The specification of the data model part in the SAM model matters

M. Aldrin[a,*], I.F. Tvete[a], S. Aanes[a], S. Subbey[b]

[a] *Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway*
[b] *Institute of Marine Research, P.O. Box 1870 Nordnes, N-5817 Bergen, Norway*

ARTICLE INFO

ABSTRACT

This paper considers a general state-space stock assessment modeling framework that integrates a population model for a fish stock and a data model. This way observed data are linked to unobserved quantities in the population model. Using this framework, we suggest two modifications to improve accuracy in results obtained from the stock assessment model SAM and similar models. The first suggestion is to interpret the "process error" in these models as stochastic variation in natural mortality, and therefore include it in the data model. The second suggestion is to consider the observed catch as unbiased estimates of the true catch and modify the observation error accordingly. We demonstrate the efficacy of these modifications using empirical data from 14 fish stocks. Our results indicate that the modifications lead to improved fits to data and prediction performance, as well as reduced prediction bias.

1. Introduction

State-space models are a popular tool for fish stock assessments (Gudmundsson, 1994; Aanes et al., 2007; Nielsen and Berg, 2014; Cadigan, 2016; Miller et al., 2016; Aeberhard et al., 2018). Most of the many variants consist of a population model for the fish stock and a data model that links observations, such as catch data and survey indices, to unobserved quantities in the population model. One such is SAM (state-space assessment model, Nielsen and Berg, 2014; Berg and Nielsen, 2016). This model, with many useful extensions, is implemented in the flexible R package SAM (https://github.com/fishfollower/SAM) within the Template Model Builder (TMB) framework (Kristensen et al., 2016), and it is used by many ICES working groups to assess fish stocks (e.g., ICES, 2017b,a, 2018).

Aldrin et al. (2019) presented the following comments about SAM:

1. The data model is in our opinion mis-specified, because the so called "process error" is ignored when the catch data are linked to the true, unknown catch in the population model by the catch equation.
2. The model formulation implies the assumption that the catch data are biased upwards compared to the true catch.
3. The expected fishing mortality rate increases by a small, yearly factor.

In their response, Nielsen and Berg (2019) stated that the difference in formulations of the catch equation simply reflects two different model choices, and which model to choose should be based on comparing model fit and prediction ability. Regarding our second concern, the authors stated that SAM assumes that the catch data are median unbiased. We agree with the authors' reply that the expected increasing fishing mortality rate was negligible in the short term. Given that the logarithm of the fishing mortality is modelled as a random walk, we recognise that our previous suggested modification has an unfortunate side effect, and this is excluded from further discussion in this manuscript.

We will therefore address the two first issues, which are not SAM specific, but are related to other assessment models as well (e.g., Miller et al., 2016). We suggest modifications, which will potentially (i) improve fit, (ii) reduce bias and improve prediction performance for catch data and survey indices, and (iii) increase estimated stock size, compared to the present data model in SAM, if the population model is the same. In addition, we believe that assessment models cannot be evaluated solely by their ability to fit or predict observed data well. Their main purpose is to estimate some unobservable quantities, including present and previous stock sizes. Therefore, it is essential to establish clear and well-defined links between the observed data and the unobservable quantities. This is possible due to the mechanistic structure of SAM and similar models.

We present a rather general stock assessment model, and relate this to SAM (Section 2). In a study on 14 fish stocks, we examine the implications of the suggested modifications in terms of the log likelihood and stock size estimate (Section 3). Furthermore, we perform a cross

---

* Corresponding author.
  E-mail address: magne.aldrin@nr.no (M. Aldrin).

Table 1
Overview of notation. For simplicity, the term "of age a" is dropped from the interpretation text, except for the first line.

Non-observable, true quantities	
Notation	Interpretation
$N_{a,y}$	True number of fish of age a at the start of year y
$C_{a,y}$	True number of fish caught during year y; fishing mortality
$D_{a,y}$	True number of fish dying from other causes during year y, natural mortality
$T_{a,y} = C_{a,y} + D_{a,y}$	True number of fish dying from all causes during year y, total mortality
$F_{a,y}$	Instantaneous fishing mortality rate during year y corresponding to $C_{a,y}$
$M_{a,y}$	Instantaneous natural mortality rate during year y corresponding to $D_{a,y}$
$Z_{a,y} = F_{a,y} + M_{a,y}$	Instantaneous total mortality rate during year y corresponding to $T_{a,y}$
$N_{a,y}(d)$	True number of fish at the end of day d of year y
$I_{a,y}(d)$	"True" survey index proportional to $N_{a,y}(d)$ Expectations from a perfect survey with full coverage

Data, observations or "preliminary" estimates, to be updated when the assessment model is estimated	
Notation	Interpretation
$\widehat{C}_{a,y}$	"Preliminary" estimate of $C_{a,y}$
$\widehat{I}_{a,y}(d)$	Observed survey index, "preliminary" estimate of $I_{a,y}(d)$
$M_{a,y}^*$	Estimate or best guess of $M_{a,y}$

validation study, predicting the catch data and survey indices (Section 3.2).

2. Models

We first present the data for analysis, followed by definition of a general stock assessment model consisting of two sub models; a population model for the fish stock and a data model that links observed data to the population model. We then describe SAM. Table 1 gives an overview of relevant variables.

2.1. Data

We consider two types of data, estimates of yearly age-specific catches and observed age-specific survey index data. Let $\widehat{C}_{a,y}$ denote an estimate of the true, but unknown, number of fish $C_{a,y}$ of age a caught during year y. Let $\widehat{I}_{a,y}(d)$ denote an observed survey index for fish of age a for a survey conducted at the end of day d of year y. It is convenient to think of a hypothetical, "true" survey index, resulting from surveys with full coverage, where the observed index estimates the true value $I_{a,y}(d)$. We assume $I_{a,y}(d)$ to be proportional to the true number of fish $N_{a,y}(d)$ of day d, defined as

$$I_{a,y}(d) = Q_a N_{a,y}(d), \tag{1}$$

where Q_a is an unknown, time-independent, proportionality constant called catchability, to be estimated. Note that, if $Q_{a+1} = Q_a$, the true index at the start of the first day in a year is exactly the same as it was at the previous year, i.e.

$$I_{a+1,y+1}(0) = I_{a,y}(365). \tag{2}$$

2.2. A general stock assessment model

The model is essentially a state space model, where the population model constitutes the state or system equations and the data model constitutes the measurement equations. This is also a hierarchical model, with sub models to be specified for certain population and data

model parts.

2.2.1. The population model

From one year to the next, fish in a cohort may be caught, die from other causes or survive. This defines the naive population model

$$N_{a+1,y+1} = N_{a,y} - T_{a,y} = N_{a,y} - C_{a,y} - D_{a,y}, \tag{3}$$

where $N_{a,y}$ and $N_{a+1,y+1}$ are respectively, the number of fish in a cohort at the start of year y, and the following year, and $C_{a,y}$ is the catch during year y, whereas $D_{a,y}$ is the number of fish that died from other causes, called natural mortality. Finally, $T_{a,y} = C_{a,y} + D_{a,y}$ is the total number of fish that died during year y. In a closed population (no migration), $D_{a,y} \geq 0$. We allow for migration, i.e. an unrestricted $D_{a,y}$. Then, the interpretation of $D_{a,y}$ and $T_{a,y}$ change, but the equations remain unaltered. For simplicity we use the term natural mortality for $D_{a,y}$ and total mortality for $T_{a,y}$. Eq. (3), with recruitment specified and possibly adjusted for a plus group, is sufficient if we only consider catch data.

