NRC Research Press

ARTICLE

Parameter uncertainty of a dynamic multispecies size spectrum model[1]

Michael A. Spence, Paul G. Blackwell, and Julia L. Blanchard

Abstract: Dynamic size spectrum models have been recognized as an effective way of describing how size-based interactions can give rise to the size structure of aquatic communities. They are intermediate-complexity ecological models that are solutions to partial differential equations driven by the size-dependent processes of predation, growth, mortality, and reproduction in a community of interacting species and sizes. To be useful for quantitative fisheries management these models need to be developed further in a formal statistical framework. Previous work has used time-averaged data to "calibrate" the model using optimization methods with the disadvantage of losing detailed time-series information. Using a published multispecies size spectrum model parameterized for the North Sea comprising 12 interacting fish species and a background resource, we fit the model to time-series data using a Bayesian framework for the first time. We capture the 1967–2010 period using annual estimates of fishing mortality rates as input to the model and time series of fisheries landings data to fit the model to output. We estimate 38 key parameters representing the carrying capacity of each species and background resource, as well as initial inputs of the dynamical system and errors on the model output. We then forecast the model forward to evaluate how uncertainty propagates through to population- and community-level indicators under alternative management strategies.

Résumé : Les modèles de spectres de tailles dynamiques sont reconnus comme offrant une approche efficace pour décrire comment les interactions basées sur la taille peuvent donner lieu à une structure de taille dans les communautés aquatiques. Ce sont des modèles écologiques de complexité moyenne représentant des solutions à des équations différentielles partielles déterminées par les processus dépendant de la taille que sont la prédation, la croissance, la mortalité et la reproduction dans une communauté d'espèces et de tailles interagissant entre elles. Pour être utiles dans la gestion quantitative des pêches, ces modèles doivent être approfondis dans un cadre statistique formel. Des travaux antérieurs ont fait appel à des valeurs moyennes de séries temporelles pour « étalonner » le modèle en utilisant des méthodes d'optimisation, ce qui a le désavantage d'occulter de l'information détaillée contenue dans les séries chronologiques. En utilisant un modèle publié de spectre de tailles multi-espèce paramétré pour la mer du Nord et comptant 12 espèces de poissons interagissant entre elles et les ressources de référence, nous avons calé le modèle sur des données de séries chronologiques en utilisant un cadre bayésien pour la première fois. Nous capturons la période de 1967 à 2010 en employant des estimations annuelles des taux de mortalité par pêche comme entrées au modèle et des séries chronologiques de données de débarquement de pêche pour caler le modèle sur les données de sortie. Nous estimons 38 paramètres clés représentant la capacité de charge de chaque espèce et les ressources de référence, ainsi que les entrées initiales du système dynamique et les erreurs sur les valeurs de sortie du modèle. Nous projetons ensuite le modèle dans le futur pour évaluer comment l'incertitude se propage par l'entremise d'indicateurs au niveau de la population et de la communauté pour différentes stratégies de gestion. [Traduit par la Rédaction]

1. Introduction

There are a number of ecological models that can be applied to answer marine management questions (Plagányi et al. 2014). An emerging class of marine ecosystem models is size spectrum models (Benoît and Rochet 2004; Law et al. 2009; Blanchard et al. 2009). Size spectrum models are models of intermediate complexity and are formulated around the McKendrick von Foerster partial differential equation. Conceptually, they are based on very simple ecological assumptions (Andersen and Pedersen 2009) about how the role of individual body size in a food web ("big individuals eat small individuals") gives rise to community abundance (and biomass) size spectra (Hartvig et al. 2011). Size-based predation leads to growth and mortality, which drive changes in the abundance of organisms along the size spectrum. Maturation is also size-dependent, and once an individual reaches maturation size, it produces offspring (Hartvig et al. 2011) that enter the model at the smaller sizes. Food for the smallest sized organisms is provided by a background community (representing phytoplankton, zooplankton, and benthos), which is modeled as an external size-structured resource that is not driven by predation but instead follows semi-chemostat logistic growth (Andersen and Pedersen 2009; De Roos et al. 2008).

Size spectrum models are increasingly being used to help us understand the structure of marine ecosystems and establish

Received 14 January 2015. Accepted 30 August 2015.

Paper handled by Associate Editor Brian Shuter.

M.A. Spence. School of Mathematics and Statistics, University of Sheffield, Sheffield, UK; Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK.
P.G. Blackwell. School of Mathematics and Statistics, University of Sheffield, Sheffield, UK.
J.L. Blanchard. Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point, TAS. 7004, Tasmania.
Corresponding author: Michael A. Spence (e-mail: m.a.spence@sheffield.ac.uk).

[1]This article is part of a special issue entitled "Size-based approaches in aquatic ecology and fisheries science: a symposium in honour of Rob Peters".

abundance baselines of marine communities and their responses to the potential effects of fishing and climate change (Benoît and Rochet 2004; Blanchard et al. 2009, 2012; Law et al. 2009; Jacobsen et al. 2013; Maury and Poggiale 2013; Woodworth-Jefcoats et al. 2013; Law et al. 2015). Several approaches exist spanning a wide range of model complexity: simple community models, trait-based models, and more detailed multispecies models (Scott et al. 2014). The generic community- and trait-based models have been used to develop theory (Benoît and Rochet 2004; Andersen and Pedersen 2009; Hartvig et al. 2011), to examine the community responses to fishing mortality and selectivity, and as a test-bed for evaluating indicators of the ecosystem effects of fishing (Rochet and Benoît 2011; Zhang et al. 2014; Law et al. 2015; Jacobsen et al. 2013).

Both size and species identity are important for fisheries management, and the development of methods to parameterize trait-based models for real multispecies fish communities has been a recent focus of research, particularly for testing indicators and management strategies at both population and community levels. Blanchard et al (2014) parameterized and calibrated a trait-based model for 12 species in the North Sea using fisheries survey data and stock assessment data to determine whether meeting management targets for exploited North Sea populations would be sufficient to meet proposed Marine Strategy Framework Directive targets for biodiversity and food web functioning (including the "large fish indicator").

Although trait-based models can be parameterized for real systems based on either the literature or statistical analyses of fisheries datasets, there are inevitably parameters that are uncertain and have to be estimated by fitting the model to data. For multispecies models to be useful for tactical management, they need to be developed and tested in a formal statistical framework (Plagányi et al. 2014). Uncertain parameters for the Blanchard et al. (2014) multispecies size spectrum model included R_{max}, the maximum recruitment for each species, and κ, the background food resource's carrying capacity. To estimate these parameters, the model was "calibrated" to time-averaged spawning stock biomass (SSB) and landings data using time-averaged fishing mortality from 1985 to 1995 by minimizing the sum of squared errors between the model and the data to find a single best parameter set. The model was cross-validated with survey data and then forced with time-varying fishing mortalities and scenarios to evaluate whether single-species F_{MSY} management targets (the fishing mortality that leads to the maximum sustainable yield) would lead to recovery in food webs and biodiversity in the North Sea. Stochasticity was incorporated in the recruitment stage. Although the model produced realistic growth rates and species size distributions, some of the time series fits to SSB and landings were poor. This is partially due to the fact that time-series data were not fully used to calibrate the model. An ideal calibration approach would enable time-series data to be more fully utilized, combined with a formal statistical framework for uncertainty.

