Version of Record: https://www.sciencedirect.com/science/article/pii/S0967064515001058
Manuscript_7b6ac58339e0b9d8c59a069bdd73bc5c

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[d]

[a]Email: Jim.Ianelli@noaa.gov , Alaska Fisheries Science Center NOAA Fisheries. Address: 7600 Sand Point Way N.E., Building 4, Seattle, Washington 98115, USA.

[b]Email: kirstin.holsman@noaa.gov , Affiliation: University of Washington JISAO/ Alaska Fisheries Science Center NOAA Fisheries. Address: 7600 Sand Point Way N.E., Building 4, Seattle, Washington 98115,USA

[c]Email: aepunt@uw.edu, University of Washington School of Aquatic and Fisheries Sciences. 1122 NE Boat St. Seattle WA 98105, USA.

[d]Email: Kerim.Aydin@noaa.gov , Alaska Fisheries Science Center NOAA Fisheries. Address: 7600 Sand Point Way N.E., Building 4, Seattle, Washington 98115, USA.

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.

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, PMFC 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. this issue, Patterson et al. 2001)

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(ABC>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, 2011; 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 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 et al. 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, 2011; Kaplan et al. 2014) or the Forage/Euphausiid Abundance in Space and Time (FEAST) model (Aydin et al. this volume) 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.

2.1 Bayesian Model Averaging (BMA)

The philosophy underlying Bayesian model averaging has been outlined by several authors (e.g. Buckland *et al.*, 1997; Durban *et al.*, 2005; Hoeting *et al.*, 1999; Kass and Raftery, 1995; Raftery *et al.*, 2005; Chimielechi and Raftery, 2011). Ideally, BMA involves fitting the range of models to the available data and computing the probability of each model given the data. This weighting of models should ideally be conducted using Bayes factors, which quantitatively evaluates the credibility of a model relative to other models (Aitkin, 1991; Kass and Raftery, 1995). However, studies have weighted alternative models, using, for example, the Deviance Information Criterion (Spiegelhalter *et al.,* 2002), Akaike's

Information Criterion (Akaike, 1973; Burnham and Anderson, 1998), and the Bayes Information Criterion (Schwartz, 1978). The latter two weighting schemes are non-Bayesian, but are relatively easy to compute in contrast to Bayes factor and DIC which require that a Bayesian analysis be conducted, which can be computationally prohibitive even for relatively simple ecosystem models (e.g. Parslow *et al.*, 2013).

Bayes factor, DIC, AIC, and BIC can only be computed if each model is fit to the same data set. If the models are fit to different data sets, weighting each model would require a more *ad hoc* approach, such as fitting the models to a subset of the data and predicting the remaining data (i.e., cross-validation). In this case, the weight assigned to each model could be proportional to the inverse of the mean square error associated with its predictions.

Given probabilities for each model, the Bayesian model averaged forecast is constructed by conducting multiple projections for each model and generating the overall forecast by selecting projections at a sampling rate proportional to the probability of the model. Table 1 summarizes an application of Bayesian model averaging in which five models are used to predict the fishing mortality and spawning biomass corresponding to maximum sustainable yield (F_{MSY} and S_{MSY}) for Atlantic cod (*Gadus morhua*; Hill et al. 2007; Brodziak and Legault 2005). The best model in Table 1 is RBH, but model RZBH is almost as likely. The model-averaged results are as expected closest to the best models, but the standard errors for the model-averaged results are larger than for either of the two best models. The weights assigned to each model in Table 1 are based on the Bayes factor. Bayes factors can be computed in this case because all of the models use the same data and the models are fairly simple.

2.2 Ensemble forecasting

Ensemble forecasting involves generating multiple projections of future system state under different choices for assumptions or parameter values. In principle, both structural and parameter uncertainty can be addressed through the use of multi-model ensembles. This approach is widely used in climate modeling where uncertainty is reflected in the accuracy of the approximations to the well-known and accepted physical principles of climate, and the inherent variability of the climate system. The climate system is chaotic, and the timing and phases of major and long-lasting fluctuations are largely unpredictable beyond time-horizons of a few years. Consequently, slightly different initial conditions for a climate model can lead to markedly different outcomes 40-50 years into the future. Whether including climate in population dynamics models has major impacts on the estimated future state of the populations under investigation depends on how the dynamics of the populations are linked to climate and the strength of the associated relationships.

Probabilities can be assigned to model configurations (the underlying model equations and the values for its parameters) or entire model configurations can be considered plausible using hindcast simulations of past conditions (e.g. Overland and Wang, 2007), although past performance is not necessarily a good indicator of success in simulating future climate (Reifen and Toumi, 2009). A'mar *et al.* (2009) based projections on six general circulation models which were selected for both their accuracy with respect to the historical data and their predictions with respect to future climate scenarios. Specifically, these six models were in the subset of models that replicated the spatial pattern and temporal characteristics of the first principal component of sea surface temperature (SST) in the North Pacific Ocean (the PDO) observed in the latter half of the twentieth century (A'mar *et al.*, 2009). It is worth noting that unlike the tactical application of ensemble forecasts used in climate systems, most fisheries applications are geared to making strategic decisions (i.e., devising tactical approaches that are robust to the plausible, yet uncertain, future outcomes).

2.3 Fisheries examples of model averaging

Model averaging in fisheries assessments is rare; the focus for fisheries management tends to be either selection of a best model or identification of harvest control rules that are robust to model selection and parameter value uncertainty (Butterworth, 2007). However, there are a few examples of where model averaging has been applied to fisheries population dynamics models and these are reviewed here.

2.3.1 Weighted model averaging

BMA has been applied to account for uncertainty regarding the form of the stock-recruitment relationship (usually Ricker vs Beverton-Holt) and the error structure (autocorrelated or not, and the distribution for the residuals) (Patterson, 1999; Brodziak and Legault, 2005). BMA was used by Brandon and Wade (2006) to account for uncertainty regarding the form of the population dynamics model underlying a stock assessment (density-dependent or non-density-dependent, and whether the stock was at its environmental carrying capacity at the start of the modeled period) in an assessment of the Bering-Chukchi-Beaufort seas stock of bowhead whales, *Balaena mysticetus*. The weights assigned to each model by Brandon and Wade (2006) were based on Bayes factor; they developed their posterior distributions for each model using the sample-importance-resample algorithm, which allowed straightforward computation of the posterior probability of each model. Wilberg and Bence (2008) used Monte Carlo simulation to show that model averaging of alternative formulations for how fishery catchability changes over time performed better than using DIC to select a “best” model.

Brodziak and Piner (2010) used BMA to integrate uncertainty due to the form of the stock-recruitment relationship (Ricker or Beverton-Holt), the extent of autocorrelation about the stock-recruitment relationship, and two values for the steepness of the stock-recruitment relationship for striped marlin (*Tetrapturus audax*) in the North Pacific. Unlike Broziak and Legault (2005), Brodziak and Piner (2010) approximated the Bayes factor using BIC.

2.3.2 “Unweighted” model averaging

The ensemble approach has been applied fairly extensively to management of groundfish off the US west coast. For example, it was applied to rebuilding strategies for cowcod (*Sebastes levis*) off southern California that were developed by Dick and Ralston (2009). They provided forecasts for a range of assessment models, each of which was conditioned on one of a set of values for the steepness of the Beverton-Holt stock-recruitment relationship. Each projection was weighted based on a pre-specified probability distribution for steepness. Hamel (2011) conducted projections to evaluate times for Pacific Ocean Perch (*Sebastes alutus*) to rebuild to the proxy for the biomass at which maximum sustainable yield is achieved, B_{MSY} for three models, given different levels of future fishing mortality and catch. Two of these models were assigned probability 0.25 and the third model was assigned a probability of 0.5 based on relative plausibility provided in Hamel and Ono (2011).

