Fisheries Research 249 (2022) 106247

Contents lists available at ScienceDirect

Fisheries Research

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

Towards best practice for specifying selectivity in age-structured integrated stock assessments

Kristin M. Privitera-Johnson [a,*], Richard D. Methot [b], André E. Punt [a]

[a] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA
[b] NOAA Fisheries, Northwest Fisheries Science Center, Seattle, WA, USA

ARTICLE INFO

Keywords:
Best practice, non-parametric selectivity
Simulation
Stock assessment
Stock synthesis

ABSTRACT

Specification of how selectivity (the combination of availability and vulnerability) is modelled in integrated stock assessments is key to avoiding bias in estimates of quantities of management interest. Many "rules of thumb" are common in the community but these have yet to be rigorously tested. This paper uses simulation to compare 12 approaches for specifying selectivity in an age-structured integrated stock assessment, including parametric and non-parametric approaches. The operating model represents a two-fishery case where selectivity for one or both fisheries can be dome-shaped and/or time-varying. The results suggest that using AIC to select among selectivity forms is not robust, including when model misspecification is absent, even though the use of model selection criteria such as AIC is common when conducting stock assessments. The use of double normal selectivity was found to be most robust to uncertainty in the true form of selectivity. Estimation of time-variation in selectivity did not lead to appreciable improvements in performance when the true time-variation was random. The double normal form performed poorly if M was estimated along with the other model parameters. Similarly, use of flexible parametric methods, such as splines, performed adequately with informative data, but poorly when the catch series exhibited low contrast and age-composition data were not available from the start of the fishery. This suggests that the best practices for selectivity will depend on knowledge of the likely information content of the data.

1. Introduction

Integrated approaches to stock assessment play a vital role in the provision of scientific advice for the management of fish and invertebrate stocks around the world. These approaches utilize multiple data sources in a single analysis. They combine three nested models—one for the population dynamics, one for the observation process, and one for the sampling distribution for the data (Maunder and Punt, 2013). The application of integrated approaches since their introduction by Fournier and Archibald (1982) now spans many stocks and jurisdictions. Punt et al. (2020) outline some best practices for developing the next generation of integrated analyses for stock assessment.

The integrated approach performs better when composition data (e.g., age-composition, length-composition, and/or conditional age-at-length data) from fishery-dependent and -independent sources are used to inform the values of the parameters of the population dynamics and observation models. The contribution of each composition datum, relative to its various sources and other types of data, relies on a weighting factor characterized as the effective sample size (Punt et al., 2021). Punt et al. (2021) demonstrated that, while higher effective sample sizes for length-composition and conditional age-at-length data resulted in smaller errors when estimating key management quantities (i.e., spawning biomass, depletion, recruitment, population model parameters, catch limit recommendations), effective sample size is often not the main factor that determines the magnitude of estimation errors. Overly simple selectivity formulations are a gross approximation to the many factors influencing a fishery's time-varying interaction with a population of fish (Maunder et al., 2014), and violations can lead to bias and imprecision when estimating quantities of management interest (Maunder and Punt, 2013).

1.1. Selectivity and estimation methods

Stewart and Martell (2014) define selectivity as length- or age-based probabilities used to link observed composition data to model predictions about population abundance-at-age/-size. Selectivity is a

* Corresponding author.
E-mail address: kpjohns@uw.edu (K.M. Privitera-Johnson).

https://doi.org/10.1016/j.fishres.2022.106247
Received 22 October 2021; Received in revised form 19 January 2022; Accepted 21 January 2022
Available online 1 February 2022

combination of two processes: availability (i.e., the probability that a fish of a specific age or size is in the same vicinity at the same time as gear deployment) and contact (or gear) selectivity (i.e., the relative probability that a fish of specific age or size is caught given it is available to the gear) (Sampson, 2014; Stewart and Martell, 2014). Fish or invertebrate life history characteristics and spatial patterns in fishing intensity can lead to different functional forms for the availability and vulnerability of a stock (Sampson and Scott, 2011). Selectivity is typically represented by three forms: monotonically decreasing such that juveniles are most selected, dome-shaped for intermediate ages, and monotonically increasing for adults (Bence et al., 1993). Analysts can select from various parametric selectivity formulations (i.e., patterns or forms): double normal, logistic, or cubic spline (see descriptions in Appendix A), among many others. Thorson and Taylor (2014) evaluated the performance of commonly used statistical models for estimating selectivity in age-structured integrated approaches: (a) parametric models, which estimate selectivity using a range of parameters, dependent on the complexity of the population dynamics model; (b) flexible parametric models, such as splines which estimate selectivity using an informative and explicit prior function, that can be updated by the available data when predicting future observations; and (c) non-parametric models, which estimate selectivity by age and penalize large fluctuations in selectivity between ages. Thorson and Taylor (2014) found that flexible and non-parametric models perform well (e.g., less bias and greater precision) when fishery selectivity does not follow a parametric function. However, these studies still leave some challenges and ambiguity regarding the best approach to use. For example, previous studies only examined a limited set of the factors that determine the performance of estimation methods and did not consider estimation of catch limits under harvest control rules.

1.2. Challenges and best practices for modelling selectivity

The model structure selected to relate selectivity and composition data impacts assessment outputs and the resulting management advice. Modelling selectivity using complex functional forms and time-varying parameters presents challenges because the model structure may: a) exceed the amount of information available in the composition data and b) require many parameters and result in unexplainable predictions, high correlations, and increased uncertainty (Hulson and Hanselman, 2014). Modelling selectivity for fishery-dependent and -independent sources is also not immune to misspecification, a common disadvantage for integrated approaches (Maunder and Punt, 2013). Misspecifying selectivity can result in data conflicts in integrated analyses (Ichinokawa et al., 2014), and Crone and Valero (2014) found that estimates of maximum sustainable yield and current biomass were sensitive to misspecification of selectivity when models are fitted to length-composition data.

Best practices for selecting an appropriate selectivity form and estimation method that minimize these challenges depends on the desired management objective (Maunder et al., 2014). Thorson and Taylor (2014) suggested flexible and non-parametric selectivity forms to diagnose and account for model misspecification. However, choosing complex and flexible selectivity patterns runs the risk of over-parameterization, confounding selectivity and recruitment parameters, and large uncertainties—a feature that is suboptimal for forecasts, yet useful for developing decision tables because uncertainty is better represented (Martell and Stewart, 2014). Assuming time-varying and complex selectivity parameterizations may result in robust assessment results for short-lived species, but not for long-lived species (Hulson and Hanselman, 2014). When there is spatial pattern in age composition data, analysts may use multiple spatially disaggregated fleets with unique selectivity to reduce bias for stocks that lack tagging data or other methods for estimating movement rates (Hurtado-Ferro et al., 2014).

1.3. Study objectives

This study investigates the relationship between selectivity assumptions and the quality of the composition data using simulation. The key questions we address in this paper are: a) what are the implications for the estimation of quantities of management interest when selectivity is misspecified, b) what are costs and benefits of forcing selectivity for at least one fleet to be asymptotic, and hence c) what is the most appropriate default assumption to make about the functional form for selectivity (parametric vs. flexible parametric, dome-shaped vs. asymptotic). We explore these questions using a two-fishery operating model and examine the sensitivity of our conclusions to data quality, which other parameters are estimated, catch history, and current stock size. We are ultimately aiming to refine the best practice guidelines that arose from the CAPAM Selectivity Workshop (Maunder et al., 2014).

We find these questions to be key points of divergence among stock assessment practitioners. In communication with 28 experienced stock assessment scientists, co-authors Methot and Punt note two divergent patterns. One is that about half of the assessment scientists believe that at least one fleet should have asymptotic selectivity in order to improve model convergence, despite the understanding that it will lead to bias if incorrect. In contrast, other assessment scientists argue that understanding of the fleet and fish ecology as driving factors could lead to dome-shaped selectivity for all fleets. Another notable point of divergence is with regard to accepting time-varying selectivity as generally a good practice, or not. Here we note that advocates of time-varying selectivity were typically experienced with data-rich assessment situations. Our study will attempt to quantitatively elucidate the veracity of these alternative approaches.

We focus on selectivity for fisheries-dependent data sources because these data sources are frequently included in stock assessment analyses. This study uses a simulation framework to evaluate the performance of a Stock Synthesis (SS3) [Methot and Wetzel (2013)] assessment when the assumptions of the data-generating model (i.e., operating model) and the model used for parameter estimation (i.e., estimation model) are altered to reflect common challenges to integrated assessment assumptions. The simulations are based on three fish life histories.

2. Methods

2.1. General structure of the simulation study

The simulation study involved using an operating model to generate pseudo data sets (200 for each of the scenarios; see Supplemental Materials for saturation test results) and applying several estimation methods to estimate quantities of management interest. The selectivity functions explored represent options most frequently used in SS3. In a review of 96 assessments conducted in SS3, 7 used both age- and length-based selectivity functions, 79 used only length-based selectivity (44 double normal, 35 logistic, and 5 cubic spline) and 10 used only age-based selectivity (double normal most common, but logistic, cubic spline, random walk, and the non-parametric form were also used) (N. Klaer, CSIRO Marine and Atmospheric Research, unpublished).

2.2. Operating model

2.2.1. Population dynamics

The operating model is a single-sex, single-area, age- and length-structured population dynamics model with two fleets, implemented using Stock Synthesis. Li et al. (2021) demonstrate that Stock Synthesis and a family of comparable models all produce indistinguishable results when configured similarly. Thus, we use Stock Synthesis as both the operating model and estimation method because the important issues which we are testing are in the model formulation and Stock Synthesis has considerable flexibility in this regard. The operating model covers 26 years, nominally “1982″ to “2017″. Table 1 lists the values for the

Table 1
Values for the biological and fishery parameters of the operating model. The values in parenthesis are the ages corresponding to the lengths concerned. An asterisk in the first column indicates that the parameter is estimated by the estimation method. A + indicates the parameter was estimated only for a subset of the simulations.

Parameter	School whiting	Tiger flathead	Blue grenadier
Plus-group age (yr)	9	20	20
Natural mortality, M (yr^{-1})* +	0.6	0.27	0.174
Length (low age, L_1) (cm)	6.7 (0)	30 (3)	9.8 (0)
Length (high age, L_2) (cm)	24.6 (L_∞)	55.9 (36)	100.3 (L_∞)
Von Bertalanffy κ (yr^{-1})	0.28	0.167	0.226
Length-at-age CV	0.08	0.106	0.125
Weight-length intercept (kg/cm^3)	0.000132	0.00000588	0.000015
Weight-length power	2.93	3.31	2.728
Length-at-50%-maturity (cm)	16	30	63.7
Maturity slope (cm^{-1})	2.0	0.25	0.261
R_0 *	Solved for (see text)		
Steepness* +	0.75	0.75	0.75
σ_R	0.35	0.6	1
Annual recruitment deviations (1982–2017) *	N(0, σ_R)	N(0, σ_R)	N(0, σ_R)
Selectivity*	Depends on the experiment		

biological parameters of the operating model.[1] These parameters are set based on the most recent Southern and Eastern Scalefish and Shark Fishery (SESSF) assessments for tiger flathead (*Neoplatycephalus richardsoni*), blue grenadier (*Macruronus novaezelandiae*), and eastern school whiting (*Sillago flindersi*) (Day, 2018a; Castillo-Jordán and Tuck, 2020; Day, 2018b, respectively). These species were selected to capture a range of longevities and variation in recruitment about the stock-recruitment relationship. Selectivity as a function of age is asymptotic for fleet 1 and dome-shaped for fleet 2 (Fig. 1a-c) for the base-case, and modelled using the flexible double-normal pattern (Supplementary Appendix A). Some of the scenarios account for time- as well as age-variation in selectivity by using a semi-parametric approach that multiplies the values for selectivity from the parametric function forms by age- and temporally-independent lognormal deviations, with a standard deviation of the logarithm of σ_S(Equation A.4). Following Xu et al. (2019), the correlation matrix for the deviations in selectivity about expected selectivity is the Kronecker product of a correlation matrix for among-age effects and a correlation matrix for among-years effects, i.e.:

$$\Omega = \Omega^1 \otimes \Omega^2 \qquad (1)$$

where $\Omega^1 = \rho_A^{|a_1-a_2|}$ and $\Omega^2 = \rho_Y^{|y_1-y_2|}$, ρ_A is the between-age correlation, and ρ_Y is the between-year correlation. The deviations apply to ages 1–5 for the longer-lived tiger flathead and blue grenadier, and ages 1–3 for the short-lived school whiting.

