NRC Research Press

ARTICLE

A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models

Haikun Xu, James T. Thorson, Richard D. Methot, and Ian G. Taylor

Abstract: Selectivity is a key parameter in stock assessments that describes how fisheries interact with different ages and sizes of fish. It is usually confounded with other processes (e.g., natural mortality and recruitment) in stock assessments and the assumption of selectivity can strongly affect stock assessment outcome. Here, we introduce a new semi-parametric selectivity method, which we implement and test in Stock Synthesis. This selectivity method includes a parametric component and an autocorrelated nonparametric component consisting of deviations from the parametric component. We explore the new selectivity method using two simulation experiments, which show that the two autocorrelation parameters for selectivity deviations of data-rich fisheries are estimable using either mixed-effect or simpler sample-based algorithms. When selectivity deviations of a data-rich fishery are highly autocorrelated, using the new method to estimate the two autocorrelation parameters leads to more precise estimations of spawning biomass and fully selected fishing mortality. However, this new method fails to improve model performance in low data quality cases where measurement error in the data overwhelms the pattern caused by the autocorrelated process. Finally, we use a case study involving North Sea herring (*Clupea harengus*) to show that our new method substantially reduces autocorrelations in the Pearson residuals in fit to age composition data.

Résumé : La sélectivité est un paramètre clé des évaluations de stocks qui décrit comment les pêches interagissent avec différents âges et différentes tailles de poissons. Elle est habituellement confondue avec d'autres processus (p. ex. mortalité naturelle et recrutement) dans les évaluations de stocks, et l'hypothèse de la sélectivité peut avoir une importante incidence sur le résultat de l'évaluation de stocks. Nous présentons une nouvelle méthode de sélectivité semi-paramétrique que nous appliquons et validons dans Stock Synthesis. Cette méthode de sélectivité comprend une composante paramétrique et une composante non paramétrique autocorrélée qui est constituée d'écarts par rapport à la composante paramétrique. Nous explorons la nouvelle méthode de sélectivité en utilisant deux expériences de simulation qui montrent que les deux paramètres d'autocorrélation pour les écarts de sélectivité de pêches pour lesquelles les données sont abondantes peuvent être estimés en utilisant des algorithmes à effets mixtes ou des algorithmes plus simples basés sur les échantillons. Quand les écarts de sélectivité d'une pêche aux données abondantes sont fortement autocorrélés, l'utilisation de la nouvelle méthode pour estimer les deux paramètres d'autocorrélation produit des estimations plus précises de la biomasse reproductrice et de la mortalité par pêche entièrement sélectionnée. Cette nouvelle méthode n'améliore toutefois pas la performance des modèles dans les cas de faible qualité des données pour lesquels l'erreur de mesurage dans les données masque le motif résultant du processus autocorrélé. Enfin, nous utilisons une étude de cas du hareng (*Clupea harengus*) de la mer du Nord pour démontrer que notre nouvelle méthode réduit sensiblement les autocorrélations dans les résidus de Pearson dans des données ajustées à la composition par âge. [Traduit par la Rédaction]

Introduction

In stock assessment models, selectivity is usually referred to as the combination of contact selectivity and population selectivity. It describes how relative fishing effectiveness varies with the age or size of the fish. Selectivity is perhaps the most crucial fisheries process that affects the age or size composition of fish we measure, as it directly controls how fisheries interact with different ages or sizes of fish (Maunder and Piner 2017). However, the exact form of selectivity is generally difficult to estimate due to the complexity of the factors that modulate selectivity (Maunder et al. 2014). Specifically, selectivity changes from year to year as a result of changes in fishing gear, fishing behavior, and spatiotemporal distribution of the fish of interest (Maunder et al. 2014; Francis 2017). Moreover, misspecifying fisheries selectivity in an assessment could result in serious consequences such as large retrospective pattern in critical population attributes such as spawning biomass (Stewart and Martell 2014), unrealistic weighting for composition data (Francis 2017), data conflict (Maunder and Piner 2017), or substantially biased model estimates (Stewart and Monnahan 2017). Therefore, selectivity is a key process in fisheries stock assessments and improving the parameterization of selectivity has been an im-

Received 10 October 2017. Accepted 29 April 2018.