Ianelli *et al.* (2011) evaluated the performance of management strategies for walleye pollock in the eastern Bering Sea. Recruitment was linked to predictions of SST from 82 Intergovernmental Panel on Climate Change (IPCC) models—SST, among other environmental factors, was found by Mueter *et al.* (2011) to be a possible factor affecting pollock recruitment. These 82 models were selected by downscaling IPCC models to the eastern Bering Sea ecosystem and using retrospective studies to identify models that performed best for this system (Wang *et al.*, 2010).

Kolody et al. (2008), Kolody (2011), and Davies *et al.* (2012) developed an ‘uncertainty grid’ for assessments of swordfish (*Xiphias gladius*) in the Indian and Pacific Oceans, and explored structural uncertainty in a balanced factorial design. The results of the assessment were presented in terms of box plots of output statistics for each level of the factors considered. Kolody *et al*. (2008) explored sensitivity to stock-recruitment steepness, mixing proportions, growth rate/maturity/mortality options, the extent of variation about the stock-recruitment relationship, selectivity constraints, and data weights. Of 768 model configurations, a set of 192 model configurations considered the "most plausible ensemble" were used to summarize stock status. This ensemble was selected using three metrics: the root mean square fit to the catch rate index, the effective sample sizes for the length-frequency data, and the difference between observed and model-predicted mean catch lengths (similar to the method of Francis 2011). Kolody (2011) assigned weights to each of the factors on which the uncertainty grid was based using auxiliary information and the quality of the fits to the data, which led to some factors, such as that recruitment is related deterministically to spawning biomass, being assigned zero weight.

Gardmark et al. (2013) evaluated a "biological ensemble modeling approach" over different ecological assumptions including climate forcing. They evaluated seven ecological models ranging from single-species to food web models and concisely distilled some key population indicators (e.g., extinction, increase from 2009 levels) under high and low fishing mortality scenarios. Importantly, they were able to highlight the relative sensitivity of biological characteristics and multispecies interactions from climate effects on the populations of interest.

An unusual form of ensemble modeling has been applied to calculate strike limits for the Bering-Chukchi-Beaufort Seas stock of bowhead whales and the Eastern North Pacific stock of gray whales, *Eschrichtius robustus*. This involves calculating strike limits from two different methods and averaging them (Punt and Donovan, 2007). The philosophy underlying this approach is that each model can be wrong some of the time so averaging model results will lead to an outcome that is never very badly wrong (but is usually somewhat wrong).

3. Application to walleye pollock, Pacific cod and arrowtooth flounder

3.1 Alternative models

Three classes of models formed the basis for the analysis (Table S.1):

1. The single species assessment models currently used by the AFSC to provide management advice for Eastern Bering Sea (EBS) walleye pollock (e.g., Ianelli *et al.*, 2012), Pacific cod (e.g. Thompson and Lauth, 2012), and arrowtooth flounder (e.g., Spies *et al.*, 2012). The assessments for these stocks are based on software developed specifically for those stocks coded using AD Model Builder (Fournier *et al.*, 2012). All of the three single species assessments have the following features in common: (a) they are fundamentally age-structured and use an annual time step; (b) estimates of annual fishing mortality rates are conditioned on the total catch (retained and discards) estimates, (c) fishery data (catch biomass and catch proportions at age) are aggregated over season*s* and areas within each year, (d) proportions at age from surveys and fisheries are fitted using estimated (or assumed) multinomial sample sizes, (e) life history parameters and environmental effects are assumed constant, and (f) survey indices (abundance or biomass) are modeled using lognormal assumptions and annually-specified observation errors (variances). Uncertainty in the projections based on these models reflects both parameter uncertainty (including selectivity), captured through MCMC sampling from the joint posterior distribution, and process error, captured through variation in recruitment about mean recruitment.
2. The Temperature-Specific Multispecies Model (MSMt, aka CEATTLE for Climate-Enhanced Age-based model with Temperature-specific Trophic Linkages and

Energetics; Holsman *et al.*, *this issue*) is an example of a "model of intermediate complexity" (Plagányi *et al.*, 2014). The implementation of MSMt for the eastern Bering Sea includes the three focal species, models natural mortality for each species and age as the sum of a pre-specified residual natural mortality and time-varying predation mortality due to the predators included in the model. Predation mortality is driven by temperature-dependent daily ration and a suitability function, which is based on observed proportions of each prey species by age in the diets of each predator species by age. Weight-at-age is also assumed to depend on temperature and varies annually. The parameters of MSMt are estimated by fitting the model to data on catch age-composition as well as survey biomass index and age-composition data. The projections of the model assume that future recruitment at age-0 is lognormal about mean recruitment. Two variants of MSMt are considered, one that includes multispecies predator-prey interactions (MSMtA), and one which assumes natural mortality is constant over time (MSMtB). The latter differs from the single species models used by AFSC in that weight-at-age in MSMtB depends on temperature and some other structural simplifications (e.g., constant fishery selectivity over time). Stochasticity in future projections based on MSMtA and MSMtB account only for process error in future recruitment.

Four climate scenarios are considered for MSMtA and MSMtB because temperature influences weight-at-age and the predation mortality function in MSMt. The first of four scenarios assumes that future temperature equals the average observed temperatures (temperature scenario 1). The next three of four temperature scenarios are based on mean summer water column temperature predicted from a ROMS model for the Bering Sea forced by three statistically downscaled global climate models (Wang et al., 2010). These temperature scenarios were:

Scenario 2	"ECHO-G version 4, T30 resolution model" (Legutke and Voss, 1999),
Scenario 3	"CCMA model" (Canadian Centre for Climate Modelling and Analysis CGCM3-t47; Flato et al., 2000, Flato and Boer 2001, Kim et al. 2002, 2003), and
Scenario 4	"MIROC 3.2" (Watanabe *et al.*, 2011, K-1 model developers, 2004)

A single realization of each of these three climate scenarios was used as plausible characterizations of atmospheric forcing and oceanic boundary conditions for the regional ROMS forecasts of the Bering Sea (2013 to 2040; Fig. 1).

3.2 Projections

Each projection of the 1,000 iterations for each model involved the forecast period (2013-2039). For the purposes of this study, projected recruitment was assumed to have the same mean and variance for each species as observed for the historical period. Additionally, population model projections included three alternative constant future catch scenarios:

1. Catches set to the mean catch over the most recent 15 years (Table 2);
2. Catches set to the maximum catch over the most recent 15 years (Table 2);
3. No future catches of any species.

4. Results

4.1 Results by model scenario

Figures 2-4 shows the time-trajectories of spawning stock biomass the three models individually and includes the effect of fishing. Also, three sets of results are shown for the MSMtA and MSMtB models, one for each climate scenario. There is relatively little difference amongst the four climate scenarios for the MSMtA and MSMtB models, although the projections including alternative future temperature scenarios (rows 2-4 in Figures 3 and 4) are more variable. This is unsurprising given the low variability shown by the data in Figure 1.

The general patterns between the two single-species models (AFSC; Figure 2 and MSMtB; Figure 4) share some qualitatively similar traits but also show some major differences. Specifically, the declines in abundance under the mean and maximum catches for pollock and Pacific cod are much greater for MSMtB. The cause of the differences between the two single-species models is unlikely to be due to temperature impacts on weight-at-age because the qualitative difference in results remains even when future temperature equals the historical mean. This is more likely due to differences in the assumptions regarding fishery selectivity in projections where MSMtB is balancing periods of selectivity shifted to be younger than the age-specific maturity schedule whereas the more recent selectivity trend is more focused on older pollock. In contrast, the single-species model uses the assumption that the most recent 5-year average selectivity-at-age is most appropriate for projection purposes. This points out that simplifications in the MSMt models' treatment of individual species are important to consider in evaluating projecting interactions and such factors should be considered when developing relative weights among models.