For given values of $N_{a+1,y+1}$, $N_{a,y}$, $C_{a,y}$ and $D_{a,y}$, there exist unique corresponding instantaneous mortality rates, $F_{a,y}$, $M_{a,y}$ and $Z_{a,y} = F_{a,y} + M_{a,y}$ for fishing, natural, and total mortality, respectively (see Appendix A). We interpret this so that fishing and natural mortality occur simultaneously with constant rates throughout a year. While this, for the moment, is not a necessary assumption, it facilitates writing the model in a more familiar way.

Replacing the numbers of fish that died, $C_{a,y}$, $D_{a,y}$ and $T_{a,y}$, by their corresponding mortality rates, Eq. (3) can equivalently be written as

$$N_{a+1,y+1} = \exp(-(F_{a,y} + M_{a,y}))N_{a,y} = \exp(-Z_{a,y})N_{a,y}. \tag{4}$$

The numbers of fish that died is uniquely given by the mortality rates and the stock size at the start of the year by the following equations:

$$T_{a,y} = N_{a,y} - N_{a+1,y+1} = (1 - \exp(-Z_{a,y}))N_{a,y}, \tag{5}$$

$$C_{a,y} = (F_{a,y}/Z_{a,y})T_{a,y}, \tag{6}$$

$$D_{a,y} = T_{a,y} - C_{a,y} = (M_{a,y}/Z_{a,y})T_{a,y}. \tag{7}$$

Eq. (6) is the so called Baranov equation (e.g., Quinn and Deriso, 1999). This equation is also valid when $Z_{a,y} < 0$, i.e. when immigration is larger than the total mortality. Furthermore, in the limit when $Z_{a,y} = 0$, then $C_{a,y} = F_{a,y}N_{a,y}$.

So far, we have made no model assumptions. Eqs. (3)–(7) are exact, except that the conversion from rates to numbers of fish ignores that these numbers should be integers. We must specify the model for the age plus group (see Appendix A), and we need stochastic sub models for recruitment, fishing and natural mortality rates. There are many sensible choices for these sub models. To enable a direct comparison with SAM, we choose only among those available in SAM, but with an additional alternative for the natural mortality model to handle non-migration. Unless otherwise stated, we assume that all normal or truncated normal variables (ε's) below are independent between years and ages.

To model recruitment into a minimum age $a \geq 0$, we use

$$N_{a^{\min},y} = N_{a^{\min},y-1}\exp(\varepsilon_{a,y}^R), \tag{8}$$

$$\varepsilon_{a,y}^R \sim N(0, \sigma_{R,a}^2). \tag{9}$$

which is a random walk process on the logarithmic scale.

The fishing mortality rate model is also a random walk process on the logarithmic scale,

$$F_{a,y} = F_{a,y-1}\exp(\varepsilon_{a,y}^F), \tag{10}$$

$$\varepsilon_{a,y}^F \sim N(0, \sigma_{F,a}^2). \tag{11}$$

We assume that the $\varepsilon_{a,y}^F$'s are positively correlated between ages within the same year, with an AR(1) correlation structure over ages as

in Nielsen and Berg (2014), but independent between years. Then, $\mathrm{corr}(\varepsilon_{a,y}^F, \varepsilon_{a',y}^F) = \rho^{|a-a'|}$, where ρ is a correlation parameter to be estimated.

We assume existence of information about the natural mortality rates, that are summarily defined by the fixed values $M_{a,y}^*$ for each age and year. These may be simply based on a best guess, or some more quantitative prior information.

For the natural mortality rate, we use the model

$$M_{a,y} = M_{a,y}^* + \varepsilon_{a,y}^M, \tag{12}$$

$$\varepsilon_{a,y}^M \sim N(0, \sigma_{M,a}^2). \tag{13}$$

$M_{a,y}^*$ is the expected natural mortality rate. Note that Nielsen and Berg (2014) prefer to reserve the term "natural mortality rate" for $M_{a,y}^*$ alone and use the term "process error" for $\varepsilon_{a,y}^M$. We interpret the sum of these two terms as the actual natural mortality rate.

Now, $M_{a,y}$ may become negative, and one can argue that $\varepsilon_{a,y}^M$ then accounts for both stochastic variations in natural mortality and for migration to or from the considered area. Then zero or negative values of $Z_{a,y}$ can occur. As mentioned above, the Baranov catch equation is still valid.

As an alternative, we also consider a model without migration, where $M_{a,y} \geq 0$, but still with $E(M_{a,y}) = M_{a,y}^*$. This is so far not possible in the SAM program. Then, we assume that $\varepsilon_{a,y}^M$ follows a truncated normal distribution, where Eq. (13) is replaced by

$$\varepsilon_{a,y}^M \sim N_{\mathrm{trunc}}(0, \sigma_{M,a}^2)\,\text{truncated at} \pm M_{a,y}^*. \tag{14}$$

Truncating symmetrically both below and above ensures that $E(M_{a,y}) = M_{a,y}^*$, and that the parameter $\sigma_{M,a}^2$ controls only the variance of $M_{a,y}$. If the truncation above were removed, $E(M_{a,y}) > M_{a,y}^*$ and $\sigma_{M,a}^2$ would control both the expectation and the variance of $M_{a,y}$.

The model presented so far describes what happens from the start of one year to the start of the next, and it is sufficient if we only have yearly catch data. If we also consider survey indices observed within years, we must model the stock size within a year. Then, we need the following assumption: Fishing and natural mortality occur simultaneously and are constant during a year. The number of fish of age a at the end of day d of year y is then

$$N_{a,y}(d) = \exp(-(d/365)Z_{a,y})N_{a,y}. \tag{15}$$

We emphasise that the population model part of SAM is defined by the Eqs. (4), (8) (one out of three options in SAM) and (9)–(13), even though Nielsen and Berg (2014) present some of these on the logarithmic scale.