Fig. 1d-g shows the four scenarios related to historical catches by area ("increasing-then-decreasing", "increasing", "constant", and "decreasing-then-increasing"). Option 1 (Fig. 1d) should be most informative regarding the parameters of the operating model and option 3 (Fig. 1f) the least informative (option 1 is the base-case for the analyses). The values for the parameter R_0 (and hence S_0) are not taken from the stock assessments but rather selected so that 2018 depletion (spawning biomass in 2018 relative to unfished spawning biomass) is 0.25 or 0.5 (see the example trajectories of spawning biomass by area and species in Fig. 1h-k).

2.2.2. *Data generation*

The data available to the estimation methods are catches (assumed known with negligible error, CV = 0.00001), catch-rate data, and age-composition data. The simulated catch-rate data are assumed to be log-normally distributed about the model expectations and to be available for all years and both fleets. Age-composition data are assumed to be multinomially distributed (as assumed by the estimation method) and all ageing is assumed to be exact. The CVs for the index data and the effective sample sizes for the age-composition data are varied (see Table 2; Section 2.4).

2.3. *Estimation method*

The estimation method is a single-sex and age-structured model implemented using Stock Synthesis. Table 1 indicates the parameters that are estimated. The values for natural mortality, the parameters of the growth curve, and steepness of the stock-recruitment relationship are assumed to be known exactly, to allow a focus on estimation of selectivity.[2]

The extent of bias-correction in the stock-recruitment relationship depends on the information content of the data regarding annual recruitment (Methot and Taylor, 2011). Methot and Taylor (2011) developed an approach for specifying temporally varying bias-correction factors for recruitment based on the results of an assessment (the "recruitment-bias-ramp" method), and this approach is applied here. The estimation method is thus applied four times to address both setting of the "recruitment bias-ramp" and the effective sample sizes for the age-composition data. This involves:

1. Applying the stock assessment with no recruitment bias-ramp and effective sample sizes equal to the nominal values (those used to generate the data).
2. Applying the algorithm of Methot and Taylor (2011) to determine the recruitment bias-ramp.
3. Applying data weighting procedures; this involves applying algorithm T1.8 of Francis (2011) to determine a multiplier for the input (stage-1) sample sizes for the age-composition data.
4. Applying Equation 19 of Xu et al. (2019) to update the value for the extent of inter-annual variation in selectivity (if selectivity is assumed to be time-varying and the extent of annual variation in selectivity is not pre-specified).
5. Refit the model given the new specifications related to data weighting.
6. Repeat steps 2, 3 and 4 three times to allow the values for multipliers to converge for most replicates.

For this study the stage-1 effective sample sizes for age composition match the number of animals whose ages were generated to construct the age-composition data.

2.4. *Scenarios*

The scenarios (see the factors considered in the analyses in Table 2) relate to the questions identified above and involve seven simulation experiments (see Table 3 for a summary of the experiments). The scenarios are not fully balanced owing to the computational demands of the calculations and the need to avoid an overly large amount of model output. The bulk of the results are shown for tiger flathead given its intermediate values for natural mortality and variation in recruitment about the stock-recruitment relationship.

2.4.1. *Experiment 1*

The first experiment examines the impact of the choice of the selectivity formulation assumed when conducting the stock assessment for the three species. The row "Estimated selectivity form" in Table 2 lists the six basic selectivity options for a total of twelve selectivity forms

[1] Taken to be those for females.

[2] Experiment 6 involves estimating natural mortality (M) and steepness (h).

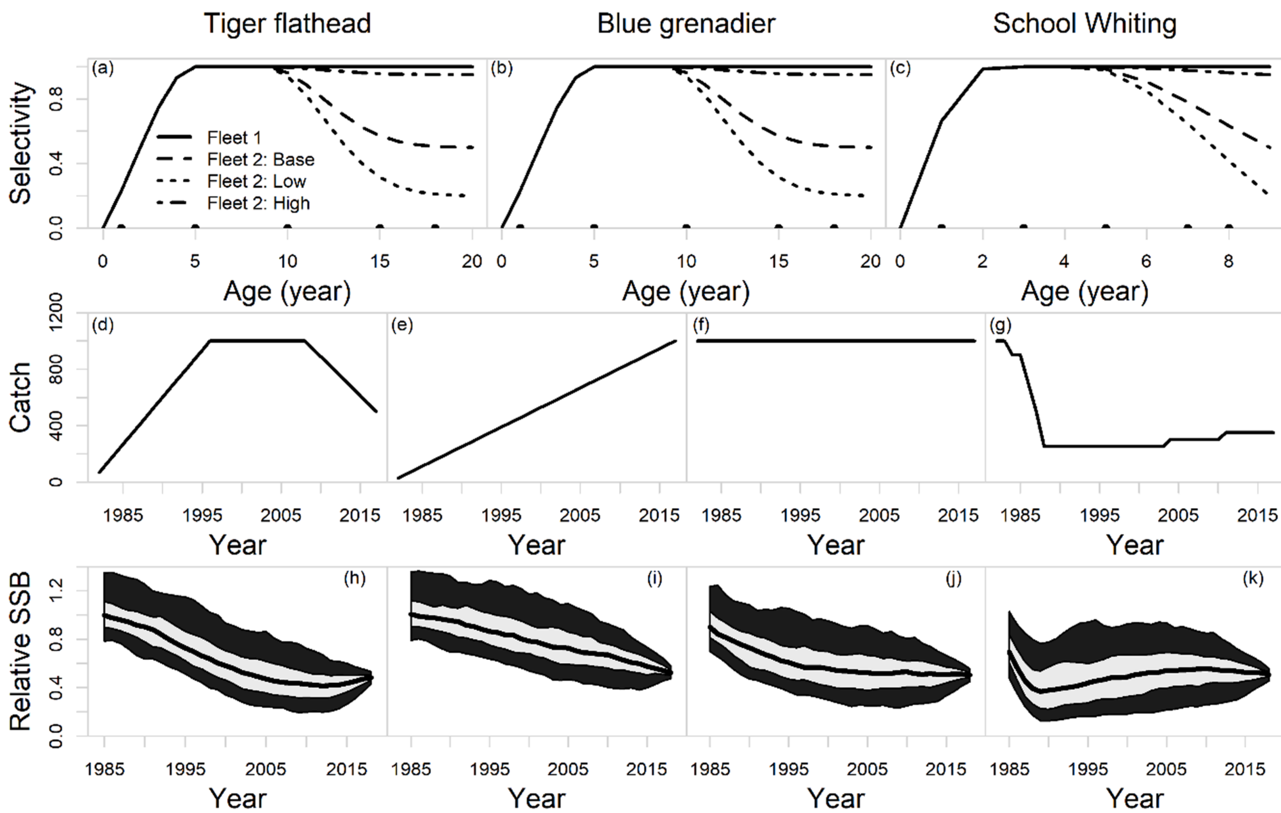

Fig. 1. Selectivity as a function of fleet and species (upper panels) and the time-trajectories of historical catch by fleet (center panels; the same catch time-series is used for all species), and time-trajectories of relative spawning biomass for tiger flathead for each time-trajectory of catch for the base level of selectivity for fleet 2 (lower panels). The results in the lower panels match the catch series in the center panels (i.e., the catch series in panel d corresponds to the biomass trajectory in panel h). The solid symbols on the x-axis in the upper panels indicate the ages for which selectivity is estimated when applying the spline estimation methods.

Table 2
Factors considered in the simulation experiments. The options indicated in bold-underline typeface represent the base-case analysis.

Factor	Options (bold is base-case)
Species	**Tiger flathead**; school whiting; blue grenadier
True selectivity form	**Fig. 1a-c (base)**; Logistic (both fleets equal to fleet 1 in Fig. 1a-c); double normal for both fleets
True time-varying selectivity	**No**; yes (σ_S=0.4; age and time correlation=0.8)
Extent of dome-ness for fleet 2	Low dome (S_A=0.2); **Median dome (S_A=0.5)**; high dome (S_A=0.95); Asymptotic
Catch series	**Increasing-then-decreasing**; increasing; constant; high initial
2018 stock size	**$0.5B_0$**; $0.25B_0$
Index CV	0.1; **0.3**; 0.5
Age-composition effective sample sizes	10; **100**; 200
First year with age-composition data	**1**; 11; 16
Estimated selectivity form	**(A) As for operating model**; (B) logistic based on double normal for both fleets; (C) spline (five knots; see Fig. 1a-c) for both fleets; (D) spline for both fleets but scaled; (E) double normal for both fleets; and (F) AIC-selected
Estimate time-varying selectivity	**No**; **yes (σ_s estimated)**; $\sigma_s = 0.2$; $\sigma_s = 0.8$
M and *h* estimated	**Neither**; *M* estimated; *h* estimated; *M* and *h* estimated

(A-F[3] with six for each of time-invariant selectivity and six with annual variation in selectivity in which σ_S is estimated using Equation 19 of Xu et al., 2019; the mathematical specifications for the options are given in Supplementary Appendix A). This experiment is based on a spawning biomass (depletion) in 2018 of 0.5 relative to unfished spawning biomass (close to the target level for SESSF species of 0.48), the increasing-then-decreasing catch series, an index CV of 0.3, and an effective sample size of 100 for the age-composition data. The remaining experiments are based on tiger flathead.

2.4.2. *Experiment 2*

Experiment 2 examines the sensitivity of the results of experiment 1 for tiger flathead to the shape of the selectivity pattern for fleet 2.

2.4.3. *Experiment 3*

Experiment 3 examines the sensitivity of the results of experiment 1 for tiger flathead to the depletion of the stock in 2018 and the catch history.

2.4.4. *Experiment 4*

Experiment 4 compares the four ways to account for time-variation in selectivity for tiger flathead and the base-case selectivity patterns (see Table 2; row "Estimating time-varying selectivity").

2.4.5. *Experiment 5*

Experiment 5 examines how estimation performance deteriorates as

[3] Selectivity formulation F involves selecting (for each replicate data set and separately for the cases in which selectivity is assumed to be time-invariant / time-dependent) a selectivity formulation using AIC.

Table 3
Summary of the seven experiments.

Factor	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5	Experiment 6	Experiment 7
Species	All three	Flathead	Flathead	Flathead	Flathead	Flathead	Flathead
True selectivity	Base-case (time-varying and time-invariant)	All logistic; Low dome for fleet 2; Median dome for fleet 2; High dome for fleet 2 (time-varying and time-invariant; double logistic for both)	Base-case (time-varying and time-invariant)	Base-case (time-varying and time-invariant)	Base-case (time-varying and time-invariant)	Base-case (time-varying and time-invariant)	Base-case (time-varying and time-invariant)
2018 depletion	0.5	0.5	0.25, 0.5	0.5	0.5	0.5	0.5
Catch series	Increasing then decreasing	Increasing then decreasing	All four	Increasing then decreasing	All four	Increasing then decreasing	Increasing then decreasing
First year with age data	1	1	1	1	1, 6, 11	1	1
Index CV	0.3	0.3	0.3	0.3	0.3	0.3	0.1; 0.3; 0.5
Age composition effective sample size	100	100	100	100	100	100	10; 100; 200
Estimated selectivity	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu et al., 2019)	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019])	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019])	A for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019]), half true σ_s and double true σ_s)	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019])	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019])	A-F for time-invariant and time-varying selectivity (Eqn 19 of Xu [et al., 2019])
Treatment of *M* and *h*	Known	Known	Known	Known	Known	Known; *M* estimated; *h* estimated; *M* and *h* estimated	Known
Number of simulation trials	3 × 2×12	10 × 12	2 × 2×4 × 12	2 × 4×6	2 × 4×4 × 12	2 × 12×4	2 × 3×3 × 12

the number of years with age-composition data is reduced. Results are shown for the four catch series given that estimation performance is expected to be sensitive to catch series.

2.4.6. Experiment 6

This experiment explores the consequences of estimating, rather than "knowing", the values for natural mortality *M*, and the steepness of the stock-recruitment relationship *h*. There are four cases: 1) *M* and *h* are known (the base case); 2) *M* is estimated and *h* is known; 3) *h* is estimated and *M* is known; and 4) both *M* and *h* are estimated.

2.4.7. Experiment 7

Experiment 7 examines the sensitivity of the results of experiment 1 for tiger flathead to changing the CV of the index of abundance, and the effective sample size for the age-composition data. The analyses are based on tiger flathead, the increasing-then-decreasing catch series, and assume that the 2018 depletion is 0.5.