The results are also markedly sensitive to whether MSMt is applied in single-species or multispecies mode (Figures 3 and 4). Specifically pollock is predicted to decline and then rebuild under all catch scenarios (including zero catch) for MSMtA (multispecies mode) whereas pollock is predicted to increase under zero catch and decline under mean and maximum catches for MSMtB (single-species mode). The difference between the single-species and multispecies predictions for cod is attributable primarily to the combined effect of cod and arrowtooth predation and pollock cannibalism in MSMtA. The abundance of Pacific cod is more robust between MSMtA and MSMtB, but the extent of decline in cod abundance is much greater under the multispecies version of the model (again reflecting the importance on the source of age-1 cod predation). The trends in biomass of arrowtooth flounder are similar between MSMtA and MSMtB for the first few years of the projection period. However, unlike the MSMtB, MSMtA predicts stability or an increasing trend in arrowtooth abundance post 2030 (Figures 3 and 5). Because the increasing trend in arrowtooth abundance is not evident in MSMtB, this result is probably a consequence of temperature effects of weight-at-age combined with a slight decrease in predation by cod (and possibly pollock).

4.2 Model averaged results

Figure 5 shows results for three model classes when results are pooled over climate scenarios for the two MSMt models. Example results of the projections, including those based on model averaging, are summarized quantitatively in Table 3 by the median and 50% and 90% intervals for spawning biomass in 2039, the last year of the projection period.

As expected, model averaging across climate scenarios (assigning equal weight to each climate scenario) confirms that the impacts of the different climate scenarios on the model outcomes are not large (Figure 5). The widths of the 90% intervals in Figure 5 for the model-averaged results for MSMtA and MSMtB are narrower than those for the individual climate models, suggesting that variation in recruitment has a larger impact on uncertainty than the choice of climate model in this case.

Figure 5 (bottom panel) and Table 3 show results when all models are averaged (with equal weight assigned to all three models). In this case, the widths of the 90% intervals are wider for the model-averaged results than for the results for each individual model, reflecting that between-model variation is greater than the variation due to climate scenario (models MSMtA and MSMtB), parameter uncertainty (AFSC single-species models) and recruitment variation (all three models).

5 Discussion

Effectively capturing uncertainty is a key focus for modern stock assessment science, and quantifying uncertainty in fisheries stock assessment models has been a focus for stock assessment scientists for decades (e.g., Patterson, 1999, Pattern *et al*., 2001, Hill *et al*., 2007, Magnusson *et al*., 2013). A full accounting for uncertainty requires adequately representing uncertainty regarding growth rates, natural mortality, the form and parameters of the stock-recruitment relationship, and how data are weighted. However, conventional approaches to quantifying uncertainty fail to capture 'model uncertainty', i.e. the uncertainty associated with the structural assumptions of a model. In general, single-species stock assessments make a small number of very strong assumptions (e.g. that natural mortality is independent of time). On the other hand, multispecies and ecosystem models typically require more numerous and more specific assumptions (e.g. that the functional relationship for predator and prey has the Holling Type II form) and commonly assume other simplifying assumptions (e.g., constant fishery selectivity). Since these model types each make somewhat different types of compromises, applying model averaging approaches (BMA or ensemble) is an appropriate way to express a broader and presumably more comprehensive range of uncertainty.

Variability in climate scenarios contribute less to overall uncertainty than recruitment variation for the MSMtA and MSMtB models (Figure 5, Table 3). However, model uncertainty is a more marked source of uncertainty than parameter uncertainty, recruitment variation, and the choice of climate scenario. It is, however, noteworthy that the impact of model uncertainty depends on the particular catch scenario under investigation. It is largest for the zero catch scenario, in particular given the impact of 'release' of Pacific cod, a major predator of pollock in the MSMtA model. The models are more consistent in their predictions when the projections are based on the mean catch and most consistent for the projections based on the maximum catch where the biomass of predators and concomitant predation mortality is lowest (and thus differences between model parameterizations of predation mortality are lowest).

Comparing alternative models has also raised another challenge in developing more "holistic" multispecies models. As noted above, simplifications in the multispecies model (e.g., constant fisheries selectivity) can introduce substantial differences in projections. For example, considering the estimated selectivity (Figure S.1) and mean body mass at age (but the same natural mortality-at-age) for pollock results in substantially different yield curves between the MSMt model and the single species model used for projections (Figure S.2). This highlights the need to compare potentially subtle fishery and demographic characteristics when comparing multispecies projections with their single-species counterparts. Nevertheless, the different ways of modeling selectivity and body mass at age are plausible and do reflect alternative hypotheses. Such choices made by modelers (lacking evidence to favor alternatives) remain an important, often unaccounted-for source of uncertainty. Ralston *et al*. (2011) characterized this type of uncertainty in single-stocks species stock assessments by the extent of different assessment variation among analysts and a similar approach could be extended to this type of model averaging.

The model forecasts were assigned equal probability in constructing the model-averaged forecasts. This was because there is no way for the hindcast and forecast skills of the three models to be compared at present. The ideal of using Bayes factor (or DIC, AIC, BIC) is infeasible in this case because although the parameters of the single-species model and MSMt are estimated by fitting them to monitoring data, each model has slightly different statistical weights and/or levels of aggregation in the data sources. In principle, each model could be weighted objectively by a cross-validation-like approach. For example, one could fit the model including data only up to 2008 and using the fitted model to predict the survey estimates of abundance for 2009, 2010, 2011, etc. given the catches that actually occurred during 2009, 2010, 2011; models that fit the observations better would obtain a higher weight.

The illustrative application of this paper was based on three models. However, there are several other models that could have been included in the application. These include alternative multispecies models such as the multispecies virtual population analysis model of Jurado-Molina and Livingston (2002), and the statistical multispecies model developed by Kinzey and Punt (2009). Other models available for the Bering Sea include an Ecosim model (Aydin *et al*., 2007), the FEAST model, the multispecies surplus production model of Mueter and Megrey (2006), and a spatially-structured model of pollock (Hulson *et al*., 2013)

Future work could involve evaluating the hindcast and forecast skill of projections based on a single model as well as on a model average of multiple models (c.f. Wilberg and Bence, 2008). This could involve fitting the model to a subset of the data and conducting projections. The skill of the modeling approach could then be evaluated in terms of the percentiles of the predicted distributions in which the actual observations occurred. Ideally, the percentiles associated with the data should be uniformly distributed over 0-100. Large numbers of observations in the upper and lower tails of the forecast distributions would suggest that uncertainty is underestimated while no or few observations in the tails would suggest that uncertainty is overestimated. The benefits of using single models or model-averaged results could also be evaluated using simulations in which a true model is defined and data typical of an actual situation generated. This approach has been used extensively to evaluate the performance of single-species stock assessment methods, but has only been applied in a limited capacity for multispecies and ecosystem models, Kinzey (2010) being a noteworthy exception).

Ultimately all approaches to applying model averaging involve subjective choices. These range from the initial choice of models to consider, along with a prior probability associated with each model. The latter is particularly a concern when many of the models are based on the same underlying philosophy. For example, the single-species assessments and MSMt, while different in several respects, make identical assumptions regarding many biological and fishery processes and cannot be considered to be totally independent. Similarly, MSMtA and MSMtB are identical except that the former allows for time-varying predation mortality while the latter does not.

The outcomes from this model averaging exercise are expressed in terms of time-trajectories of spawning output given a time-series of catches. However, the primary use of stock assessments is to define whether overfishing is taking place and whether the stock is in overfished stock, which, given the way fisheries management advice is provided in the US, requires a way to define the management reference points S_{MSY} and F_{MSY}. These reference points are well-defined for the single-species assessments (although the precision of the estimates even from single-species models can be poor). However, there are several alternative ways to define these reference points for multispecies and ecosystem models (Moffit *et al*., this issue). Holsman *et al*. (this issue) illustrates how S_{MSY} and F_{MSY} can be calculated for a range of definitions for S_{MSY} and F_{MSY} for three species included in the analyses of this paper. Ultimately, model averaging could be used to compute ensemble distributions for stock status relative to reference points if probabilities could be assigned to each of the definitions for S_{MSY} and F_{MSY}.