It is important to report uncertainty associated with model-derived research findings when used for advising policy makers and environmental managers (Harwood and Stokes 2003). Uncertainty can be separated into four main types: parameter uncertainty, structural or model uncertainty, residual variation, and data uncertainty. Parameter uncertainty comes from uncertain knowledge about parameters (Li and Wu 2006); structural uncertainty is uncertainty associated with the model itself caused by simplifications, uncertain processes, or even numerical approximations; residual variation is the uncertainty caused by demographic and environmental stochasticity (Kennedy and O'Hagan 2001; Vernon et al. 2010), and data uncertainty is often referred to as measurement or observation error. This can be transferred to the parameters but can be propagated through to the model output and can be caused by sampling biases or errors in data collection (Harwood and Stokes 2003).

Table 1. The species used in the model and their data sets used to estimate the parameters, as well as the fishing mortality in 2010 (F_{2010}) and at the maximum sustainable yield (F_{MSY}), as shown in Blanchard et al (2014).

i	Species	Name	Landings	F_{2010}	F_{MSY}
1	*Sprattus sprattus*	Sprat	1967–2010	0.31	0.2
2	*Ammodytes marinus*	Sandeel	1983–2010	0.36	0.2
3	*Trisopterus esmarkii*	Norway pout	1983–2010	0.42	0.2
4	*Limanda limanda*	Dab	1967–2010	0.14	0.2
5	*Clupea harengus*	Herring	1967–2010	0.12	0.25
6	*Eutrigla gurnardus*	Gurnard	1967–2010	0.10	0.2
7	*Solea solea*	Sole	1967–2010	0.34	0.22
8	*Merlangius merlangus*	Whiting	1990–2010	0.27	0.2
9	*Pleuronectes platessa*	Plaice	1967–2010	0.24	0.25
10	*Melanogrammus aeglefinus*	Haddock	1967–2010	0.23	0.3
11	*Pollachius virens*	Saithe	1967–2010	0.38	0.19
12	*Gadus morhua*	Cod	1967–2010	0.68	0.3

Parameter uncertainty has not been formally explored in dynamic size spectrum models, although some work has been done with a length-based multispecies model (Thorpe et al. 2015) and an age-based model (Tsehaye et al. 2014). To improve the utility of multispecies size spectrum models for supporting fisheries management, the parameter, model, and data uncertainty need to be quantified. Here we further investigate the model of Blanchard et al. (2014) (for the model description, see the Supplementary material[2]) using a Bayesian framework, a more realistic error model, and an improved estimation strategy to assess uncertainty from parameters and the data and demonstrate how this uncertainty can be included in evaluating multispecies effects of fisheries management scenarios.

2. Methods

In this section, we describe the model parameters, their prior distributions, and how the model outputs can be related to the observed data in a probabilistic way. We then describe the steps used to sample from the posterior distributions using a Markov chain Monte Carlo (MCMC) algorithm (Gelman et al. 2013) (see the Supplementary material[2] for details).

Uncertain parameters

In the multispecies model, there are a number of uncertain parameters to estimate. For the inputs $R_{max.i}$, where i represents the species as described in Table 1, we specify priors in terms of $\psi_i = \log R_{max.i}$ for $i = 1, \ldots, 12$ and $\psi_0 = \log\kappa$ taking $\psi_i \sim U(\cdot|\alpha_i, \beta_i)$, where $\alpha_i < \beta_i$. So the prior densities for $R_{max.i}$ and κ are $p(R_{max.i}|\alpha_i, \beta_i)$ and $p(\kappa|\alpha_0, \beta_0)$, respectively, where

$$p(x|\alpha, \beta) = \begin{cases} \dfrac{\exp(x)}{\beta - \alpha} & \text{if } \exp(\alpha) \le x \le \exp(\beta) \\ 0 & \text{otherwise} \end{cases}$$

For the present analysis, we represent identical priors for each species by $\alpha_i = 0$ and $\beta_i = 50$ for $i = 0, \ldots, 12$, which means that they are not very constraining.

The dynamic model requires a "spin-up" period in which the fishing mortality, F_i, is fixed, so that the model reaches a steady state before the fishing mortality is varied and output is collected in 1967, the first year of the empirical time series. It is not obvious what the fishing mortality should be while the model is in the spin-up period so we have added the spin-up fishing mortality as an additional parameter to estimate for each of the 12 species,

[2]Supplementary material is available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2015-0022.

Table 2. The uncertain parameters.

Parameters	Also	Units	Prior	Notes
$\psi_{1:12}$	$\log R_{max}$	$\log(m^{-3}{\cdot}g^{-1}{\cdot}year^{-1})$	$U(\cdot\|0, 50)$	Log of the maximum recruitment for each species
ψ_0	$\log \kappa$	$\log(g^{\lambda-1}{\cdot}vol^{-1})$	$U(\cdot\|0, 50)$	Log of carrying capacity of resource spectrum
$F_{1:12}$		$year^{-1}$	Half-normal$(\cdot\|0, (1.824)^2)$	Fishing mortality during the spin-up period for each species
ρ		$year^{-1}$	$\exp(\cdot\|1/0.34)$	Fishing mortality for Norway pout in 2005
$\sigma^2_{1:12}$		Unitless	Inv-Gamma$(\cdot\|0.0001, 0.0001)$	Standard deviation of the error on the log landing

Note: $\psi_{0:12}$, $F_{1:12}$, and ρ are needed to run the model, and σ^2 is the error between the model output and the observed landings.

$[F_i]_{i=1}^{12}$. The spin-up period is used to run the model into the best-fitting stationary states before the fishing mortality is varied. It does not make sense for F_i to be negative so we decided on

$$F_i \sim \text{Half-normal}(\cdot|0, (1.824)^2)$$

for $i = 1, \ldots, 12$.

We used the same fishing mortalities as Blanchard et al. (2014) based on stock assessments (www.ices.dk) for the 12 species from 1967 to 2010. According to these inputs, the fishing mortality for Norway pout in 2005 was 0. This is inconsistent with the fact that there were landings in that year. To estimate this, we have added the fishing mortality of Norway pout in 2005 as another parameter, ρ. We assumed that the zero value was likely due to a rounding error for Norway pout so we used an informative prior on ρ such that

$$\rho \sim \exp\left(\cdot\left|\frac{1}{0.34}\right.\right)$$

We elicited (see e.g., O'Hagan et al. 2006) these values using expert knowledge from JLB by examining the 50th percentiles of the distributions and then confirming the priors graphically. Table 2 summarizes the uncertain parameters and their prior distributions.

Likelihood

The model was fit to landings data, $\boldsymbol{Y}$ (in tonnes), from stock assessments (www.ices.dk) for the years shown in Table 1 using a Bayesian framework. For an introduction to Bayesian statistics, see McCarthy (2007); for a more detailed review of the area, see Gelman et al. (2013). If the modeled landings, assumed to be the same as the catches (i.e., discards are ignored), were expressed as $M(\boldsymbol{\theta})$, where the unknown parameters are defined as $\boldsymbol{\theta}$ and the other inputs are implicit in $M(\cdot)$, then we assumed

$$\log \boldsymbol{Y} = \log M(\boldsymbol{\theta}) + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon}$$

where Σ's off-diagonal elements are 0 and diagonal elements are σ_i^2 ($i = 1, \ldots, 12$) and $\boldsymbol{\epsilon}$ is a vector of standard normals (Nielsen and Berg 2014; Tsehaye et al. 2014).

