are required to provide recommendations for overfishing limits, OFLs, and Acceptable Biological Catches, ABCs, as well as evaluate whether a stock is subject to overfishing or is in an overfished state. For most major stocks, these recommendations are based on the outcomes of quantitative stock assessment methods, which involve fitting population dynamics models to monitoring data collected during fishing and surveys. For stocks managed by the North Pacific and Pacific Fishery Management Council (NPFMC, 2012; PFMC, 2011), the stock assessments are based on single-species models that typically ignore the impacts of time-varying predation mortality.

Most stock assessments involve pre-specifying the values for some of the parameters of the population dynamics model (e.g., the rate of natural mortality, M, fecundity as a function of length or age, and the survey catchability coefficient), making structural assumptions (e.g. vulnerability for a given fleet is a time-varying logistic function of length, recruitment is related to spawning stock size according to the Beverton–Holt form of the stock-recruitment relationship), choosing the data sets used when fitting the model (e.g., should fishery catch rate data be used or ignored given uncertainties regarding the relationship between catch rate and abundance), and assigning statistical weights to different assessment data components. Although model fits to data may be similar, the results of stock assessments can be highly sensitive to parameter values and choices regarding model structure (e.g., Myers et al., 1994; Taylor and Stephens, 2013; Holsman et al., 2016; Patterson et al., 2001)

* Corresponding author.
E-mail addresses: Jim.Ianelli@noaa.gov (J. Ianelli), kirstin.holsman@noaa.gov (K.K. Holsman), aepunt@uw.edu (A.E. Punt), Kerim.Aydin@noaa.gov (K. Aydin).

http://dx.doi.org/10.1016/j.dsr2.2015.04.002
0967-0645/Published by Elsevier Ltd.

In general, fisheries management advice (and hence OFLs and ABCs) is based on a single "best" model (and hence set of assumptions), and uncertainty is quantified about that model conditioned on its assumptions being correct. Typically, uncertainty is quantified using asymptotic methods, bootstrapping, or Bayesian methods (Magnusson et al., 2013). However, many sources of uncertainty are ignored when applying these methods, so the measures of uncertainty reported to managers usually underestimate the true amount of uncertainty (Ralston et al., 2011; Punt et al., 2012). The difference between the OFL and the ABC for a stock (the "buffer") is meant to reflect the amount of scientific uncertainty. ABCs are often set so that the probability that the ABC exceeds the true OFL equals a selected value, P^* (where $P^* < 0.5$), i.e. $P(\mathrm{ABC} > \mathrm{OFL}) = P^*$ (Prager et al., 2003; Shertzer et al., 2008; Prager and Shertzer, 2010). However, the true probability that the ABC exceeds the OFL will be larger than the P^* estimate if uncertainty is underestimated. This would occur if the uncertainty associated with assumptions regarding model structure were ignored. Here we propose an example on how an EBFM approach could be used from multiple alternative ecosystem models to provide a better accounting of structural uncertainties.

The use of multispecies and ecosystem models for fisheries management is generally considered to be a key component of Ecosystem Based Fisheries Management (EBFM) (Marasco et al., 2007; Plagányi, 2007). However, similar to single-species stock assessment methods, projections based on two ecosystem models (or variants of one ecosystem model with alternative assumptions) often reflect uncertainty about model structure and assumptions regarding values for pre-specified parameters. For example, Kaplan et al. (2013) evaluated the impacts of depleting forage species in the California Current ecosystem using Atlantis (Fulton et al., 2004, 2011a, 2011b; Horner et al., 2010) and Ecopath-with-Ecosim (Christensen and Walters, 2004; Field et al., 2006). However, the results from these two ecosystem models differed markedly and increased the uncertainty about whether reducing forage species abundance would have a negative or positive effect on some ecosystem components. In another study, Kinzey and Punt (2009) showed that the results of a multispecies stock assessment were sensitive to the choice of the relationship between predation mortality and the density of predators and prey. The multispecies models examined by Kinzey and Punt (2009) predicted that Pacific cod (*Gadus macrocephalus*) in the Aleutian Islands could have been increasing or decreasing prior to 1990 depending on this relationship. This illustrates that assumptions about functional responses can affect predictions in critical ecosystem components. Regarding reference points, including trophic interactions in models can have large impacts, especially for key prey species (Collie and Gislason, 2001).