2.5. The performance metrics

The model outputs (management quantities) used to evaluate the implications of different sampling schemes are:

- the natural mortality rate (if estimated);
- the steepness of the stock-recruitment relationship (if estimated);
- the selectivity at age 20 (tiger flathead and blue grenadier) and at age 9 (school whiting) for the two fleets in the last year with catches (2017);
- unfished spawning biomass (SSB_0);
- the spawning biomass at the start of 2018 (one year beyond the end of the assessment);
- the depletion (spawning biomass at the start of 2018 relative to expected unfished spawning biomass); and
- the value of the Recommended Biological Catch (RBC) for 2018 based on the 20–35-48 harvest control rule used (Fulton et al., 2019) as the basis for decision making for the SESSF (Supplementary Fig. S1).

The results of the simulations are quantified by the median (over 200 replicates) of the absolute values of the relative differences between the "true" (operating model) and estimated values of the quantities of management interest (Median Absolute Relative Errors, MAREs).

The above statistics quantify how well the parameters and quantities of management interest are estimated. The results are also summarized in terms of coverage probability, i.e., the probability that the 90% confidence intervals estimated by inverting the Hessian matrix for the parameters and applying the delta method for derived quantities, include the true value of a parameter or derived variable.

3. Results

3.1. Experiment 1

3.1.1. Tiger flathead

Fig. 2 summarizes the performances of the five alternative forms for selectivity and the AIC-selected approach for tiger flathead when selectivity is time-invariant in the operating model, and the estimation methods make the same assumption (see Supplementary Figs 2a-c for results when the estimation model includes time-varying selectivity and when selectivity is time-varying in the operating model). Apart from the model based on logistic selectivity for both fleets, all estimation methods provide close to unbiased estimates (median relative errors < 5%) of

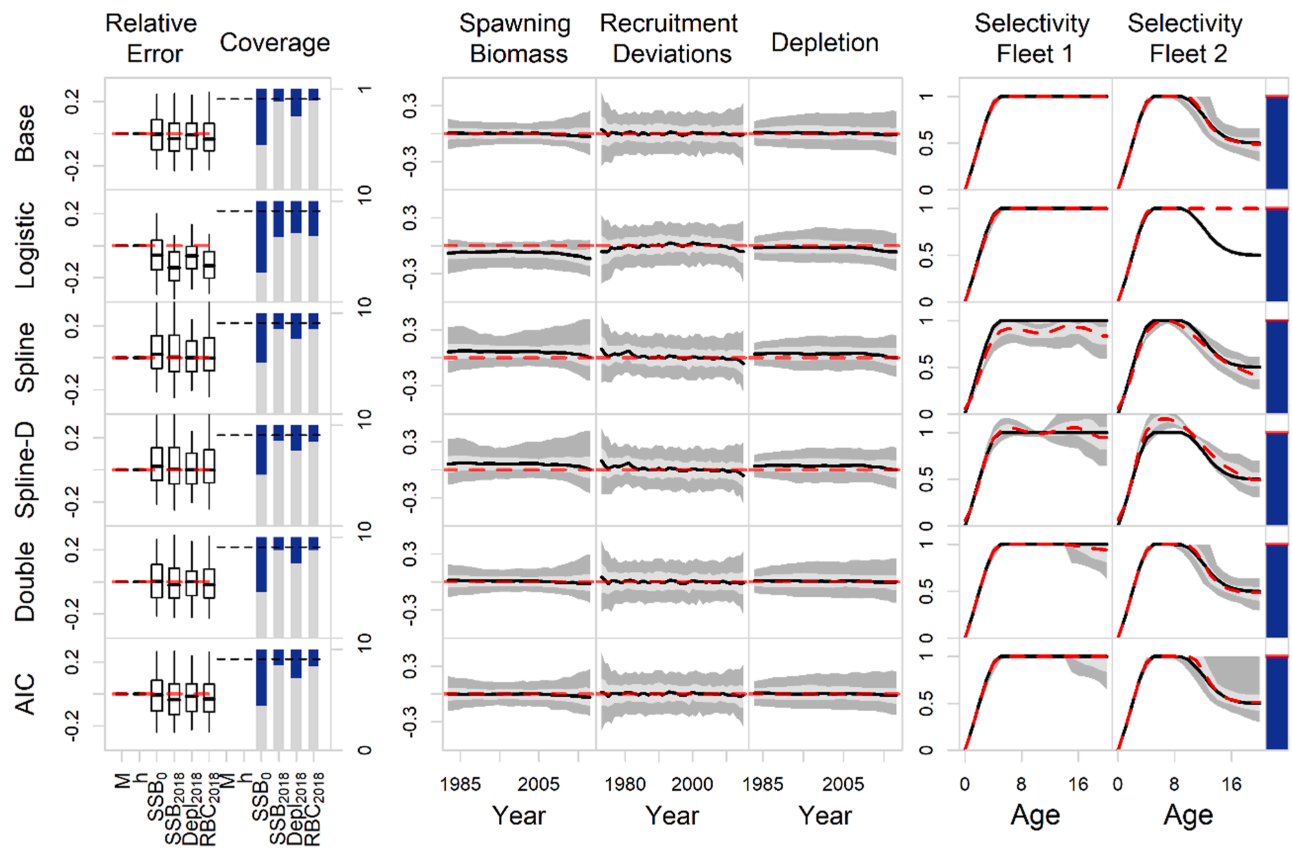

Fig. 2. The results for tiger flathead when the operating model and estimation method assume time-invariant selectivity. Relative error distributions (boxes 50% of distributions; bars extend to 90% intervals) for the model outputs and the coverage probability of 90% confidence intervals for these outputs (the gray bars indicate the proportion of simulations for which the true value is in the estimated 90% confidence intervals and the dotted line is placed at 0.9) (1st column), relative error distributions for spawning biomass, the deviations in recruitment about the stock-recruitment relationship, and depletion (columns 2–4; black line median, dashed red line no error, light shading 50% intervals, dark shading 90% intervals), the distributions for the estimates of selectivity, with the operating model selectivity indicated by the black lines, and final column indicates the proportion of replicates that led to a positive definite Hessian matrix (solid red line indicates 100%; Experiment 1 results for Species 2 and 3 lead to some non-positive definite Hessian matrices [see Supplementary Materials]).

spawning biomass[4], the deviations in recruitment about the stock-recruitment relationship, and depletion (spawning biomass relative to unfished spawning biomass). The two spline-based approaches lead to wider among-simulation intervals for spawning biomass and some evidence for bias. The estimates of SSB_0, SSB_{2018}, $Depl_{2018}$, and RBC_{2018} are also close to unbiased except for those produced when selectivity is assumed to logistic for both fleets (Fig. 2; Supplementary Figs 2a-c). The estimates of selectivity are least biased and most precise, as expected, for the base selectivity form because this form matches the operating model exactly, and most biased (almost as expected) when both fleets are assumed to have logistic selectivity. The cases in which the functional form for selectivity allows selectivity to be dome-shaped for both fleets (spline, spline-D, and double) perform adequately, but there are some cases when selectivity for fleet 1 is estimated to be dome-shaped rather than asymptotic (Fig. 2, right panels).

For the operating model-estimation combination in Fig. 2, the correct selectivity form is selected in 36% of simulations, with AIC selecting logistic selectivity for both fleets in 16% of simulations, and both fleets having dome-shaped double normal selectivity in 46% of simulations (Fig. 3a) [the equivalent values for time-varying selectivity in the operating model are 25%, 27% and 37% Fig. 3b], indicating that AIC can select both more and less complex selectivity formulations and does not always correctly select the correct form even when there are 'good' data and no model mis-specification. The spline forms for selectivity were almost never selected by AIC when the operating model and estimation method had time-invariant selectivity, even though they led to the estimates with the lowest absolute relative errors for management quantities in 15% and 1% of the replicates (Fig. 3a,b). In contrast, spline selectivity was selected in 12% of simulation replicates when the estimation method allowed for time-varying selectivity (Fig. 3b). The results in Fig. 3 are not sensitive to the whether the operating model allows for time-varying selectivity (Supplementary Fig. S3).

Table 4 compares the selectivity forms for tiger flathead for the four key quantities of management interest in terms of MAREs.[5] The base form usually leads to the lowest MAREs (although Fig. 3a,b shows that assuming this form does not always lead to estimates of the lowest absolute relative errors for each simulation replicate), with the double normal form often performing equally well (or even better). Logistic selectivity performs poorest with MAREs that are more than 120% larger than for the base form depending on the management quantity, and whether the operating model and estimation method allow for time-varying selectivity. The two spline forms perform very similarly but often poorer than the base form. The AIC-selected form does not perform as well as the base or double normal forms and performs amongst the poorest when the operating model includes time-varying selectivity. Consequently, the AIC-selected form is not a focus for the remaining

[4] Natural mortality and the steepness of the stock-recruitment relationship are set to true values for these analyses so no relative error distributions are provided.

[5] A difference of 10% are considered substantial enough to warrant attention.

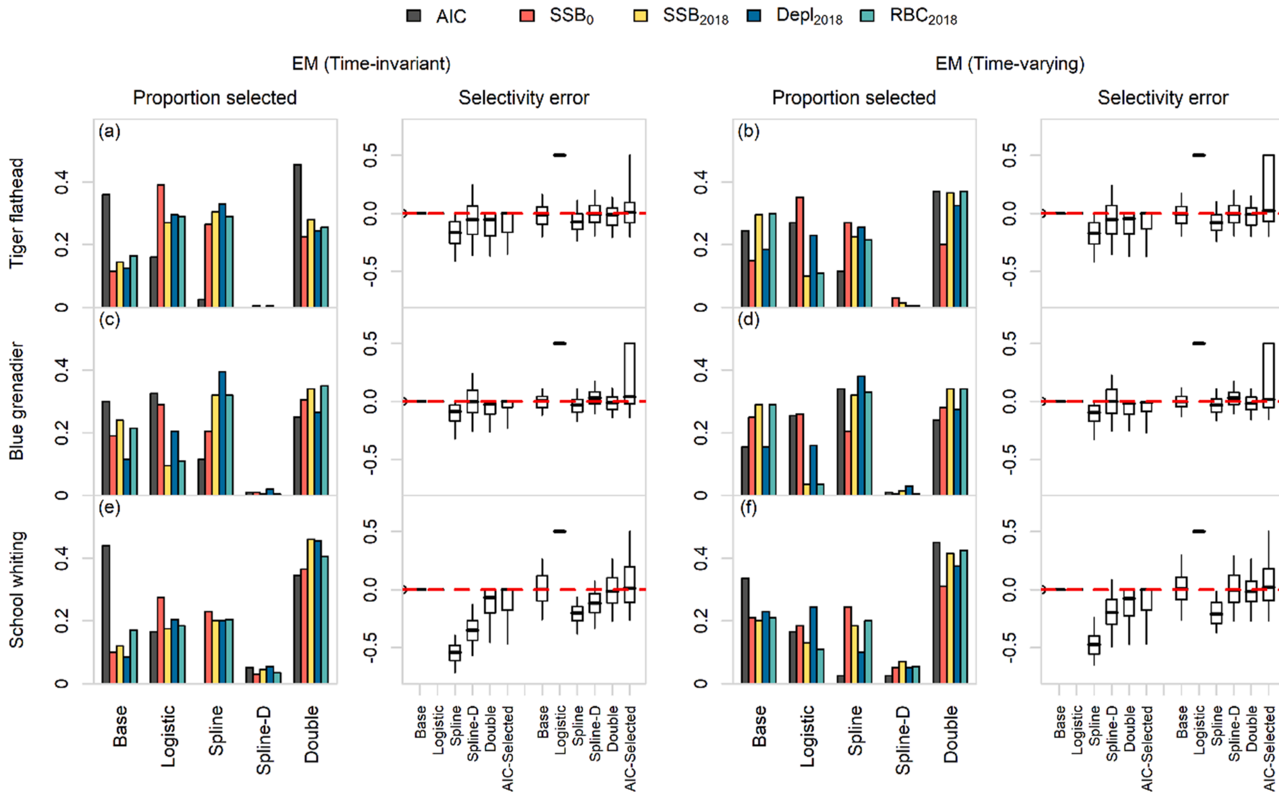

Fig. 3. Frequency that a selectivity form was selected by AIC and that led to the lowest absolute relative errors by model output (colors), and the relative error distributions for the estimates of selectivity for the oldest age in 2017 by fleet when there is time-invariant selectivity in the operating model. Results are shown in the upper panels for tiger flathead, in the center panels for blue grenadier, and in the lower panels for school whiting. The results in the first two columns for each species pertain to when selectivity is time-invariant in the estimation method while those in the right two columns pertain to when selectivity is time-varying in the estimation method. For the selectivity error plots, the first set of six represent time-invariant selectivity in the operating model and the second set of six represent time-varying selectivity in the operating model.

analysis of results. Including time-varying selectivity in the operating model leads to larger relative errors for the base form when the estimation method has time-invariant selectivity, but allowing for time-varying selectivity in the estimation method generally did not improve estimation performance, particularly for SSB_0 and SSB_{2018} (Table 4).