We suggest that model uncertainty can be as large, or even exceed, many of the types of uncertainties considered routinely in stock assessments. Use of model averaging can quantify the range of outcomes from multiple models and better characterize uncertainty. Given that ABCs and OFLs are often reduced based on scientific uncertainty, accounting for model uncertainty can inform buffers between OFLs and ABCs and hence provide an improved ability to achieve fishery goals such as avoiding overfishing and preventing stocks from becoming overfished.

Acknowledgements

This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA10OAR4320148, Contribution No. 2229. Excellent suggestions from Ian Stewart, Mike Sigler, and two anonymous reviewers helped improve the earlier draft. Thanks also for support from NOAA's Integrated Ecosystem Assessment program. This work benefitted from the support of colleagues at BEST-BSIERP Bering Sea Project (this is publication number 152) and the North Pacific Research Board (publication number 530).

References

A'mar, Z.T., Punt, A.E., Dorn., M.W., 2009. The evaluation of two management strategies for the Gulf of Alaska walleye pollock fishery under climate change. *ICES Journal of Marine Science* 66: 1614-1632.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle, in: 2nd International Symposium on Information Theory. Academiai Kiado, Budapest, pp. 267–281.

Aitkin, M. 1991. Posterior Bayes factors. *Journal of the Royal Society Series* B 53: 111–142.

Aydin, K., Gaichas, S., Ortiz, I., Kinzey, D., Friday, N., 2007. A Comparison of the Bering Sea, Gulf of Alaska, and Aleutian Islands Large Marine Ecosystems Through Food Web Modeling. NOAA Tech Memo NMFS AFSC 178, 298 pp.

Brandon, J.R., Wade, P.R., 2006. Assessment of the Bering-Chukchi-Beaufort Seas stock of bowhead whales using Bayesian model averaging. *Journal of Cetacean Research and Management* 8: 225–239.

Brodziak, J., Legault, C.M., 2005. Model averaging to estimate rebuilding targets for overfished stocks. *Canadian Journal of Fisheries and Aquatic Sciences* 62: 544–562.

Brodziak, J., Piner, K., 2010. Model averaging and probable status of North Pacific striped marlin, *Tetrapturus audaz. Canadian Journal of Fisheries and Aquatic Sciences* 67: 793–805.

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

Burnham, K.P., Anderson, D.R., 1998. Model Selection and Multimodel Inference: a Practical Information-theoretical Approach. Springer Verlag, New York. 353 pp.

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. *Reports of the International Whaling Commission* 46: 637–640.

Butterworth, D.S., 2007. Why a management procedure approach? Some positives and negatives. *ICES Journal of Marine Science* 64: 613–617.

Christensen, V., Walters, C.J., 2004. Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling* 172: 109–130.

Chimielechi, R.M., Raftery, A.E., 2011. Probabilistic visibility forecasting using Bayesian Model Averaging. *Monthly Weather Review* 139: 1626–1636.

Collie, J.S. Gislason, H., 2001. Biological reference points for fish stocks in a multispecies context. Canadian Journal of Fisheries and Aquatic Sciences, Vol. 58, No. 11, p. 2167-2176.

Davies, N., Hoyle, S., Hampton, J., 2012. Stock assessment of striped marlin (*Kajikia audax*) in the southwest Pacific Ocean. WCPFC document WCPFC-SC8-2012/SA-WP-05. 90 pp.

Dick, E.J., Ralston, S., 2009. Cowcod Rebuilding Analysis. Pacific Fishery Management Council, 7700 NE Ambassador Place, Portland, OR 97220. 32 pp.

Durban, J.W., Elston, D.A., Ellifrit, D.K., Dickson, E., Hammond, P.S., Thompson, P.M., 2005. Multisite mark-recapture for cetaceans: population estimates with Bayesian model averaging. *Marine Mammal Science* 21: 80–92.

Flato, G.M., Boer, G.J., Lee, W.G., McFarlane, N.A., Ramsden, D., Reader, M.C., Weaver, A.J., 2000. The Canadian Centre for Climate Modelling and Analysis global coupled model and its climate. Climate Dynamics 16, 451-467

Flato, G.M. and G.J. Boer, 2001: Warming Asymmetry in Climate Change Simulations. Geophys. Res. Lett., 28, 195-198.

Field, J.C., Francis, R.C. and K. Aydin. 2006. Top-down modelling and bottom-up dynamics: linking a fisheries-based ecosystem model with climate hypotheses in the Northern California Current. *Progress in Oceanography* 68**:** 238–270.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* 68: 1124–1138.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* 27: 233–249.

Fulton, E.A., Smith, A.D.M., and Johnson, C.R. 2004. Effects of spatial resolution on the performance and interpretation of marine ecosystem models. *Ecological Modelling* 176: 27–42.

Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R.J., Smith, A.D.M., and D.C. Smith. 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish and Fisheries* 12: 171–188.

Fulton, E.A., Smith, A.D.M., Smith, D.C. and I.E. van Putten. 2011. Human behavior: the key source of uncertainty in fisheries management. *Fish and Fisheries* 12: 2–17.

Gardmark, A., Lindegren, M., Neuenfeldt, S., Blenckner, T., Heikinheimo, O., Muller-Karulis, B., Niiranen, S., Tomczak, M.T., Aro, E., Wikstrom, A., Mollmann, C., 2013. Biological ensemble modeling to evaluate potential futures of living marine resources. Ecological Applications 23(4):742-754.

Hamel, O.S., 2011. Rebuilding analysis for Pacific Ocean Perch in 2011. Pacific Fishery Management Council, 7700 NE Ambassador Place, Portland, OR 97220. 15 pp.

Hamel, O.S., Ono, K., 2011. Stock Assessment of Pacific Ocean Perch in Waters off of the U.S. West Coast in 2011. Pacific Fishery Management Council, 7700 NE Ambassador Place, Portland, OR 97220. 165 pp.

Hill, S.L., Watters, G.M., Punt, A.E., McAllister, M.K., LeQuere, C., Turner, J., 2007. Model uncertainty in the Ecosystem Approach to Fisheries. *Fish and Fisheries*. 8: 315–336.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian Model Averaging: a tutorial. *Statistical Sciences* 14: 382–417.

Holsman K. K., Ianelli, J., Aydin, K., Punt, A. E., Moffitt, E.A., *This issue.* Comparative biological reference points estimated from temperature-specific multispecies and single species stock assessment models. Deep Sea Res II 00:00-00.

Horner, P.J., Kaplan, I.C., Marshall, K.N., Levin, P.S., Harvey, C.J., Hermann, A.J., Fulton, E.A., 2010. Design and parameterization of a spatially explicit ecosystem model of the Central California Current. NOAA Technical Memorandum NMFS-NWFSC-104, NOAA, USA. 140 pp

Hulson, P-J,F., Quinn, T.J. II, Hanselman, D.H., Ianelli, J.N., 2013. Spatial modeling of Bering Sea walleye pollock with integrated age-structured assessment models in a changing environment. *Canadian Journal of Fisheries and Aquatic Sciences* 70: 1402–1416.

Ianelli, J.N., Honkalehto, T., Barbeaux, S., Kotwicki, S., Aydin, K., Williamson, N., 2012. Pg 51-155 in Stock Assessment and Fishery Evaluation Report for the Groundfish Resources for the Bering Sea/Aleutian Islands Regions. North Pacific Fishery Management Council, 605 West 4th Ave, Suite 306, Anchorage AK 99501.