2.2.2. The data model

The data model relates observations to the true, unknown, quantities in the population model. If we assume that the catch data $\widehat{C}_{a,y}$ are (mean-)unbiased estimates of the true catch (i.e. that $E(\widehat{C}_{a,y}) = C_{a,y}$) and log-normally distributed, the data model for the catch can be written as

$$\widehat{C}_{a,y} = C_{a,y}\exp(\varepsilon_{a,y}^C), \tag{16}$$

$$\varepsilon_{a,y}^C \sim N(-1/2\sigma_{C,a}^2, \sigma_{C,a}^2). \tag{17}$$

We could alternatively, assume $\widehat{C}_{a,y}$ to be a median-unbiased estimate of the true catch (median$(\widehat{C}_{a,y}) = C_{a,y}$). The corresponding expectation of $\varepsilon_{a,y}^C$ should be 0, i.e. Eq. (17) should be replaced by

$$\varepsilon_{a,y}^C \sim N(0, \sigma_{C,a}^2). \tag{18}$$

This implies that the catch data are mean-biased upwards by a factor $\exp(1/2\sigma_{C,a}^2)$ compared to the true catch, due to the nature of the log-normal distribution. Hence, when the model is fitted to data, the average of the estimated catches, will tend to be lower than the average of the catch data. Furthermore, to assume that the age-specific catch

data are median-unbiased, implies that the corresponding observation for the total catch in a year, $\widehat{C}_y = \sum_a \widehat{C}_{a,y}$, is assumed to be both mean-biased and median-biased upwards compared to the true total catch.

The assumptions about the expected values of $\varepsilon_{a,y}^C$ will not only influence the fitted catch values, but also the estimates of the population sizes $N_{a,y}$. Unfortunately, whether we should use Eq. (17) or Eq. (18) cannot be guided by their fit to data, and their likelihood values will be exactly equal if $\sigma_{C,a}$ is equal for all ages. Instead, the choice between these alternatives depends on our belief about how the catch data relates to the true catch, which further depends on how the catch data are constructed. Usually, the yearly catch of age data are point estimates based on more detailed data, including for instance age data from samples of fish. Hence the optimal model choice depends on the properties of these point estimates. As an example, Norwegian catch of age estimates are often based on the Bayesian ECA (estimating catch of age) model (Hirst et al., 2004, 2005, 2012). This model reports the posterior means, which we believe are more mean-unbiased than median-unbiased.

The data model for the survey index can be written

$$\widehat{I}_{a,y}(d) = I_{a,y}(d)\exp(\varepsilon_{a,y}^I), \tag{19}$$

$$\varepsilon_{a,y}^I \sim N(-1/2\sigma_{I,a}^2, \sigma_{I,a}^2), \tag{20}$$

if we assume that the observed survey index is an unbiased estimate of the true survey index, or with Eq. (20) replaced by

$$\varepsilon_{a,y}^I \sim N(0, \sigma_{I,a}^2) \tag{21}$$

if we assume it is median-unbiased. The choice between these two is usually not of practical relevance, since the expected value is confounded with the catchability. The contrary is true, however, if one uses the model to predict an observed index $\widehat{I}_{a,y}(d)$, for instance in a cross validation experiment as in Section 3.2.

2.3. The data model in SAM

In SAM, catch observations are modelled as

$$\widehat{C}_{a,y} = C_{a,y}^{\mathrm{SAM}}\exp(\varepsilon_{a,y}^C), \tag{22}$$

where

$$C_{a,y}^{\mathrm{SAM}} = (F_{a,y}/(F_{a,y} + M_{a,y}^*))(1 - \exp(-(F_{a,y} + M_{a,y}^*)))N_{a,y}, \tag{23}$$

$$= (F_{a,y}/Z_{a,y}^*)(1 - \exp(-Z_{a,y}^*))N_{a,y},$$

is the definition of catch in SAM. Here, $Z_{a,y}^* = F_{a,y} + M_{a,y}^*$, so $\varepsilon_{a,y}^M$, the variation in natural mortality or the process error, is ignored here. The quantity $C_{a,y}^{\mathrm{SAM}}$ is different from the true catch, which explains why we refer to the data model in SAM as "mis-specified". These quantities are related by

$$C_{a,y}^{\mathrm{SAM}} = (Z_{a,y}/Z_{a,y}^*)[(1 - \exp(-Z_{a,y}^*))/(1 - \exp(-Z_{a,y}))]C_{a,y}. \tag{24}$$

The corresponding number of fish that dies from other causes is

$$D_{a,y}^{\mathrm{SAM}} = (N_{a,y} - N_{a+1,y+1}) - C_{a,y}^{\mathrm{SAM}} = T_{a,y} - C_{a,y}^{\mathrm{SAM}} \tag{25}$$

$$= [(1 - \exp(-Z_{a,y})) - (F_{a,y}/Z_{a,y}^*)(1 - \exp(-Z_{a,y}^*))]N_{a,y}.$$

Note the lack of similarity between the expressions for $C_{a,y}^{\mathrm{SAM}}$ and $D_{a,y}^{\mathrm{SAM}}$, as opposed to the definitions of $C_{a,y}$ and $D_{a,y}$ in Eqs. (6) and (7).

The SAM data model for the survey index is

$$\widehat{I}_{a,y}(d) = I_{a,y}^{\mathrm{SAM}}(d)\exp(\varepsilon_{a,y}^I), \tag{26}$$

where

$$I_{a,y}^{\mathrm{SAM}}(d) = Q_a \exp(-(d/365)Z_{a,y}^*)N_{a,y} \tag{27}$$

$$= \exp[(d/365)\varepsilon_{a,y}^M]I_{a,y}(d)$$

is SAM's definition of a survey index, where $\varepsilon_{a,y}^{M}$ is ignored here as well. The value of this index at the end of a year differs from its value for the same cohort at the start of the next year, i.e.

$$I_{a+1,y+1}^{\mathrm{SAM}}(0) \neq I_{a,y}^{\mathrm{SAM}}(365), \tag{28}$$

even if $Q_{a+1} = Q_a$, as opposed to the true survey index, given in Eq. (2).

The corresponding stock size at the end of a year is

$$N_{a,y}^{\mathrm{SAM}}(365) = \exp(-Z_{a,y}^{*})N_{a,y}, \tag{29}$$

whereas the stock size for the same cohort at the start of the following year is

$$N_{a+1,y+1} = \exp(-\varepsilon_{a,y}^{M})N_{a,y}^{\mathrm{SAM}}(365). \tag{30}$$

Therefore, our interpretation of SAM is this: Until the end of a year, there is a constant fishing and natural mortality that occur simultaneously. Then in addition, at the eve of a new year, fish may either die due to additional natural mortality, or migrate to, or from the area.

Finally, in SAM, the expectation of $\varepsilon_{a,y}^{C}$ and $\varepsilon_{a,y}^{I}$ are set to 0. Hence observations in SAM are assumed to be median-unbiased, but mean-biased, compared to the corresponding true values.

2.3.1. Estimation

We estimate unknown quantities in the model from the observed data by maximum likelihood as in SAM, using the TMB software for optimisation. We obtain estimates of the model parameters of the fishing and natural mortalities and of the population sizes. We also obtain updated estimates of the catches $C_{a,y}$ and of the hypothetical indices $I_{a,y}(d)$, here termed $\widetilde{C}_{a,y}$ and $\widetilde{I}_{a,y}(d)$ to distinguish them from the observations (regarded as preliminary estimates) $\widehat{C}_{a,y}$ and $\widehat{I}_{a,y}(d)$.