The coverage probabilities are less than the nominal level for all quantities, with lowest coverage probability for SSB_0 and the highest for SSB_{2018} and RBC_{2018} (Fig. 2; Supplementary Fig. S2a-c).

3.1.2. Blue grenadier and school whiting

The estimation methods are generally able to converge with positive definite Hessian matrices, except when the assessment is based on the double form for blue grenadier. The results for blue grenadier and school whiting are qualitatively the same as those for tiger flathead (Figs. 3c-f; 4; Supplementary Fig. S4; Table 4). However, the MAREs are generally higher for blue grenadier and lower for school whiting and the two spline forms lead to noticeably better estimates for 2018 depletion than other forms for blue grenadier and poorer estimates for school whiting.

3.2. Experiment 2

Allowing for a stronger dome in selectivity in the operating model ("Low dome" in Fig. 5 and Table 5) leads to larger MAREs when selectivity is assumed in the estimation method to be logistic for both fleets (MAREs $>$ 3 times that for the base form in some cases). In contrast, logistic selectivity in the estimation method performs as well as, and often better than the base form for the "High dome" and "Both logistic" operating models, and the spline and spline-D forms perform almost as well as "logistic" when selectivity is time-invariant in the estimation method. As before, the "double" form in the estimation method performs as well as (or slightly ($<$5% lower MARE) better) than base (and "logistic"). The two spline forms again perform between the base/double and logistic. The relative errors are negative for the base and logistic forms if selectivity is double normal for both fleets in the operating model (Fig. 5), with the consequence that the MAREs are higher for this operating model than for the remaining operating models. In contrast, the spline and spline-D forms are able to detect that the two fleets have dome-shaped selectivity (Fig. 6, right panels) and the two spline forms and double normal perform better than the base form for the operating models with double normal selectivity. For this operating model, the AIC-selected formulation is predominantly "double" when selectivity is time-invariant in the estimation model, which is the correct formulation (Fig. 7g,h,o,p).

Overall, the double form with time-invariant selectivity in the estimation method leads to the best performance when selectivity is double-normal in the operating model and is never poorest. This is not the case for the other forms that could allow for flexible selectivity (spline, spline-D and AIC-selected).

3.3. Experiment 3

The relative performances of the various selectivity forms are generally robust to the depletion of the population in 2018 (0.25 or 0.5) and the pattern of historical catches (Table 6; Supplementary Table 1). However, the MAREs are generally lower when depletion is 0.25 (the exceptions being depletion in 2018, for which the MARE is lower when depletion is 0.5 and when the catch history follows the "high initial" pattern). Although logistic selectivity is usually outperformed by base selectivity, the magnitude to which this is the case is lower for a depletion of 0.25 than a depletion of 0.5, most likely because there are

Table 4
Median Absolute Relative Errors (MAREs) for experiment 1 for tiger flathead, blue grenadier and school whiting. The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

	Estimation method (time-invariant selectivity)						Estimation method (time-varying selectivity)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
Tiger flathead (time-invariant selectivity in the OM)												
SSB_0	0.092	1.19	1.16	1.16	1.10	1.11	1.10	1.23	1.06	1.06	1.12	1.12
SSB_{2018}	0.099	1.50	1.06	1.06	0.98	1.03	0.95	2.22	1.42	1.42	1.02	1.45
$Depl_{2018}$	0.083	1.24	1.12	1.12	0.98	1.01	1.08	1.74	1.40	1.40	0.95	1.39
RBC_{2018}	0.094	1.38	1.05	1.05	0.97	1.05	0.97	2.16	1.45	1.45	1.00	1.48
Tiger flathead (time-varying selectivity in the OM)												
SSB_0	0.103	1.05	0.99	0.99	1.07	1.05	0.88	1.06	0.95	0.95	0.91	1.04
SSB_{2018}	0.120	1.24	1.04	1.04	0.99	1.05	0.99	1.92	1.10	1.10	0.96	1.32
$Depl_{2018}$	0.092	1.14	1.04	1.04	0.97	1.02	1.06	1.65	1.21	1.21	1.07	1.29
RBC_{2018}	0.110	1.28	1.11	1.11	1.08	1.12	1.01	1.94	1.18	1.18	1.01	1.36
Blue grenadier (time-invariant selectivity in the OM)												
SSB_0	0.190	1.34	1.01	1.08	1.09	0.98	0.84	1.40	0.96	0.96	0.89	0.91
SSB_{2018}	0.144	2.26	1.08	1.08	1.13	1.32	0.94	3.15	1.07	1.07	1.05	1.34
$Depl_{2018}$	0.145	1.53	0.75	0.76	0.92	1.02	1.38	1.95	0.89	0.89	1.16	1.21
RBC_{2018}	0.136	2.25	1.06	1.10	1.05	1.32	1.08	3.15	1.08	1.08	1.06	1.40
Blue grenadier (time-varying selectivity in the OM)												
SSB_0	0.194	1.23	0.98	0.98	1.04	0.89	0.91	1.36	0.91	0.92	0.97	0.97
SSB_{2018}	0.166	1.78	0.97	0.97	1.03	1.10	0.98	2.63	0.95	0.96	0.95	1.25
$Depl_{2018}$	0.140	1.37	0.77	0.79	1.01	0.96	1.39	2.00	0.86	0.87	1.23	1.22
RBC_{2018}	0.156	1.76	1.02	0.99	1.00	1.19	1.09	2.70	1.05	1.05	1.06	1.33
School whiting (time-invariant selectivity in the OM)												
SSB_0	0.062	1.03	1.05	1.14	0.92	1.11	1.01	1.22	1.01	0.97	1.04	1.03
SSB_{2018}	0.094	1.09	1.17	1.20	1.00	1.10	1.24	1.89	1.86	1.65	1.26	1.40
$Depl_{2018}$	0.080	1.03	1.17	1.20	0.96	0.97	1.22	1.49	1.76	1.59	1.17	1.18
RBC_{2018}	0.066	1.22	1.23	1.30	0.98	1.09	1.31	2.01	1.91	1.62	1.31	1.42
School whiting (time-varying selectivity in the OM)												
SSB_0	0.068	1.03	1.02	1.08	0.99	1.01	0.91	1.11	1.04	0.98	0.96	1.04
SSB_{2018}	0.100	1.13	1.34	1.41	0.99	1.08	1.24	1.76	1.76	1.50	1.30	1.29
$Depl_{2018}$	0.094	1.07	1.17	1.20	1.10	1.08	1.16	1.31	1.53	1.36	1.18	1.10
RBC_{2018}	0.078	1.11	1.24	1.31	0.98	1.06	1.06	1.72	1.70	1.56	1.25	1.27

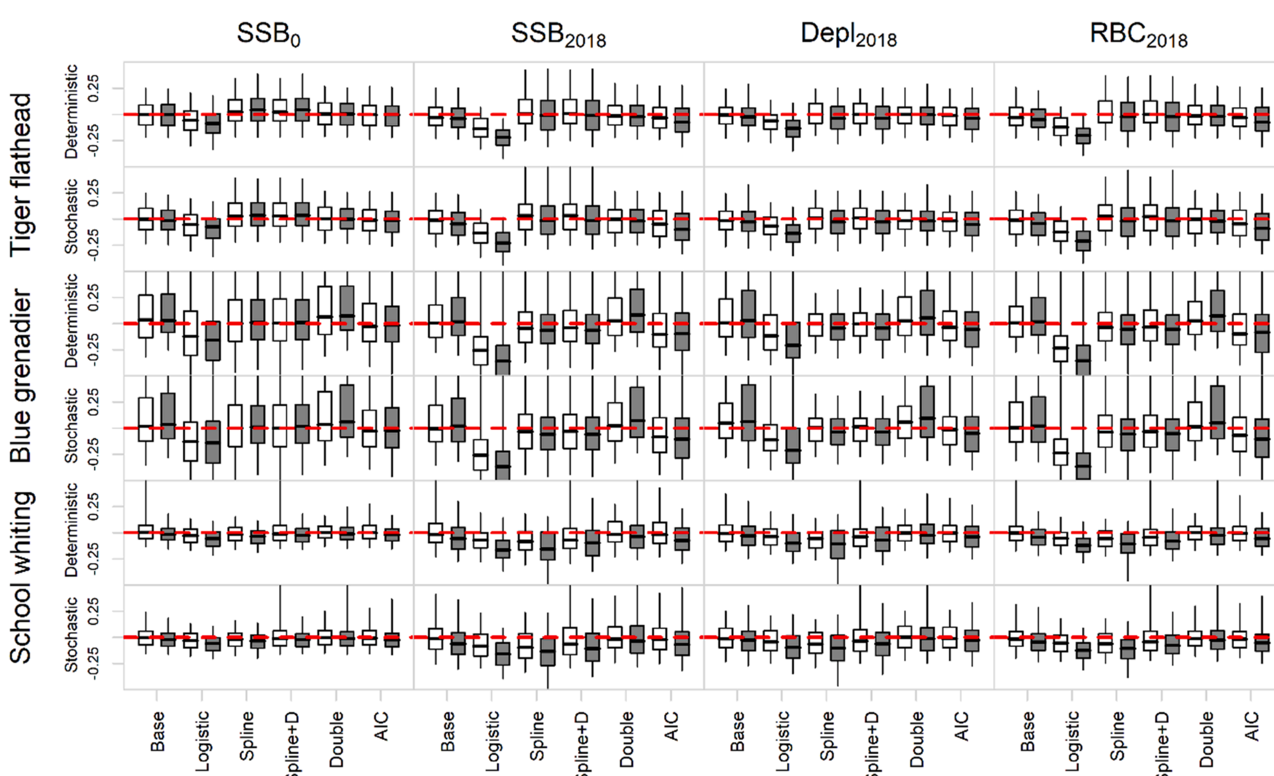

Fig. 4. Relative error distributions for experiment 1. The rows show results by species and when selectivity is time-invariant or time-varying in the operating model. The open and shaded box plots represent results for time-invariant / time-varying selectivity in the estimation method.