Ianelli, J.N., Hollowed, A.B., Haynie, A.C., Mueter, F.J., Bond, N.A., 2011. Evaluating management strategies for eastern Bering Sea walleye pollock (*Theragra chalcogramma*) in a changing environment. *ICES Journal of Marine Science* 68: 1297–1304.

Jurado-Molina, J. Livingston, P., 2002. Climate-forcing effects on trophically linked groundfish populations: implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences* 59: 1941–1951.

Kaplan, I.C., Brown, C.J., Fulton, E.A., Gray, I.A., Field, J.C. and A.D.M. Smith. 2013. Impacts of depleting forage species in the California Current. *Environmental Conservation* 40 (4): 380–393.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. *Journal of the American Statistical Association* 90: 773–795.

K-1 model developers, 2004. K-1 coupled model (MIROC) description, K-1 technical report, 1, H. Hasumi and S. Emori (eds.), Center for Climate System Research, University of Tokyo, 34pp.

Kinzey, D. 2010. Multispecies stock assessment with predator-prey interactions. Ph.D. thesis, University of Washington, 123 pp.

Kinzey, D., Punt, A.E., 2009. Multispecies and single-species models of fish population dynamics: comparing parameter estimates. *Natural Resource Modelling* 22: 167–204.

Kim, S.J., Flato, G.M., Boer, G.J., McFarlane, N.A., 2002. A coupled climate model simulation of the Last Glacial Maximum, Part 1: transient multi-decadal response. Climate Dynamics, 19, 515-537.

Kim, S.J., Flato, G.M., Boer, G.J., 2003. A coupled climate model simulation of the Last Glacial Maximum, Part 2: approach to equilibrium Climate Dynamics, 20, 635-661.

Kolody, D. 2011. An age-, sex- and spatially-structured stock assessment of the Indian Ocean swordfish fishery 1950-2009, including special emphasis on the south-west region. IOTC Document IOTC-2011-WPB-17. 75 pp.

Kolody, D., Campbell, R., Davies, N., 2008. A MULTIFAN-CL stock assessment of southwest Pacific swordfish 1952-2007. WCPFC document WCPFC-SC4-2008/SA-WP-6 (Rev1). 90 pp.

Legutke S., Voss, R., 1999. ECHO-G, the Hamburg atmosphere-ocean coupled circulation model. *DKRZ technical report* 18, DKRZ, Hamburg.

Magnusson, A., Punt, A.E., Hilborn, R., 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. *Fish and Fisheries* 14: 325–342.

Marasco, R.J., Goodman, D., Grimes, C.B., Lawson, P.W., Punt, A.E., Quinn T.J. II, 2007. Ecosystem-based fisheries management: some practical suggestions. *Canadian Journal of Fisheries and Aquatic Sciences* 64: 928–939.

Millar, C., Jardim, E., 2015. Model averaging to streamline the stock assessment process. ICES Journal of Marine Science, 72(1): 93–98.

Mueter, F.J. and Megrey, B.A., 2006. Using multispecies surplus production models to estimate ecosystem-level maximum sustainable yields. *Fisheries Research* 81: 189–201.

Mueter, F.J., Bond, N.A., Ianelli, J.N., Hollowed, A.B., 2011. Expected declines in recruitment of walleye pollock (*Theragra chalcogramma)* in the eastern Bering Sea under future climate change. *ICES Journal of Marine Science* 68: 1284–1296.

Myers, R.A., A.A. Rosenberg, P.M. Mace, N. Barrowman Restrepo, V.R., 1994. In search of thresholds for recruitment overfishing. ICES Journal of Marine Science. 51(2): 191-205.

North Pacific Fishery Management Council. 2012. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea / Aleutian Islands Regions. North Pacific Fishery Management Council, 605 West 4th Ave, Anchorage, AK 99501.

Overland, J.E., Wang, M., 2007. Future climate of the North Pacific Ocean. *EOS Transactions of the American Geophysical Union* 88: 182.

Pacific Fishery Management Council. 2011. Pacific Coast Groundfish Fishery Management Plan for the California, Oregon and Washington Grounfish Fishery. 7700 NE Ambassador Place NE, Portland, OR 97220

Parslow, J., Cressie, N., Campbell, E.P., Jones, E., Murray, L., 2013. Bayesian learning and predictability in a stochastic nonlinear dynamical model. *Ecological Applications* 23:679-698.

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.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., Punt, A.E., Restrepo, V, Skagen, D.W., Stefansson, G., 2001. Estimating uncertainty in fish stock assessment and forecasting. *Fish and Fisheries* 2:125-157.

Plagányi, É.E. 2007. Models for an ecosystem approach to fisheries. FAO Fish. Tech. Pap. 477. Rome, FAO. 108 pp.

Plagányi, É.E., Punt, A.E., Hillary, R., Morello, E.B., Thébaud, O., Hutton, T., Pillans, R.D., Thorson, J.T., Fulton, E.A., Smith, A.D.M., Smith, F., Bayliss, P., Haywood, M., Lyne, V. and P.C. Rothlisberg. 2014. Models of intermediate complexity for ecosystem assessment to support tactical management decisions in fisheries and conservation. *Fish and Fisheries* 15: 1-22.

Prager, M.H. and K.W. Shertzer. 2010. Deriving Acceptable Biological Catch from the Overfishing Limit: Implications for Assessment Models. *North American Journal of Fisheries Management* 30: 289–294.

Prager, M.H., Porch, C.E., Shertzer, K.W., Caddy, J.F., 2003. Targets and limits for management of fisheries: a simple probability-based approach. *North American Journal of Fisheries Management* 23: 349–361.

Punt, A.E., Donovan, G., 2007. Developing management procedures that are robust to uncertainty: Lessons from the International Whaling Commission. *ICES Journal of Marine Science* 64: 603–612.

Punt, A.E., Siddeek, M.S.M., Garber-Yonts, B., Dalton, M., Rugolo, L., Stram, D., Turnock, B.J., Zheng, J., 2012. Evaluating the impact of buffers to account for scientific uncertainty when setting TACs: Application to red king crab in Bristol Bay, Alaska. *ICES Journal of Marine Science* 69:624–634.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* 133:1155–1174.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J., Conser, R.J., 2011. An approach to quantifying scientific uncertainty in stock assessment. *Fishery Bulletin* 109: 217–231.

Reifen, C., Toumi, R., 2009. Climate projections: Past performance no guarantee of future skill? *Geophysical Research Letters* 36:L13704.

Ricker, W.E., 1954. Stock and Recruitment. *Journal of the Fisheries Research Board of Canada* 11: 559–623.

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

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of Royal Society B* 64:583–639.

Spies, I., Wilderbuer, T.K., Nichol, D.G., Aydin, A., 2012. Arrowtooth flounder. Pp 895-965 in Stock Assessment and Fishery Evaluation Report for the Groundfish Resources for the Bering Sea/Aleutian Islands Regions. North Pacific Fishery Management Council, 605 West 4th Ave, Suite 306, Anchorage AK 99501.

Schwartz, G., 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461–464.

Taylor, I.G., Stephens, A., 2013. Stock assessment of shortspine thornyhead in 2013. Pacific Fishery Management Council, 7700 NE Ambassador Place, Portland, OR 97220.

Thompson, G.G., Lauth, RR., 2012. Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands area. pp 245-254 in Stock Assessment and Fishery Evaluation Report for the Groundfish Resources for the Bering Sea/Aleutian Islands Regions. North Pacific Fishery Management Council, 605 West 4th Ave, Suite 306, Anchorage AK 99501.

Wang, M., Overland, J.E., Bond, N.A., 2010. Climate projections for selected large marine ecosystems. *Journal of Marine Systems* 79:258–266.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., Kawamiya, M., 2011. MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev. 4:845–872.

Wilberg, M.J., Bence, J.R., 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fisheries Research* 93:212–221.