All of the variance parameters, σ_i^2, had independent inverse-gamma prior distributions defined as

$$\sigma_i^2 \sim \text{Inv-Gamma}(\cdot|0.0001, 0.0001)$$

for $i = 1, \ldots, 12$.

The simulation model is a solution of partial differential equations (PDEs) that is intractable and is approximated by discretizing both time and size (Hartvig et al. 2011). The year is divided into intervals of length (δt), and the PDEs are estimated at these points. Initially we experimented with $\delta t = 1$, the same value used by Blanchard et al (2014), i.e., the PDEs were estimated every year, and we found that the likelihood surface was very unstable and that often made a large difference to the model output. As δt decreases, the numerical estimation becomes more accurate. Changing δt, we found that the estimate stabilized at around $\delta t = 1/4$. However, as we decreased δt, the model took longer to run so we had the classic problem of efficiency versus accuracy.

Exploration of the parameter space

The model output, $M(\boldsymbol{\theta})$, from the 26-dimensional input space is not smooth, even with a low value of δt. It contains many local minima that an MCMC chain would get stuck in, and the quality of the fit, as measured by likelihood or posterior density, varies by many orders of magnitude. Thus a standard MCMC algorithm would be unable to fully explore the parameter space in any reasonable time. To overcome this, our strategy involved first carrying out an extensive search of the space, followed by local optimization, and then a parallel tempering algorithm (Swendsen and Wang 1986).

Our initial search could have been carried out by selecting completely random points in the parameter space. However, in view of the computational costs, we instead used a more efficient design for the selection of the points, Latin hypercube sampling (LHS) (McKay et al. 1979). For efficiency, we also carried out these exploratory runs with $\delta t = 1/2$.

In the first round, we used LHS to sample 50 000 parameter sets and evaluated the model at each of these, setting all of the σ^2s to 1, which is effectively using the sum of squared errors as a measure of how good a parameter set was.

We then performed a second round of LHS around each of the 10 best points found in round 1. For each top-ten point ($\theta_1, \ldots, \theta_{26}$), we applied LHS on the Cartesian product, $j = 1, \ldots, 26$, of the parameter intervals

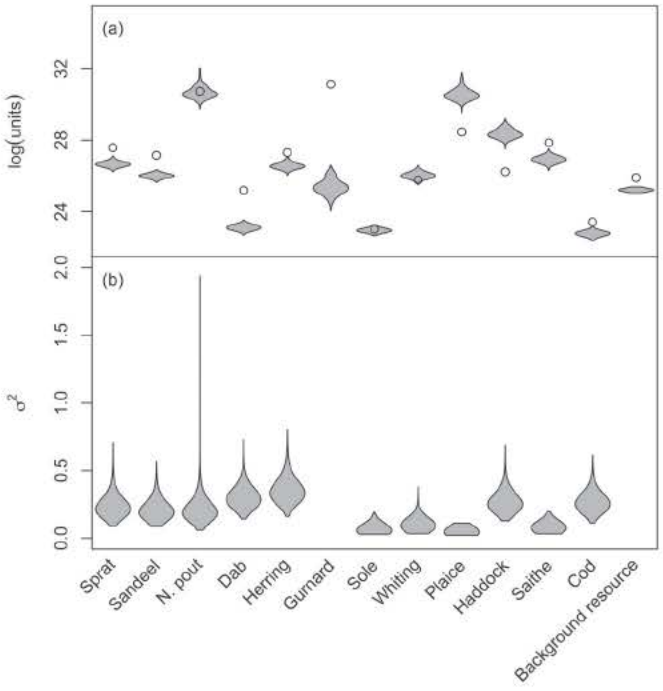
$$\left(P_j^{-1}(\max\{P_j(\theta_j) - \epsilon, 0\}), P_j^{-1}(\min\{P_j(\theta_j) + \epsilon, 1\})\right)$$

where $P_j(\cdot)$ is the prior cumulative distribution function of parameter j, and we took $\epsilon = 0.025$.

From the best 49 points from the second round, plus the point representing the parameters that Blanchard et al (2014) found, we then optimized using a Nelder–Mead algorithm (Nelder and Mead 1965) to find 50 high local maxima, capping the number of model runs to keep the computational effort down. This gave us 50 candidate points, fitting the data much better than randomly selected starting points, and we applied the Metropolis-within-Gibbs algorithm described in the Supplementary material,[2] running 50 chains starting from these local maxima, to explore their neighborhoods in the parameter space, using $\delta t = 1/4$ for accuracy and allowing $\sigma^2_{1:12}$ to vary.

We took the best five points and performed parallel tempering starting from these points (see the Supplementary material[2] for details). From the parallel tempering, we found that two of these Metropolis-within-Gibbs runs identified a region fitting so much better than any others that effectively all of the posterior probability was associated with these two runs. The quality of the fit, and the posterior probability, associated with each of the other regions of the parameter space was so low in comparison that they had essentially no effect on the parameter estimates or un-

Fig. 1. (*a*) The marginal posterior distribution for $\psi_{0:12}$ with the estimates from Blanchard et al. (2014) being the points. The units are $m^{-3}{\cdot}g^{-1}{\cdot}year^{-1}$ for $\psi_{1:12}$ and $g^{\lambda-1}{\cdot}vol^{-1}$ for ψ_0. (*b*) The estimates of the error parameters for all of the parameters except Gurnard, which is very uncertain and has a mean of 3.05 and variance of 0.75. The order of the species is that of their asymptotic size.



certainties. The fit is also, of course, very much better than would be found by a naïve random search; some further detail is given in the Discussion.

To explore the consequences of alternative management strategies, we sampled 2500 parameter sets from the posterior distribution, and for each set, we ran the model until 2010 and then projected the model to 2050 under two contrasting scenarios: (1) a status-quo scenario in which each species fishing mortality is held at 2010 levels, F_{2010}, and (2) a single-species F_{MSY} scenario suggested by ICES using the values shown in Table 1. To evaluate the uncertainty associated with population, we estimated

$$\frac{B_{F_{\text{scenario}}}}{B_{F_0}}$$

where B is the total spawner biomass with the fishing mortality set to either F_{MSY} or F_{2010} divided by the SSB at the baseline, F_0, where the fishing mortality is 0 for the whole of the simulation (including the spin-up period). We also estimated the large fish indicator (LFI), the proportion of biomass of demersal fish that are >40 cm in length, for each of the three fishing scenarios and the slope of the community size spectrum for demersal fish as described in the Supplementary material.[2]

3. Results

The results in this section are based on running the final MCMC chain from the previous section for 60 000 iterations and discarding the first 10 000 as burn-in.

Posterior distributions

We found that the marginal posteriors for the recruitment parameters are unimodal; summaries are shown in Fig. 1 using violin plots (Hintze and Nelson 1998) and in Table 3.

Table 3. The means and standard errors (SE) of the marginal posterior distributions of $\psi_{0:12}$ and $\sigma^2_{1:12}$ rounded to three decimal places.