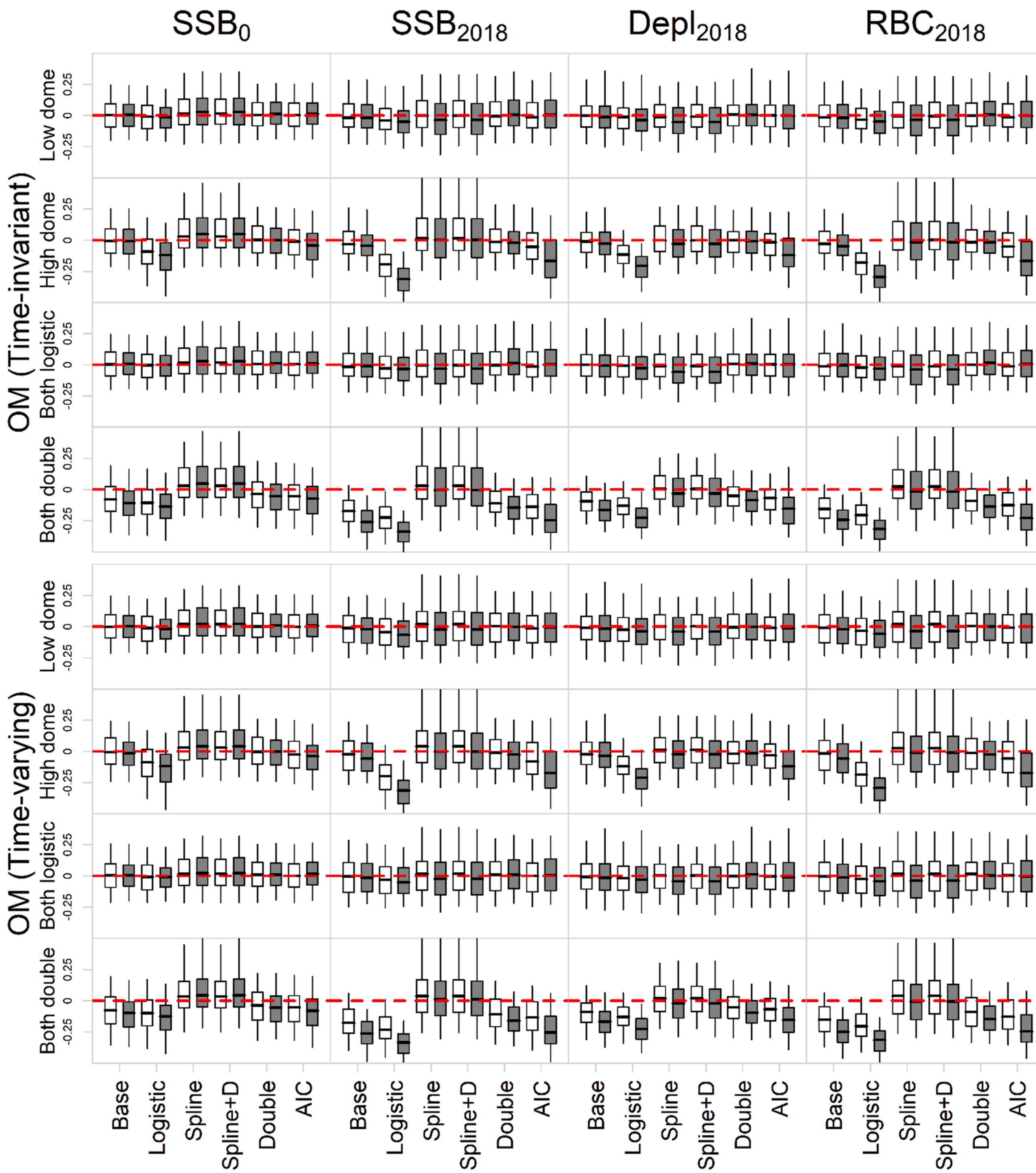

Fig. 5. Relative error distributions for experiment 2. The rows show results for tiger flathead for four of the operating model selectivity forms and whether selectivity is time-invariant or time-varying in the operating model. The open and shaded box plots represent results for time-invariant / time-varying selectivity in the estimation method.

far fewer old animals when depletion is 0.25. The MAREs are higher for the 'constant', 'high initial' and particularly the 'increasing' catch series compared to the 'increasing and decreasing' catch series, which is not surprising given that these series would be considered to be less informative that the "increasing-then-decreasing" series (Table 6; Supplementary Table 1).

3.4. Experiment 4

The effects of fixing rather than estimating the values for the extent to which selectivity varies over time (σ_S) depend on the "true" value for σ_S. Not unexpectedly, performance is about the same for σ_S= 0.2 and not allowing for stochastic selectivity in the estimation method (Table 7). However, setting σ_S= 0.8 leads to generally poorer performance for the base, logistic and spline selectivity patterns when selectivity is time-invariant in the operating model (Table 7). The ability to estimate

Table 5
Median Absolute Relative Errors (MAREs) for experiment 2. The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

(a) Time-invariant selectivity in the operating model												
	Estimation method (time-invariant selectivity)						Estimation method (time-varying selectivity)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
Base												
SSB_0	0.092	1.19	1.16	1.16	1.10	1.11	1.10	1.23	1.06	1.06	1.12	1.12
SSB_{2018}	0.099	1.50	1.06	1.06	0.98	1.03	0.95	2.22	1.42	1.42	1.02	1.45
$Depl_{2018}$	0.083	1.24	1.12	1.12	0.98	1.01	1.08	1.74	1.40	1.40	0.95	1.39
RBC_{2018}	0.094	1.38	1.05	1.05	0.97	1.05	0.97	2.16	1.45	1.45	1.00	1.48
Low dome												
SSB_0	0.103	1.25	1.05	1.05	1.02	1.09	0.98	1.36	1.04	1.04	1.00	1.15
SSB_{2018}	0.098	2.00	1.27	1.26	0.98	1.17	1.00	3.20	1.57	1.57	0.98	1.96
$Depl_{2018}$	0.087	1.38	1.16	1.16	0.96	1.14	1.03	2.37	1.38	1.38	0.96	1.77
RBC_{2018}	0.097	1.88	1.22	1.22	0.94	1.12	0.96	3.06	1.50	1.50	0.94	1.92
High dome												
SSB_0	0.093	0.98	1.13	1.10	0.95	0.98	0.94	0.96	1.05	1.04	0.97	0.94
SSB_{2018}	0.095	1.00	1.05	1.04	1.04	1.05	1.01	1.00	1.40	1.40	1.05	1.14
$Depl_{2018}$	0.086	0.97	1.09	1.08	0.99	1.00	1.02	1.14	1.28	1.28	0.99	1.09
RBC_{2018}	0.087	1.01	1.16	1.15	1.08	1.08	1.15	1.06	1.54	1.54	1.14	1.24
Both logistic												
SSB_0	0.092	1.06	1.10	1.10	0.94	1.04	0.91	0.91	1.05	1.05	0.96	0.90
SSB_{2018}	0.091	1.09	1.09	1.09	0.99	1.10	1.04	1.08	1.43	1.43	1.02	1.17
$Depl_{2018}$	0.085	1.00	1.08	1.08	1.00	1.01	1.03	1.12	1.33	1.33	1.02	1.10
RBC_{2018}	0.092	1.00	1.09	1.09	0.99	1.02	1.03	1.11	1.40	1.40	1.01	1.12
Both double normal (base dome)												
SSB_0	0.119	1.09	0.97	0.97	0.97	1.00	1.08	1.28	0.94	0.99	0.93	1.09
SSB_{2018}	0.180	1.26	0.69	0.69	0.72	0.88	1.45	1.87	0.85	0.85	0.86	1.42
$Depl_{2018}$	0.105	1.28	0.99	0.99	0.85	0.93	1.62	2.15	1.15	1.15	1.09	1.67
RBC_{2018}	0.161	1.28	0.66	0.66	0.70	0.87	1.51	1.96	0.94	0.93	0.89	1.48
(b) Time-varying selectivity in the operating model												
	Estimation method (time-invariant selectivity)						Estimation method (time-varying selectivity)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
Base												
SSB_0	0.103	1.05	0.99	0.99	1.07	1.05	0.88	1.06	0.95	0.95	0.91	1.04
SSB_{2018}	0.120	1.24	1.04	1.04	0.99	1.05	0.99	1.92	1.10	1.10	0.96	1.32
$Depl_{2018}$	0.092	1.14	1.04	1.04	0.97	1.02	1.06	1.65	1.21	1.21	1.07	1.29
RBC_{2018}	0.110	1.28	1.11	1.11	1.08	1.12	1.01	1.94	1.18	1.18	1.01	1.36
Low dome												
SSB_0	0.109	1.10	1.09	1.09	1.02	1.04	0.85	1.30	0.96	0.96	0.89	0.97
SSB_{2018}	0.127	1.57	1.07	1.07	1.00	1.06	0.96	2.49	1.13	1.13	0.95	1.66
$Depl_{2018}$	0.094	1.38	1.10	1.13	0.94	1.06	1.04	2.25	1.16	1.16	1.05	1.63
RBC_{2018}	0.118	1.58	1.07	1.07	1.04	1.13	1.00	2.49	1.18	1.18	0.95	1.80
High dome												
SSB_0	0.095	0.98	1.04	1.04	1.03	0.99	0.92	0.86	1.03	1.03	0.97	0.93
SSB_{2018}	0.114	0.97	1.03	1.03	1.00	0.99	0.96	1.06	1.20	1.20	0.96	1.04
$Depl_{2018}$	0.095	0.97	1.03	1.03	1.02	1.03	1.11	1.14	1.19	1.19	1.09	1.16
RBC_{2018}	0.111	1.00	1.11	1.11	1.03	1.01	0.98	1.07	1.16	1.16	0.95	1.07
Both logistic												
SSB_0	0.097	0.97	1.04	1.04	1.05	0.98	0.88	0.84	1.03	1.03	0.90	0.90
SSB_{2018}	0.120	0.87	0.96	0.96	0.96	0.94	0.96	0.97	1.14	1.14	0.93	1.04
$Depl_{2018}$	0.101	0.95	0.95	0.95	0.95	1.00	1.01	1.06	1.15	1.15	1.03	1.13
RBC_{2018}	0.114	1.00	1.04	1.04	1.00	1.00	0.97	0.99	1.15	1.15	0.99	1.10
Both double normal (base dome)												
SSB_0	0.112	1.13	1.05	1.05	1.01	1.01	1.13	1.32	0.99	0.99	0.94	1.13
SSB_{2018}	0.180	1.29	0.77	0.77	0.74	0.87	1.46	1.86	0.78	0.78	0.92	1.48
$Depl_{2018}$	0.108	1.25	0.94	0.94	0.82	0.86	1.65	2.12	1.05	1.05	1.05	1.51
RBC_{2018}	0.156	1.32	0.84	0.84	0.82	0.96	1.61	2.02	0.87	0.87	0.98	1.59

depletion is much poorer when σ_S= 0.8 in the estimation methods compared to estimating σ_S or setting it 0.2 when σ_S= 0.4 in the operating model, and the logistic form generally performs more poorly when σ_S= 0.8 in the estimation method.

3.5. Experiment 5

With few exceptions, not having age-composition data from the start of the fishery leads, as expected, to poorer estimation performance. However, the degradation in estimation performance is not uniform across the selectivity forms. In particular, the performances of the two spline forms degrade the most when data are only available from year 6 and then year 11 (Table 8; Supplementary Table 2), with the effects most marked for the 'constant' catch series (Table 8). The ability of the double form to estimate 2018 depletion and to a lesser extent 2018 spawning biomass and 2018 RBC also gets poorer with fewer data. The ability to estimate the management quantities does not deteriorate as much when there are no data for the earliest years of the fishery when selectivity is logistic for both fleets compared to the other forms. However, logistic selectivity is still outperformed by the base and double forms.

3.6. Experiment 6

Experiment 6 explores the ability to estimate M, h and selectivity simultaneously, and the impact on the ability to estimate quantities of management interest. The MAREs for the management quantities for the base form are notably larger when M is estimated and the operating model has time-invariant selectivity (Table 9a). However, the effects on

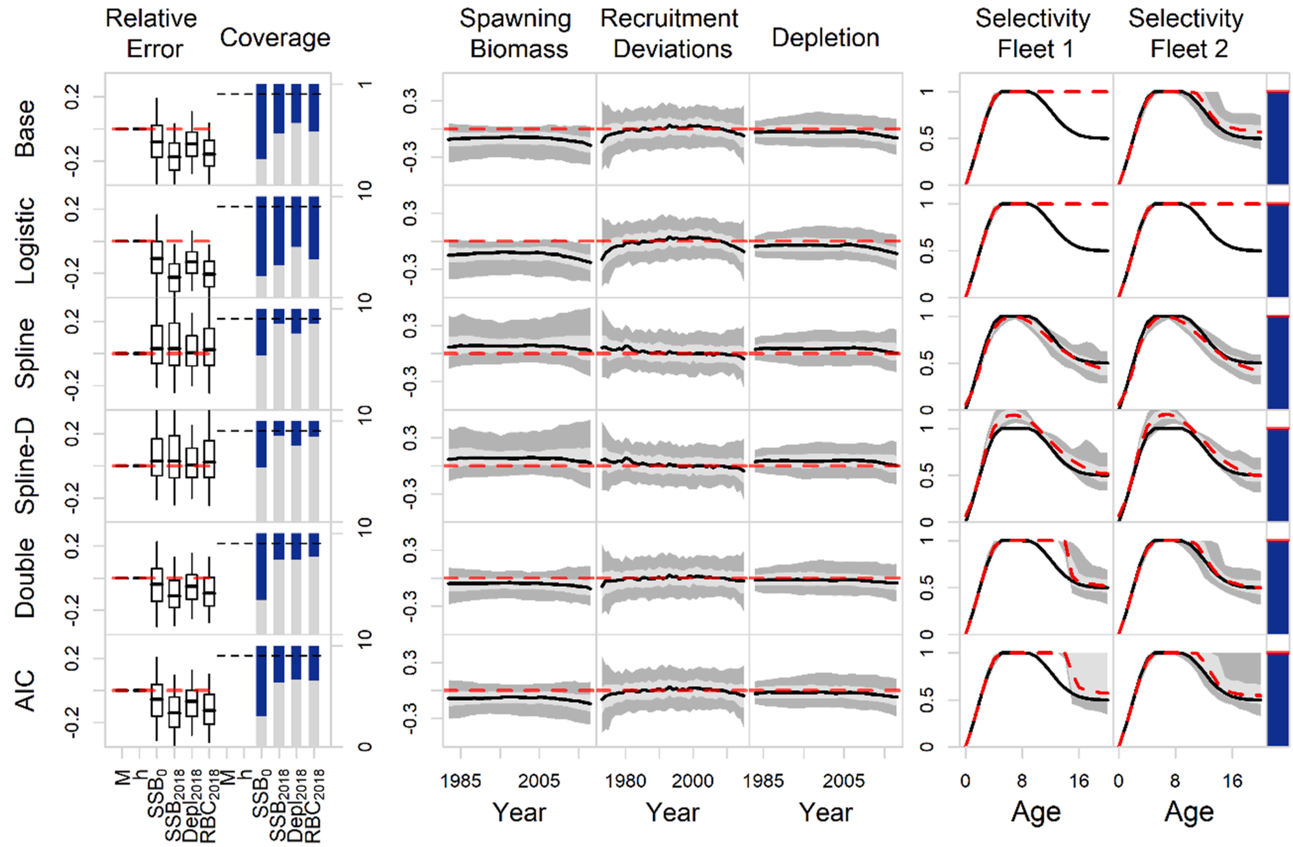

Fig. 6. As for Fig. 2, except that the results pertain to experiment 2 when the operating model has double-normal selectivity for both fleets.