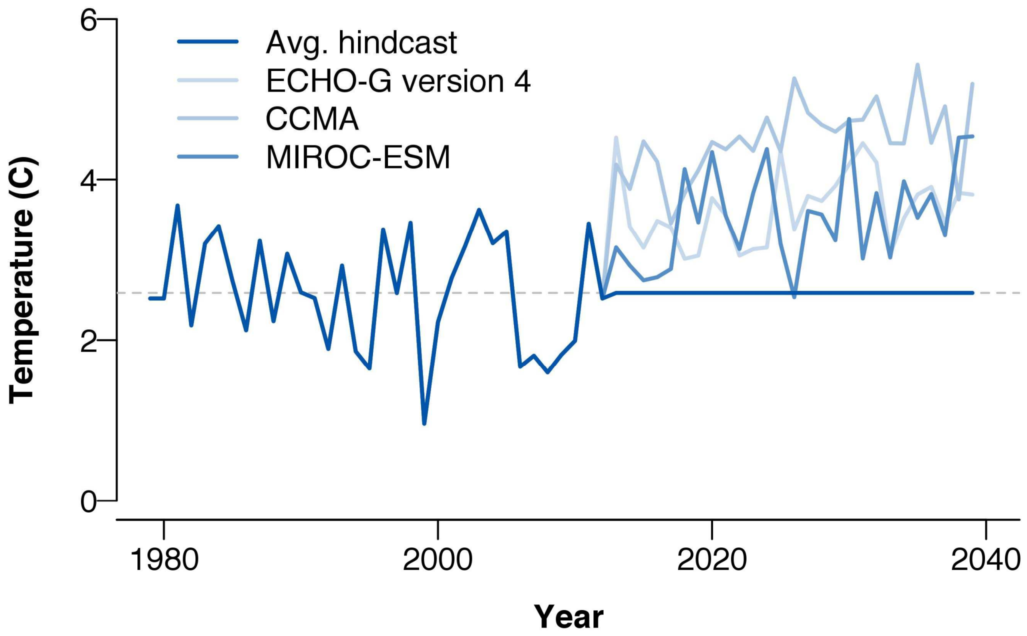

Figure 1. The four future temperature time-series on which the MSMt projections are based. The constant temperature is the average over time for the "hindcast" (dashed line).

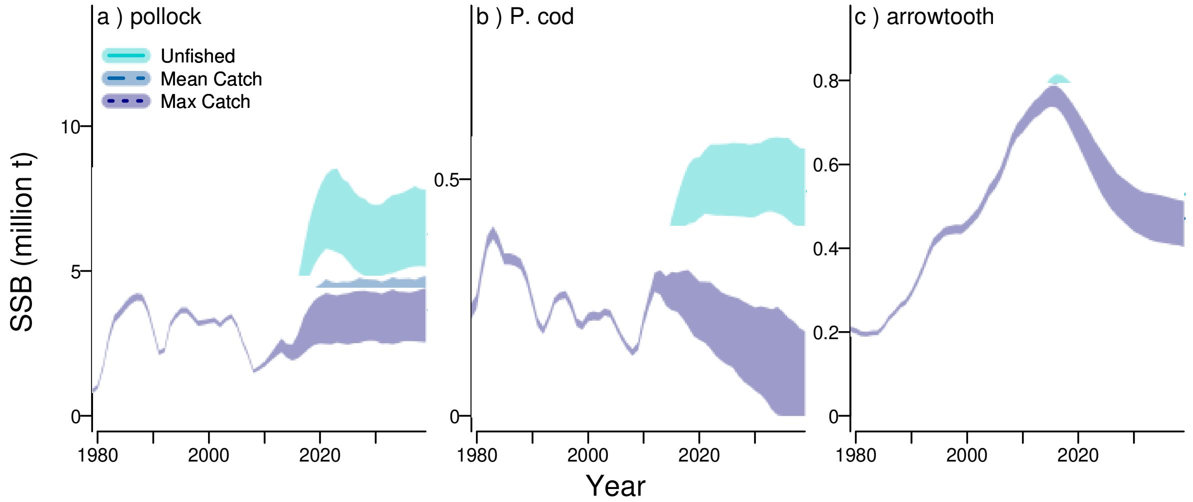

Figure 2. Time-trajectories of spawning stock biomass for walleye pollock, Pacific cod and arrowtooth flounder for three catch series when the projections are based on the AFSC single-species model (zero catch: solid line; dashed line: mean catch; dotted line: maximum catch). The bold lines are distribution medians, the light shaded areas (zero catch: blue-green; light-blue: mean catch; darker blue: maximum catch) contain 50% of the distributions.

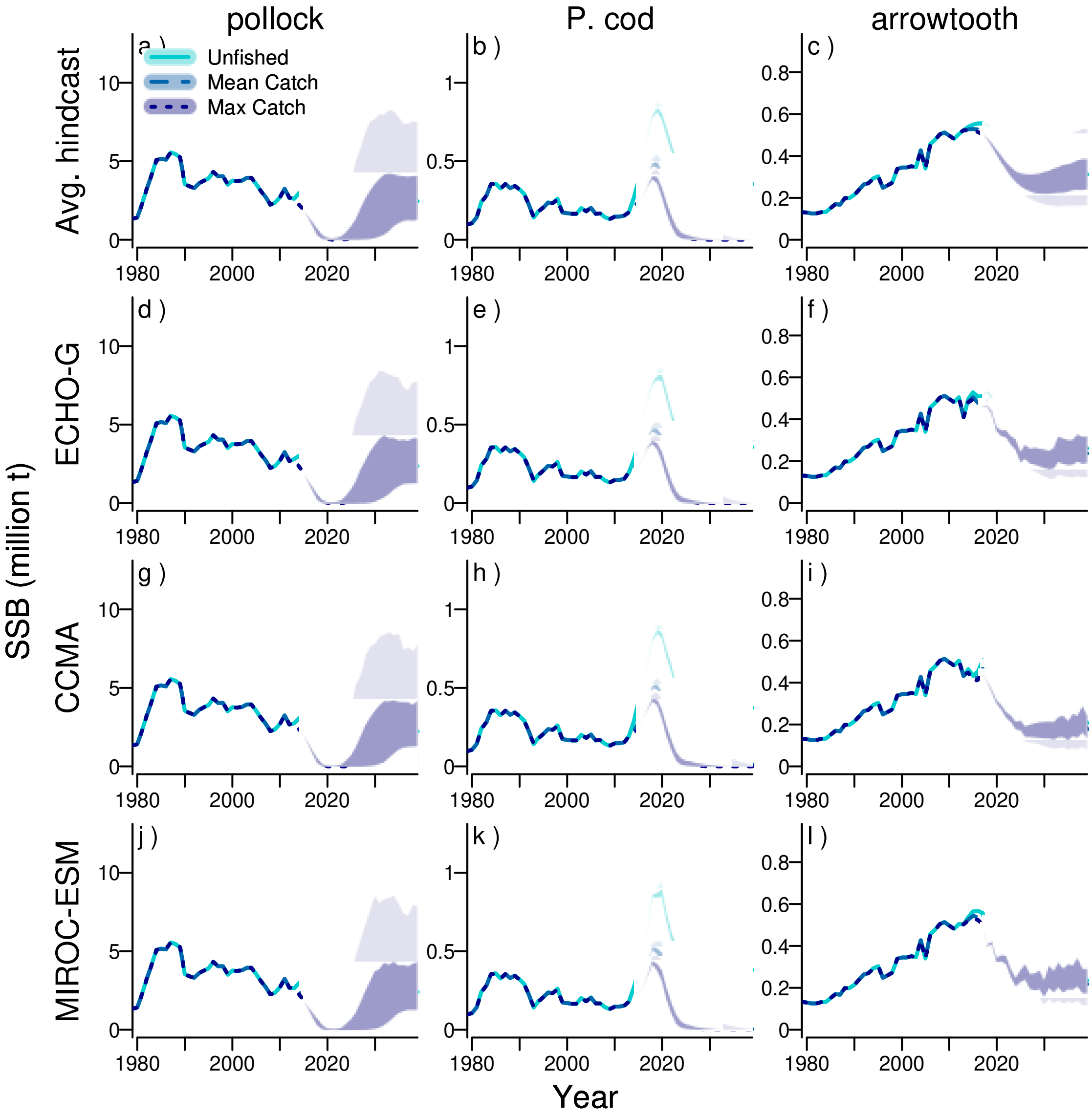

Figure 3. Time-trajectories of spawning stock biomass for walleye pollock, Pacific cod and arrowtooth flounder (columns) for three catch series when the projections are based on the MSMtA model. The results for each temperature scenario are shown as rows: average of hindcast values (a-c), ECHO-G (d-f), CCMA (g-i), and MIROC-ESM (j-l). The bold lines are distribution medians, the light shaded areas contain 50% of the distributions and the dark shaded areas contain 90% of the distributions.

