Fisheries Research 207 (2018) 95–109

Contents lists available at ScienceDirect

Fisheries Research

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

Pushing the limits of a data challenged stock: A size- and age-structured assessment of ling (*Molva molva*) in Icelandic waters using Gadget

Bjarki Þór Elvarsson, Pamela J. Woods, Höskuldur Björnsson, Jamie Lentin, Guðmundur Thordarson

Marine and Freshwater Research Institute, Skúlagata 4, 121 Reykjavik, Iceland

ARTICLE INFO

Handled by A.E. Punt

Keywords:
Stock-assessment
Uncertainty
Bootstrap
Database
Gadget
Fish population dynamics
Nonlinear models
Correlated data
Bootstrapping
Data-limited species

ABSTRACT

In recent years a greater emphasis has been placed on developing management strategies that prevent overexploitation. Harvest control rules (HCRs) have therefore, in many places, been developed and implemented. Commonly these HCRs are developed for stocks that are assessed using age-structured models, and various platforms exist to evaluate their performance and analyze various sources of bias for that particular class of models. Many stocks, however, cannot be assessed reliably using classical age-structured methods due to data limitations (gaps in data series, unreliable age readings, etc.). One such stock is the common ling (*Molva molva*) in Icelandic waters. Availability of data on the stock dynamics, in particular age data for both survey and commercial samples, has been a limiting factor when assessing the stock. When modeling stocks such as this, data limitations need to be considered, and how associated uncertainty is propagated both through the assessment and into the advice. In this study, ling was assessed using the size- and age-structured model Gadget after synthesizing all available data. Having limited age data available causes high uncertainty in the model fitting process, especially in estimating growth. However, including this key uncertainty in the assessment allowed the subsequent management strategy evaluation to take it into account directly while deriving common management reference points and estimating uncertainties in stock status and other derived quantities. Uncertainty was estimated using a specialized bootstrap for disparate data sets that mimics the sampling process. The process of assimilating data for the assessment model and the bootstrap procedure was performed using a specialized database program, MFDB, ensuring that the whole process is reproducible.

1. Introduction

In recent years there has been a call for sustainable management of fisheries. This is reflected in a number of common multinational resolutions on the governance of marine ecosystems (e.g. UN, 2002; Parliament, 2008). In particular the European Union has stipulated that all fish stock should be managed according to maximum sustainable yield principle (Anon, 2002a,b). To ensure that these objectives are reached, management plans that restrict fishing effort have been proposed and implemented (e.g. see Annex 1 of ICES, 2013b). Typically these plans include some form of a harvest control rule (HCR) based on the available data (e.g. see Baldursson et al., 1996; Butterworth and Punt, 1999).

Evaluating fisheries management plans is not a trivial undertaking. The HCR is often simulation tested using an operating model which is based on knowledge of the population dynamics (discussed by Butterworth and Punt, 1999, and references therein) and industry governance. ICES (2013a) provides guidelines on how to conduct these simulations, and ICES (2017b) specifically describes how to derive management reference points necessary to implement an HCR in European waters. For many species, the information typically needed for traditional age-based assessments is lacking, leaving little data available to inform general productivity and stock structure. This is true for many of the stocks assessed by multinational bodies such as ICES (e.g. see ICES, 2014a). For example, some age-based methods do not allow for years of missing data (e.g. Shepherd, 1999, and other VPA-based methods). According to ICES, the lack of data to produce an assessment, and subsequently quantitative forecasts, warrants a classification as data limited (ICES, 2012). Without age composition data, variants of the surplus production model (as described by Pella and Tomlinson, 1969) are commonly applied (Carruthers et al., 2014). These approaches allow for the analytical estimation of reference points while being based on fairly limited data. For example, a popular expansion of this approach was developed by Pedersen and Berg (2017), which has been rapidly applied in a variety of cases (e.g., ICES, 2017a). Surplus production methods, however, are known to fail in situations where

E-mail address: bjarki.elvarsson@hafogvatn.is (B.Þ. Elvarsson).

https://doi.org/10.1016/j.fishres.2018.06.005
Received 8 March 2018; Received in revised form 6 June 2018; Accepted 7 June 2018

little contrast is available in the survey and catch time series or such contrast is only exhibited as a constant decrease (i.e., "one-way trip"). In contrast, age-structured models are still able to capture some information from such scenarios (Magnusson and Hilborn, 2007) as there is contrast in cohort strength.

Age-based methods are therefore preferred, and considered a standard in stock assessment the advisory process, because age transfers important information into a stock assessment model: it allows for inference of the time scale on which population dynamics occur, by supplying information on the growth of individuals and how it translates into growth of the population. Age in combination with weight- and maturity-at age is used to calculate the rate at which spawning stock biomass is generated, which can in turn be used to detect recruitment impairment due to low spawning stock biomass within a stock–recruitment relationship. At the same time, however, inclusion of faulty age-related information can lead to bias. Age-based methods assume that age is known perfectly with no error, a false assumption in many cases (Reeves, 2003; Yule et al., 2008; Treble et al., 2008). Ageing error can then cause bias in a number of age-based processes within age-based stock assessment methods, since a variety of data included are a discretized by age (e.g., numbers-, catch-, maturity-, selectivity- and weight-at age). Quite often the age determination involves the processing of sagital otoliths and/or a study of length distributions to infer a cohort structure (as discussed by Jobling, 2002 and references therein). The ageing process is for many species, such as cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*), fairly straightforward. Both species are highly abundant, many facets of their life cycle well known and they are of considerable commercial value. However, even for highly studied data-rich species, inconsistencies in ageing can bias stock assessments (e.g., Baltic cod among other species Bertignac and De Pontual, 2007; Koenigs et al., 2013; Henríquez et al., 2016; Hüssy et al., 2016), as can undetected changes in growth rates over time (e.g., Icelandic haddock ICES, 2017c). Fishing-induced changes in growth can also bias age-based assessments, spurring the development of "platoons" in age-structured models (Taylor and Methot, 2013; Akselrud et al., 2017). Finally, for many species, information on age is simply hard to obtain. This may be due to lack of hard parts that show year rings, inconclusive otolith readings, or difficulties/inconsistencies in age data collection (Treble et al., 2008). As a result, even when otoliths are available, translating continuously deposited bone tissue (i.e., rings) into discrete annual growth measures (i.e., age) typically require sound and validated methods.

In response to the common need for stock assessment models to both handle data limitations as well as propagate error appropriately throughout the assessment by integrating various steps of analysis into a single stock assessment model, integrated stock assessment models, such as the Gadget model presented here, have increased in popularity over the last decades (Maunder and Punt, 2013). The ICES classification of data-limited is often a misnomer as there may be a wealth of other information, such as size composition data, on the species than that which is directly applicable to standard assessment models. Were a stock to be evaluated under alternate criteria, it may not be considered data limited due to the existence of at least some compositional data in addition to survey indices and catch (Ralston et al., 2011; Berkson and Thorson, 2014; Carruthers et al., 2014). For example, the lack of reliable age data on ling in Icelandic waters is reminiscent of assessments of many invertebrate stocks (e.g. see Punt et al., 2013, 2016, and similar papers). The stock assessment of ling in Icelandic waters presented here is therefore more analogous to size-structured assessments, as historically little information has been collected from commercial samples, particularly on age.

Size-structured models that also track age, so that data on growth may be used to supplement them, are commonly referred to as size- and age-structured models (Punt et al., 2017), and are most commonly implemented as integrated models. In size- and age-structured models, the data and model predictions have two attributes: a length-group bin and an age-group bin. As a result, when parameterized such that growth, maturation, and selection processes are only dependent on size, size-structured models are a special case of size- and age-structured models. However, common implementations of age-based models (e.g., Stock Synthesis Methot, 2013 or MULTIFAN-CL Fournier et al., 1998) are often not a special case of size- and age-structured models, due to the need to apply a summarised effect of growth, maturity, and/or selectivity (when these are size-based processes) to all individuals (regardless of length) within an age bin (see Punt et al., 2017, for an example). Size- and age-based models, such as those developed using Gadget (Begley and Howell, 2004) or CASAL2 (Doonan et al., 2016), offer alternative methods to assess the stock status combining compositional data if and when available.

The trade-off for using size- and age-structured models comes in the form of a reduction in computational efficiency, due to the higher dimensionality of the model (Punt et al., 2017). But in the case of data limited species, the resulting benefits may be well worth the cost. Even if there is little or no information available on age, other size-structured biological information may be available that can provide insights into the stock dynamics. In terms of management, the inclusion of even very limited length data may improve estimate on how much the stock can reasonably be harvested without severely depleting the stock (Wetzel and Punt, 2011).

The goal of this study is to demonstrate how a size- and age-based model (i.e., Gadget) can be suitable for stock assessment by providing an appropriate means to propagate error, especially age-related error, into a management strategy evaluation of harvest control rules. This framework is especially valuable where data are limited, such as in the case for ling (*Molva molva*) in Icelandic waters, as standard age-based methods are likely to misrepresent age-related uncertainty. Gadget is a statistical modelling and simulation framework that allows the creation of a multi–species, multi–fleet, multi–stock, size- and age-structured simulation model. Originally outlined by Stefansson and Palsson (1998) Gadget is a conceptual continuation of the work described by Gavaris (1988) and Bogstad et al. (1997) and is implemented as a computer program (Begley, 2005). We present simulations where observation uncertainty (and to a certain extent structural error) is projected forward using a specialised spatial bootstrap approach described by Elvarsson et al. (2014). Robust data handling is also essential for this line of work; therefore, a specialised database system, MFDB (Lentin, 2014), is also presented which builds upon concepts of database design that particularly suit the needs of stock assessment and ecosystem studies, as described by Kupca (2006). This database procedure is used in conjunction with a specialised R package, Rgadget (Elvarsson and Lentin, 2018), that allow rapid and reproducible model building within the Gadget framework. Although the simulation procedure described here is applied to a single-species assessment, it can be generalized to a wider set of models, e.g. multi-species, multi-stock, or multi-fleet models, as implemented in the Gadget framework.

2. Materials and methods

The first step of this study details a data challenged stock assessment using Gadget, after synthesizing the available data on the population dynamics of ling. The second step extends the assessment model by setting up a projection model in which precautionary biomass reference points were derived (first set of projections). In the final step, the projection model was used as the operating model on which a management strategy evaluation (MSE) was based, in which the application of simple harvest control rule was simulated (second set of projections).