3. Empirical study based on 14 fish stocks

We consider the model alternatives given in Section 2.2, divided into two experiments, with the natural mortality rate being unrestricted in the first experiment and non-negative in the other. For both, we explore three data model specifications. The log likelihoods are comparable because all the models have the same number of parameters and use the same type data distribution. We also consider four measures of prediction performance based on cross validation.

3.1. Data on 14 fish stocks

The data encompass (i) the single data set used in Nielsen and Berg (2014), (ii) two data sets from the ICES Arctic Fisheries Working Group 2017 report (ICES, 2017b), and (iii) eleven data sets at stock-assesment.org (2019) with status "final" on the 19th of October 2018,

excluding duplicate data sets (Table 2). We use the catch series and one survey index series from each dataset, and ignore all other data. The assessments are only for model comparison, and not stock assessment suggestions.

3.2. Cross validation

The main purpose of stock assessment is to estimate the number of fish and the spawning stock biomass. Model evaluation is challenging since the true values of these parameters are unknown. We could simulate from a model, and use the simulated data as the "truth", but this would favour models close to the "true" simulation model. An alternative is to investigate how well the models can predict the observed catch and index data. Hopefully, a method that predicts well is also suitable for fish abundance estimation.

We perform cross validation for each data set, leaving out both catch data and indices for one year at a time, i.e. splitting the data into training and test data. We estimate the model on the training data, and predict the left out catch and index data. Data for each year are left out once, except for the first year, which is always included in the training data. We calculate two root-mean-square measures of prediction performance for catch, one per age and year (CV^{Cay}), and another for total catch in a year (CV^{Cy}), and two corresponding measures for the indices (CV^{Iay}, CV^{Iy}):

$$\mathrm{CV}^{\mathrm{Cay}} = \sqrt{(1/n^{\mathrm{Cay}}) \sum_{y} \sum_{a} (\widehat{C}_{a,y} - \widetilde{C}_{a,y,(-y)})^2}, \tag{31}$$

$$\mathrm{CV}^{\mathrm{Cy}} = \sqrt{(1/n^{\mathrm{Cy}}) \sum_{y} \left(\sum_{a} \widehat{C}_{a,y} - \sum_{a} \widetilde{C}_{a,y,(-y)} \right)^2}, \tag{32}$$

$$\mathrm{CV}^{\mathrm{Iay}} = \sqrt{(1/n^{\mathrm{Iay}}) \sum_{y} \sum_{a} (\widehat{I}_{a,y} - \widetilde{I}_{a,y,(-y)})^2}, \tag{33}$$

$$\mathrm{CV}^{\mathrm{Iy}} = \sqrt{(1/n^{\mathrm{Iy}}) \sum_{y} \left(\sum_{a} \widehat{I}_{a,y} - \sum_{a} \widetilde{I}_{a,y,(-y)} \right)^2}, \tag{34}$$

where n is the number of elements in each sum, and the subscript $(-y)$ means that catch and index data for year y was not used for estimation.

3.3. Set up for two experiments

In Experiment 1, the basic model is SAM given in Section 2 with unrestricted $M_{a,y}$ (includes possible migration), here termed S. We define two alternatives with different data models:

Table 2

Overview of data sets used. The name of the dataset is given if the data source is stockassesment.org (2019). The name of the survey index is given if several indices were available. The minimum and plus ages are the ones used in the models. To save computation time, we have ignored catch data before 1961 for North-East Atlantic cod.

Short name	Data source	Area	Species	Years catch	Years index	Min. age plus age	Survey index
NSC	Nielsen and Berg (2014)	North Sea	Cod	1963–2011	1983–2012	1–7	
CC	ICES (2017b)	Coast of Norway	Cod	1984–2016	1995–2016	2–10	Table T26, p. 69
NEAC	ICES (2017b)	North-East Atlantic	Cod	1961–2016	1981–2017	3–15	Table A3, p. 180
BW	BW_2018	Widely distributed	Blue whiting	1981–2018	2004–2018	1–10	
NSSOLE	sole2024_newidx	North Sea	Sole	1984–2017	2004–2017	1–9	Fisherman
FSAITHE	sam-tmb-fsaithe-2017-01	Faroe Plateau	Saithe	1961-2017	1994–2018	3–15	Spring
FCOD	sam-tmb-fcod-2017-01	Faroe Plateau	Cod	1959-2017	1996–2017	1–10	Summer
NSW	NSwhiting_2018	North Sea	Whiting	1978–2017	1983–2017	0-8	IBTS-Q1
ICEEGCOD	codEastNWWG2018	Iceland/East Greenland	Cod	1973–2017	1982–2017	1–10	WH
FHADDOCK	sam-tmb-fhaddock-2017-01	Faroe Plateau	Haddock	1957–2017	1996–2017	1–10	Summer
BSHERRING	WBSS_mf_004_CB_corrCF	Baltic Sea	Herring	1991–2016	1991–2016	0–8	HERAS
BSPLAICE	PLE2123_WGBFAS2017_Final_run	Baltic Sea	Plaice	1999–2016	1999–2017	1–10	IQ IBTS + BITS
MACKEREL	WGWIDE2017.V2	Widely distributed	Mackerel	1980–2016	2010–2017	0–12	Swept-idx
BSCOD	WBcod_2017	Baltic Sea	Cod	1994–2016	2001–2016	0-7	SD2224w_4Q

1. M (Modified catch and index equations): Include $\varepsilon_{a,y}^M$ in the catch and index data models, i.e. replace Eqs. (22) and (26) by Eqs. (16) and (19).

2. Mu (Modified plus Unbiased observations): Also change $E(\varepsilon_{a,y}^C)$ and $E(\varepsilon_{a,y}^I)$ from 0 to $-1/2\sigma_{C_a}^2$ and $-1/2\sigma_{I,a}^2$, respectively, i.e. assume the observations to be unbiased with respect to the corresponding true values.

While Experiment 2 is similar to Experiment 1, the natural mortality rate $M_{a,y}$ is non-negative, and modelled by a truncated normal distribution given by Eq. (14).

We estimate models for the 14 fish stocks, both for the complete data set and sub-data sets in the cross validation analyses. To enable reasonably good model fits for the different fish stocks, we apply a common, yet flexible, model structure for all stocks and for experiments. Consequently, the model choice might not be optimal for every data set.

The standard deviations for the fishing mortality rates, $\sigma_{F,a}$, are equal for all ages. The logarithms of the standard deviations $\sigma_{C,a}$ and $\sigma_{I,a}$ and the catchability parameters Q_a are allowed to vary smoothly over ages by quadratic functions of age. For instance, the standard deviations of catch is given by

$$\sigma_{C,a} = \exp(\beta_0^{\sigma_C} + \beta_1^{\sigma_C}(a - \bar{a}) + \beta_2^{\sigma_C}(a - \bar{a})^2), \tag{35}$$

where $\bar{a}$ is the average age, and the β's are parameters to be estimated. This formulation allows the standard deviations and catchabilities to be age-specific and smooth, but with only three parameters for each equation. This flexible, yet parsimonious, formulation is useful when we fit many data sets automatically. We add a penalty term of the form $(\beta_1^{\sigma_C})^2 + (\beta_2^{\sigma_C})^2$ for each of the three quadratic terms to the log likelihood to ensure identifiability and numerical stability This shrinks the age-specific standard deviations and catchabilities slightly towards an average value for all ages.