These considerations imply that alternative model formulations should be based on plausible working hypotheses and assigning model weights or prior probabilities (given the a priori likelihood of the specified model). Ideally, within-model estimation uncertainty would further contribute to statistical inference of the combined multiple-model results. Results typically include projections of population size under alternative harvest control rules or catch scenarios as well as specific outputs such as OFLs and ABCs. Model averaging allows diverse, yet plausible, model results to collectively be used to guide management, and can provide estimates of uncertainty derived from both data fit (as is the case with individual models) as well as model structure and assumptions. It allows the uncertainty regarding which model is correct to be reflected in the advice used for management rather than simply selecting a single "best" model and ignoring the others.

Here we provide a brief review of the multi-model inference for fisheries assessment applications, focusing in particular on two alternative ways to implement model averaging for EBFM. We then use model averaging to integrate the results from three classes of model (single-species, temperature-specific single-species, temperature-specific multispecies) for three scenarios regarding future catch in the eastern Bering Sea in terms of impacts on the spawning stock biomass of walleye pollock (*Gadus chalcogrammus*), Pacific cod and arrowtooth flounder (*Atheresthes stomias*).

2. Overview of model averaging

This study focuses on practical approaches for model averaging and contrasts weighted versus unweighted methods. For the weighted approach, we focus on a Bayesian Model Averaging (BMA) and categorize unweighted methods as "ensemble" forecasting. Burnham and Anderson (2002) detail a number of alternatives, e.g., weighting models using AIC and others contrast approaches including frequentist weights (Millar and Jardim, 2015). For our purposes, BMA requires that estimates of the posterior probability of each candidate model be available. This probability needs to be derived by fitting the model to available data. However, the probability of the model given the data cannot be derived for all models (e.g. dynamic ecosystem models) such as Atlantis (Fulton et al., 2004, 2011a, 2011b; Kaplan et al., 2013) or the Forage/Euphausiid Abundance in Space and Time (FEAST) model (Aydin et al., 2016) because they cannot be formally fitted to data. It is consequently impossible to apply BMA or methods which weight models based on other metrics of model fit such as AIC weights in many situations. When this is the case, posterior probability distributions can be approximated by "envelopes of plausibility" derived from ensemble/Monte Carlo runs of each model where each run is based on a different (yet plausible) set of parameters, with the probability assigned to each model based on expert judgment (i.e. the "Delphi method"), a process which we refer to as "ensemble" forecasting. Butterworth et al. (1996) proposed the following four-level scheme to assign 'plausibility ranks' to the hypotheses underlying alternative models that could be used to weight models when "ensemble" forecasting is conducted:

1. how strong is the basis for the hypothesis in the data for the species or region under consideration;
2. how strong is the basis for the hypothesis in the data for a similar species or another region;
3. how strong is the basis for the hypothesis for any species; and
4. how strong or appropriate is the theoretical basis for the hypothesis?

For the population dynamics models typical of fisheries management, BMA and ensemble forecasting fundamentally involve making projections. Each model can be projected multiple times (the outcomes will differ if there are multiple parameter choices for each model or the projections account for future stochasticity due to recruitment variability for example). The results of model averaging can be summarized by the overall mean or median of some quantity of management or scientific interest (the median is used here), the spread of results, and by individual trajectories. The mean of the projections is a "best estimate", but simply showing the median trajectory loses the advantage of conducting multiple forecasts, namely to characterize uncertainty. Ianelli et al. (2011) summarized the results of projections for multiple models by illustrating intervals containing 50% and 80% of the combined outcomes over future climate scenarios to illustrate the overall uncertainty. They also showed a subset of individual trajectories to characterize the nature of year-to-year variability.