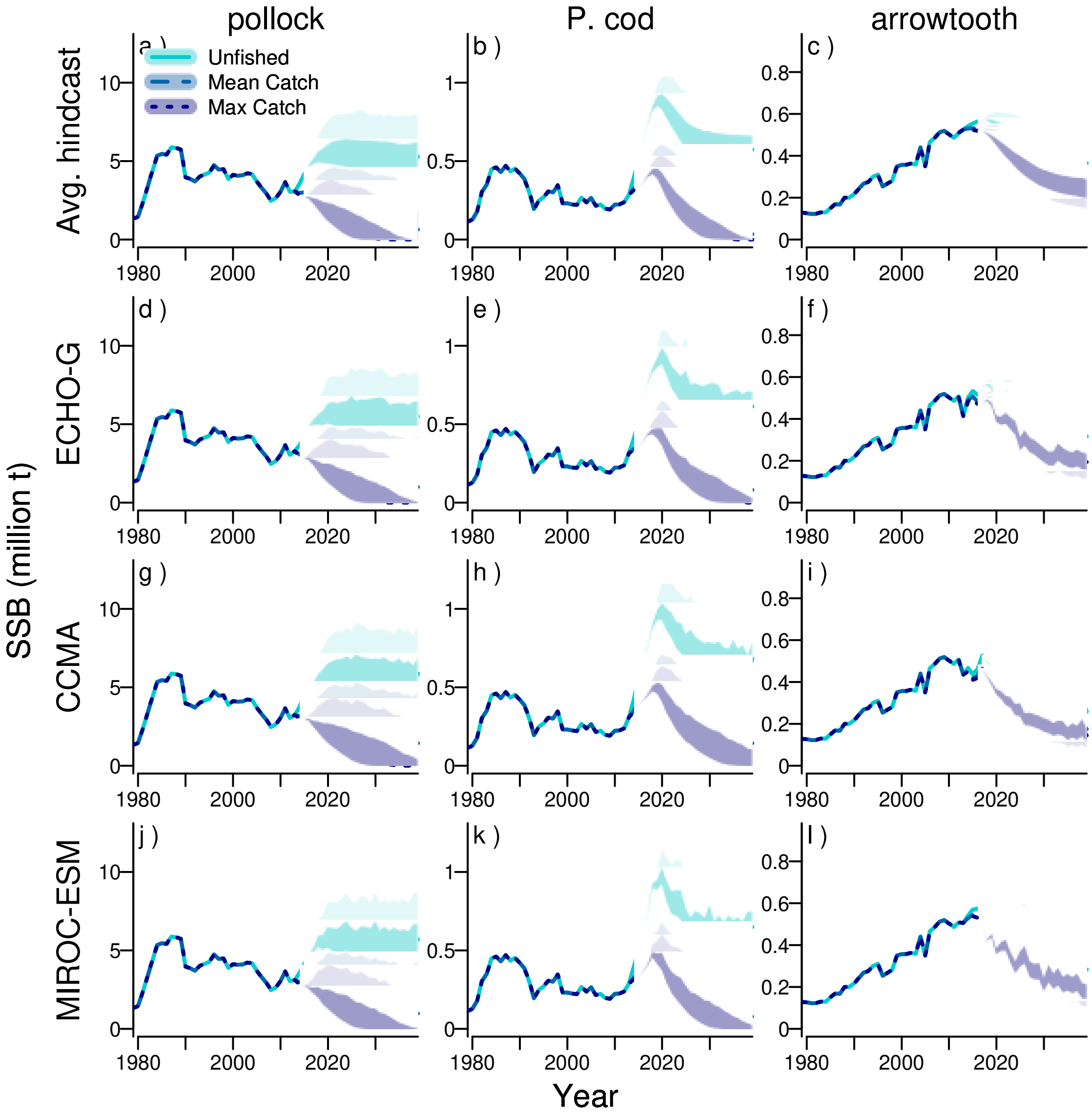

Figure 4. Time-trajectories of spawning stock biomass for walleye pollock, Pacific cod and arrowtooth flounder (columns) for three catch series when the projections are based on the MSMtB model. The results for each temperature scenario are shown as rows: average of hindcast values (a-c), ECHO-G (d-f), CCMA (g-i), and MIROC-ESM (j-l). The bold lines are distribution medians, the light shaded areas contain 50% of the distributions and the dark shaded areas contain 90% of the distributions.