We do no direct comparison between the two experiments, because the main focus is the impact of the data model specification, and not on population sub-models. The experimental setups with data and computer code are given in Appendix B.

3.4. Results for Experiment 1

In fitting the full data sets, the log likelihood improves for 75% of the data sets by applying model M instead of S, i.e. instead of a "process" error at the end of each year, we apply a stochastic variation in natural mortality and migration (Fig. 1, panel (a). The potential gain in log likelihood is larger than the potential loss. When we consider model Mu, the log likelihoods do not change much compared to model M. This is as expected, as these bias adjustments for the catch and index data

are confounded with other model factors, and the potential value of these changes cannot be judged by the likelihood. Most of the gain in applying Mu instead of S is hence achieved by changing from model S to M. The total gain in the log likelihoods by applying model Mu instead of S has no systematic relationship with the estimated size of the standard deviation $\sigma_{M,a}$ in Mu (Fig. 1, panel (b). The estimates of σ_a^M are slightly larger in model Mu than in model S (Fig. 1, panel (c).

Panel (a) in Fig. 2 shows the relative (to model Mu) changes in the cross validated root mean squared errors (RMSEs) CV^{Cay} for predictions of age-and-year specific catch, when changing from one model to another. The RMSEs tend to decrease slightly with model M instead of S, and further with model Mu. By applying model Mu instead of S, we tend to get a moderate improvement in the RMSEs CV^{Cay}. When we consider the RMSEs CV^{Cy} for predictions of year specific total catch (panel b), the tendency is even clearer, with an improvement for most data sets when considering model Mu instead of S. Finally, the relative bias in predictions for each model is given in panel (c). The predictions for models S and M are biased downwards (too low), as expected.

The lower panels in Fig. 2 correspond to the upper ones, but for the indices. The effects of applying model M or Mu instead of S are similar to those for the catch.

We find that applying model Mu tends to result in a better fit and improved catch predictions compared to applying model S. Does this give an important change in what is the focus of the stock assessment, to estimate the stock size? To address this, we calculate the relative changes in the averaged estimated number of fish in the stocks, by applying models M and Mu instead of S, using all data in each data set.

The estimated stock size tends to be slightly lower using model M instead of S (Fig. 3). Considering model Mu instead of M gives an increase in stock size estimates, at most 10% increase. Overall, the estimated stock sizes increase for all data sets by applying model Mu instead of S.

For illustration, we give some detailed results for the Iceland and East Greenland Cod (*Gadus morhua*) stock. This stock has the highest estimated standard deviation for the natural mortality ($\sigma_{M,a} = 0.28$ for all ages) and the highest standard deviation for the catch data for the average age ($\beta_0^{\sigma_C} = 0.49$). Both models S and M have some predictive capacity (Fig. 4). Considering model S, both predicted and fitted catch is in average lower than the observed catch, as expected.

We also conduct a simulation test (Deroba et al., 2015) for model Mu based on its fitted model for this fish stock, and estimate the stock sizes, catch and survey indices well (see Appendix A).

3.5. Results for Experiment 2

Figs. 5–7 summarise results of Experiment 2. Applying model M instead of S has much less effect now. This is natural, as the stochastic
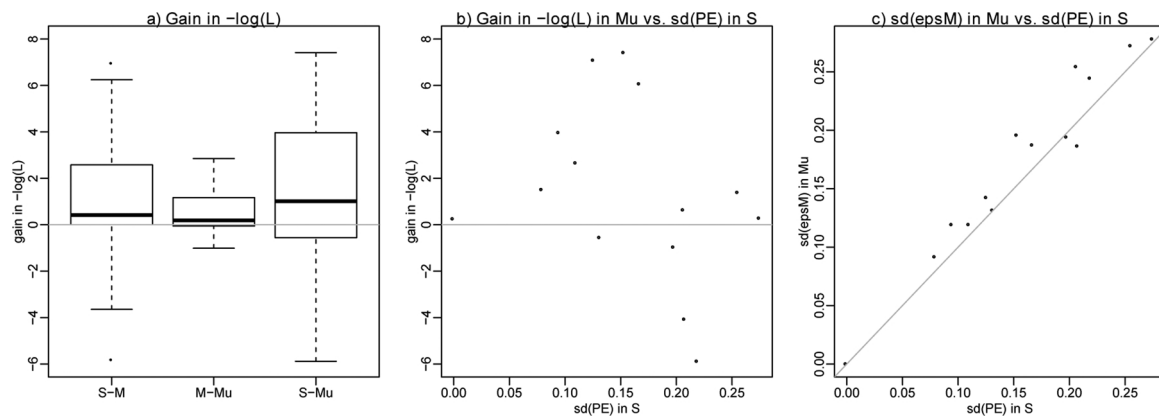