	ψ		σ^2	
Species	Mean	SE	Mean	SE
Sprat	26.659	0.124	0.236	0.070
Sandeel	26.008	0.091	0.208	0.062
Norway pout	30.684	0.326	0.212	0.085
Dab	23.108	0.126	0.304	0.071
Herring	26.556	0.145	0.355	0.084
Gurnard	25.381	0.422	3.049	0.861
Sole	22.948	0.087	0.059	0.14
Whiting	26.034	0.158	0.099	0.035
Plaice	30.562	0.307	0.046	0.011
Haddock	28.375	0.252	0.269	0.067
Saithe	26.920	0.177	0.078	0.019
Cod	22.767	0.125	0.268	0.065
Background resource	25.210	0.056		

Many of the posterior distributions of the fishing mortality parameters, $F_{1:12}$, were not too dissimilar to their respective prior distributions, others were more concentrated (see the Supplementary material[2]).

The variance parameters describe the estimated distribution of the error around the observed landings. These were close to zero (Fig. 1), suggesting that the modeled landings captured the observed landings reasonably well on average. This was particularly the case for sole, whiting, plaice, and saithe. The model was particularly poor at estimating gurnard landings; the error parameter for gurnard is omitted from Fig. 1 because it is too big to plot on the same scale.

The posterior mean fishing effort for Norway pout in 2005 was about 0.019, confirming our suspicion that there may have been a rounding error in either the landings or fishing mortality for that species.

Time-series model output

A comparison of the observed time series of the landings to the model output (Fig. 2) showed that the model does a reasonable job of fitting the dynamics of the data. We more formally assessed how well the model fit the dynamics of the landings by calculating the values of σ_i^2 relative to the variabilities of their respective landings. Figure 3 shows the posterior distribution of σ_i^2/ι_i, with ι_i being the unbiased estimate of the variance of the landings,

$$\frac{1}{n-1}\sum_{t=1967}^{2010}\left(Y_i^{(t)}-\overline{Y_i}\right)^2$$

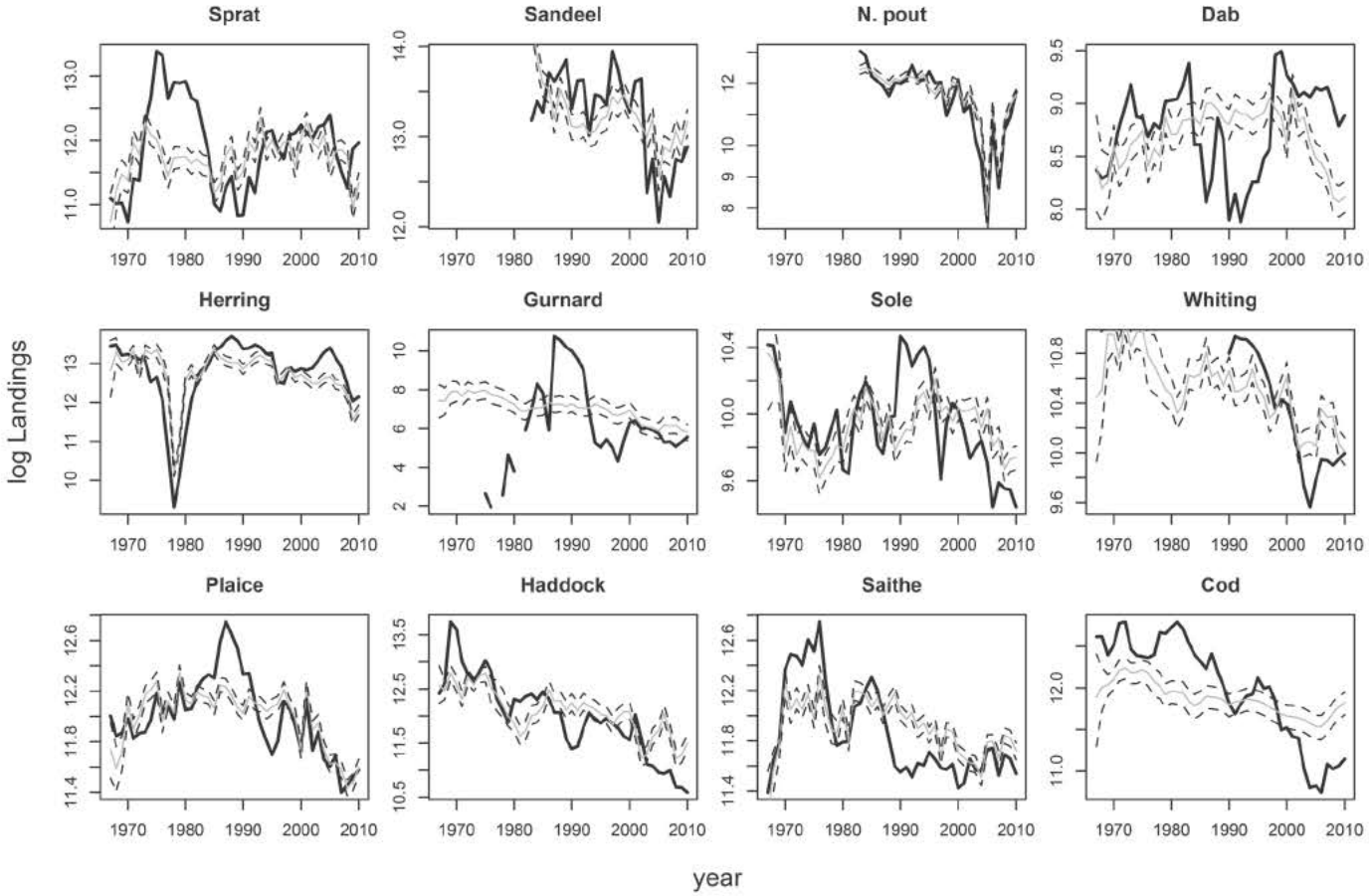
where $\overline{Y_i}$ is the mean landings for species i. We found lowest values of relative variance (meaning best fit) for sprat, Norway pout, and plaice. Higher values of relative variances were for gurnard and dab, implying poorer fits. Figure 4 shows the model output for SSB for nine of the species and compared it with single-species stock assessments (www.ices.dk). This comparison is not intended to evaluate goodness of fit but rather to examine differences between our model predictions with the single-species model outputs. We found lower SSB for most of the species except for sandeel, Norway pout, and herring compared with single-species assessments. The temporal trends in SSB were broadly similar.

Scenarios

We simulated the model forward to 2050 under the two scenarios described in the Methods, but the model was almost in a steady state by 2020. The results of these forecasts are shown in Fig. 5.

Under the status-quo scenario, sprat, sandeel, and cod were the most depleted, with the spawner biomass of sandeel and sprat

Fig. 2. Runs of the model with parameters sampled from the posterior distribution. The grey line shows the median model output, the dashed lines are the 5th and 95th percentiles for the model output, and the thick black line is the observed landings.



ranging from 0.368 to 0.394 and from 0.307 to 0.317 of their respective unexploited spawner biomasses. Cod is the most depleted, ranging from 0.133 to 0.094, with a 0.02 probability of being less than 0.1 of its unexploited biomass, which has been used as a threshold for collapse. Several species have a high chance of being higher than unexploited biomass due to the much lowered biomass of cod resulting in prey release. Under the single-species F_{MSY} scenario, these species have higher probability of being closer to their unexploited values. Plaice, saithe, and cod were the most depleted, ranging from 0.324 to 0.496, 0.438 to 0.476, and 0.486 to 0.514 of their respective unexploited values.