2.1. Ling in Icelandic waters

Ling (*Molva molva*) is a demersal fish found in the Northeast Atlantic, with the main spawning grounds observed south of Iceland, by the Faroe islands and in the Norway Sea, representing different stocks

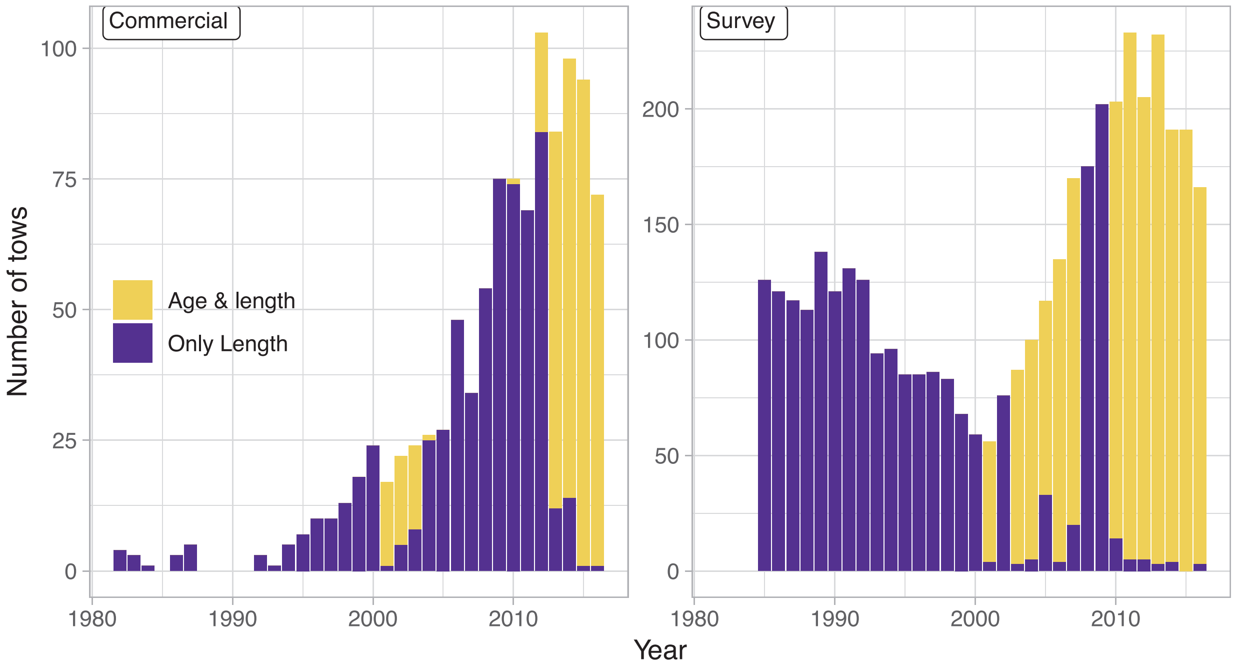

Fig. 1. The number of tows in which biological samples have been taken of ling in Icelandic waters, from both the survey and commercial fisheries operations.

biologically and for assessment purposes. The assessment presented here is based on ling from Icelandic waters, where it is fairly abundant in the southern part of the Icelandic continental shelf area (ICES, 2014a). Ling is caught at depths 15–1000 m, while the bulk of the catches are between 100 and 400 m. It feeds mainly on fish (the Northern herring *Clupea harengus*, various flatfish, haddock, and cod) and benthic invertebrates. Spawning occurs in May and June at depths of 150–300 m near the southern edge of the Icelandic shelf.

Historical catch series of ling date as far back as the cod fishery in Iceland. However, due to relatively low abundance in catches compared to other species, very few biological samples are available despite consistent surveys (Fig. 1). Relatively few commercial samples are available prior to beginning of the century. In more recent years, an increase in survey abundance has been observed, which has coincided with increased catches.

Ling in Icelandic waters is mainly caught by three types of gear: longlines, gillnets and trawls. Most ling are caught by longlines (65% in 2009–2011). The fishery for ling in Icelandic waters has not changed substantially in recent years, although as the proportion caught by longlines has increased, the proportion of ling caught by gillnets has decreased from 20–30% in 2000–2001 to 3–8% in 2008–2011. Catches in trawls have remained more constant, around 20%. This pattern reflects broad changes in the commercial fleet operations in Icelandic waters (MFRI, 2017). Ling is mainly caught as bycatch in cod or haddock fisheries which are not highly seasonal.

2.2. Operating model

To encompass the available knowledge on ling and fishing operations, a population dynamics model describing the main stock dynamics and interaction with fleets was set up. The model, whose components are illustrated in Fig. 2, was essentially a single species model (i.e., multi-species features were not used) with stock components defined for immature and mature portions (reflecting a single stock in the biological or spatial sense). Four fleets targeted ling, three gear–based commercial fleets plus the survey, allowing for differences in selectivity and resulting temporal changes in catch composition due to changes in fleet composition. In a typical Gadget model the simulated quantity is the number of individuals, $N_{a,l,s,y,t}$, at age $a = a_{\min} \dots a_{\max}$, in a length-group l (representing lengths ranging between $l_{\min}$ and $l_{\max}$ in length-group ranges of Δl), in stock component s where $s = 0$ represents the immature stock component and $s = 1$ represents the mature stock component, at year y which is divided into timesteps of $t = 1 \dots t_{\max}$ (usually $t_{\max} = 4$, reflecting quarters). The time step length is denoted Δt (usually in units y^{-1}). Population numbers are governed by the following equations:

$$
\begin{aligned}
N_{a,l,s,y,t+1} &= \textstyle\sum_{l'} G_{l'}^{l}\Big[(N_{a,l',s,y,t} - \sum_f C_{f,a,l',s,t})e^{-M_a \Delta t} + A_{a,l',l,s,y,t}\Big] && \text{if } t < t_{\max}\\
N_{a,l,s,y+1,t=1} &= \textstyle\sum_{l'} G_{l'}^{l}\Big[(N_{a-1,l',s,y,t=t_{\max}} - \sum_f C_{f,a-1,l',s,t=t_{\max}})e^{-M_{a-1}\Delta t} && \text{if } a_{\min} < a < a_{\max}\\
&\quad + A_{a-1,l',l,s,y,t=t_{\max}}\Big] &&\\
N_{a,l,s,y+1,t=1} &= \textstyle\sum_{l'} G_{l'}^{l}\Big[(N_{a,l',s,y,t=t_{\max}} - \sum_f C_{f,a,l',s,y,t=t_{\max}} + &&\\
&\quad N_{a-1,l',s,y,t=t_{\max}} - \sum_f C_{f,a-1,l',s,y,t=t_{\max}})e^{-M_a \Delta t}\Big] && \text{if } a = a_{\max}
\end{aligned}
\tag{1}
$$

where $G_{l'}^{l}$ is the proportion in length-group l that has grown $l - l'$ length units (e.g., cm) in Δt. $C_{f,a,l',s,y,t}$ denotes the catches by fleet $f \in \{S, L, G, T\}$, i.e. the fleets representing the survey (S),[1] longliners (L), gillnetters (G), and trawlers (T). M_a is the natural mortality at age a and $A_{a,l',l,s,y,t}$ denotes the numbers of immature fish at length l' that have matured (i.e., "moved" from the immature to the mature stock component) as they grew to length l. Throughout the model description, l and l' are used either to reflect two separate length groups, or the midpoints of these length-group intervals, depending on the context.

2.2.1. Growth

Growth in length is modeled as a two-stage process. First, an average length increment over Δt is modeled as a size-dependent process using a parametric growth function, such as a Von Bertanlanffy function as employed in this study:

$$\Delta l = (l_\infty - l)(1 - e^{-k\Delta t}) \tag{2}$$

where l_∞ is the maximum asymptotic length and k is the annual growth rate. Second, dispersion around the growth increment $(l - l')$ is modeled according to a beta–binomial density (as described by Stefánsson,

[1] The survey fleet catches are given a nominal catch (e.g., 1 kg per year) to allow for survey age and length distribution predictions.

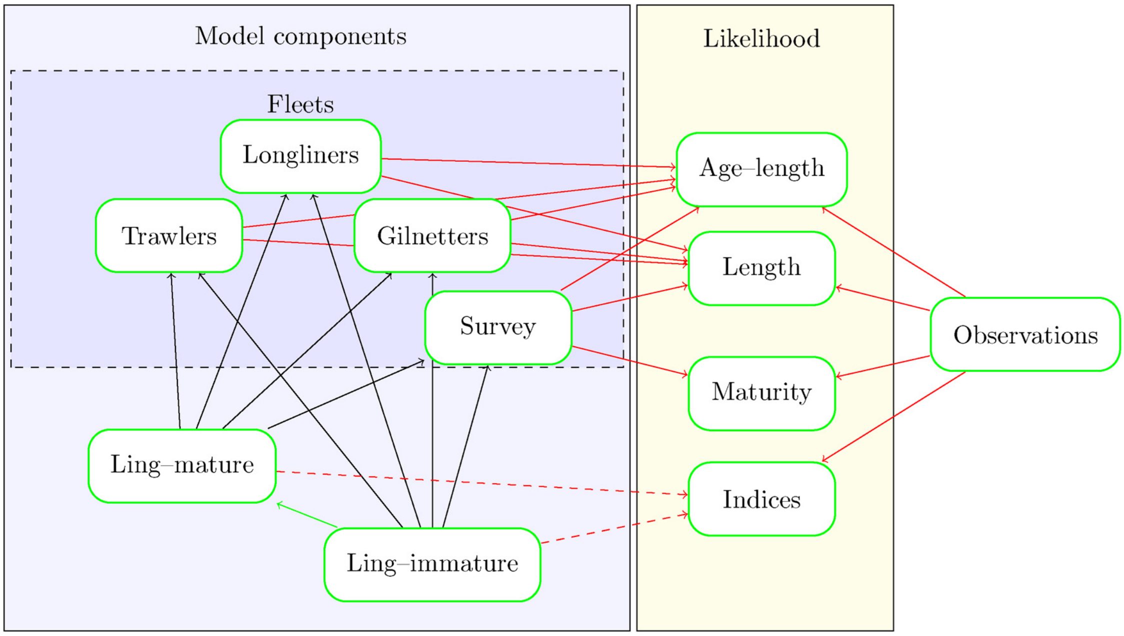

Fig. 2. Schematic description of the Gadget model for ling. Lines indicate flow from one model component to the other. Black lines indicate consumption by predators (fleets), red lines the modelled predictions/observations sent to the likelihood and green lines movement between immature and mature stock components. Dashed red lines indicate that predictions of survey indices were not adjusted by survey selectivity; instead, catchability was estimated. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

2005):

$$G_{l'}^{l} = B(\alpha, \beta, n) \tag{3}$$

with a mean $\alpha\beta = \Delta l$ and α is subject to

$$\alpha = \frac{\beta \Delta l}{n - \Delta l} \tag{4}$$

with n fixed as the maximum length group steps over which a fish can grow in a particular timestep, and β estimated. Note that inherent in this approach is that the growth increment is always positive, which requires a slightly different interpretation of the k and L_∞ parameters compared to what one would expect from a more traditional Von Bertanlanffy age-based model (i.e., L_∞ is rather like an upper boundary than a mean maximum expected length).

The weight at length-group l, $W_{s,l}$, is calculated according to the following length–weight relationship[2]:

$$W_{s,l} = \mu_s l^{\omega_s} \tag{5}$$

2.2.2. Recruitment and initial abundance

The number of recruits each year, R_y, is estimated within the model as a fixed effect (separate parameters per year) and recruitment enters the population according to:

$$N_{a_{\min},l,s=0,y,t'} = R_y p_l \tag{6}$$

where t' denotes the recruitment time-step and p_l is the proportion in length-group l that is recruited. The proportion p_l is determined by a normal density with mean length set to the initial length l_0, which corresponds with the recruitment age 1 according to Eq. (2) and variance at the recruitment age a' ($\sigma_{a'}^2$). [3]

A similar formulation of initial abundance in numbers is used for older age groups in length-group l:

$$N_{a \neq a_{\min},l,s,y=1982,t=1} = \nu_{a,s} q_{a,l} e^{-(F_0 + M_a)} \tag{7}$$

where $\nu_{a,s}$ is the initial number at age a in the initial year of stock s scaled by the initial mortality $F_0 + M_a$ and $q_{a,l}$, the proportion at length-group l which is determined by a normal density with a mean predicted by the growth model in Eq. (2) and a variance σ_a^2. The initial fishing mortality, F_0, is used here for numerical purposes to allow for sensible starting values and parameter boundaries when estimating $\nu_{a,s}$.

2.2.3. Maturation

Maturation is modeled as unidirectional movement from the immature ($s = 0$) to the mature ($s = 1$) stock component:

$$A_{a,l',l,s,y,t} = \begin{cases} N_{a,l',s'=0,y,t} \times m_{l'}^{l} & \text{if } s = 1 \text{ and } t > 1 \\ -N_{a,l',s'=0,y,t} \times m_{l'}^{l} & \text{if } s = 0 \text{ and } t > 1 \end{cases} \tag{8}$$

where $m_{l'}^{l}$, the proportion of immatures that mature after growing from length l' to l, is defined as:

$$m_{l'}^{l} = \begin{cases} \frac{\lambda(l - l')}{1 + e^{-\lambda(l - l_{50})}} & \text{if a} < a_{\text{maxmat}} \\ m_{l'}^{l} = 1 & \text{if a} = a_{\text{maxmat}} \end{cases} \tag{9}$$

where l_{50} is the length at 50% maturity and estimated along with λ, and a_{maxmat} reflects a set upper bound for the maturity ogive to apply, after which all fish are mature.