H. Xu.* School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98105, USA.
J.T. Thorson† and I.G. Taylor. Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA.
R.D. Methot. NOAA Senior Scientist for Stock Assessments, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA.
Corresponding author: Haikun Xu (email: hkxu@iattc.org).
*Present address: Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

†James T. Thorson currently serves as an Associate Editor; peer review and editorial decisions regarding this manuscript were handled by Verena Trenkel.

portant research topic in the stock assessment community (Maunder et al. 2014).

Selectivity is commonly estimated using a simple parametric function that is constant over time. In practice, however, selectivity often deviates from this pattern and deviations are autocorrelated in both time and age (Sampson and Scott 2012). The deviations in selectivity can be autocorrelated for various reasons. For example, they can be autocorrelated across age because some age-specific factors (e.g., spatial distribution) that influence selectivity variation are likely to be similar between adjacent age groups. In fact, autocorrelated selectivity variation among age groups has been found in some European fisheries stock assessments conducted using a state-space model that allows process error in selectivity (Nielsen and Berg 2014). In addition, they can be autocorrelated across years, as some factors (e.g., fishing gear and behavior) that influence selectivity variation are likely to be similar between adjacent years. A unique characteristic of age-structured models is that the population signal in a specific age group and year will propagate to the next age group 1 year later. The propagation can link the two dimensions (age and year) together (Butterworth et al. 2003), potentially leading to two-dimensional autocorrelated variation in selectivity.

No model formulation to date, however, can account for the among-age and among-year autocorrelations in selectivity deviations simultaneously. In a previous study by Martell and Stewart (2014), a bicubic spline penalty term was included in the objective function to smooth the variation in selectivity over both age and time, but the levels of variation and smoothness in selectivity associated with the penalty term had to be prespecified on an ad hoc basis. Stock Synthesis (Methot and Wetzel 2013), a widely used stock assessment package in the United States and worldwide, assumes that age or size composition data follow a multinomial distribution that allows only negative and weakly correlated residuals between two age bins (Francis 2011). A logistic-normal distribution, an alternative likelihood for composition data, was recently advocated by Francis (2014) due to its capacity of accounting for positive among-age autocorrelation in residuals for composition data. However, the logistic-normal distribution cannot deal with the among-year autocorrelation in residuals for composition data (Francis 2017; Thorson et al. 2017). Except for developing a more appropriate likelihood form to replace the multinomial distribution, adding process errors to selectivity under the increasingly popular mixed-effect model framework (Gelman 2005) is another feasible method to deal with autocorrelated residuals for composition data (Francis 2017).

In this paper, we introduce a new semi-parametric selectivity method to account for two-dimensional autocorrelated deviations in selectivity. More specifically, we calculate the selectivity of a fishery as the product of a parametric form and a deviation term that is treated as a process error. For the selectivity deviation term, we adapt a multivariate likelihood function that allows the model to estimate and account for the among-age and among-year autocorrelations in selectivity deviations simultaneously. We evaluated the performance of the new semi-parametric selectivity method in estimating spawning biomass (SB) and fully selected fishing mortality (*F*), which in this study is defined as the fishing mortality at fully selected ages (i.e., the expected selectivity at age is 1). In the first section, two simulation experiments were conducted to answer three key questions. (1) How well can the levels of selectivity variation and autocorrelation be estimated? (2) When selectivity of the fishery is highly autocorrelated among ages and years, does accounting for selectivity autocorrelation result in less biased and more accurate estimates of SB and *F* than ignoring autocorrelation or variation in fishery selectivity? (3) If the answer to the question above is true, how does the improvement by accounting for selectivity autocorrelation change under various levels of selectivity autocorrelation?