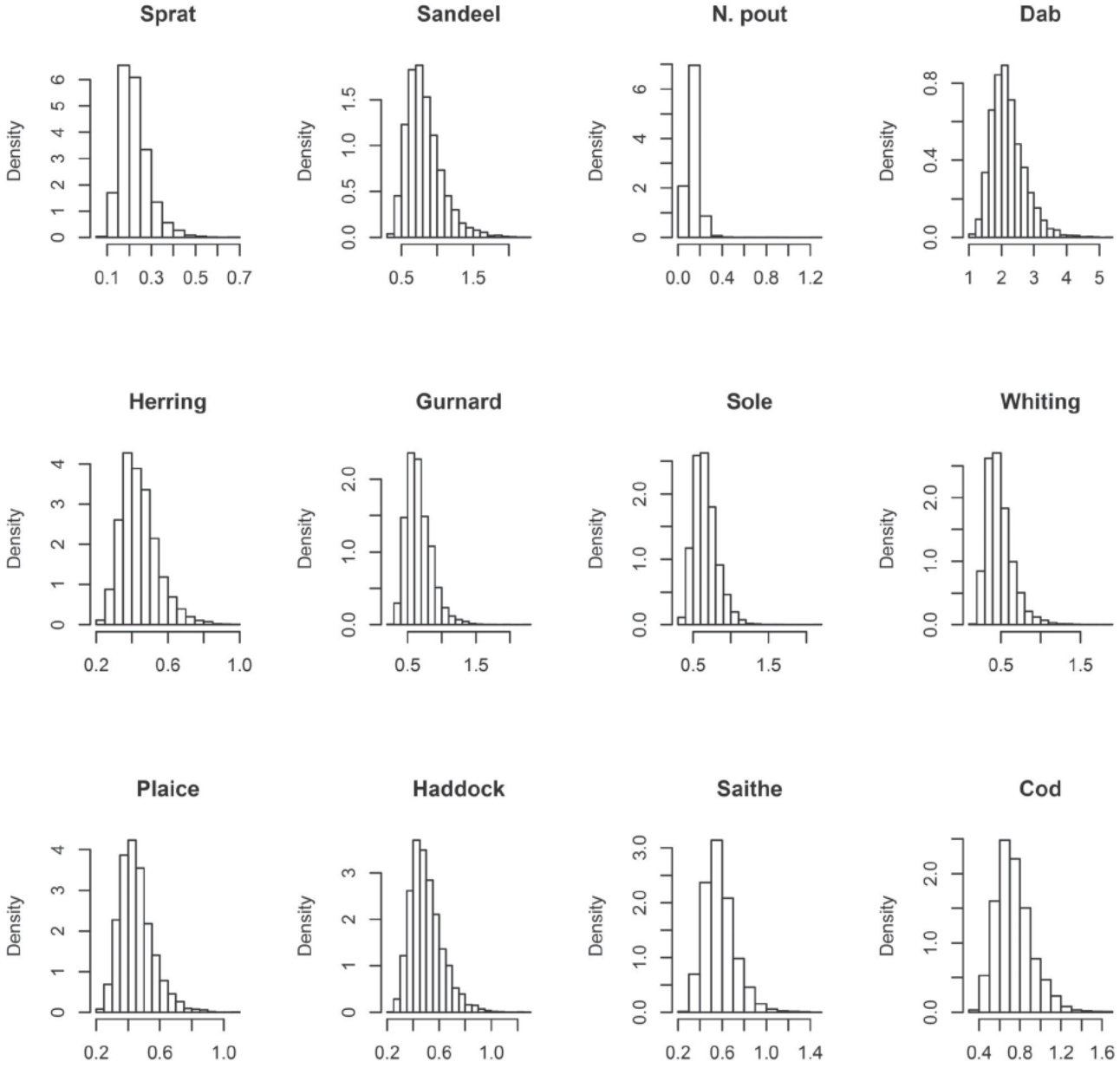
The uncertainty is higher for some species such as haddock under the status-quo (standard error is 0.155) and plaice under F_{MSY} (standard error is 0.028) than others such as sandeel under the status-quo (standard error is 0.002) and cod under F_{MSY} (standard error is 0.004). Consistent with the findings of Blanchard et al (2014), the LFI did not differ under the two fishing scenarios (the median is 0.385 and 0.380 under the status-quo and F_{MSY}, respectively), whereas the F_{MSY} scenario gave a much shallower size spectrum slope (the median is about −2.12) than the status-quo (the median is about −2.35) for all parameter sets.

4. Discussion

An ecosystem approach to fisheries management requires tools that can evaluate the risks of fisheries management actions on both target and nontarget species. Although extensive work on model uncertainty has been carried out through simulation approaches such as management strategy evaluation, a wide range of ecosystem and multispecies models being used to support ecosystem advice rely on projections from single best-fitting parameter sets, ignoring parameter uncertainty, and are considered to be strategic or "big picture" rather than of tactical use to support management decision (Plagányi et al. 2014). Robust estimates of uncertainty in model parameters are also important for reporting results of management scenarios to policy makers (Harwood and Stokes 2003). Few attempts have been made to explicitly address parameter uncertainty in more complex models (Thorpe et al. 2015), and this study is the first to develop such a framework for multispecies size spectrum models. Multispecies size spectrum models are still in their infancy in fisheries and fall into the strategic category. Our methods demonstrate how this class of models can be developed further using a Bayesian framework. The key advantage, as illustrated here through two simple fisheries scenarios, is that it is possible to make probabilistic statements of scenario outcomes that enable more informed assessments of risk.

Fisheries landings data are often assumed to not contain error but in reality contain high uncertainty due to misreporting and discarding. Here, we treated the landings data as uncertain, assuming the model and data uncertainty result in independent Gaussian errors on the log scale. In addition to quantifying the uncertainty around the modelled landings, we also estimated variance parameters of the Gaussian errors for each species in the model. These parameters take into account the data uncertainty and the residual variability and can be interpreted as how well, on average, the model does at recreating the observations. A small value of σ_i^2 means that, on average, the model recreates the landings of species i well. If all of these parameters are the same, then the likelihood of the observations is related in a simple way to the sum-of-squares metric used by Blanchard et al. (2014). If the vari-

Fig. 3. The posterior distribution for σ_i^2/ι_i, where ι_i is an unbiased estimate of the variance of the observed log landings for species i.



ance parameters are not equal, the appropriate metric becomes the weighted sum of squares, with a lower value of σ_i^2 implying that observations on species i should be more highly weighted. We found that these variances clearly do differ across species and may be a possible reason why the points found by Blanchard et al (2014) are not in the posterior distribution (see Fig. 1). Other reasons may be that Blanchard et al. fitted their model with time-averaged SSB and landings data, whereas we account for temporal dynamics over the 1967–2011 period for landings only; we also fitted the model using $\delta t = 1/4$ instead of $\delta t = 1$. Blanchard et al.'s choice of $\delta t = 1$ succeeded in fitting the equilibrium behaviour of the model, which is largely unaffected by δt, to the time-averaged data; their fitted parameter values were shown to capture time-averaged size distributions and growth rates from survey data well. However, for fitting to time-varying data or using time-varying (fishing mortality) inputs to predict time series, it is necessary to describe the dynamics of the model in more detail and hence to use a smaller value of δt. This does affect the likelihood surface and potentially the parameter estimates. Our experiments with fitting to SSB as well as landings (Spence 2015) did not give parameters close to Blanchard et al.'s (despite starting one of the MCMC chains there). Blanchard et al. also used a penalty function for species that went extinct when fishing mortality was zero; however, because all of the species coexist with both fitted parameter sets, this is unlikely to have a substantial effect on the precise parameter estimates. Overall, it seems that the differences in species weighting, (i.e., value of σ^2) and the choice of δt are most likely to account for the parameter differences.