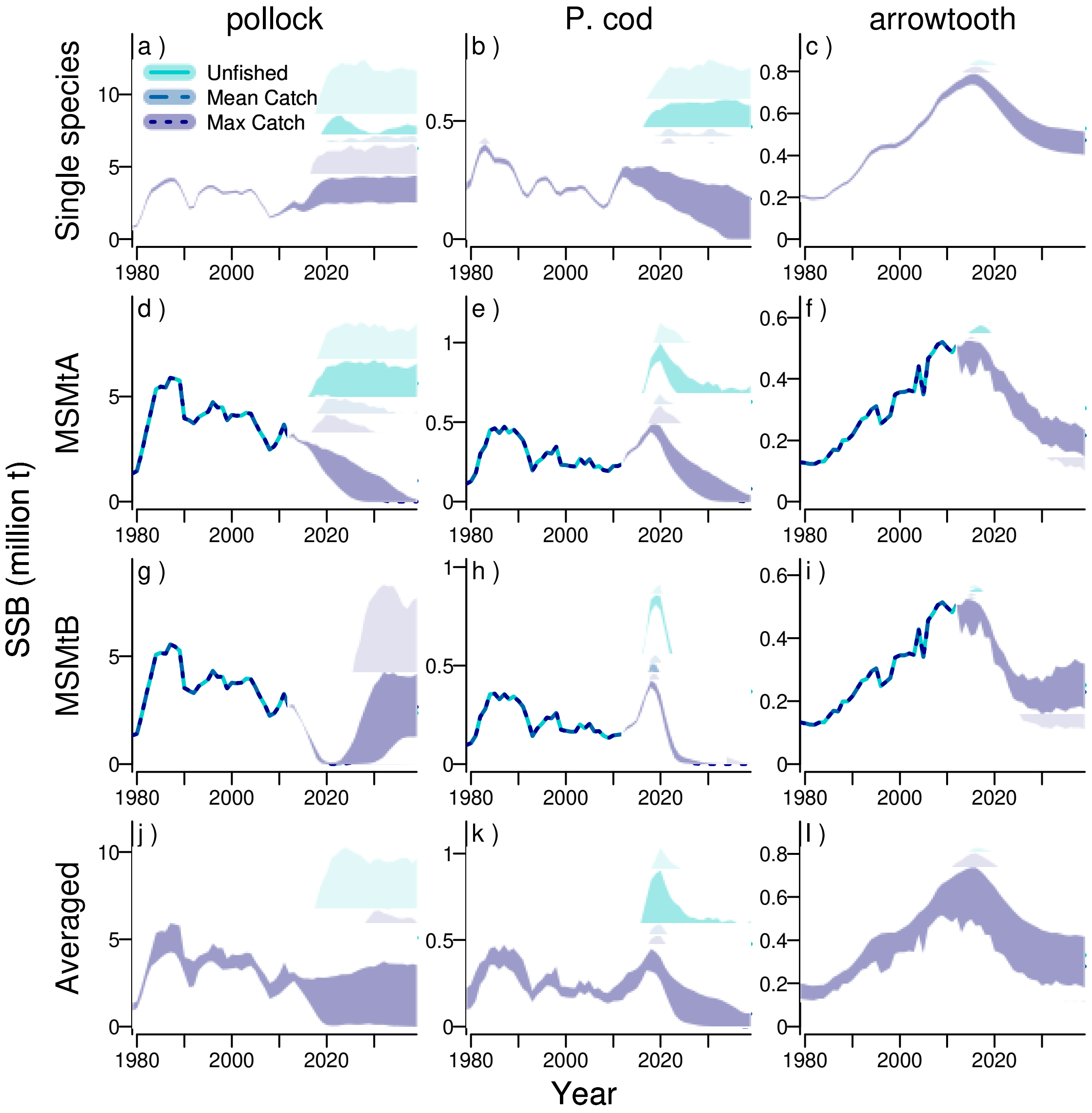

Figure 5. Model averaged results (over climate scenarios) for time-trajectories of spawning stock biomass for walleye pollock, Pacific cod and arrowtooth flounder for three catch series. The bold lines are distribution medians, the light shaded areas contain 50% of the distributions and the dark shaded areas contain 90% of the distributions.

Tables

Table 1. Spawning stock biomass (S_{MSY}: thousands of metric tons) and fishing mortality rate (F_{MSY}: per year) associated with MSY for Georges Bank Atlantic cod (*Gadus morhua*) based on five stock-recruitment models (Brodziak and Legault, 2005). Standard deviations are given in parentheses.

Model	**Posterior Probability**	S_{MSY}	F_{MSY}
RBH	0.34	193.7 (26.2)	0.21 (0.03)
RABH	0.15	176.1 (39.1)	0.23 (0.05)
RZBH	0.33	188.7 (33.6)	0.22 (0.02)
RZABH	0.16	172.7 (34.6)	0.23 (0.03)
SRK	0.01	87.5 (57.4)	0.69 (0.01)
Model Average		184.7 (38.2)	0.23 (0.06)
80% credibility intervals		(135.8, 233.6)	(0.15, 0.31)

RBH, informative recruitment priors with uncorrelated Beverton-Holt; RABH, informative recruitment priors with autocorrelated Beverton-Holt; RZBH, informative recruitment and steepness priors with uncorrelated Beverton-Holt; RZABH, informative recruitment and steepness priors with autocorrelated Beverton-Holt; SRK, informative slope at origin priors with uncorrelated Ricker (Ricker 1954, modified from Hill *et al.*, 2007).

Table 2 Catches (t) used in the projections.

Stock	Mean catch over 1998-2012	Maximum catch over 1998-2012
Pollock	1,226,280	1,490,900
Pacific cod	191,938	220,134
Arrowtooth flounder	13,458	17,737

Table 3 Percentiles of the distributions for the 2039 estimated spawning stock biomass for the individual models and for the model averaged results.

(a) a) Zero catch scenario

		Pollock					Pacific cod					Arrowtooth flounder				
Model	Climate Scenario	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%
MSMtB	Average	3,671	4,566	5,281	6,208	8,039	398	498	575	662	815	275	322	365	414	498
	ECHO-G	3,827	4,764	5,509	6,471	8,391	425	532	615	708	871	237	279	317	358	434
	CCMA	4,093	5,089	5,876	6,906	8,959	471	591	682	785	969	192	227	258	292	355
	MIROC-ESM	3,963	4,931	5,690	6,696	8,678	449	562	649	747	920	211	250	284	321	390
MSMtA	Average	1,525	2,011	2,461	3,103	4,683	258	310	354	405	497	217	265	313	368	470
	ECHO-G	1,426	1,903	2,359	3,017	4,649	261	313	357	410	510	181	221	261	308	395
	CCMA	1,359	1,784	2,226	2,859	4,342	275	329	376	432	531	142	174	206	242	309
	MIROC-ESM	1,500	1,975	2,412	3,056	4,566	279	334	381	438	537	163	199	237	278	352
Single species		4,042	5,144	6,269	7,806	11,616	322	398	474	564	725	412	477	529	588	691
MSMtB	Averaged	3,830	4,810	5,620	6,538	8,450	432	544	628	730	902	212	260	304	357	445
MSMtA	Averaged	1,458	1,894	2,378	3,022	4,616	265	324	369	421	518	161	206	251	308	411
All	Averaged	1,631	2,927	5,088	6,556	9,653	294	381	478	611	812	178	254	331	476	617

b) Mean catch scenario

Model	Climate Scenario	Pollock					Pacific cod					Arrowtooth flounder				
		Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%
MSMtB	Average	0	1	644	1,919	3,794	0	0	33	121	279	183	229	272	319	406
	ECHO-G	0	55	996	2,251	4,159	0	6	83	176	342	148	188	225	266	341
	CCMA	0	340	1,447	2,663	4,737	0	46	143	240	423	110	144	175	208	271
	MIROC-ESM	0	41	989	2,306	4,277	0	4	78	178	355	133	170	204	240	309
MSMtA	Average	145	1,609	2,773	4,160	7,149	0	0	2	36	139	174	239	299	376	507
	ECHO-G	104	1,453	2,642	4,091	7,216	0	0	5	51	161	132	187	237	306	422
	CCMA	44	1,312	2,532	3,929	7,130	0	0	14	69	181	91	134	175	229	322
	MIROC-ESM	105	1,545	2,797	4,257	7,404	0	0	5	53	170	121	172	219	278	381
Single species		2,031	2,844	3,651	4,820	6,973	0	61	170	261	437	357	421	471	531	632
MSMtB	Averaged	0	66	993	2,240	4,311	0	5	81	184	360	129	176	216	266	353
MSMtA	Averaged	66	1,428	2,669	4,129	7,216	0	0	5	52	167	116	172	229	306	446
All	Averaged	0	1,134	2,636	3,867	6,380	0	2	73	189	378	127	198	279	427	561

c) Maximum catch scenario

Model	Climate Scenario	Pollock					Pacific cod					Arrowtooth flounder				
		Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%	Low 5%	Low 25%	Med. 50%	Up 75%	Up 95%
MSMtB	Average	0	1	644	1919	3794	0	0	33	121	279	183	229	272	319	406
	ECHO-G	0	55	996	2251	4159	0	6	83	176	342	148	188	225	266	341
	CCMA	0	340	1447	2663	4737	0	46	143	240	423	110	144	175	208	271
	MIROC-ESM	0	41	989	2306	4277	0	4	78	178	355	133	170	204	240	309
MSMtA	Average	145	1609	2773	4160	7149	0	0	2	36	139	174	239	299	376	507
	ECHO-G	104	1453	2642	4091	7216	0	0	5	51	161	132	187	237	306	422
	CCMA	44	1312	2532	3929	7130	0	0	14	69	181	91	134	175	229	322
	MIROC-ESM	105	1545	2797	4257	7404	0	0	5	53	170	121	172	219	278	381
Single species		2031	2844	3651	4820	6973	0	61	170	261	437	357	421	471	531	632
MSMtB	Averaged	0	66	993	2240	4311	0	5	81	184	360	129	176	216	266	353
MSMtA	Averaged	66	1428	2669	4129	7216	0	0	5	52	167	116	172	229	306	446
All	Averaged	0	1134	2636	3867	6380	0	2	73	189	378	127	198	279	427	561