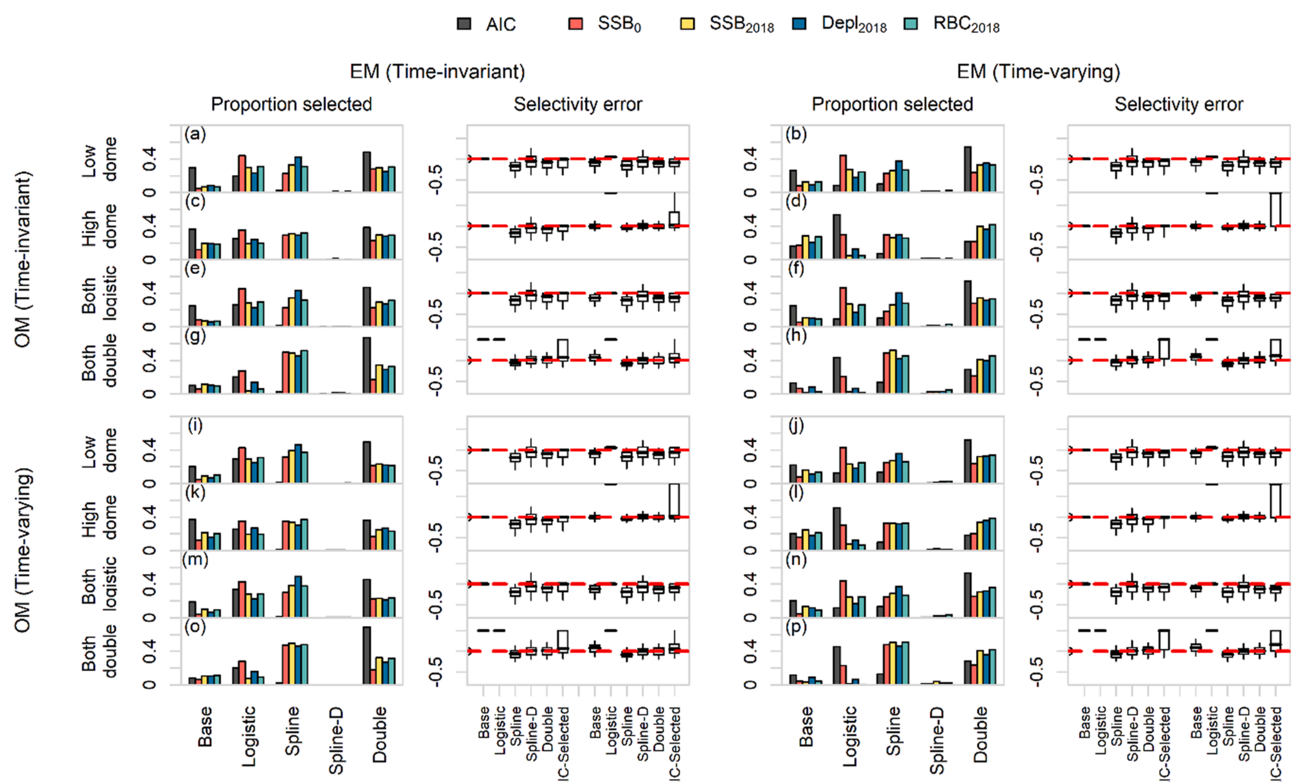

Fig. 7. As for Fig. 3, except the results pertain to experiment 2.

Table 6
Median Absolute Relative Errors (MAREs) for experiment 3 when selectivity is time-varying in the operating model and in the estimation method (results are shown for all combinations of whether selectivity is time-varying or not in Supplementary Tables 1). The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

	$Depl_{2018} = 0.25$						$Depl_{2018} = 0.5$ (base)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
Increasing-then-decreasing catch series												
SSB_0	0.075	1.13	1.09	1.09	1.04	1.03	0.092	1.19	1.16	1.16	1.10	1.11
SSB_{2018}	0.091	1.16	1.13	1.13	1.00	1.03	0.099	1.50	1.06	1.06	0.98	1.03
$Depl_{2018}$	0.095	1.04	1.14	1.15	0.99	0.98	0.083	1.24	1.12	1.12	0.98	1.01
RBC_{2018}	0.087	1.19	1.23	1.24	1.00	1.04	0.094	1.38	1.05	1.05	0.97	1.05
Increasing catch series												
SSB_0	0.085	1.12	1.11	1.10	1.00	0.98	0.121	1.25	1.25	1.21	1.05	1.08
SSB_{2018}	0.117	1.30	1.50	1.46	1.07	1.12	0.138	1.84	1.40	1.39	1.00	1.09
$Depl_{2018}$	0.114	0.99	1.11	1.08	0.96	0.95	0.090	1.44	1.16	1.16	0.96	1.08
RBC_{2018}	0.107	1.29	1.50	1.46	1.04	1.12	0.135	1.72	1.41	1.36	1.00	1.08
Constant catch series												
SSB_0	0.081	0.98	1.05	1.03	0.98	0.98	0.108	1.13	1.25	1.25	1.01	1.01
SSB_{2018}	0.098	1.16	1.30	1.29	0.99	0.99	0.119	1.55	1.33	1.33	1.00	1.05
$Depl_{2018}$	0.104	1.11	1.17	1.16	0.96	1.14	0.090	1.34	1.23	1.23	0.97	1.15
RBC_{2018}	0.094	1.15	1.22	1.19	1.00	1.00	0.112	1.49	1.25	1.25	1.01	1.07
High initial catch series												
SSB_0	0.083	0.95	1.06	1.06	1.00	1.02	0.087	1.35	1.23	1.25	1.05	1.14
SSB_{2018}	0.096	1.12	1.35	1.25	1.08	1.10	0.111	1.56	1.62	1.47	0.97	1.15
$Depl_{2018}$	0.113	1.02	1.16	1.12	1.00	1.00	0.093	1.26	1.39	1.28	0.95	1.27
RBC_{2018}	0.094	1.09	1.28	1.25	1.04	1.08	0.108	1.36	1.47	1.37	0.94	1.01

Table 7
Median Absolute Relative Errors (MAREs) for experiment 4. The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

	σ_S fixed at 0.2 in the estimation method						σ_S fixed at 0.8 in the estimation method					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
No stochastic selectivity in the OM												
SSB_0	0.091	1.21	1.18	1.18	1.12	1.15	1.01	1.40	1.13	1.13	1.04	1.15
SSB_{2018}	0.099	1.51	1.07	1.07	0.99	1.06	1.20	3.06	1.52	1.52	1.15	1.96
$Depl_{2018}$	0.083	1.23	1.12	1.12	0.99	1.05	1.40	2.65	1.67	1.67	1.32	1.82
RBC_{2018}	0.091	1.43	1.07	1.07	1.01	1.09	1.26	3.22	1.76	1.76	1.30	2.01
Stochastic selectivity in the OM												
SSB_0	0.101	1.07	1.00	1.00	1.08	1.07	0.86	1.17	0.93	0.93	0.90	1.08
SSB_{2018}	0.120	1.26	1.05	1.05	1.02	1.06	1.04	2.47	1.21	1.21	1.00	1.39
$Depl_{2018}$	0.092	1.14	1.07	1.05	0.96	1.02	1.39	2.28	1.48	1.48	1.25	1.72
RBC_{2018}	0.112	1.28	1.10	1.09	1.05	1.14	1.17	2.67	1.37	1.37	1.04	1.61

the performance of the base form of estimating M when the operating model has time-varying selectivity and estimating h for both variants of the operating model are relatively smaller (Table 9b). The performance of the logistic form generally improves when M and h are estimated (contrast Figs. 2 and 8), likely because it frees up some parameters to account for mis-specification, with the consequence that there are cases when assuming the logistic form of selectivity for both fleets leads to the lowest MAREs. The performances of the two spline forms deteriorate markedly when M is estimated and to a lesser (but still to a substantial) extent when h is estimated. The performance of the double form also deteriorates when M and h are estimated, often leading to the base form performing best overall.

Estimation of M is best when the estimation method assumes the base form (Table 9; Fig. 8), with the MAREs for the remaining forms often twice as large as those for the base form. Steepness is poorly estimated in all cases, possible because Experiment 6 used a depletion in the final year of only 0.5, but the ability to estimate h is poorer when one of the spline forms is assumed.

3.7. Experiment 7

Reducing the effective sample size for the age-composition data from 100 to 10 leads, not surprisingly, to markedly poor estimation performance and higher MAREs (Supplementary Table 3). In contrast, the benefits in terms of lower MAREs are not as obvious when the effective sample sizes are doubled. This is perhaps not unexpected given that the variances of the age-composition proportions are increased 10-fold in one case and reduced by half in the other. Decreasing the CV of index of abundance from 0.3 to 0.1 leads to lower MAREs unless the effective sample size is also decreased from 100 to 10. The effects of changing the effective sample size on the relative performances of the selectivity forms are generally insensitive to the index CV and the effective sample size for the age-composition data. The exception to this general result is when selectivity is logistic for both fleets, in which case, reducing the effective sample size for the age-composition data leads to better performance (and often performance that is not very different from that of the best-performing selectivity forms). This improvement in performance can again be attributed to freeing up some parameters to account for mis-specification.

4. Discussion

The model structure assumptions used to relate selectivity and composition data in this study influenced the bias and precision of parameter estimates for the underlying population dynamics model and, thus, the resulting quantities of management interest. Natural mortality and selectivity are confounded due to the population dynamics, with implications for the estimation of derived outputs, such as stock size and the catch limits corresponding to a harvest control rule. As expected, estimation performance is generally best when the correct functional

Table 8
Median Absolute Relative Errors (MAREs) for experiment 5. The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