In spite of these differences, the consequences of the two fisheries scenarios explored resulted in similar qualitative outcomes. Under the status-quo scenario, both models showed agreement in cod being the most heavily depleted with the same smaller bodied species (herring, whiting, plaice, and haddock) having biomasses higher than their unexploited biomass. The latter is a feature that would not emerge from single-species models that ignore food web interactions. Under the F_{MSY} scenario, both models also show that cod biomass and the community size spectrum slope return closer to their unexploited levels, whereas the large fish indicator

Fig. 4. Log spawning stock biomass (SSB) of the model with parameters sampled from the posterior distribution. The grey line shows the median output, the dashed lines are the 5th and 95th percentiles for the model output, and the think black line is the log SSB estimates from stock assessments.

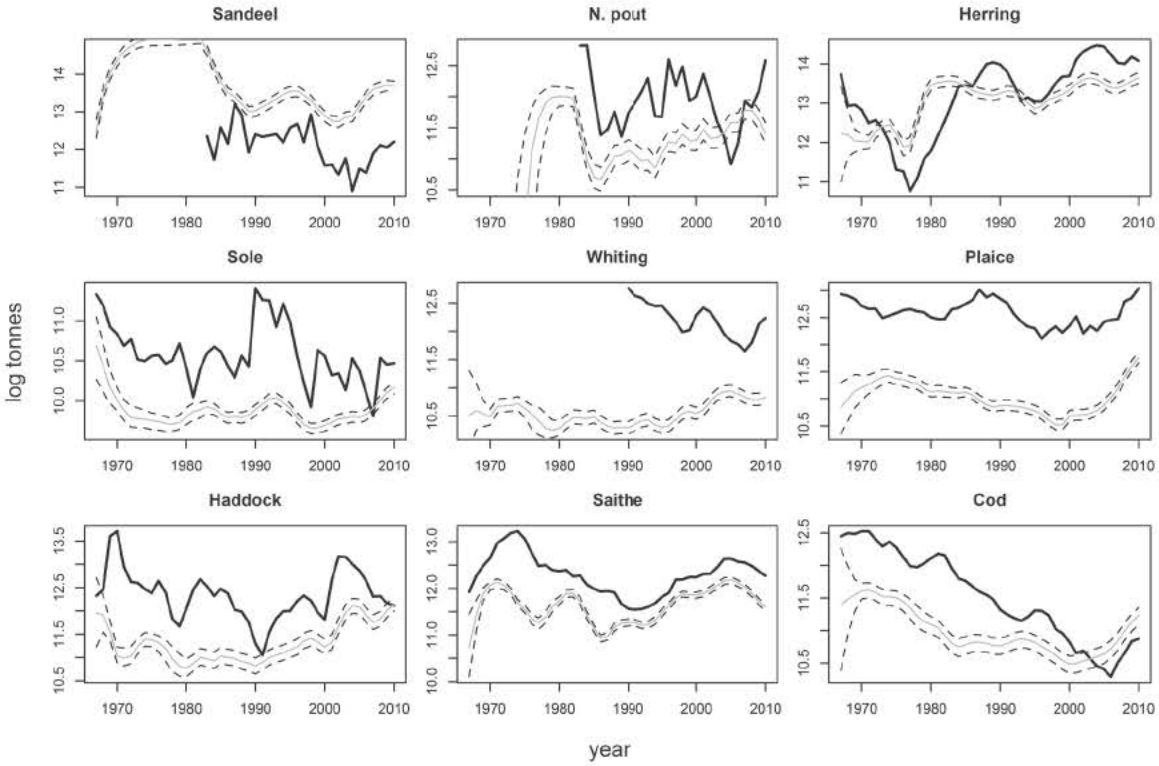
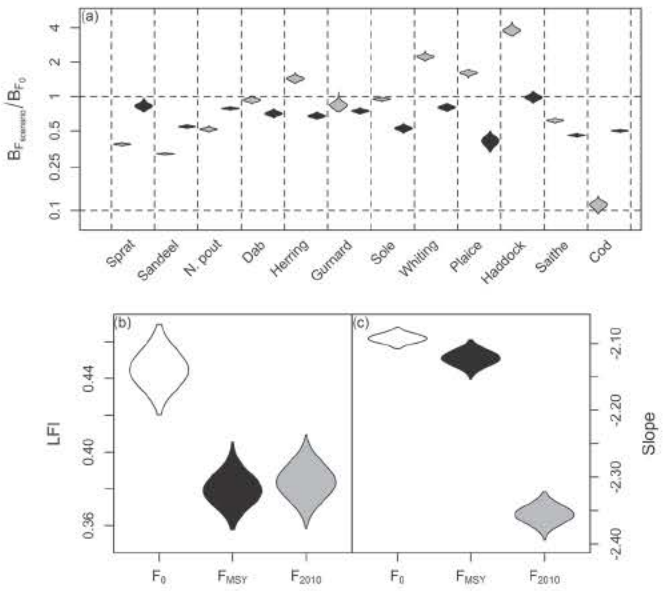


Fig. 5. The forecast for 2020: (*a*) the spawning stock biomass (*B*) with the fishing mortality at that of 2010 (F_{2010}; grey) and at maximum sustainable yield (F_{MSY}; black) divided by the spawning stock biomass when the fishing mortality is 0 (F_0) for the whole of the simulation; (*b*) the large fish indicator (LFI) for the fishing mortality equal to 0 (F_0; white), maximum sustainable yield (F_{MSY}; black), and that of 2010 (F_{2010}; grey); (*c*) the same but for the community size spectrum slope.



does not. The major advantage of the approach shown here is that the scenarios account for the range of likely parameters (as opposed to a single parameter set), enabling a probability distribution of the model outcomes formally linked to the parameter uncertainty.

Thorpe et al. (2015) use a multispecies length-structured model and show a stronger correlation between the response of the size spectrum slope and the large fish indicator than reported here. There are a few reasons that could explain this difference. First, the Thorpe et al. (2015) model differs from ours in terms of the dynamics. The model used here contains more complex dynamical feedbacks; the growth process is food-dependent and the dynamics are governed by a system of partial differential equations, whereas growth is nondynamic and with discrete time dynamics in the models used in Thorpe et al. (2015). It is worth noting that Thorpe et al. (2015) reported higher power of the size spectrum slope to detect a change over a 5- or 15-year fishing scenario compared with the LFI. Second, the species composition between models and inclusion in the calculation of the community metrics differed. Here, demersal species only were used to calculate community metrics (in keeping with empirical analyses; Fung et al. (2012)), and from further experiments, we found that the LFI is more sensitive to species subsetting than the slope of the community size spectrum.

We are not limited to forecasting the SSB, LFI, and size spectrum but can make forecasts, with robust measures of uncertainty, of any indicator that the model is able to predict. In Fig. 4, we compared the model output and the SSB from single-species stock assessments. Stock assessments use landings and survey data to estimate fishing mortalities and predict SSBs for each species separately, with different underlying assumptions across models. We used fishing mortalities from stock assessments as inputs to the multispecies model and fitted it to landings data. Because of the fundamental differences between single- and

multi-species models, we a priori expected SSB predictions to differ from single-species SSB estimates. The multispecies model predicts an overall higher SSB for sandeel than the single-species model, reflecting the need to meet predation requirements of larger fish in the model. With the exception of herring, lower SSBs were evident for several species, which is a result of the higher and explicit dynamically changing predation mortality present in the multispecies model.