2.2.4. Fleet operations

Fleet-specific selectivity functions were modeled as

$$S_f(l) = \frac{1}{1 + e^{(-b_f(l - l_{50,f}))}} \tag{10}$$

Annual landings data were fleet- and time-step-specific ($L_{f,y,t}$), and assimilated into the model as direct removals of biomass by first simulating the expected composition of catches based on Eq. (11).[4] The expected compositions of catches in biomass ($C_{f,a,l,s,y,t}^{B}$) were determined by multiplying landings ($L_{f,y,t}$) by the proportion of the total

[2] Gadget also allows for time-varying weight–length relationships, but this was not considered beneficial in this setting.

[3] Initial length l_0 used here was parameterized to have a similar effect as t_0 in a typical von Bertalanffy growth model with a one-to-one mapping between the two.

[4] Other functional forms, referred to as ‘suitability’, are defined in Gadget.

biomass available to be fished at a given age, length and stock component (immature/mature). The proportion was defined by biomass at a specific value of a, l, and s in the numerator, as available to the fleet (i.e., multiplied by selectivity S_f in Eq. (11), divided by the sum of all biomass fishable by the fleet (i.e., values summed over the full range s', l', and a' in the denominator). Catches in numbers are then defined by dividing by stock- and length-specific weights ($C_{f,a,l,s,y,t}$):

$$C^B_{f,a,l,s,y,t} = L_{f,y,t}\frac{S_f(l)N_{a,l,s,y,t}W_{l,s}}{\sum_{s'}\sum_{l'}\sum_{a'}S_f(l')N_{a',l',s',y,t}W_{l',s'}}$$

$$C_{f,a,l,s,y,t} = \frac{C^B_{f,a,l,s,y,t}}{W_{l,s}} \tag{11}$$

2.3. Observation model

The model simulation begins in 1982, with a maximum age set as a plus group to 15 (at which ≈1% of fish were minimally aged). Maturity data were only available for aged fish and sparse for older fish. Therefore, all ling were set to mature at 10 ($a_{maxmat} = 10$), as 90% aged 10 or older were mature, to simplify estimation of the maturity give applied to younger fish. Recruitment to the immature stock component occurs at age 3, at the end of the 1st quarter. Fully selected ages include age 12 and above. The length range in the model was between 20 and 160, in 4 cm length intervals. The reported reference biomass was the biomass of fish larger than or equal to 75 cm, denoted $B_{75\text{cm}+,y}$. An overview of the data sets and model parameters used in the model study is shown in Tables 2 and 3 respectively. In formulations below it is assumed that the compositional data are sampled at random, both from the fishery and surveys, as this is how the sampling protocol in Icelandic waters is set up for ling.[5]

The bulk of the biological samples of ling comes from the Icelandic groundfish survey (described by Pálsson et al., 1997) which is conducted in March every year. The survey started in 1985 with the aim of measuring the changes in the biomass of key fish species in Iceland, cod being the primary focus. Ling is typically caught south of Iceland in the survey, with the average number of ling caught in a given year around 420 individuals in 131 tows (see Fig. 3). In the survey, all fish lengths are measured and every fourth fish (or a minimum of 5 and maximum of 25 per sampling station, although this upper limit is rarely reached) is further processed for biological information such as age, maturity, sex and organ weight. In the most recent years a considerable increase in the catches of ling has been observed (Fig. 1), resulting in a substantial increase in the number of samples from commercial catches despite similar sampling effort.

With a model such as the one used here that compiles and aggregates various sources of data, accurate and efficient data processing is imperative to mimimize potential errors, increase transparency in the model-fitting process, and facilitate exploration of different model configurations. To this end, all data used in this study were processed through as specialised database system. This database system was interfaced using an R package, MFDB (Lentin, 2014), where all data relevant to single- or multi-species stock assessment in Iceland were stored in a minimally aggregated manner. Data were imported to the system from various sources such as biological institutional databases, landing statistics and logbooks. Similarly MFDB has aggregation and export routines that generate input files for the likelihood components in the desired file format needed for Gadget. This allowed for the age and length aggregation of the data to be adjusted whenever needed, thereby speeding the model exploration and development process. In addition to MFDB, all model settings, estimation (including the iterative reweighting described below) and output processing was conducted using a specialised R package, Rgadget (Elvarsson and Lentin, 2018).

[5] Other forms of likelihoods are implemented in Gadget that can be used to address other types of sampling, e.g. length-stratified sampling of age or maturity.

Table 1
Length aggregation of survey indices used for fitting the model.

Name	Min	Max
si.20–50	20	52
si.50–60	52	60
si.60–70	60	72
si.70–80	72	80
si.80–90	80	92
si.90–100	92	100
si.100–160	100	160

The combination of these two R packages, which are freely available but still under development, allowed the whole modelling process to be reproducible. Model scripts are available at https://github.com/fishvice/gadget-models/06-ling.

2.3.1. Survey indices

Survey indices from the Icelandic groundfish survey were used to fit the model. The survey abundance indices were aggregated into seven length intervals (Table 1, Fig. S1 in supplementary material) and defined as the total number of fish caught in a survey within a certain length interval. Eight or 12 cm intervals were used for the indices (due to the 4 cm length–bin model structure), except the smallest and the largest length intervals. These intervals were enlarged to avoid getting zero values in the bootstrap replicates described below.

For each length range g the survey index is compared to the modeled abundance at year y and time-step t using:

$$l^{\text{SI}}_g = \sum_y\sum_t(\log I_{g,y} - (\log q_g + b_g\log\widehat{N_{g,y,t}}))^2 \tag{12}$$

where

$$\widehat{N_{g,y,t}} = \sum_{l\in g}\sum_a\sum_s N_{a,l,s,y,t},$$

I_{gy} refers to survey index summed within indicated length range, q_g is the associated catchability, and b_g controls the shape of the power function that relates the index to absolute abundance (i.e., $b_g = 1$ indicates linearity).

2.3.2. Fleet data

Length frequency distributions from survey and commercial catches were compared to predictions as proportions using

$$l^{\text{LD}}_f = \sum_y\sum_t\sum_l(\pi_{f,l,y,t} - \hat{\pi}_{f,l,y,t})^2 \tag{13}$$

where f denotes the fleet from which data was sampled, and the proportions

$$\pi_{f,l,y,t} = \frac{\sum_{a'}\sum_{s'}O_{f,a',l,s',y,t}}{\sum_{a'}\sum_{l'}\sum_{s'}O_{f,a',l',s',y,t}}$$

and

$$\hat{\pi}_{f,l,y,t} = \frac{\sum_{a'}\sum_{s'}C_{f,a',l,s',y,t}}{\sum_{a'}\sum_{l'}\sum_{s'}C_{f,a',l',s',y,t}}$$

represent the observed and modeled proportions respectively. Length frequency proportions for a given fleet (f), year (y) and time-step (t) combination were formed as the numbers of fish at a particular length (l), but summed over ages (a') and stock components (s'), divided by numbers summed over all possible lengths (l'), ages (a') and stock components (s'). The choice of the form of likelihood follows the suggestion from Taylor et al. (2007), as the assumption of a multinomial likelihood is inappropriate due to the correlation structure in length composition data (e.g. see Hrafnkelsson and Stefánsson, 2004; Babak et al., 2007). Data on age-length and maturity proportions were compared to model predictions in a similar manner.

Table 2
Overview of data used in the observation model. Survey indices were calculated from the length distributions and are disaggregated ("sliced") into seven groups (Table 1). Number of data points refers to aggregated data used as inputs in the Gadget model and represents the original data set. Weight groups are those used during iterative re-weighting. All data can obtained from the Marine and Freshwater Research Institute, Iceland.

Origin	Time-span	Length group size	Num. data-points	Likelihood-function	Weight group
Age–length distributions:					
Bottom trawl	All quarters, 2001–2016	4 cm	946	See Eq. (13)	comm
Gillnet	All quarters, 2001–2016	4 cm	449	See Eq. (13)	comm
March survey	2nd, 2001–2016	4 cm	935	See Eq. (13)	aldist.igfs
Longline	All quarters, 2001–2016	4 cm	1291	See Eq. (13)	aldist.lln
Length distributions:					
Bottom trawl	All quarters, 1982–2016	4 cm	1440	See Eq. (13)	comm
Gillnet	All quarters, 1982–2016	4 cm	693	See Eq. (13)	comm
March survey	2nd, 1985–2016	4 cm	928	See Eq. (13)	ldist.igfs
Longline	All quarters, 1994–2016	4 cm	2129	See Eq. (13)	ldist.lln
Ratio of immature: mature by length group:					
March survey	2nd, 1990–2016	8 cm	680	See Eq. (13)	matp.igfs
Survey indices:					
March survey	1st, 1985–2016	20–52 cm	32	See Eq. (12)	sind1
March survey	1st, 1985–2016	52–60 cm	32	See Eq. (12)	sind1
March survey	1st, 1985–2016	60–72 cm	32	See Eq. (12)	sind1
March survey	1st, 1985–2016	72–80 cm cm	32	See Eq. (12)	sind2
March survey	1st, 1985–2016	80–92 cm	32	See Eq. (12)	sind2
March survey	1st, 1985–2016	92–100 cm	32	See Eq. (12)	sind2
March survey	1st, 1985–2016	100–160 cm	32	See Eq. (12)	sind2

Table 3
Overview of parameters in the operating model.

Description	Notation	Estim.	Comments	Eq.
Natural mortality	M_a	No	0.15	(1)
Growth function	k, L_∞, l_0	Yes	Maximal growth rate, asymptotic length, and initial length (at age 1).	(2)
Growth dispersion	β, n	Yes, No	$n = 15$ max length group increase in a time step	(3)
Selectivity	$b_f, l_{50,f}$	Yes	Fleet-specific	(10)
Maturity ogive	λ, l_{50}	Yes		(8)
Variance in length at recruitment	σ_a^2	Yes	For $a = 3$, $y \in [1982, 2016]$	(6)
Variances around initial mean lengths	σ_a^2	No	For $a \neq 3$, $y = 1982$, based on length distributions obtained in the survey.	(7)
Number of recruits	R_y	Yes	Determines numbers at $a = 3$ for $y \in [1982, 2016]$.	(6)
Initial abundances	$\nu_{s,a}$	Yes	For $a \neq 3$, $y = 1982$	(7)
Survey catchability	q_g, b_g	Yes	Intercepts and slopes term in a log–linear relationships of indices with abundances. The slope b_g is fixed to 1 for all indices but si.20–50 and si.50–60.	(12)
Length–weight relationship	μ_s, ω_s	No	Based on lengths and weights obtained in the survey	(5)
Scalars	$R_c, I_{c,s}, F_0$	Yes	Scaling coefficients for recruitment, initial numbers at age, and initial fishing mortality (applied to all age groups)	(7)

2.3.3. Iterative re–weighting

The objective function used in the modeling process combines Eqs. (12)–(13) with the following formula:

$$l^T = \sum_g w_{gf}^{\mathrm{SI}} l_{g,S}^{\mathrm{SI}} + \sum_{f \in \{S,T,G,L\}} (w_f^{\mathrm{LD}} l_f^{\mathrm{LD}} + w_f^{\mathrm{AL}} l_f^{\mathrm{AL}}) + w^M l^M \tag{14}$$

where SI, LD, AL and M refer to number (SI = survey index) and proportion data types (LD = length distribution, AL = age-length, M = maturity; see previous Sections 2.3.1 and 2.3.2) and w reflects the weight assigned to each likelihood components.