	Estimation method (time-invariant selectivity)				Estimation method (time-varying selectivity)							
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
Increasing-then-decreasing catch series												
Data from year 1												
SSB_0	0.092	1.19	1.16	1.16	1.10	1.11	1.10	1.23	1.06	1.06	1.12	1.12
SSB_{2018}	0.099	1.50	1.06	1.06	0.98	1.03	0.95	2.22	1.42	1.42	1.02	1.45
$Depl_{2018}$	0.083	1.24	1.12	1.12	0.98	1.01	1.08	1.74	1.40	1.40	0.95	1.39
RBC_{2018}	0.094	1.38	1.05	1.05	0.97	1.05	0.97	2.16	1.45	1.45	1.00	1.48
Data from year 6												
SSB_0	0.098	1.19	1.33	1.33	1.05	1.08	0.96	1.16	1.22	1.22	0.98	1.03
SSB_{2018}	0.101	1.44	1.25	1.25	1.02	1.16	1.09	2.04	1.62	1.62	1.04	1.52
$Depl_{2018}$	0.083	1.15	1.08	1.08	0.94	1.07	1.18	1.69	1.60	1.60	1.01	1.49
RBC_{2018}	0.097	1.33	1.18	1.18	1.01	1.10	1.08	1.88	1.67	1.67	1.07	1.54
Data from year 11												
SSB_0	0.098	1.15	1.43	1.43	0.99	1.10	0.96	1.14	1.32	1.32	1.00	1.11
SSB_{2018}	0.094	1.43	1.58	1.59	1.00	1.21	1.22	2.05	2.29	2.25	1.15	1.53
$Depl_{2018}$	0.079	1.16	1.34	1.38	0.99	1.10	1.20	1.49	2.12	2.12	1.14	1.42
RBC_{2018}	0.091	1.31	1.40	1.42	0.99	1.16	1.04	1.79	2.30	2.27	1.19	1.45
Constant catch series												
Data from year 1												
SSB_0	0.108	1.13	1.25	1.25	1.01	1.01	0.81	1.19	1.24	1.24	0.99	1.02
SSB_{2018}	0.119	1.55	1.33	1.33	1.00	1.05	0.92	2.28	1.53	1.53	0.95	1.48
$Depl_{2018}$	0.090	1.34	1.23	1.23	0.97	1.15	1.06	1.96	1.24	1.24	0.99	1.64
RBC_{2018}	0.112	1.49	1.25	1.25	1.01	1.07	0.91	2.24	1.48	1.48	1.01	1.46
Data from year 6												
SSB_0	0.104	1.15	1.96	1.96	1.07	1.06	0.91	1.22	1.79	1.79	1.05	1.19
SSB_{2018}	0.123	1.41	2.01	2.01	1.01	1.08	0.94	1.92	2.00	2.00	0.91	1.33
$Depl_{2018}$	0.090	1.17	1.64	1.64	0.98	1.09	1.11	1.69	1.89	1.89	1.14	1.35
RBC_{2018}	0.115	1.33	1.84	1.84	1.04	1.05	0.95	1.86	2.05	2.08	0.95	1.27
Data from year 11												
SSB_0	0.109	1.09	260.66	250.07	0.98	1.01	0.88	1.17	286.96	289.59	0.89	1.21
SSB_{2018}	0.117	1.45	401.32	389.27	1.02	1.22	1.01	1.90	428.62	434.68	1.12	1.60
$Depl_{2018}$	0.090	1.06	6.34	6.26	0.93	1.02	1.12	1.55	6.01	6.01	1.10	1.49
RBC_{2018}	0.111	1.36	355.26	343.17	1.01	1.18	1.01	1.77	363.27	372.70	1.09	1.47

form is chosen for selectivity-at-age by fleet. However, the consequences of choosing a functional form that is not sufficiently flexible (i.e., assuming a logistic form when selectivity is actually dome-shaped) are more severe than assuming a more flexible form (e.g., a double normal form or a spline). In general, forcing at least one fleet to have asymptotic selectivity does not lead to markedly more robust estimation performance, with the exception being when there is limited contrast in the biomass time-series or the sample sizes for age-composition are low. Considerations for determining the most appropriate default assumptions about the functional form of selectivity are outlined below.

As expected, the ability to estimate selectivity and hence quantities of management interest depended on the quality of the data, with performance deteriorating markedly with lower effective sample sizes for the age-composition data and improving with a more precise index of abundance. Indeed, incorrectly assuming logistic selectivity was often the best performing approach for low effective sample sizes or shortened data time series. Better performance with a good index of abundance is not unexpected given that most of the management quantities are related to absolute abundance. Perhaps less expected, allowing for stochastic variation in selectivity did not markedly improve estimation performance, even when selectivity for young ages was time-varying in the operating model. However, the authors recommend when time-varying selectivity is estimated, it is best to estimate the extent of time-variation (e.g., following Xu et al., 2019) because setting the extent of time-variation larger than the true extent led to markedly poorer estimation performance in this study (Table 7). We also expect that other management quantities, particularly recent recruitment estimates and longer-term catch projections, will be more sensitive to time-varying selectivity.

The spline forms are the most general of those considered in this paper and performed adequately when there were data from the start of the fishery and the most informative catch series was used. However, these forms could lead to very biased estimates of selectivity and hence management quantities on occasion. The biases in estimates from the spline form increased when: a) M was estimated along with the other parameters of the model (Fig. 8), b) data were not available from the start of the fishery and c) the catch series was uninformative. Thus, while they allow for considerable flexibility, care is needed when implementing spline-based selectivity patterns for all the fleets in an assessment (see below). There was also little difference in results between standard and spline-D forms.

The base form used in this study implemented the most general interpretation of the practice of assuming that at least one selectivity pattern is asymptotic. In general, this assumption matched the specifications of the operating model. The exception was when selectivity for both fleets was actually domed-shaped. AIC often selected the double normal form when this form was correct (Fig. 7). However, the performance of the double normal form deteriorated when: a) data were not available from the start of exploitation, b) M was estimated along with selectivity (thereby increasing confounding) and c) the less informative catch series was used. Thus, no clear best practice guideline emerges. In general, assuming at least one fleet has asymptotic selectivity is robust to lack of informative data, but typically leads to bias. Use of model selection criteria can sometimes identify that selectivity is really dome-shaped for all fleets, but this is not fully reliable.

The authors anticipated that use of model selection methods could lead to improved estimation performance because the analyses of this paper included the true selectivity form in the set of estimation methods. However, this was not the case. AIC selected both forms that were too complicated (e.g., dome-shaped when the true form was asymptotic) and too simple (e.g., logistic when the true form was dome-shaped). The exact reasons for this are unclear but are probably related to the over-parameterization for the more complex forms (*sensu* Monnahan et al., 2020) and the fact that the models are not formed as state-space models

Table 9
Median Absolute Relative Errors (MAREs) for experiment 6. The leftmost "Base" column reports the MAREs and, for the remaining columns, the MARE for the cases concerned is divided by the MARE for "Base".

	Estimation method (time-invariant selectivity)						Estimation method (time-varying selectivity)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
(a) Time-invariant selectivity in the OM												
Base												
SSB_0	0.092	1.19	1.16	1.16	1.10	1.11	1.10	1.23	1.06	1.06	1.12	1.12
SSB_{2018}	0.099	1.50	1.06	1.06	0.98	1.03	0.95	2.22	1.42	1.42	1.02	1.45
$Depl_{2018}$	0.083	1.24	1.12	1.12	0.98	1.01	1.08	1.74	1.40	1.40	0.95	1.39
RBC_{2018}	0.094	1.38	1.05	1.05	0.97	1.05	0.97	2.16	1.45	1.45	1.00	1.48
Estimate *M*												
M	0.023	2.76	31.58	31.56	1.28	1.87	1.03	3.15	31.71	31.71	1.55	3.02
SSB_0	0.096	1.08	34.65	34.65	1.03	1.03	1.10	1.00	37.75	37.75	1.04	1.41
SSB_{2018}	0.111	0.84	24.48	24.48	1.06	1.02	1.13	0.87	28.06	28.06	1.22	1.44
$Depl_{2018}$	0.086	0.98	2.67	2.67	1.03	1.10	1.08	1.03	2.58	2.58	1.22	1.34
RBC_{2018}	0.122	1.16	3.04	3.01	1.15	1.08	1.12	1.06	2.72	2.72	1.20	1.44
Estimate Steepness												
h	0.333	1.00	1.31	1.31	0.99	1.00	1.00	1.00	1.64	1.64	1.00	1.00
SSB_0	0.082	1.55	3.54	3.54	1.27	1.41	1.07	1.81	5.93	5.93	1.17	1.64
SSB_{2018}	0.097	1.41	1.62	1.62	0.98	1.08	1.00	1.97	1.96	1.95	1.02	1.49
$Depl_{2018}$	0.083	0.98	2.50	2.50	1.12	1.13	1.12	1.13	4.27	4.27	1.25	1.24
RBC_{2018}	0.090	1.36	1.44	1.44	1.07	1.17	1.09	1.94	2.07	2.02	1.07	1.47
Estimate *M* and steepness												
M	0.027	2.11	27.45	27.46	1.41	1.50	0.99	1.00	27.47	27.48	1.50	2.96
h	0.333	1.00	1.56	1.52	1.00	1.00	1.00	1.00	2.20	2.20	1.00	1.00
SSB_0	0.114	0.98	34.89	35.06	0.94	1.02	1.05	1.31	46.75	47.04	1.00	1.74
SSB_{2018}	0.112	0.89	29.48	29.48	1.03	1.00	1.12	1.71	38.72	38.72	1.21	1.65
$Depl_{2018}$	0.087	1.15	2.61	2.61	1.00	1.11	1.04	1.07	2.91	2.91	1.07	1.46
RBC_{2018}	0.120	1.12	3.01	3.02	1.16	1.05	1.15	1.45	2.55	2.57	1.26	1.63
	Estimation method (time-invariant selectivity)						Estimation method (time-varying selectivity)					
	Base	Logistic	Spline	Spline-D	Double	AIC	Base	Logistic	Spline	Spline-D	Double	AIC
(b) Time-varying selectivity in the OM												
Base												
SSB_0	0.103	1.05	0.99	0.99	1.07	1.05	0.88	1.06	0.95	0.95	0.91	1.04
SSB_{2018}	0.120	1.24	1.04	1.04	0.99	1.05	0.99	1.92	1.10	1.10	0.96	1.32
$Depl_{2018}$	0.092	1.14	1.04	1.04	0.97	1.02	1.06	1.65	1.21	1.21	1.07	1.29
RBC_{2018}	0.110	1.28	1.11	1.11	1.08	1.12	1.01	1.94	1.18	1.18	1.01	1.36
Estimate *M*												
M	0.026	2.45	28.32	28.32	1.11	1.84	0.99	2.99	28.49	28.49	1.39	2.95
SSB_0	0.096	0.95	35.15	35.15	1.01	1.01	1.02	0.94	39.58	39.58	1.03	1.47
SSB_{2018}	0.123	0.95	24.28	24.28	0.99	1.01	0.96	0.88	27.93	27.93	1.07	1.33
$Depl_{2018}$	0.082	1.12	2.68	2.68	1.01	1.12	1.13	1.19	2.94	2.94	1.21	1.34
RBC_{2018}	0.127	1.20	2.91	2.91	1.07	1.13	1.01	1.06	2.66	2.66	1.28	1.36
Estimate Steepness												
h	0.333	1.00	1.29	1.29	1.00	1.00	1.00	1.00	1.64	1.64	1.00	1.00
SSB_0	0.096	1.34	2.60	2.60	1.07	1.17	0.95	1.56	4.96	4.96	1.05	1.42
SSB_{2018}	0.121	1.10	1.26	1.26	0.98	0.99	0.96	1.65	1.68	1.69	0.97	1.28
$Depl_{2018}$	0.099	0.81	2.00	2.00	0.99	0.92	0.99	0.96	3.37	3.37	1.04	1.21
RBC_{2018}	0.111	1.19	1.28	1.28	0.99	1.04	1.01	1.61	1.76	1.79	1.01	1.35
Estimate *M* and steepness												
M	0.028	2.14	26.82	26.82	1.31	1.70	0.99	1.00	26.59	26.59	1.58	3.03
h	0.333	1.00	1.51	1.51	1.00	1.00	1.00	1.00	2.20	2.19	1.00	1.00
SSB_0	0.110	1.00	37.65	37.65	0.98	1.07	1.09	1.36	49.49	49.68	1.06	1.69
SSB_{2018}	0.120	0.95	27.18	27.18	0.99	0.95	1.09	1.66	35.85	35.85	1.14	1.63
$Depl_{2018}$	0.102	1.16	1.97	1.97	0.93	0.98	1.02	0.93	2.51	2.51	0.91	1.36
RBC_{2018}	0.124	1.22	2.96	2.96	1.06	1.09	1.11	1.45	2.61	2.63	1.29	1.48

so the effective number of parameters is not correctly calculated. This could be explored further in future studies based on simpler selectivity patterns of models such as SAM (e.g. Berg and Nielsen, 2016) that are based on state-space estimation.

For the analyses of this paper, the double form was the most robust to the true form of selectivity (i.e., it performed better when it was the true selectivity form and not much poorer when it was more complicated than needed). An exception to this general result was when *M* was also estimated, presumably due to the confounding between selectivity and natural mortality in age-composition data.

4.1. Caveats and future work

The absolute values for the MAREs should be interpreted with some caution because the simulations herein make assumptions that are likely to be violated in practice. Specifically, apart from the forms for selectivity by fleet, the population dynamics and observation models were correctly specified. Moreover, many parameters (e.g., fecundity and growth) were assumed known and were known exactly for most analyses (i.e., natural mortality and stock-recruitment steepness). The assumption that growth is known exactly would be more consequential had size- rather than age-composition data been available for parameter estimation and selectivity was assumed to be a function of size rather than age. The true forms for selectivity were all fairly simple. No scenarios considered cases where a fleet is a combination of metiers (i.e., a group of vessels that target a specific set of species using a particular gear during a specific time period or within a specific area) with very different selectivity patterns. Such cases may result in an overall selectivity pattern that is multi-modal—as has been observed for tropical tuna assessments. Future work should also consider a constantly increasing selectivity pattern, often relevant for hook-dominated reef fish fisheries.