In reality, the North Sea was not in a steady state in 1967, which could be a reason why we do not fit the dynamics of the landings well for all of the species (as indicated by larger values of σ_i^2/ι_i). Instead of restricting the spin-up period to the set of steady states, we could look at all possible states of the model before the dynamical fishing mortality was added to the model. This may be difficult to do in practice. Another possible reason for some of the poorer fits is that we are assuming that landings and catches are equivalent. For some species, there is likely to be a systematic difference between these two due to discards, e.g., gurnard.

The trend of the model simulations is the same for most of the possible parameter values that make up the posterior distribution, i.e., throughout the posterior, we overestimate the landings at one time and always underestimate the landings at another. Further experiments (for details, see Spence 2015) show that this is a feature of the model and is not sensitive to the parameter estimates. However, rather than assuming that the errors are independent and identically distributed, we could re-model the error structure so that the errors are correlated through time, possibly using an autoregressive model of order 1 (AR1; see, for example, Brockwell and Davis (2002)). We believe that this would improve the representation of the errors.

Figure 1 and Table 3 show no systematic pattern between the estimated maximum recruitment and asymptotic size as suggested in Andersen and Pedersen (2009) and Andersen and Beyer (2015). It is believed that $R_{\max}$ changes over time, possibly due to changes in habitat and temperature that have occurred in the North Sea (Bigg et al. 2008). We could include dynamic changes in $R_{\max}$ by including it as the hidden state in a state–space model (see, e.g., Rabiner 1989). This approach could also be used to estimate other useful parameters and even the model inputs (such as the fishing mortality) for each year.

We have used a carefully designed strategy involving Latin hypercube sampling, numerical optimization, and parallel tempering methods to explore a complex likelihood surface over a large parameter space as thoroughly and efficiently as possible. The high dimension of the space means that naïve methods would perform very poorly or be completely infeasible. For example, a simple systematic search with all combinations of two levels of each parameter would require 2^{26} or 67 108 864 runs of the model, and numerical integration over the parameter space would require even more. Numerical optimisation, with or without derivative information, and MCMC applied in isolation would be hampered by the many local maxima, although it is worth noting that our MCMC algorithm performs well locally, and so there is little to be gained by varying the details of the sampler. One way of improving the posterior distribution would be to use more informative priors. This could be done by eliciting the parameters (O'Hagan et al. 2006) or using simpler, more tractable models to produce priors (e.g., the single-species model of Andersen and Beyer (2015)).

As it stands, our overall strategy gives an enormous improvement over the results of even a relatively efficient single-stage Latin hypercube search. The best point out of the 50 000 sampled in the first round of our search had a log-likelihood of −13 790.19, and in the MCMC round, the best point from the sampled posterior had a log-likelihood of −322.08. Thus the likelihood itself is higher by a factor greater than 10^{5000}. As an informal interpretation, this means that the latter point represents a model that, using a simple model selection criterion such as the AIC, would be preferred statistically even if it involved thousands of extra parameters (whereas in fact it uses none). This leads us to believe that the method described here gives a good estimate of the posterior distribution and certainly much better parameter estimates and uncertainties than in previous work (Blanchard et al. 2014) or in what would be obtained with standard methods.

Our analysis allows for parameter uncertainty and for observation error. As it stands, it does not allow for the effects of structural uncertainty due to imperfections or limitations of the model itself. That could be handled by adding a discrepancy term, $\delta(\cdot)$, (Kennedy and O'Hagan 2001) to the formulation under "Likelihood"

$$\log \boldsymbol{Y} = \log M(\boldsymbol{\theta}) + \delta(\boldsymbol{\theta}) + \Sigma^{\frac{1}{2}}\boldsymbol{\epsilon}$$

Note that this is likely to have a similar effect to allowing for autocorrelation in the observation errors, as outlined above. The discrepancy term is used to allow for structural uncertainties. Such uncertainties are often caused by simplifications in the model, e.g., the dynamic model fitted here did not model discards.

Another source of uncertainty in predictions is stochasticity in the model, not addressed here because the model that we use is deterministic. With a stochastic model such as that of Andersen and Pedersen (2009), the principles of our approach would remain the same, but the details would differ. Instead of MCMC, we would need to use approximate Bayesian computation (ABC) (Tavaré et al. 1997; Beaumont 2010); the inclusion of observation errors means that a so-called exact ABC (Wilkinson 2013) or likelihood-free MCMC (Wilkinson 2010) could be used. This approach would retain the key advantages of the analysis described here: proper allowance for parameter and observation and uncertainty, and its propagation through to predictions. More generally, this Bayesian predictive framework can be applied to a wide variety of models and ecosystems. The range of computational tools to permit this in practice is constantly increasing; Spence (2015) gives some recent examples. As an alternative to formalizing the discrepancy within a single model, a promising approach is to consider a number of distinct models collectively, forming a multimodel ensemble. This can improve understanding of the strengths and weaknesses of individual models and potentially give better predictions and assessments of uncertainty overall. We are at present working on an ensemble that includes the current model as one of its members by considering discrepancy shared between models and specific to each model as used in climate modelling (e.g., Chandler 2013).

Further work on model uncertainty with size spectrum and other ecosystem models will enable multispecies forecasts to be reported to decision makers in a manner that is comparable to single-species decision tables. This would help further develop the use of formal risk assessment in ecosystem approaches to fisheries management, which has been fairly limited to date but is a burgeoning area of research (Plagányi et al. 2014).

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (grant EP/I000917/1, National Centre for Statistical Ecology) and by the Natural Environment Research Council and Department for Environment, Food and Rural Affairs (grant number NE/L003279/1, Marine Ecosystems Research Programme). We would also like to thank Toni Ingolf Gossmann, Christopher Griffiths, Abigail Marshall, Beth Mindel, Philipp Neubauer, and an anonymous reviewer for their useful comments on earlier versions of the manuscript.

References

Andersen, K.H., and Beyer, J.E. 2015. Size structure, not metabolic scaling rules, determines fisheries reference points. Fish Fish. 16(1): 1–22. doi:10.1111/faf.12042.

Andersen, K.H., and Pedersen, M. 2009. Damped trophic cascades driven by fishing in model marine ecosystems. Proc. R. Soc. B Biol. Sci. 227: 795–802. doi:10.1098/rspb.2009.1512.

Beaumont, M.A. 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41: 379–406. doi:10.1146/annurev-ecolsys-102209-144621.

Benoît, E., and Rochet, M.J. 2004. A continuous model of biomass size spectra governed by predation and the effects of fishing on them. J. Theor. Biol. 226: 9–21. doi:10.1016/S0022-5193(03)00290-X. PMID:14637050.

Bigg, G.R., Cunningham, C.W., Ottersen, G., Pogson, G.H., Wadley, M.R., and Williamson, P. 2008. Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. Proc. R. Soc. B Biol. Sci. 275(1631): 163–173. doi:10.1098/rspb.2007.1153.

Blanchard, J.L., Jennings, S., Law, R., Castle, M.D., McCloghrie, P., Rochet, M.J., and Benoît, E. 2009. How does abundance scale with body size in coupled size-structured food webs? J. Anim. Ecol. 78: 270–280. doi:10.1111/j.1365-2656.2008.01466.x. PMID:19120607.