In the second section, we implemented the new semi-parametric selectivity method in Stock Synthesis and evaluated the performance of the new Stock Synthesis feature using the real data set for North Sea herring (*Clupea harengus*) as a case study.

Methods and data

We construct a simulation–estimation model to examine the performance of our new semi-parametric selectivity method in an age-structured population dynamics model. The two simulation experiments are conducted based on the age-structured simulation–estimation package CCSRA by Thorson and Cope (2015). This package uses Template Model Builder (TMB) (Kristensen et al. 2015) to implement mixed-effect parameter estimation, so we can compare estimation performance when implementing either sophisticated (mixed-effect) or simplified (penalized likelihood) approaches to estimating semi-parametric selectivity.

We first introduce our new semi-parametric age- and time-varying selectivity approach and describe a simulation experiment that consists of an operating model (OM) simulating the true population dynamics, a sampling model (SM) generating data from the true population dynamics, and an estimation model (EM) fitting to the generated data. The performance of the EM in simulation experiments is evaluated by comparing the estimates the EM provides with the corresponding true values generated by the OM.

We then describe our implementation of the semi-parametric age- and time-varying selectivity option in the common Stock Synthesis software (Methot and Wetzel 2013) and a case study involving data for North Sea herring. We use this case study to compare the performance of the estimation model where age- and time-varying selectivity is ignored, estimated separately for each age and year, or estimated using our proposed semi-parametric age and time selectivity smoother.

Operating model in simulation experiments

The OM and EM in the simulation experiment are both age structured and have the same population dynamics. The abundance of fish in age a and year t ($t \in \{1, 2, \ldots, T\}$) is

$$(1) \qquad N_{a,t} = \begin{cases} R_t & \text{if} \quad a = 0 \\ N_{a-1,t-1}\exp(-S_{a,t-1}F_{t-1} - M) & \text{if} \quad a \in \{1, 2, \ldots, A-1\} \\ N_{A-1,t-1}\exp(-S_{A,t-1}F_{t-1} - M) + N_{A,t-1}\exp(-S_{A,t-1}F_{t-1} - M) & \text{if} \quad a = A \end{cases}$$

where R_t is recruitment in year t, $S_{a,t-1}$ is fishery selectivity in age a and year $t-1$, and M is natural mortality rate that is the same for all ages and years. $A = 15$ is the plus group for both the OM and EM. Recruitment is from the steepness parameterization of the bias-corrected Beverton–Holt stock–recruit function and is assumed to have stochastic lognormal deviations:

$$(2) \qquad \ln(R_t) \sim \mathrm{N}\left(\ln\left(\frac{4hR_0\mathrm{SB}_t}{\mathrm{SB}_0(1-h) + \mathrm{SB}_t(5h-1)}\right) - \frac{\sigma_R^2}{2}, \sigma_R^2\right)$$

while specifying that the magnitude of variability in recruitment $\sigma_R = 0.4$ and unfished recruitment $R_0 = 10^9$. Steepness (h) quantifies

Table 1. Comparison of the parameters for the two types of life history (fast and slow) investigated in the two simulation experiments.

Parameter name	Symbol	Pacific hake	Pacific sardine
Natural mortality rate	M	0.386·year^{-1}	0.552·year^{-1}
Length at age 0	L_0	1 cm	1 cm
Asymptotic maximum length	L_{inf}	90 cm	30 cm
Von Bertalanffy growth coefficient	k	0.20·year^{-1}	0.30·year^{-1}
Log maximum annual spawner per spawner	LMARR	2	1
Age at 50% selection in the fishery	S_{50}	5.44	3.55
Rate of change in selectivity at age	S_{slope}	1	1
Age at maturity	a_{mat}	5.44	3.55
Steepness of the Beverton–Holt SR function	h	0.83	0.55

Note: We used Pacific hake (*Merluccius productus*) and Pacific sardine (*Sardinops sagax*) to represent the fast and slow types of life history, respectively.