Weights are assigned according to the iterative weighting heuristic introduced by Stefánsson (2003), and subsequently implemented by Taylor et al. (2007). Essentially this heuristic is a inverse variance approach where the component variances are based on multiple model runs in which a particular component has been excessively emphasized (Elvarsson and Lentin, 2018). To avoid issues related to overfitting individual length-grouped survey indices, likelihood components were grouped and weighted together, thus ensuring that more than one survey index per year were available within a weighted group (Table 2).

2.3.4. Uncertainty estimation via bootstrapping

To estimate the uncertainty in the model parameters and derived quantities, a specialised boostrap for disparate datasets was used (described by Elvarsson et al., 2014). The approach was based on spatial subdivisions, which are illustrated in Fig. 3, that are considered to be i.i.d. This assumption of identical and independent distribution is of course not always true, but the bootstrap approach does roughly mimic the sampling process. When performing resampling, the compositional data within each subdivision were aggregated, and then summarised over the subdivision selected in the bootstrap. For each bootstrap replicate, the operating model (described in the previous sections) was fitted the same way as for the original dataset base run. One hundred bootstrap replicates were used, following the advice of Elvarsson et al. (2014), who considered 100 to be a good balance between computing time and accuracy in parameter variance estimates. The bootstrap approach was implemented within MFDB (Lentin, 2014), where the user defines the spatial subdivisions and the bootstrap procedure generates the desired number of replicate input files.

2.4. Stock status and derivation of biological reference points

Estimates of stock status were used to derive biomass and effort-related reference points according the guidelines described by ICES (2013a). Generally, ICES derives reference points based on fishing

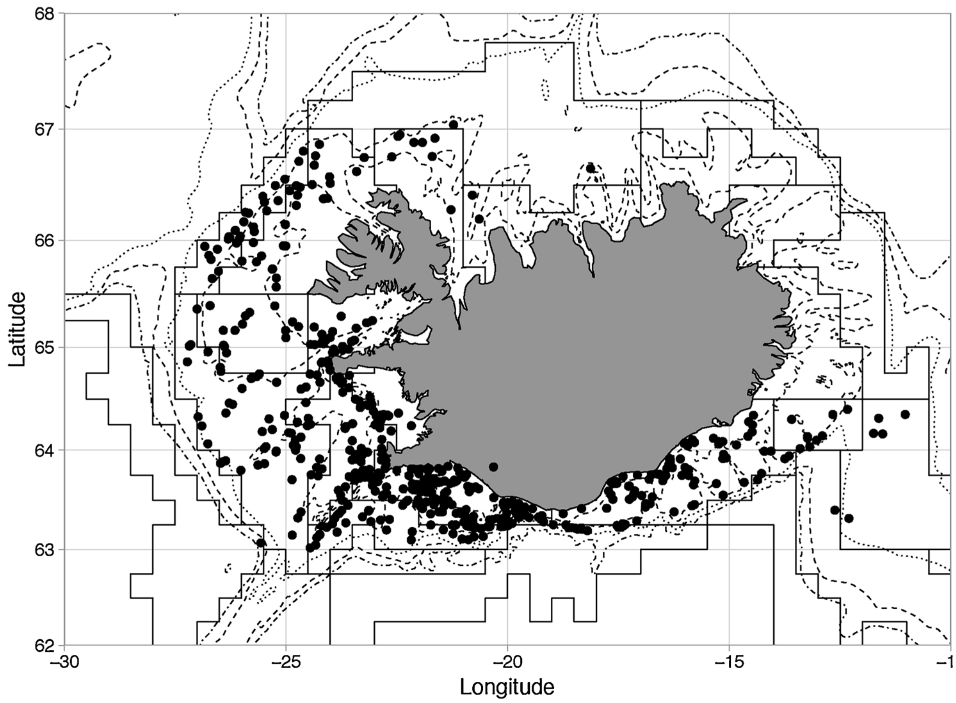

Fig. 3. Locations of ling catches (points) by commercial and survey fleets in 2015 relative to the spatial subdivision (solid lines) on the Icelandic continental shelf area. The dotted lines represent the depth contours.

mortality, but for some stocks, the reference points are determined in terms of harvest rate. These harvest rates relate the amount of catches to a reference biomass (e.g. spawning stock biomass [SSB], or biomass of fish larger than a minimum size or older than a minimum age [B_{ref}], which is B_{75cm+} for ling). For ling, a management plan was suggested in terms of the harvest rate rather than fishing mortality (see next section and ICES, 2017d); however, we present results in terms of both. All biomass reference points were calculated in terms of SSB, whereas B_{ref} was only used within the harvest control rule.

The effective annual fishing mortality at age and time step t was calculated according to the following equation:

$$F_{a,s,y,t} = \frac{-\log(1.0 - \frac{C_{a,s,y,t}}{N_{a,s,y,t}})}{\Delta t} \quad (15)$$

where $C_{a,s,y,t} = \Sigma_{f,l}C_{f,a,l,s,y,t}$ and $N_{a,s,y,t} = \Sigma_l N_{a,l,s,y,t}$. For ling, the reported F_y is the average $F_{a,y}$ for fully recruited ages ($a = 5+$).

Harvest rate in terms of the reference biomass is calculated as:

$$H_y = \frac{C_y}{B_{\text{ref},y}} \quad (16)$$

where $C_y = \Sigma_{f,a,l,s,t}C_{f,a,l,s,y,t}$ and $B_{ref,y} = \Sigma_{a,l,s,t}N_{a,l,s,y,t}W_{s,l}$.

Within the ICES advisory framework, biological reference points are used to guide the creation of harvest control rules and are used in other advisory protocols. These reference points include B_{lim}, B_{pa}, and the harvest rates or fishing mortalities that result in biomasses equal to these reference points at equilibrium (H_{lim} and H_{pa} or F_{lim} and F_{pa}, respectively). B_{lim} is defined as “a deterministic biomass limit below which a stock is considered to have reduced reproductive capacity” (ICES, 2017b, p. 1). B_{pa} is defined as “a stock status reference point above which the stock is considered to have full reproductive capacity, having accounted for estimation uncertainty” (ICES, 2017b, p. 1). ICES standard procedures (ICES, 2017b) were used to derive these quantities.

2.5. Evaluation of a harvest control rule

An initial set of simulations were used to select a candidate harvest rate that maximized yield with low risk of achieving low biomass levels (i.e., crossing B_{pa}), and to define other biological reference points defined in the previous section. This candidate harvest rate and reference points were then implemented within a harvest control rule in a second set of simulations used to evaluate its performance as a management strategy. In each set of simulations, the effect of uncertainty was evaluated by projecting all bootstrap data sets and comparing equilibrium status to reference points. Each set of simulations included (1) bootstrap sampling of the stock data to replicate observation error and estimation of model parameters based on that bootstrap replicate (i.e., a simulated stock assessment, described previously), (2) setting a TAC either (a) corresponding to a range of harvest rates (0–0.7, initial simulations), or (b) using a harvest control rule that implements a chosen candidate harvest rate with scaling of the harvest rate below a trigger biomass level (management strategy evaluation), (3) forward simulation of the stock based on estimated parameters and process error (described below), and (4) fishing by all fleets according to their estimated selectivities. Overall proportions of catches were allotted among fleets based on the last three years in the time series of actual catches. Each year, TACs were filled exactly.

Stock structure has been discussed in an ICES working group (ICES, 2007), which concluded that ling should be assessed as a single stock unit. However, the spatially based bootstrap approach employed here to some degree accounts for errors in the assumption that spatial structure does not exist by quantifying uncertainty based on spatial resampling. As a result, this uncertainty includes observation error alongside error generated from unknown spatial structure. Bias resulting from illegal landings and discards by Icelandic fishing vessels are considered to be negligible (MRI, 2013). The largest source of error outstanding was assumed to be process error, in particular variation in recruitment and assessment error.

Ling do not show a relationship between spawning stock and recruitment (ICES, 2017d), so recruitment in projections was drawn from the historical distribution using a block-bootstrap, with each block including a randomly drawn starting year and six consecutive years. The six-year block length was arbitrary but chosen to ensure some autocorrelation in recruitment could be captured despite having a relatively

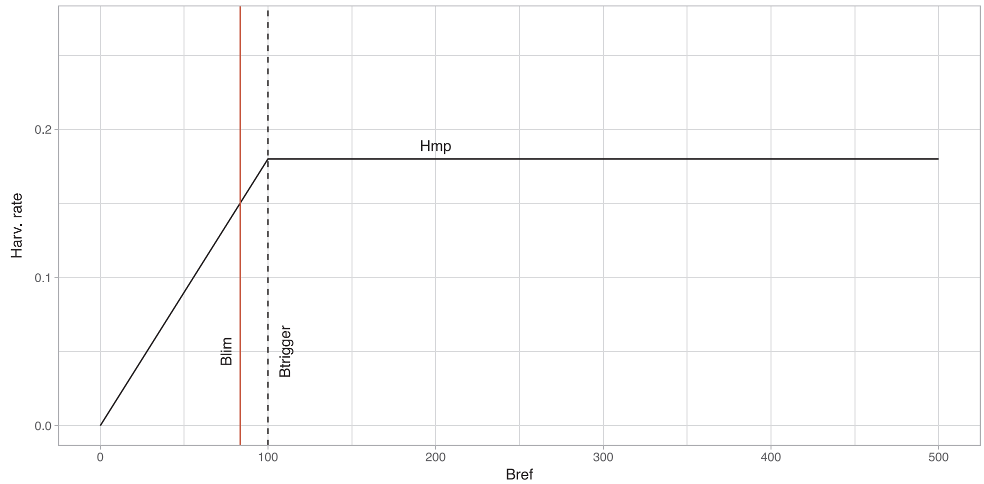

Fig. 4. Graphical presentation of the proposed management rule. The black solid line indicates the harvest rate as a function of the B_y^{ref}.

short recruitment time series from which random blocks could be generated. To account for autocorrelated corrections in retrospective plots, which is a commonly observed pattern believed to be caused by inconsistencies in survey indices (ICES, 2014b, 2015), autocorrelation was also introduced in assessment error. Assessment errors were emulated as:

$$\hat{B}_{\text{ref},y} = e^{E_y} B_{\text{ref},y} \tag{17}$$

where $B_{ref,y}$ is the reference biomass and $E_y = \sigma(\rho\epsilon_{y-1} + \sqrt{1-\rho^2}\,\epsilon_y)$ the assessment error. Observation error was reflected by σ in Eq. (17), whereas autocorrelation between assessment years was described by ρ and error $\epsilon_y \sim N(0, 1)$. Observation error in σ was set in forward projections by using the CV generated among bootstrap replicates in initial projections. The autocorrelation in assessment error ρ was set to 0.8, which was perceived as the upper limit to potential correlation (similar default values used by ICES are $F_{CV} = 0.212$ and autocorrelation $F_{phi} = 0.423$, see ICES, 2014b, 2015).

Formation of the harvest control rule follows standard ICES methodology. The evaluation framework can be classified as a simulation without assessment feedback, as it is assumed that the simulation within the operating model represents the true stock dynamics. The harvest control rule then allocates catches to the fleets using a simple scalar applied to the estimated reference biomass (Fig. 4):

$$\text{TAC}_{y+1} = \begin{cases} H\frac{\text{SSB}_y}{B_{\text{trigger}}}\hat{B}_{\text{ref},y} & \text{if } \text{SSB}_y < B_{\text{trigger}} \\ H\hat{B}_{\text{ref},y} & \text{if } \text{SSB}_y \geq B_{\text{trigger}} \end{cases} \tag{18}$$

In the model, as in practice, the reference biomass used to calculate the TAC is observed during the fishing year prior to when the TAC takes effect. Reference biomass is observed at the end of the first quarterly timestep of year y during the assessment procedure, but this TAC applies from the fourth timestep of year y through the first three timesteps of year $y + 1$. This two-quarter time lag results in a slight mismatch between expected and realized harvest rates, as reference biomass will change slightly over the time lag. Finally, fishing the entire TAC was implemented with portions of the TAC spread to ages and lengths according fleet suitabilities evenly across timesteps (Eq. (11)).