Fig. 1. Experiment 1: Changes in $-\log$ likelihood and in $\sigma_{M,a}$ comparing model alternatives. Panel (a) Box plots of difference in $-\log$ likelihood by changing from models S to M, from M to Mu and in total from S to Mu. Panel (b) Difference in $-\log$ likelihood by changing from model S to Mu vs. the estimated value of the standard deviation $\sigma_{M,a}$ in model Mu. (c) Estimated standard deviation $\sigma_{M,a}$ in model Mu vs. the corresponding estimates in model S.
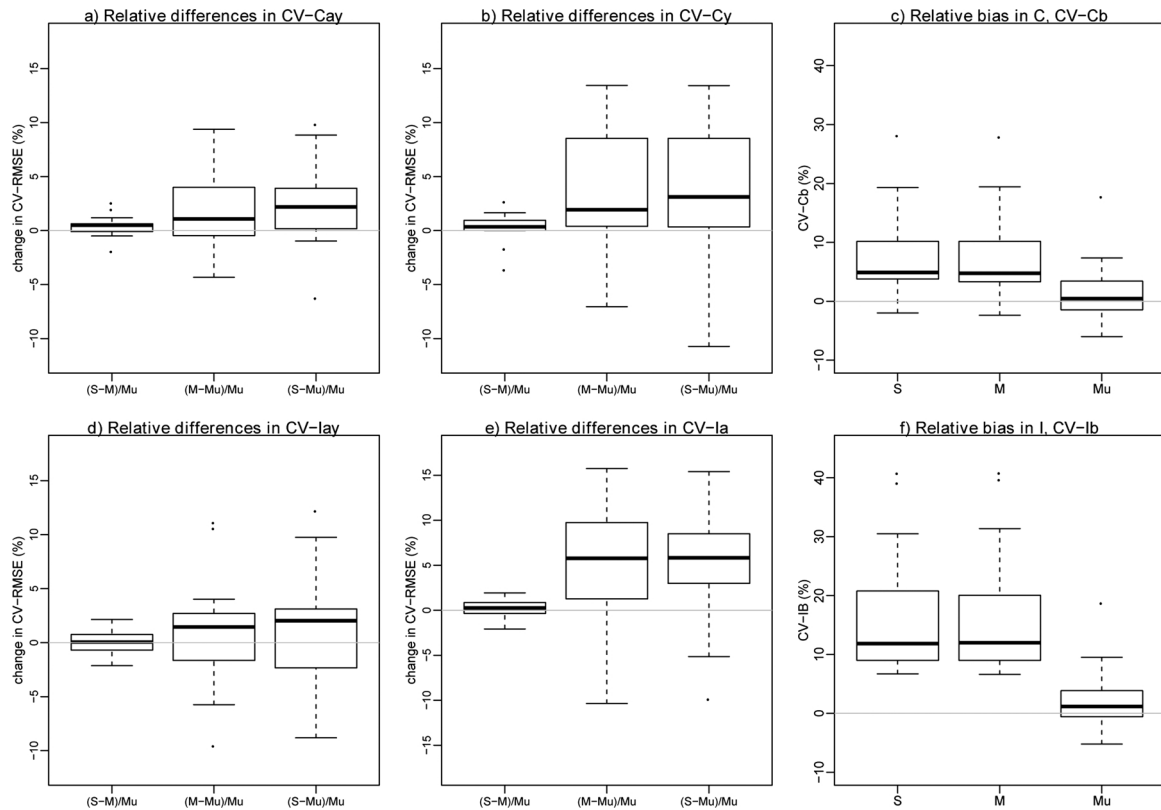
Fig. 2. Experiment 1: Box plots with relative cross-validated errors in catch and indices. Panel (a) Percentage changes in age-and-year specific catch prediction errors CV^{Cay}, by changing from models S to M, from M to Mu and in total from S to Mu, relative to the prediction error for Mu. Panel (b) Corresponding percentage changes in year specific catch prediction errors CV^{Cy}. Panel (c) Corresponding percentage changes in prediction bias. Panel (d) As panel (a), but for age-and-year specific index prediction errors CV^{Iay}. Panel (e) As panel (b), but for year specific index prediction errors CV^{Iy}. Panel (f) As panel (c), but for bias in indices.
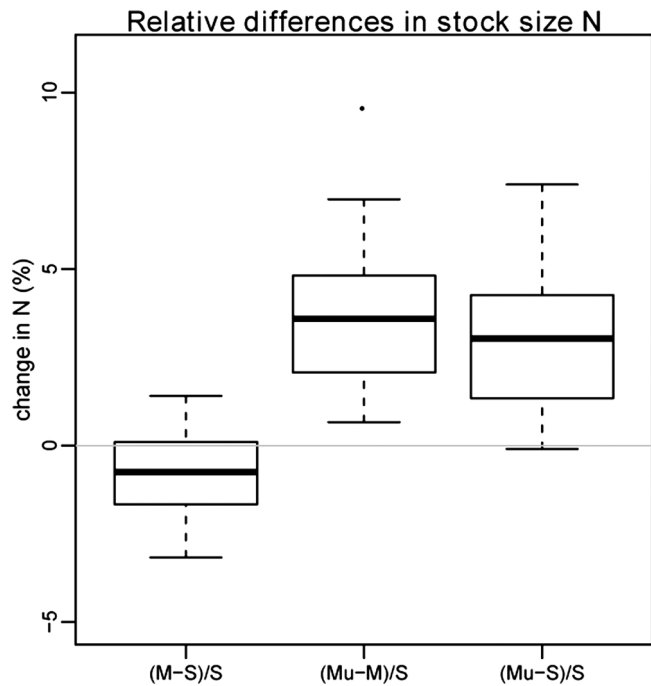
Fig. 3. Experiment 1: Percentage change in estimated average total stock size by changing from model S to M, from model M to Mu and in total from model S to Mu.

variation of the natural mortality or the "process error" is restricted. There is still a tendency for a small gain in the likelihoods and slightly lower estimated stock size with model M vs. S, with RMSEs being

almost identical. Still, the effect of applying model Mu instead of M is similar to the corresponding results in Experiment 1. Hence, the effect of applying model Mu instead of model S is also similar to the effect in Experiment 1.

4. Conclusions

We present a general stock assessment model, composed of a population model for a fish stock, and a data model that links observed data to the unobserved quantities in the population model.

We focus on catch and survey index data, but other types of data may also be considered. Several variants of this model structure exist in the literature. We focus on the popular model of Nielsen and Berg (2014), and suggest some modifications for improvement, some of which have been previously used by others. Both Aanes et al. (2007) and Cadigan (2016) use a stochastic natural mortality (then with a multiplicative lognormal error) instead of a "process error" at the end of the year, and Aanes et al. (2007) assume that the catch observations are unbiased, i.e. they used a lognormal error term with mean 1. In the empirical study of 14 fish stock datasets, our suggested modifications of the data model tend to improve the fit and the prediction performance, and reduce the prediction bias. Finally, we show that the modifications jointly have an impact on the estimated stock size, and that the two modifications alone have opposite effects on the stock size estimates. We made a prototype software with computer code for the two experiments, to be downloaded from Appendix B. Our hope is that our modifications will be included in the SAM package, as we regard this package well developed with many useful features, and we see no reasons for an alternative software package.