Blanchard, J.L., Jennings, S., Holmes, R., Harle, J., Merino, G., Allen, J.I., Holt, J., Dulvy, N.K., and Barange, M. 2012. Potential consequences of climate change for primary production and fish production in large marine ecosystems. Philos. Trans. R. Soc. B Biol. Sci. 367(1605): 2979–2989. doi:10.1098/rstb.2012.0231.

Blanchard, J.L., Andersen, K.H., Scott, F., Hintzen, N.T., Piet, G., and Jennings, S. 2014. Evaluating targets and trade-offs among fisheries and conservation objectives using a multispecies size spectrum model. J. Appl. Ecol. 51(3): 612–622. doi:10.1111/1365-2664.12238.

Brockwell, P., and Davis, R. 2002. Introduction to time series and forecasting. 2nd ed. Springer, New York.

Chandler, R.E. 2013. Exploiting strength, discounting weakness: combining information from multiple climate simulators. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 371(1991). doi:10.1098/rsta.2012.0388.

De Roos, A.M., Schellekens, T., Van Kooten, T., Van De Wolfshaar, K., Claessen, D., and Persson, L. 2008. Simplifying a physiologically structured population model to a stage-structured biomass model. Theor. Popul. Biol. 73(1): 47–62. doi:10.1016/j.tpb.2007.09.004. PMID:18006030.

Fung, T., Farnsworth, K.D., Reid, D.G., and Rossberg, A.G. 2012. Recent data suggest no further recovery in North Sea Large Fish Indicator. ICES J. Mar. Sci. 69: 235–239. doi:10.1093/icesjms/fsr206.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. 2013. Bayesian data analysis. 3rd ed. Chapman and Hall.

Hartvig, M., Andersen, K.H., and Beyer, J.E. 2011. Food web framework for size-structured populations. J. Theor. Biol. 272: 113–122. doi:10.1016/j.jtbi.2010.12.006. PMID:21146543.

Harwood, J., and Stokes, K. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends Ecol. Evol. 18(12): 617–622. doi:10.1016/j.tree.2003.08.001.

Hintze, J.L., and Nelson, R.D. 1998. Violin plots: a box plot – density trace synergism. Am. Stat. 52(2): 181–184. doi:10.1080/00031305.1998.10480559.

Jacobsen, N.S., Gilsason, H., and Andersen, K.H. 2013. The consequences of balanced harvesting of fish communities. Proc. R. Soc. B Biol. Sci. 281(1775). doi:10.1098/rspb.2013.2701.

Kennedy, M.C., and O'Hagan, A. 2001. Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B Stat. Methodol. 63(3):425–464. doi:10.1111/1467-9868.00294.

Law, R., Plank, M.J., James, A., and Blanchard, J.L. 2009. Size-spectra dynamics from stochastic predation and growth of individuals. Ecology, 90: 802–811. doi:10.1890/07-1900.1. PMID:19341149.

Law, R., Kolding, J., and Plank, M.J. 2015. Squaring the circle: reconciling fishing and conservation of aquatic ecosystems. Fish Fish. 16(1): 160–174. doi:10.1111/faf.12056.

Li, H., and Wu, J. 2006. Uncertainty analysis in ecological studies. *In* Scaling and uncertainty analysis in ecology: methods and applications. *Edited by* J. Wu, K.B. Jones, H. Li, and O.L. Loucks. Springer. pp. 43–64.

Maury, O., and Poggiale, J.C. 2013. From individuals to populations to communities: a dynamic energy budget model of marine ecosystem size-spectrum including life history diversity. J. Theor. Biol. 324: 52–71. doi:10.1016/j.jtbi.2013.01.018. PMID:23395776.

McCarthy, M.A. 2007. Bayesian methods for ecology. Cambridge University Press.

McKay, M.D., Beckman, R.J., and Conover, W.J. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2): 239–245. doi:10.1080/00401706.1979.10489755.

Nelder, J.A., and Mead, R. 1965. A simplex method for function minimization. Comput. J. 7: 308–313. doi:10.1093/comjnl/7.4.308.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state–space models. Fish. Res. 158: 96–101. doi:10.1016/j.fishres.2014.01.014.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., and Rakow, T. 2006. Uncertain judgements: eliciting experts' probabilities. John Wiley and Sons.

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 Rothlisberg, P.C. 2014. Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. Fish Fish. 15(1): 1–22. doi:10.1111/j.1467-2979.2012.00488.x.

Rabiner, L.R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, 77(2): 257–286. doi:10.1109/5.18626.

Rochet, M.J., and Benoît, E. 2011. Fishing destabilizes the biomass flow in the marine size spectrum. Proc. R. Soc. B Biol. Sci. 279(1727): 284–292. doi:10.1098/rspb.2011.0893. PMID:21632631.

Scott, F., Blanchard, J.L., and Andersen, K.H. 2014. mizer: an R package for multispecies, trait-based and community size spectrum ecological modelling. Methods Ecol. Evol. 5(10): 1121–1125. doi:10.1111/2041-210X.12256. PMID:25866613.

Spence, M.A. 2015. Statistical issues in ecological simulation models. Ph.D. thesis, University of Sheffield.

Swendsen, R.H., and Wang, J.S. 1986. Replica Monte Carlo simulation of spin-glasses. Phys. Rev. Lett. 57(21): 2607–2609. doi:10.1103/PhysRevLett.57.2607. PMID:10033814.

Tavaré, S., Balding, D.J., Griffiths, R.C., and Donnelly, P. 1997. Inferring coalescence times from DNA sequence data. Genetics, 145(2): 505–518. PMID:9071603.

Thorpe, R.B., Le Quesne, W.J.F., Luxford, F., Collie, J.S., and Jennings, S. 2015. Evaluation and management implications of uncertainty in a multi-species size-structured model of population and community responses to fishing. Methods Ecol. Evol. 6(1): 49–58. doi:10.1111/2041-210X.12292. PMID:25866615.

Tsehaye, I., Jones, M.L., Bence, J.R., Brenden, T.O., Madenjian, C.P., and Warner, D M. 2014. A multispecies statistical age-structured model to assess predator–prey balance: application to an intensively managed Lake Michigan pelagic fish community. Can. J. Fish. Aquat. Sci. 71(4): 627–644. doi:10.1139/cjfas-2013-0313.

Vernon, I., Goldstein, M., and Bower, R.G. 2010. Galaxy formation: a Bayesian uncertainty analysis. Bayesian Anal. 5(4): 619–670. doi:10.1214/10-BA524.

Wilkinson, D.J. 2010. Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology. *In* Bayesian statistics 9. *Edited by* J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West. Oxford University Press.

Wilkinson, R.D. 2013. Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error. Stat. Appl. Genet. Mol. Biol. 12(2): 129–141. PMID:23652634.

Woodworth-Jefcoats, P.A., Polovina, J.J., Dunne, J.P., and Blanchard, J.L. 2013. Ecosystem size structure response to 21st century climate projection: large fish abundance decreases in the central North Pacific and increases in the California Current. Global Change Biol. 19(3): 724–733. doi:10.1111/gcb.12076. PMID:23504830.

Zhang, L., Hartvig, M., Knudsen, K., and Andersen, K.H. 2014. Size-based predictions of food web patterns. Theor. Ecol. 7(1): 23–33. doi:10.1007/s12080-013-0193-5.