3. Results

3.1. Parameter estimates

Parameter searches were bounded, but final estimates rarely rested at the boundaries. Most distributions were spread around the base run value shown in Fig. 5. Two parameters related to growth l_0 and $\sigma_{a=3}$, recruitment length and length variance, were estimated at the boundary in a substantial portion of the bootstrap trials, and therefore effectively fixed. Growth dispersion for older fish (not including initial condition parameters) was determined by a combination of the β and n parameters, the latter of which is fixed and can be set low enough to be constraining if needed. Where n is set to a reasonable maximum, as in this case, it does not constrain growth estimation. The maximum length estimate L_∞ appeared to be effectively fixed as the bulk of the bootstrap replicates estimated it to be at or close to the length of the largest fish observed; however, it did not hit parameter boundaries. Although the other two parameters were ill-determined, they had little effect on the estimated biomass and fishing mortality. Estimates of recruitment and initial population parameters are shown in supplementary material (Fig. S2). In no instance did they hit the boundaries and the estimates showed a fairly symmetric spread. The greater spread of recruitment estimates in and after 2013 revealed large uncertainty, as is to be expected given little data on those cohorts that can help inform recruitment. The base run follows the median of the bootstrap estimates closely, indicating little discernable bias.

3.1.1. Abundance indices

For all survey index length groups the fit to the abundance indices was fairly good as shown in Fig. 6. The mode of catchability was centered around the 60–70 cm survey index length group and started to taper off for larger fish. For the larger length groups, the model tended to predict a lower increase in abundance than was observed. The assumption of a slope of 1 (i.e. a linear relationship between index and abundance) for the larger length groups appeared to be valid while combined with estimating the slope for the smallest two groups, as it improved the fit to the indices.

3.2. Compositional data and selectivity

Selected model fits to various compositional dataset are illustrated in supplementary material (Fig. S3). The model compares data to predictions at each time step if and when the data are available. In general, the model appeared to replicate well the observed size and maturity proportions. The estimated selectivity curves for model fleets are shown in supplementary material (Fig. S4). along with the median and 5–95% interquantile range from bootstrap runs. The longline and trawl fleet selectivities were not substantially different, with respective median values of l_{50} at 76.91 cm (range 72.93–78.28) and 75.37 cm (range 69.96–77.55). The l_{50} in the survey was estimated to be slightly lower, 68.28 cm, but with higher uncertainty (range 54.18–86.92 cm). The l_{50} of the gillnet fleet was considerably higher: 101.62 cm (97.97–101.62).

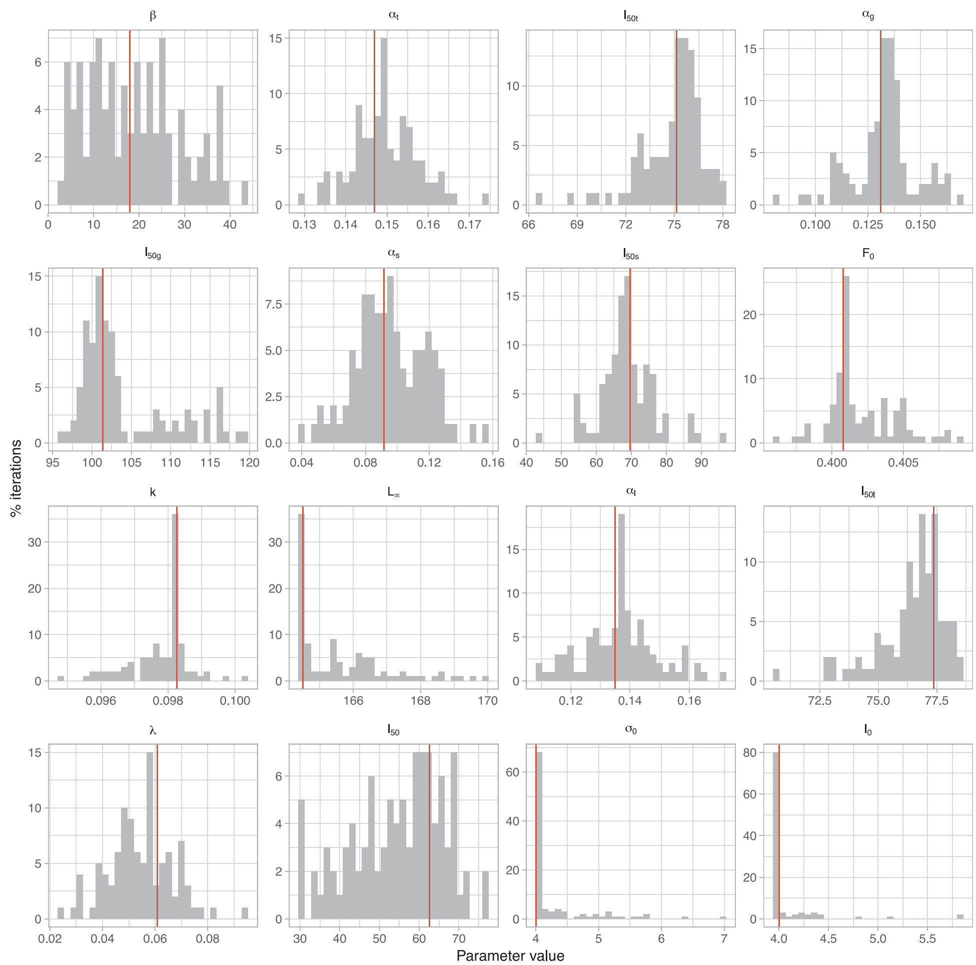

Fig. 5. Histogram of parameter estimates from 100 bootstrap samples. The red line indicates the estimate from the base run. Parameter descriptions are in Table 3. Note that the panel boundaries are set relative to the spread of the bootstrap estimates and do not illustrate parameter bounds. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

3.3. Stock status

The model predicted that total biomass was currently declining from its peak level in 2014. Similarly, reference biomass and the spawning stock biomass (SSB) was starting to decline but remained at a level considerably higher than previously observed (Fig. 7). The SSB reached its lowest point at 9.93 kt in 1991. The bootstrap runs indicated that uncertainty about population estimates has been increasing in recent years, as expected in light of the increase in the variance of the survey indices and fewer data points related to those cohorts. The coefficient of variation (CV) of the population estimates from the bootstrap runs were on average around 0.15 for SSB, 0.15 for reference biomass ($>$ 75 cm), and 0.16 for recruitment (Fig. 7). However, for the assessment year, the CVs of SSB and reference biomass were around 0.25 (ranging 0.1–0.26 historically) and 0.28 (historically ranging 0.09–0.28). From the model the fishery appears to mainly target the mature population. The total biomass caught of immature fish was estimated to have varied between 12.4% in 1982 to 26.5% in 2011 and in 2016, with a current estimate of 12.6%.

3.4. Derivation of biological reference points

Initial simulations validated the base run as unbiased, so derivation of the reference points proceeded from the base run. ICES technical guidelines ICES (2017b) indicate that B_{lim} should be chosen by examination of the SSB–Recruitment scatterplot (Fig. S5 in supplementary material). Ling's stock–recruitment relationship showed that ling has historically had a relatively narrow dynamic range of SSB and no evidence of impaired recruitment. ICES guidelines suggest that in this situation where a low dynamic range in SSB has been observed, B_{lim} cannot be estimated from these data and that the lowest observed SSB during that period (i.e. SSB(1992) = 9.93 kt from the baseline model) is an appropriate value at which to set B_{pa}. A proxy for B_{lim} could then be calculated based on the inverse of the standard conversion factor from B_{lim} to B_{pa}, $e^{\sigma * 1.645}$ where σ is a constant equal to 0.2 (a standard procedure within ICES, see (ICES, 2017b), for further details. Therefore, a proxy for B_{lim} was set at $B_{pa}/e^{1.645*0.2} = 9.93/1.4 = 7.09$ kt. The harvest rates H_{pa}, H_{lim} and H_{msy} were then set as the rates that result respectively in B_{pa}, B_{lim} and B_{msy} (biomass at which maximum sustainable yield is achieved at equilibrium).

There was a large spread in yield and SSB for different values of H around the median shown in Fig. 8, as a result of high variation in productivity. The first set of projections showed maximum of the median yield rested close to the harvest rate of 0.24 (see Fig. 8). The limit harvest rate, H_{lim}, resulting in 50% long–term probability of SSB $> B_{lim}$, was estimated at 0.56 (an equivalent F of 0.70). The precautionary harvest rate, H_{pa}, fell at a higher rate than H_{msy}, as H_{pa} is intended to reflect an upper limit of harvest rates that avoids a risk of the harvest rate surpassing H_{lim}, thereby risking a drop in stock biomass below B_{lim}.

The range of harvest rates considered for the harvest control rule lie at the flattest portion of the peak in yield, ranging 0.18–0.3 (Fig. 8). Projections of bootstrap runs from these harvest rates suggested high variability in biomass and catches in projections, mainly as a result of

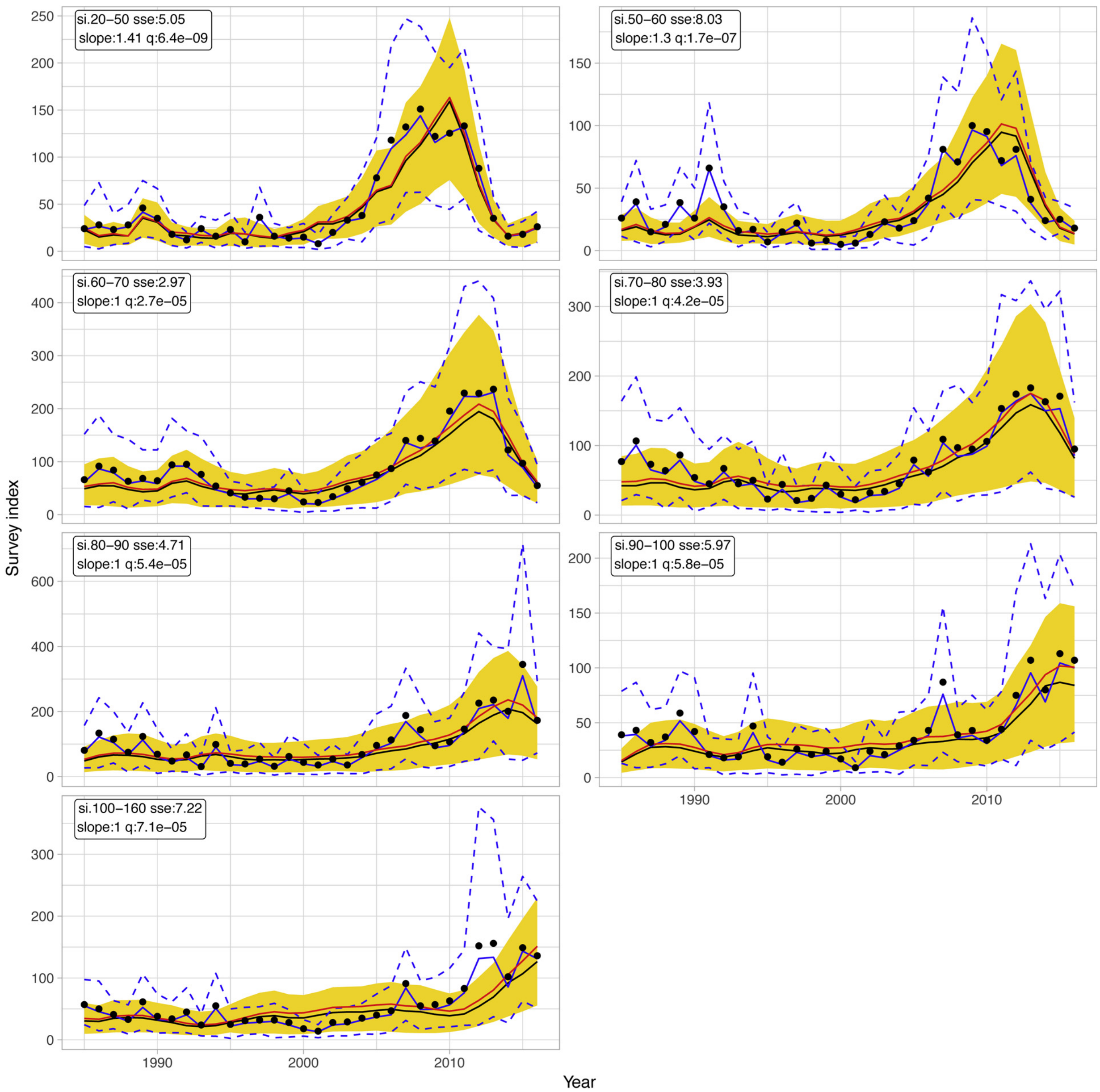

Fig. 6. Comparison of the distribution of bootstrap length-aggregated abundance index data with survey indices predicted from the simulated management procedure, as well as baseline model results. The red lines and yellow shaded areas represent the median and 5–95% interquantile range of the predicted indices resulting from simulations, respectively. The blue solid and dotted lines are the median and 5–95% interquantile range of the bootstrap data. The base model is represented by original index data (black points) and predicted indices from the base model (black line). Length ranges of the survey indices follow the 'si.' label, and sse indicates the standard error of the survey index. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