We use an additive model for the natural mortality rate. Since $F_{a,y}$ is modelled by a multiplicative model, it is natural to also consider
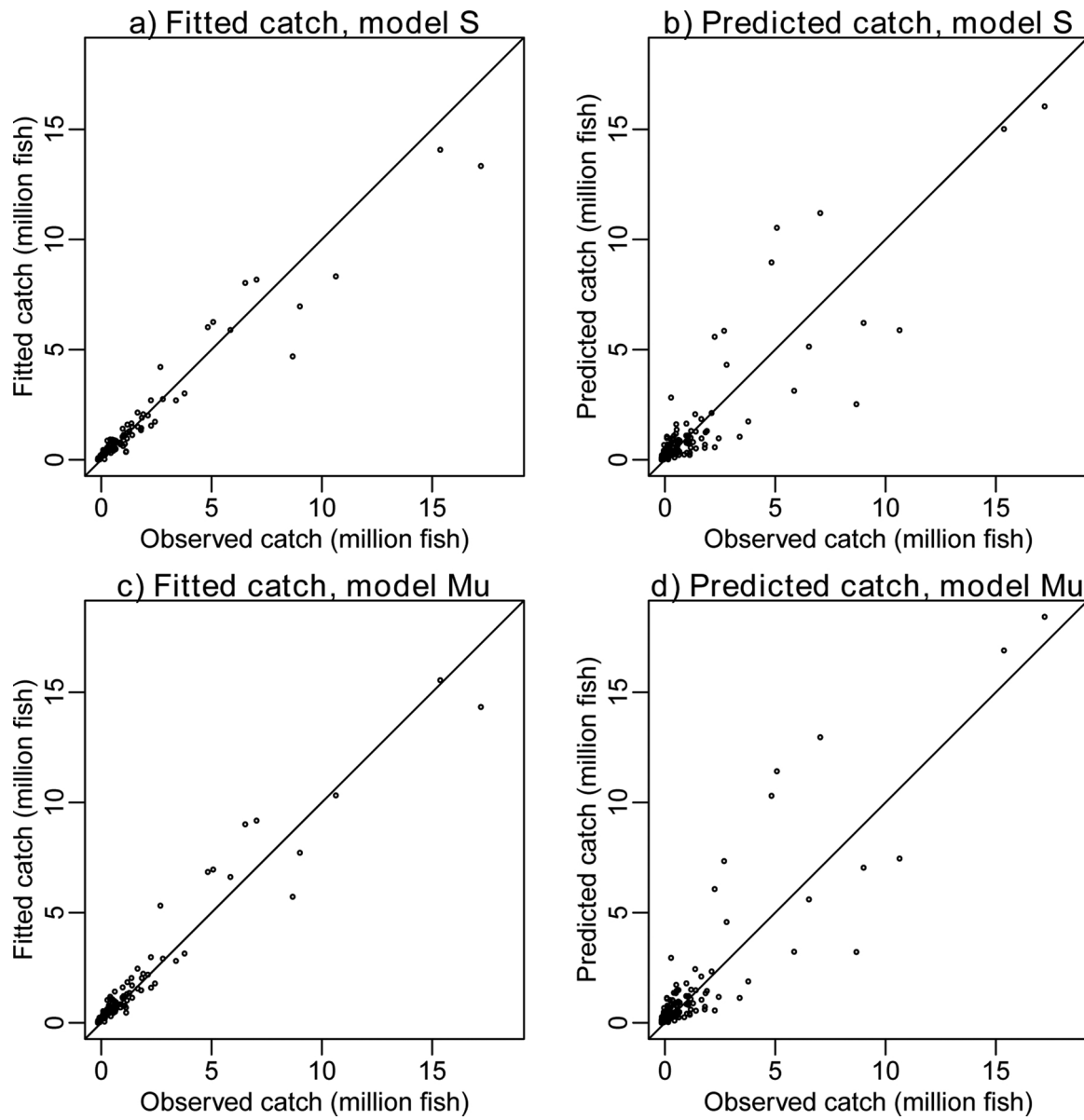
Fig. 4. Experiment 1: Fitted, predicted (by cross validation) and observed catch for models S and Mu. Panel (a) Fitted vs. observed catch for model S. Panel (b) Predicted vs. observed catch for model S. Panel (c) Fitted vs. observed catch for model Mu. Panel (d) Predicted vs. observed catch for model Mu.
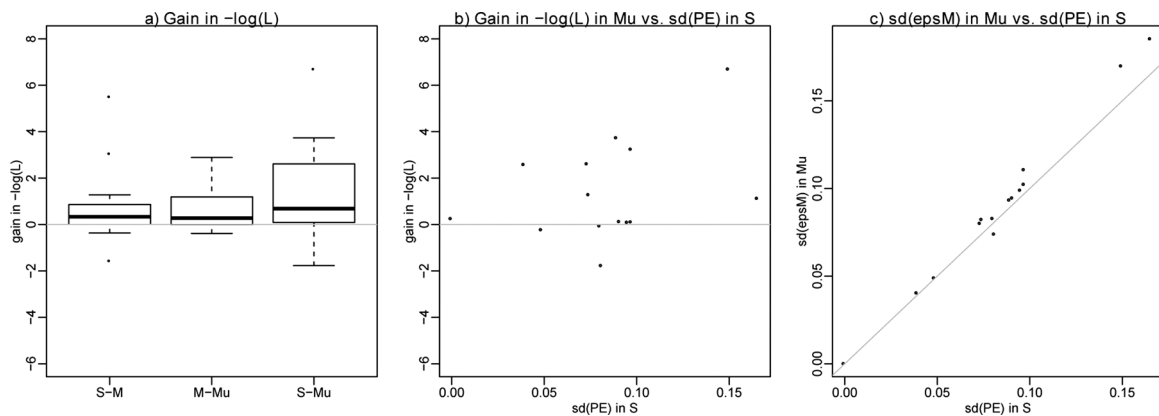

Fig. 5. Experiment 2: Changes in $-\log$ likelihood and in $\sigma_{M,a}$ comparing model alternatives. This figure corresponds to Fig. 1.
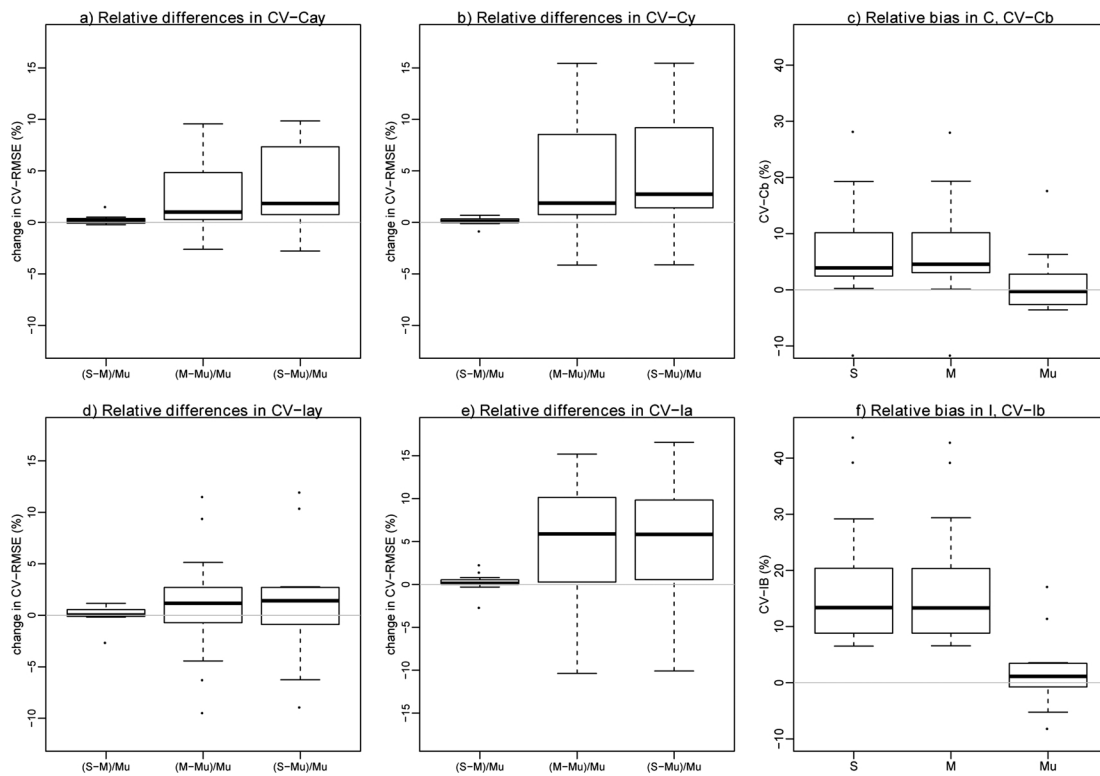
Fig. 6. Experiment 2: Box plots with relative cross-validated errors in catch and indices. This figure corresponds to Fig. 2.
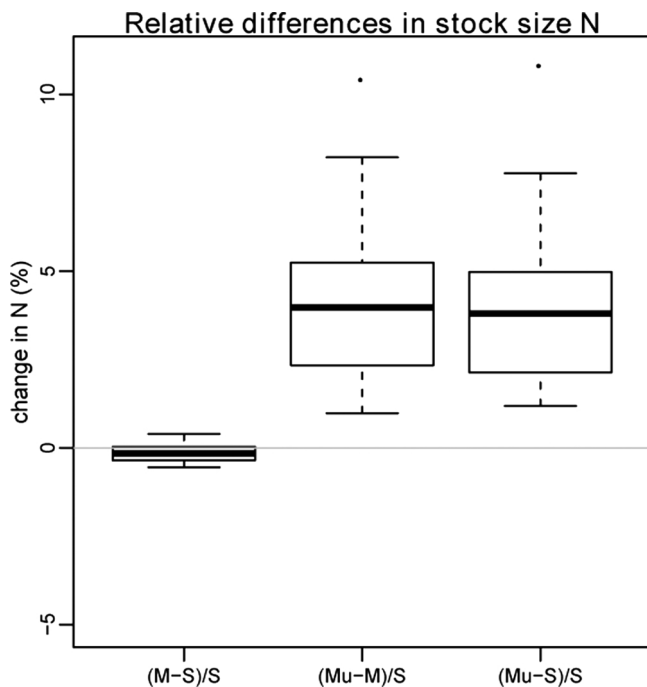