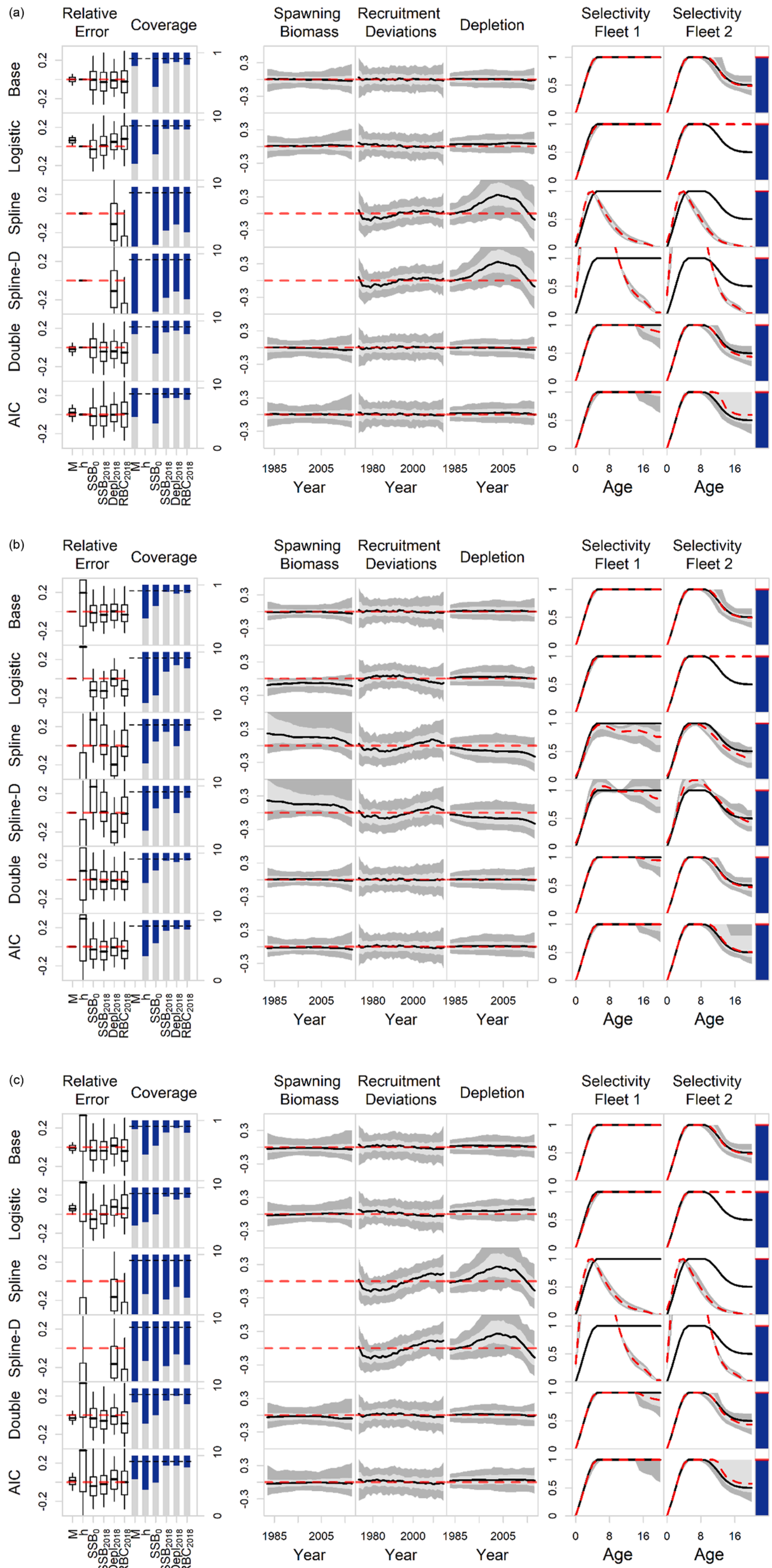

Fig. 8. a As for Fig. 2 but for experiment 6, in which *M* is estimated and *h* is fixed., Fig. 8b. As for Fig. 2 but for experiment 6, in which *M* is fixed and *h* is estimated., Fig. 8c. As for Fig. 2 but for experiment 6, in which *M* and *h* are estimated.

The results suggest that the best approach to specifying selectivity in an assessment depends to some extent on species' life history (e.g. the performance of the spline approaches for school whiting compared to that for tiger flathead and blue grenadier). The three life histories considered in this paper cover many of those typically encountered when conducting stock assessments. However, a complete analysis should examine very long-lived species, such as rockfishes (*Sebastes spp*) and orange roughy (*Hoplostethus atlanticus*), and particularly very short-lived species as such anchovies (*Engraulis spp*). Future work could examine the reasons (e.g., difference in the value of *M* or number of ages) that lead to changes in performance among species life history.

This study focused on the MARE over the entire time series of data. Future work could focus on performance at the end of the time series, including retrospective bias (Carvalho et al., 2021) and longer projections of future catch.

This paper investigated the ability to estimate parameters and quantities of management interest (i.e., summarizing results in terms of MAREs). Thus, the results provide guidance of best practices, which pertain to evaluating stock status (e.g., is a stock above a target or limit reference point). However, the primary reason for conducting stock assessments is often to support application of harvest control rules and hence achievement of management goals. An investigation of the impact of modelling different selectivity formulations based on a closed-loop simulation framework (i.e., management strategy evaluation (MSE); *sensu* De le Mare, 1986; Cook, 1999, Smith et al., 1999; Punt et al., 2016) would better address these management goals.

4.2. Conclusions

Here, the authors outlined support for known, novel and future work best practices for specifying selectivity in age-structured integrated stock assessments. As expected, stock life history and the information content of the data (e.g., effective sample size) should guide selectivity assumptions. Regarding the latter, for less informative data sets, and when there is a poor understanding of fishery and stock dynamics, it is best to ignore time-variation in selectivity and force at least one fleet to have asymptotic selectivity, while recognizing that results can be biased towards high mortality, low biomass. The assumption that one fleet has asymptotic selectivity can be relaxed for data rich stocks with informative catch series and reliable age composition data over many years. As for novel best practices, we recommend only using the spline-based selectivity patterns when there is considerable contrast in catches (and when there is contrast, both spline methods perform well). We also recommend using the methods from Xu et al. (2019) when assuming time-varying selectivity, with a caveat to analysts to think critically about whether the extent of time-variation in the model may be larger than the "truth" (at least, what best available science can support) and the implications for estimation performance. In terms of practices in need of future work, the authors also caution that model selection approaches (i.e., AIC) did not perform consistently in this study. We recommend that analysts consider multiple selectivity options, examine residual patterns and consider including alternative assumptions in the form of decision tables.

CRediT authorship contribution statement

Kristin Privitera-Johnson: Formal analysis; Visualization; Writing – original draft. **Richard Methot:** Conceptualization; Funding acquisition; Methodology; Software; Supervision; Writing – original draft; Writing – review & editing. **André Punt:** Conceptualization; Formal analysis; Funding acquisition; Methodology; Software; Supervision; Writing – original draft; Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Eric Williams (SEFSC), Kiva Oken (NWFSC), and an anonymous reviewer are thanked for their comments on earlier versions of this paper. This publication was partially funded by CSIRO and the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CICOES) under NOAA Cooperative Agreement NA20OAR4320271, Contribution No. 022-1174.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fishres.2022.106247.

References

Bence, J.R., Gordoa, A., Hightower, J.E., 1993. Influence of age-selectivity surveys on the reliability of stock synthesis assessments. Can. J. Fish. Aquat. Sci. 50, 827–840.
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.
Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I.G., Wetzel, C.R., Doering, K., Johnson, K.F., Methot, R.D., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fish. Res. 240, 105959.
Castillo-Jordán, C., Tuck, G., 2020. Blue grenadier (*Macruronus novaezelandiae*) stock assessment based on data up to 2017 base case. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2018 and 2019. Part 1, 2018. Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere, Hobart, pp. 314–341, 526 p.
Cook, J.G., 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms? ICES J. Mar. Sci. 56, 797–810.
Crone, P.R., Valero, J.L., 2014. Evaluation of length- vs. age-composition data and associated selectivity assumptions used in stock assessments based on robustness of derived management quantities. Fish. Res. 158, 165–171.
Day, J., 2018a. Tiger flathead (*Neoplatycephalus richardsoni*) stock assessment using data to 2015. Pp 443 - 512. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2016 and 2017. Part 1, 2016. Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere Flagship, Hobart, p. 629.
Day, J., 2018b. School whiting (*Sillago flindersi*) stock assessment based on data up to 2016. Pp 588 - 663. In: Tuck, G.N. (Ed.), Stock Assessment for the Southern and Eastern Scalefish and Shark Fishery 2016 and 2017. Part 2, 2017. Australian Fisheries Management Authority and CSIRO Oceans and Atmosphere, Hobart, p. 837 (p).
De le Mare, W., 1986. Simulation studies on management procedures. Rep. Int. Whal. Comm. 36, 429–449.
Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.
Fournier, D., Archibald, C.P., 1982. A general theory for analysing catch at age data. Can. J. Fish. Aquat. Sci. 39, 1195–1207.
Fulton, E.A., Punt, A.E., Dichmont, C.M., Harvey, C.J., Gorton, R., 2019. Ecosystems say good management pays off. Fish. Fish. Oxf. 20, 66–96.
Hulson, P.F., Hanselman, D.H., 2014. Tradeoffs between bias, robustness, and common sense when choosing selectivity forms. Fish. Res. 158, 63–73.
Hurtado-Ferro, F., Punt, A.E., Hill, K.T., 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fish. Res. 158, 102–115.
Ichinokawa, M., Okamura, H., Takeuchi, Y., 2014. Data conflict caused by model mis-specification of selectivity in an integrated stock assessment model and its potential effects on stock status estimation. Fish. Res. 158, 147–157.
Li, B., Shertzer, K.W., Lynch, P.D., Ianelli, J.N., Legault, C.M., Williams, E.H., Methot, R. D., Brooks, E.N., Deroba, J.J., Berger, A.M., Sagarese, S.R., Brodziak, J.K.T., Taylor, I.G., Karp, M.A., Wetzel, C.R.M., Supernaw, M., 2021. A comparison of 4 primary age-structured stock assessment models used in the United States. Fish. Bull. 119, 149–167.
Martell, S.J.D., Stewart, I.J., 2014. Toward defining good practices for modelling time-varying selectivity. Fish. Res. 158, 84–95.
Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.
Maunder, M.N., Crone, P.R., Valero, J.L., Semmens, B.X., 2014. Selectivity: theory, estimation, and application in fishery stock assessment models. Fish. Res. 158, 1–4.
Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. 68, 1744–1760.
Methot, 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.
Monnahan, C.C., Branch, T.A., Thorson, J.T., Stewart, I.J., Szuwalski, C.S., 2020. Overcoming long Bayesian run times in integrated fisheries stock assessments. ICES J. Mar. Sci. 76, 1477–1488.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., M. Haddon, M., 2016. Management Strategy Evaluation: Best Practices. Fish. Fish. 17, 303–334.

Punt, A.E., Dunn, A., Elvarsson, B.P., Hampton, J., Hoyle, S., Maunder, M.N., Methot, R. D., Nielsen, A., 2020. Essential features of the next-gen integrated assessment: a perspective. Fish. Res. 229, 105617.

Punt, A.E., Tuck, G.N., Day, J., Burch, P., Thomson, R.B., Bessell-Browne, P., 2021. The impact of alternative age-length sampling schemes on the performance of stock assessment methods. Fish. Res. 238, 105904.

Sampson, D.B., 2014. Fishery selection and its relevance to stock assessment and fishery management. Fish. Res. 158, 5–14.

Sampson, D.B., Scott, R.D., 2011. A spatial model for fishery age-selection at the population level. Can. J. Fish. Aquat. Sci. 68, 1077–1086.

Smith, A., Sainsbury, K., Stevens, R., 1999. Implementing effective fisheries-management systems – management strategy evaluation and the Australian partnership approach. ICES J. Mar. Sci. 56, 967.

Stewart, I.J., Martell, S.J.D., 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. 158, 40–49.

Thorson, J.T., Taylor, I.G., 2014. A comparison of parametric, semi-parametric, and non-parametric approaches to selectivity in age-structured assessment models. Fish. Res. 158, 74–83.

Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2019. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Can. J. Fish. Aquat. Sci. 76, 268–285.