high recruitment variability (Fig. 9). Individual projections likewise showed high interannual variability, indicating consistency in the effect of high recruitment variability across bootstrap runs. Removing the effect of recent high recruitment (years 2010–2016) on recruitment projections to test a more conservative recruitment option had little effect (< 1%) on the estimated value of the harvest rate reference points in projections.

3.5. Evaluation of a harvest control rule

There was essentially no difference in catch levels taken within the range $H_{mp} = \{0.18, 0.3\}$, but there was a substantial decrease in equilibrium SSB as harvest rates increased. Therefore it was concluded that $H_{mp} = 0.18$ would be employed for management purposes, chosen as a precautionary level within the context of the recent drop in recruitment and SSB levels. The second set of simulations that implemented H_{mp} within the full HCR as a MSE revealed that SSB never dropped below $B_{trigger}$ in projections (likely due to ling's relatively high stock status in Iceland); therefore, results are the same as those obtained in the first set of simulations for $H_{mp} = 0.18$ (Fig. 9).

4. Discussion

The management strategy evaluation of ling in Icelandic waters presented here represents an empirical case study in which a size- and age-structured stock assessment was fitted with limited age data and subsequently used for development of a harvest control. Overall, the Gadget model presented here on ling in Icelandic waters captures the general trends in the data, and in spite of minor mis-fits the model has been used to present advice to managers (ICES, 2017d), which were taken in the form of the harvest control rule with $H_{mp} = 0.18$. Although ICES usually uses H_{msy} to set harvest rates, which would have been 0.24

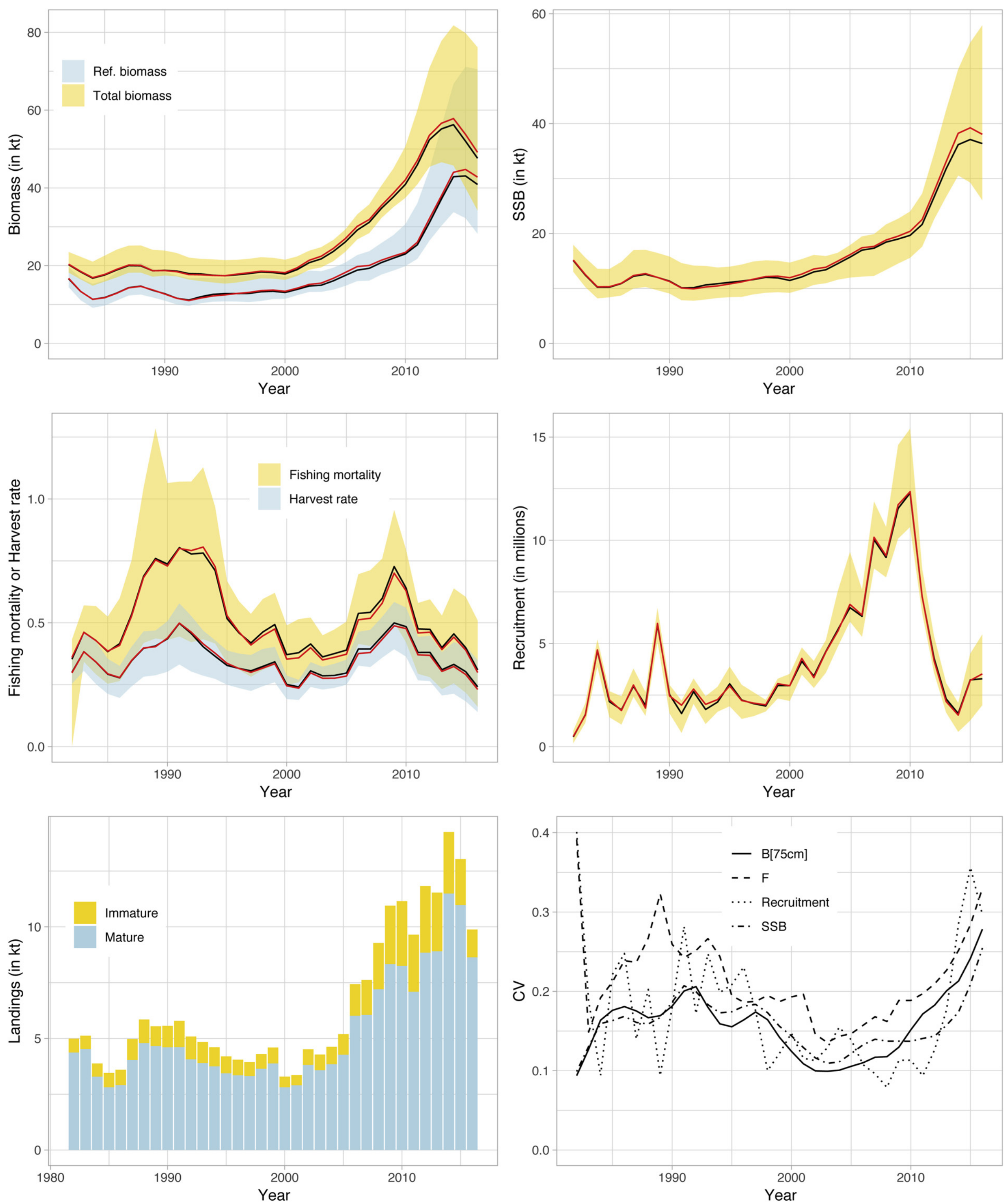

Fig. 7. Estimates of total and reference biomass, the latter including only ling larger than 75 cm (a), spawning stock biomass (b), harvest rate and fishing mortality (age 5⁺) (c), recruitment (d), immature and mature ling landings (e), and coefficients of variation (CVs) for key quantities (f). For a–d, the distribution of bootstrap estimates is represented by a black line (median) and yellow or blue shaded area (5–95% interquantile range). The red line shows estimates from the baseline run. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

in this case, a more conservative measure was instead chosen with stakeholder approval because the difference in resulting catch between the two levels was very little while the difference in spawning stock biomass was substantial, and recent years indicated that spawning stock biomass may be beginning to decrease. In addition, catch-quota balancing regulations implemented within Iceland allow for the possibility of surpassing TACs to a certain extent, yielding some additional risk of stock reductions (Woods et al., 2015). As a result, the more conservative harvest control rule was also deemed acceptable by the industry.

In complex integrated models that attempt to estimate many parameters using diverse data sets of varying quality, it is expected that there may be problems with estimability of some parameters and fit to some data-sets. The automatic iterative re-weighting procedure implemented in Gadget (Taylor et al., 2007; Elvarsson and Lentin, 2018) aids the consistency and effectiveness of finding a best-fit model (as

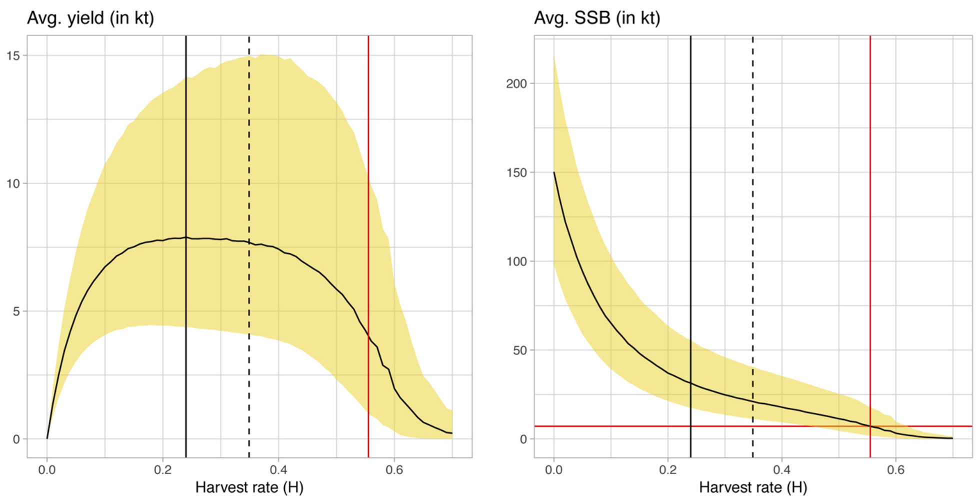

Fig. 8. Distribution of equilibrium densities of yield (left) and SSB (right) curves as a function of H, as estimated from bootstrap data sets. The black solid curves with yellow shaded regions indicate the median and 5–95% interquantile ranges, respectively. Red lines indicate H_{lim} (vertical) and B_{lim} (horizontal). Black vertical lines indicate H_{pa} (dashed) and H_{msy} (solid). (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

assessed visually) while giving the user some control over the process by enabling grouping of likelihood components. In the case of ling in Icelandic waters, the main difficulty inherent in the data was the rapid increase in the survey index values in recent years and the resulting large CV of these indices from that period. This problem, however is unlikely to change with the modeling framework, and given the decrease in the last five years of the smaller length groups in the data series, this may resolve itself in coming years.

Most parameters were well-defined except for some growth parameters, which had little influence on model results, and selectivities related to gillnets. Catches from gillnets have been decreasing in recent years, so data are limited from this fleet. For example, catches in gillnets were around one-third to one half of ling catches in 2000–2001, but have now decreased to around 7% of Icelandic commercial catches in 2016 due to changes in fleet dynamics and regulations. Therefore, it is likely that the gillnet selectivity model is misspecified due to unaccounted for changes through time, but these problems are of minor importance in the current fishery and this likely has little effect on model results.