Fig. 7. Experiment 2: Percentage change in estimated average total stock size. This figure corresponds to Fig. 3.

multiplicative models for $M_{a,y}$ in the case without migration, since $F_{a,y}$ and $M_{a,y}$ enters the population model (4) symmetrically. We experiment with a lognormal random variation for the natural mortality rate, and the effect on the estimated stock sizes becomes larger. But we experienced convergence problems for some data sets. The specification and selection of models for $F_{a,y}$ and $M_{a,y}$ should hence be subject for further research.

Funding

Authors' contributions

Magne Aldrin: Conceptualization, Methodology, Software, Formal analysis, Writing – Original Draft
Ingunn Fride Tvete: Software, Data Curation, Writing – Review & Editing
Sondre Aanes: Conceptualization, Software, Writing – Review & Editing
Sam Subbey: Conceptualization, Writing – Review & Editing

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Additional details

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fishres.2020.105585.

References

Aanes, S., Engen, S., Sæther, B.E., Aanes, R., 2007. Estimation of the parameters of fish stockdynamics from catch-at-age data and indices ofabundance: can natural and fishing mortality be separated? Can. J. Fish. Aquat. Sci. 64, 1130–1142. https://doi.org/10.1139/f07-074.

Aeberhard, W., Flemming, J., Nielsen, A., 2018. Review of state-space models for fisheries science. Annu. Rev. Stat. Appl. 5, 215–235. https://doi.org/10.1146/annurev-statistics-031017-100427.

Aldrin, M., Aanes, S., Subbey, S., 2019. Comments on incongruous formulations in the sam (state-space assessment model) model and consequences for fish stock assessment. Fish. Res. 210, 224–227. https://doi.org/10.1016/j.fishres.2018.08.001.

Berg, C., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J. Mar. Sci. 73, 1788–1797. https://doi.org/10.1093/icesjms/fsw046.

Cadigan, N., 2016. A state-space stock assessment model for northern cod, including under-reporting catches and variable natural mortality rates. Can. J. Fish. Aquat. Sci. 73, 296–308. https://doi.org/10.1139/cjfas-2015-0047.

Deroba, J., Butterworth, D., Methot Jr., R., De Oliveira, J., Fernandez, C., Nielsen, A., Cadrin, S., Dickey-Collas, M., Legault, C., Ianelli, J., Valero, J., 2015. Simulation testing the robustness of stock assessment models to error: some results from the ices strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72, 19–30. https://doi.org/10.1093/icesjms/fst237.

Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. J. R. Stat. Soc, C – Appl. Stat. 43, 117–126. https://doi.org/10.2307/2986116.

Hirst, D., Aanes, S., Storvik, G., RB, H., Tvete, I., 2004. Estimating catch at age from market sampling data by using a bayesian hierarchical model. J. R. Stat. Soc, C – Appl. Stat. 53, 1–14. https://doi.org/10.1111/j.1467-9876.2004.00422.x.

Hirst, D., Storvik, G., Aldrin, M., Aanes, S., Huseby, R., 2005. Estimating catch-at-age by combining data from different sources. Can. J. Fish. Aquat. Sci. 62, 1377–1385. https://doi.org/10.1139/F05-026.

Hirst, D., Storvik, G., Rognebakke, H., Aldrin, M., Aanes, S., Vølstad, J., 2012. A bayesian modelling framework for the estimation of catch-at-age of commercially harvested fish species. Can. J. Fish. Aquat. Sci. 69, 2064–2076. https://doi.org/10.1139/cjfas-2012-0075.

ICES, 2017a. Report of the Arctic Fisheries Working Group (AFWG), 19–25 April 2017. Technical Report ICES CM 2017/ACOM:06. ICES, Copenhagen, Denmark, pp. 493. http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2017/AFWG/01%20AFWG%20Report%20of%20the%20Arctic%20Fisheries%20Working%20Group.pdf.

ICES, 2017b. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 30 August–5 September 2017. Technical Report ICES CM 2017/ACOM:23. ICES, Copenhagen, Denmark. 994 pp., http://www.ices.dk/sites/pub/PublicationReports/ExpertGroupReport/acom/2017/WGWIDE/01ReportoftheWorkinggrouponWidelyDistributedStocks(WGWIDE).pdf.

ICES, 2018. Report of the Working Group on Assessment of Demersal Stocks in the North Sea and Skagerrak (2017), 26 April–5 May 2017. Technical Report ICES HQ. ICES CM 2017/ACOM:21. ICES, Copenhagen, Denmark. 1077 pp., https://www.ices.dk/sites/pub/PublicationReports/ExpertGroupReport/acom/2017/WGNSSK/01WGNSSKReport2017.pdf.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B., 2016. Tmb: automatic differentiation and laplace approximation. J. Stat. Softw. 70, 1–21. https://doi.org/10.18637/jss.v070.i05.

Miller, T., Hare, J., Alade, L., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern new England yellowtail flounder. Can. J. Fish. Aquat. Sci. 73, 1261–1270. https://doi.org/10.1139/cjfas-2015-0339.

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

Nielsen, M., Berg, C., 2019. Response to: Comments on incongruous formulations in the sam (state-space assessment model) model and consequences for fish stock assessment. Fish. Res. 210, 228–229. https://doi.org/10.1016/j.fishres.2018.09.008.

Quinn, T., Deriso, R., 1999. Quantitative Fish Dynamics. Oxford University Press, Oxford.

stockassesment.org, 2019. https://www.stockassessment.org/login.php (accessed 20.10.19).