With a lack of consistent and reliable age data, as was the case in this study, fitting an integrated size- and age-structured models was an appealing alternative to many age-based methods, which could have required extrapolations of age structure from the survey to commercial fleets, as well as back in time to periods lacking sufficient age data. Following such procedures could introduce bias or autocorrelation into model estimates by supplying erroneous information regarding somatic growth. Meanwhile, using simpler production-based methods may prevent the introduction of bias related to somatic growth processes, but may not be as sensitive to stock dynamics (Magnusson and Hilborn, 2007) and would disregard the valuable information on stock dynamics contained within compositional data (Wetzel and Punt, 2011; Ono et al., 2015). Species that routinely lack age data, such as invertebrate species, are often assessed using size-structured models, similar to the one presented here, sometimes with tag-recapture data informing the time-scale of population dynamics (Punt et al., 2013, 2016). However, recent studies have suggested that the utility of size-structured models is not limited to species with no age data. For example, a simulation study using a data-rich scenario indicated that size- and age-structured models tended to outperform age-structured models in terms of model fit and relative error of management reference points. It was concluded that this likely stemmed from a greater influence of the population dynamics mis-specification (i.e., age-based versus size-based dynamics) rather than mis-specification of growth, which was evident in the size-based model (Punt et al., 2017). Within an integrated age-structured modeling framework, even limited amounts of additional length-frequency data can notably improve the assessment, as Wetzel and Punt (2011) observed in simulations. Furthermore, utilizing the greatest resolution available to discretize size data improves model performance, even in age-structured models (Monnahan et al., 2016; Szuwalski, 2016). Given the technical constraints in discretizing age data to a finer resolution than annual rings, it would not be surprising if size-structured models are able to outperform age-structured models simply based on their ability to track population dynamics at a higher resolution, not to mention the greater availability of data with which to parameterize such models, given the relative ease of collecting size data in comparison to age data. There is however a balance to be sought as higher resolution scale may introduce unwanted noise (Vandermeer, 1978) and it may increase computing time considerably.

Nonetheless, at present size- and age-structured models are rarely used for management advice. Size data alone do not yield a time scale onto which population dynamics can be mapped. Tracking both age and size in a population dynamics model can be highly computationally demanding (Punt et al., 2016), and size-structured stock assessment modeling packages are few and far between (but see CASAL2 Doonan et al., 2016, for an exception). Integrated age-structured methods that incorporate length frequency data, such as Stock Synthesis (Methot, 2013) and MULTIFAN–CL (Methot, 2013), work sufficiently well for many of stock assessment needs at hand. Scientists also often use similar packages regionally, due to "historical and current culture of stock assessment practice" (sensu Dichmont et al., 2016, p. 449). This regional effect of stock assessment culture undoubtedly results in part from the investment necessary to learn and/or teach on-site colleagues how to use individual stock assessment models, many of which are highly technical to run and diagnose, while sufficient numbers of qualified analysts necessary for the assessment and review process lag behind (Dichmont et al., 2016; Maunder and Piner, 2015). Therefore, the utility of size- and age-structured models in contemporary stock assessment is highly limited by the lack of widely used size- and age-structured modeling packages (Dichmont et al., 2016). Furthermore, as integrated models are increasingly developed to incorporate more complex structures and diverse data sets, there likewise increases the opportunity for data conflicts and hence a need for diagnostic procedures (Maunder and Piner, 2017; Carvalho et al., 2017).

The stock assessment model used in this study to investigate the effects of the various management decisions on the population dynamics of ling in Icelandic waters is no exception to these rules, but has some significant advantages as well. As with other integrated assessment methods, some advantages include that diverse data sources can be used to inform the same processes (e.g., size-composition and tagging data to inform growth), multiple processes are used to form predictions to which data are compared (e.g., both selectivity and growth parameterizations affect size compositions), uncertainty can be more

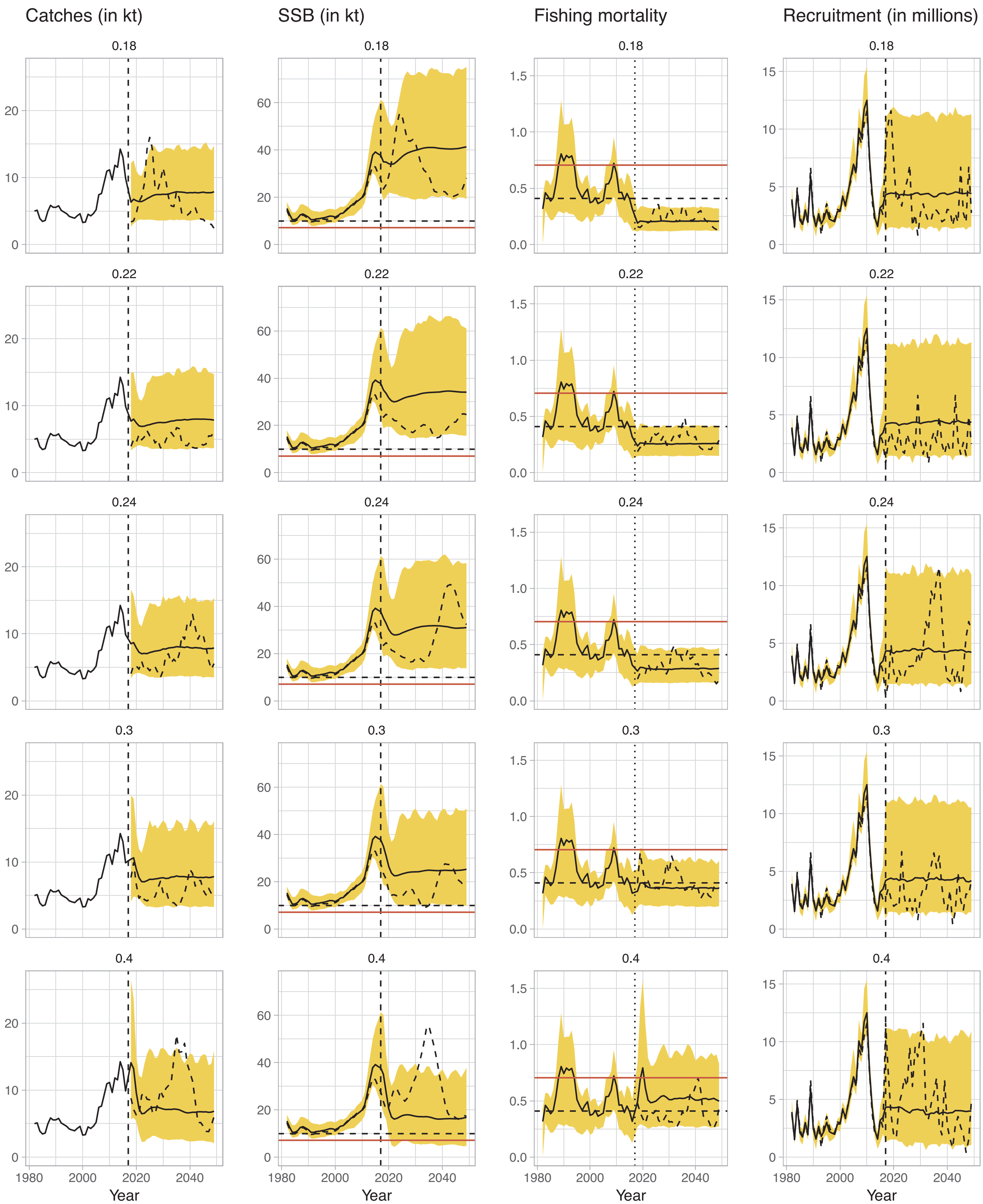

Fig. 9. Projected catches, spawning stock biomasses, and fishing mortalities resulting from select target harvest rates tested in initial simulations. The first panel row also represents results obtained in the MSE (second set of simulations), as 0.18 was implemented as the harvest rate in the HCR and SSB never dropped below $B_{trigger}$ in projections. The black solid lines and yellow shaded regions represent the median and 5–95% interquantile ranges, respectively, of estimates obtained from bootstrap replicates. Two examples of these replicates are shown as line trajectories to show variability among runs (solid vs. dashed). Projection years begin with the vertical dotted line. Horizontal red and black dashed lines represent the limit and pa reference points respectively for SSB and F. (For interpretation of the references to color in this legend, the reader is referred to the web version of the article.)

accurately modeled and propagated throughout the modeling process, and the effects of assumptions or data sets on model results are more easily compared within a single framework (Maunder and Punt, 2013). Gadget can use various types of data comparisons that can be included in the objective function, including size distributions, age-length keys or distributions, survey indices by length or age (as either abundance or biomass), catch per unit effort data, mean length and/or weight at age, tagging data and stomach content data. Some implemented options for functional relationships, such as maturation, include dependencies on both age and size, thereby allowing for dependency on only age or size by setting some parameters to 0 (thereby apparently setting it apart from CASAL2 Doonan et al. (2016), for example). Growth and selectivity are both size-based, thereby avoiding the difficulties of fitting age-structured models to size composition data when the growth curve includes a dependency only on age (Francis, 2016). Importantly, this ability to handle length data directly means that age data are not a requirement: they can be sparse or nonexistent. However, to obtain a well-specified model, some information regarding the time scale of somatic growth is desirable (e.g., tagging data). As in other integrated approaches, a maximum likelihood approach can be used to find the best fit to a weighted sum of the data sets (Begley and Howell, 2004).

Because most processes estimated within Gadget are size-based (e.g., except aging and natural mortality), the model naturally reduces to a size-structured model, in which no age data are necessary for fitting the model. The same cannot be said for implementing an age-structured model within Gadget, as growth and selectivity are currently not implementable as age-based. However, an additional feature of Gadget drastically increases its flexibility, which is an ability to construct multiple "stocks" (with migration among them) and specifying any parameters as stock-specific and time-varying (or not). Such flexibility not only allows for more intuitive parameterizations of stock components, such as differences between spatial segments, males and females, or mature versus immature (the last of which was implemented in this study), but could also allow for less common implementations, such as the addition of growth platoons or cohort structure. For example, although Gadget is essentially a size-based model, all size data could be removed from the model by implementing age-based 'stocks' that migrate annually and unidirectionally, fixing length-based growth parameters (which would have no bearing on the fitting procedure) and implementing constant (or even time-varying) selectivity for each 'stock'. As far as the authors are aware, this level of flexibility within a package is otherwise only available in CASAL2 (Doonan et al., 2016). In addition, because Gadget historically developed as an ecosystem simulation framework (Stefánsson and Pálsson, 1997), it has additional functionality with multi-species linkages via consumption models.

However, like many other implementations (Maunder and Punt, 2013; Dichmont et al., 2016), Gadget has limitations in its current implementation. For example, it does not currently support models with random effects, which would be useful when estimating a stock-–recruitment relationship for a stock, it has a relatively long optimisation time (Maunder and Punt, 2013, as it is not based on automatic differentiation methods), and efforts at improving the accessibility of Gadget by streaming model input, diagnostics, and simulation via R statistical framework (R Core Team, 2017) are underway but not yet completed (Elvarsson and Lentin, 2018). Due to issues related to the properties of length distribution data discussed by Hrafnkelsson and Stefánsson (2004) and Babak et al. (2007), model comparisons using Aikaike information criteria (as suggest by Punt et al., 2017) are not currently considered feasible. Nonetheless, Gadget has been used as a basis for management in a variety of cases, including tusk (*Brosme brosme*) in Icelandic waters and hake (*Merluccius merluccius*) west of the Iberian peninsula. If Gadget is to be used more widely, further validation of its structural properties through simulation studies are desirable, similar to the series that has developed through the wide usage of Stock Synthesis (Methot, 2013), as listed by Dichmont et al. (2016).

Software initiatives that enable reproducible research have in recent year become extremely popular, both with on-line hosting that allow easy version control such as Github and Bitbucket, and literate programming tools like knitr (Xie, 2014) and rmarkdown (Allaire et al., 2017). In the spirit of these initiatives MFDB and Rgadget were developed which simplify the modelling process by abstracting complicated data aggregation and model settings away from the user. While these packages are in constant development they are considered to be at a stage useful for the general modeler. In the development of these packages the developers have attempted to follow recent trends in with the R community collectively referred to as "tidyverse" (Wickham, 2017) by allowing the modeler to build a model in a sequence of simple stages, each defining an attribute to the model. The authors hope that these tools will encourage further model developments using the Gadget and other size- and age-structured frameworks.

Acknowledgements

We would like to thank Carmen Fernandez former vice chair of ICES ACOM and Alfonso Perez Rodriguez researcher at Wageningen University who reviewed this work at an earlier stage.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 613571 – MareFrame.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fishres.2018.06.005.

References

Akselrud, C.I.A., Punt, A.E., Cronin-Fine, L., 2017. Exploring model structure uncertainty using a general stock assessment framework: the case of Pacific cod in the eastern Bering sea. Fish. Res. 193, 104–120.
Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., 2017. rmarkdown: Dynamic Documents for R. R Package Version 1.8.
Anon, 2002a. Council Regulation (EC) No. 2371/2002 of 20 December 2002 on the Conservation and Sustainable Exploitation of Fisheries Resources Under the Common Fisheries Policy. Technical Report. Council of the European Communities, Brussels.
Anon, 2002b. Communication from the Commission on the Reform of the Common Fisheries Policy: Roadmap. Technical Report. Commission of the European Communities, Brussels.
Babak, O., Hrafnkelsson, B., Palsson, O.P., 2007. Estimation of the length distribution of marine populations in the Gaussian-multinomial setting using the method of moments. J. Appl. Stat. 34 (8), 985–991.
Baldursson, F., Danielsson, A., Stefansson, G., 1996. On the rational utilization of the Icelandic cod stock. ICES J. Mar. Sci. J. Conseil 53 (4), 643.
Begley, J., Howell, D., 2004. An Overview of Gadget, The Globally Applicable Area-Disaggregated General Ecosystem Toolbox, ICES CM. pp. 1–15.
Begley, J., 2005. Gadget User Guide. Marine Research Institute, Reykjavík, Iceland. 90 Marine Research Institute Report 120.
Berkson, J., Thorson, J.T., 2014. The determination of data-poor catch limits in the United States: is there a better way? ICES J. Mar. Sci. 72 (1), 237–242.
Bertignac, M., De Pontual, H., 2007. Consequences of bias in age estimation on assessment of the northern stock of European hake (*Merluccius merluccius*) and on management advice. ICES J. Mar. Sci. 64 (5), 981–988.
Bogstad, B., Hauge, K., Ulltang, Ø., 1997. MULTSPEC – a multi-species model for fish and marine mammals in the Barents Sea. J. Northw. Atl. Fish. Sci. 22, 317–341.
Butterworth, D., Punt, A., 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. J. Conseil 56 (6), 985–998.
Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J., 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. 153, 48–68.
Carvalho, F., Punt, A.E., Chang, Y.-J., Maunder, M.N., Piner, K.R., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fish. Res. 192, 28–40.
Dichmont, C.M., Deng, R.A., Punt, A.E., Brodziak, J., Chang, Y.-J., Cope, J.M., Ianelli, J.N., Legault, C.M., Methot Jr., R.D., Porch, C.E., et al., 2016. A review of stock assessment packages in the United States. Fish. Res. 183, 447–460.
Doonan, I., Large, K., Dunn, A., Rasmussen, S., Marsh, C., Mormede, S., 2016. Casal2: New Zealand integrated population modelling tool. Fish. Res. 183, 498–505.
Elvarsson, B.T., Lentin, J., 2018. Rgadget: R Package for Gadget Models. R Package Version 0.5.
Elvarsson, B., Taylor, L., Trenkel, V., Kupca, V., Stefansson, G., 2014. A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets. Afr. J. Mar. Sci. 36 (1), 99–110.
Fournier, D.A., Hampton, J., Sibert, J.R., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to south pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55 (9), 2105–2116.
Francis, R.C., 2016. Growth in age-structured stock assessment models. Fish. Res. 180, 77–86.
Gavaris, S., 1988. An adaptive framework for the estimation of population size. CAFSAC Res. Doc 88 (29), 12.
Hüssy, K., Radtke, K., Plikshs, M., Oeberst, R., Baranova, T., Krumme, U., Sjöberg, R., Walther, Y., Mosegaard, H., 2016. Challenging ICES age estimation protocols: lessons learned from the eastern Baltic cod stock. ICES J. Mar. Sci. 73 (9), 2138–2149.
Henríquez, V., Licandeo, R., Cubillos, L.A., Cox, S.P., 2016. Interactions between ageing error and selectivity in statistical catch-at-age models: simulations and implications for assessment of the Chilean Patagonian Tooth fish fishery. ICES J. Mar. Sci. 73 (4),

1074–1090.
Hrafnkelsson, B., Stefánsson, G., 2004. A model for categorical length data from groundfish surveys. Can. J. Fish. Aquat. Sci. 61 (7), 1135–1142.
ICES, 2007. Report of the Working Group on the Biology and Assessment of Deep-Sea Fisheries Resources (WGDEEP), 8–15 May, 2007, Copenhagen, Denmark. ICES CM 2007/ACFM:20. Technical Report. International Council for the Exploration of the Seas.
ICES, 2012. Report of the Workshop on the Development of Assessments Based on Life History Traits and Exploitation Characteristics (WKLIFE), 13–17 February 2012, Lisbon, Portugal. ICES CM 2012/ACOM:36. Technical Report. International Council for the Exploration of the Seas.
ICES, 2013a. General Context of ICES Advice. Technical Report. ICES.
ICES, 2013b. Report of the Workshop on the on Guidelines for Management Strategy Evaluations (WKGMSE), 21–23 January 2013, Copenhagen, Denmark. ICES CM 2013/ACOM:39. Technical Report. International Council for the Exploration of the Seas.
ICES, 2014a. Report of the Working Group on the Biology and Assessment of Deep-Sea Fisheries Resources (WGDEEP), 4-11 April, 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:17. Technical Report. International Council for the Exploration of the Seas.
ICES, 2014b. Report of the Joint ICES Myfish Workshop to Consider the Basis for FMSY Ranges for All Stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. Technical Report. International Council for the Exploration of the Seas.
ICES, 2015. Report of the Workshop to Consider FMSY Ranges for Stocks in ICES Categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. Technical Report. International Council for the Exploration of the Seas.
ICES, 2017a. ICES. 2017. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 19–26 April 2017, Copenhagen, Denmark. Technical Report. International Council for the Exploration of the Seas.
ICES, 2017b. 12.4.3.1 Technical Guidelines – ICES Fisheries Management Reference Points for Category 1 and 2 Stocks. Technical Report. International Council for the Exploration of the Seas.
ICES, 2017c. Report of the North Western Working Group (NWWG), 27 April–4 May 2017, 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:08. Technical Report. International Council for the Exploration of the Seas.
ICES, 2017d. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. Technical Report. International Council for the Exploration of the Seas.
Jobling, M., 2002. Environmental factors and rates of development and growth. Handb. Fish Biol. Fish. 1, 97–122.
Koenigs, R.P., Bruch, R.M., Kamke, K.K., 2013. Impacts of aging error on walleye management in the Winnebago system. N. Am. J. Fish. Manag. 33 (5), 900–908.
Kupca, V., 2006. Dst^2 Data Warehouse. Technical Report. Marine Research Institute, Reykjavík, Iceland.
Lentin, J., 2014. MFDB: MareFrame DB Querying Library. R Package Version 6.0-0.
Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish Fish. 8 (4), 337–358.
Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72 (1), 7–18.
Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fish. Res. 192, 16–27.
Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.
Methot Jr., R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.
MFRI, 2017. Ástand nytjastofna sjávar og ráðgjöf (e. State of Marine Stocks and Advice), 2017, Reykjavik, Iceland. Technical Report. Marine and Freshwater Research Institute, Iceland.
Monnahan, C.C., Ono, K., Anderson, S.C., Rudd, M.B., Hicks, A.C., Hurtado-Ferro, F., Johnson, K.F., Kuriyama, P.T., Licandeo, R.R., Stawitz, C.C., et al., 2016. The effect of length bin width on growth estimation in integrated age-structured stock assessments. Fish. Res. 180, 103–112.
MRI, 2013. Mælingar á brottkasti Þorsks og ýsu (e. Measurements of Discards of Cod and Haddock), 2013, Reykjavik, Iceland. Technical Report. Marine Research Institute, Iceland.
Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., et al., 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. 72 (1), 31–43.
Pálsson, Ó.K., Steinarsson, B.Æ, Jónsson, E., Guðmundsson, G., Jénsson, G., Stefánsson, G., Björnsson, H., Schopka, S.A., 1997. Icelandic Groundfish Survey. ICES CM 1997/Y:29.
Parliament, E., 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive). Off. J. Eur. Union L. 164, 19–40.
Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. Fish Fish 18 (2), 226–243.
Pella, J.J., Tomlinson, P.K., 1969. A Generalized Stock Production Model. Inter-American Tropical Tuna Commission.
Punt, A.E., Huang, T., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70 (1), 16–33.
Punt, A.E., Haddon, M., McGarvey, R., 2016. Estimating growth within size-structured fishery stock assessments: what is the state of the art and what does the future look like? Fish. Res. 180, 147–160.
Punt, A.E., Akselrud, C.A., Cronin-Fine, L., 2017. The effects of applying mis-specified age- and size-structured models. Fish. Res. 188, 58–73.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J.D., Conser, R.J., 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fish Bull. 109 (2), 217–232.
Reeves, S.A., 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. ICES J. Mar. Sci. 60 (2), 314–328.
Shepherd, J., 1999. Extended survivors analysis: an improved method for the analysis of catch-at-age data and abundance indices. ICES J. Mar. Sci.: J. Conseil 56 (5), 584–591.
Stefánsson, G., Pálsson, Ó., 1997. BORMICON. Boreal Migration and Consumption Model. Marine Research Institute Report, vol. 58 223 pp.
Stefánsson, G., 2003. Issues in multispecies models. Nat. Resour. Model. 16 (4), 415–437.
Stefánsson, G., 2005. Growth Models in Population Simulations. Final Report: dst^2: Development of Structurally Detailed Statistically Testable Models of Marine Populations, Volume 118 of Technical Report. Marine Research Institute, Reykjavik, pp. 73–83.
Stefansson, G., Palsson, O., 1998. A framework for multispecies modelling of Arcto-boreal systems. Rev. Fish Biol. Fish. 8 (1), 101–104.
Szuwalski, C., 2016. Biases in biomass estimates: the effect of bin width in size-structured stock assessment methods. Fish. Res. 180, 169–176.
Taylor, I.G., Methot Jr., R.D., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. 142, 75–85.
Taylor, L., Begley, J., Kupca, V., Stefansson, G., 2007. A simple implementation of the statistical modelling framework Gadget for cod in Icelandic waters. Afr. J. Mar. Sci. 29 (2), 223–245.
Treble, M.A., Campana, S.E., Wastle, R.J., Jones, C.M., Boje, J., 2008. Growth analysis and age validation of a deepwater arctic fish, the Greenland halibut (*Reinhardtius hippoglossoides*). Can. J. Fish. Aquat. Sci. 65 (6), 1047–1059.
UN, 2002. Johannesburg Declaration on Sustainable Development. Report. United Nations.
Vandermeer, J., 1978. Choosing category size in a stage projection matrix. Oecologia 32 (1), 79–84.
Wetzel, C.R., Punt, A.E., 2011. Performance of a fisheries catch-at-age model (stock synthesis) in data-limited situations. Mar. Freshw. Res. 62 (8), 927–936.
Wickham, H., 2017. Tidyverse: Easily Install and Load the 'Tidyverse'. R Package Version 1.2.1.
Woods, P.J., Bouchard, C., Holland, D.S., Punt, A.E., Marteinsdóttir, G., 2015. Catch-quota balancing mechanisms in the Icelandic multi-species demersal fishery: are all species equal? Mar. Policy 55, 1–10.
Xie, Y., 2014. knitr: a comprehensive tool for reproducible research in R. In: Stodden, V., Leisch, F., Peng, R.D. (Eds.), Implementing Reproducible Computational Research. Chapman and Hall/CRC ISBN 978-1466561595.
Yule, D.L., Stockwell, J.D., Black, J.A., Cullis, K.I., Cholwek, G.A., Myers, J.T., 2008. How systematic age underestimation can impede understanding of fish population dynamics: lessons learned from a lake superior cisco stock. Trans. Am. Fish. Soc. 137 (2), 481–495.