the magnitude of density dependence in recruitment and SB_t is spawning biomass in year t:

$$\text{(3)} \qquad SB_t = \sum_{a=0}^{A} w_a m_a N_{a,t}$$

where w_a and m_a are mass at age and maturity at age, respectively. The OM includes one fishery fleet and the catch (in numbers) in age a and year t is calculated from the Baranov catch equation:

$$\text{(4)} \qquad C_{a,t} = N_{a,t}\frac{S_{a,t}F_t}{S_{a,t}F_t + M}(1 - e^{-S_{a,t}F_t - M})$$

The total catch (in biomass) in year t is the sum of element-wise product of $C_{a,t}$ and w_a and is assumed known without error. The initial abundance at age is assumed to be lognormally distributed with mean derived from an approximately unfished state:

$$\text{(5)} \qquad \ln(N_{a,1}) \sim \mathrm{N}\left(\ln\left(R_0 e^{-aM}\right) - \frac{\sigma_R^2}{2}, \sigma_R^2\right)$$

The OM has one fishery, the selectivity (S) of which in age a and year t is

$$\text{(6)} \qquad S_{a,t} = \frac{1}{1 + e^{-S_{\mathrm{slope}}(a - S_{50})}} \times e^{\varepsilon_{a,t}}$$

where $\varepsilon_{a,t}$ is simulated as a two-dimensional (2D) AR(1) process:

$$\text{(7)} \qquad \mathrm{vec}(\boldsymbol{\varepsilon}) \sim \mathrm{MVN}\left(\mathbf{0}, \sigma_S^2\mathbf{R}_{\mathrm{total}}\right)$$

such that the first multiplicand in eq. 6 represents a parametric selectivity form $\left(\frac{1}{1 + e^{-S_{\mathrm{slope}}(a - S_{50})}}\right)$, while the second multiplicand ($e^{\varepsilon_{a,t}}$) represents nonparametric deviations from this parametric form. We call this product of nonparametric and parametric components a "semi-parametric" age- and time-varying selectivity model (Shelton et al. 2014; Thorson and Taylor 2014). We fix the standard deviation of selectivity (σ_S) at 0.4 for representing a case with a moderate level of selectivity variation. The correlation matrix $\mathbf{R}_{\mathrm{total}}$ is equal to the kronecker product ($\otimes$) of the two correlation matrices for the among-age and among-year AR(1) processes:

$$\text{(8)} \qquad \mathbf{R}_{\mathrm{total}} = \mathbf{R} \otimes \tilde{\mathbf{R}}$$

$$\text{(9)} \qquad \mathbf{R}_{a,\tilde{a}} = \rho_a^{|a-\tilde{a}|}$$

$$\text{(10)} \qquad \tilde{\mathbf{R}}_{t,\tilde{t}} = \rho_t^{|t-\tilde{t}|}$$

where ρ_a and ρ_t are the among-age and among-year AR(1) coefficients, respectively. When both of which are zero, $\mathbf{R}$ and $\tilde{\mathbf{R}}$ are two identity matrices and their Kronecker product, $\mathbf{R}_{\mathrm{total}}$, is also an identity matrix. The 2D selectivity random field was simulated using the "mvrnorm" function in the MASS package in R (R Core Team 2015). Considering that estimating $\varepsilon_{a,t}$ is usually difficult for the youngest and oldest age groups due to most limited composition data, we assume that $\varepsilon_{a,t} = \varepsilon_{2,t}$ for $a < 2$ and $\varepsilon_{a,t} = \varepsilon_{7,t}$ for $a > 7$. Thus, vec($\boldsymbol{\varepsilon}$) in eq. 7 is defined as

$$\text{(11)} \qquad \mathrm{vec}(\boldsymbol{\varepsilon}) = (\varepsilon_{2,1}, \ldots, \varepsilon_{2,T}, \varepsilon_{3,1}, \ldots, \varepsilon_{3,T}, \ldots \ \ldots, \varepsilon_{7,1}, \ldots, \varepsilon_{7,T})'$$

Following Thorson and Cope (2015), F_t was derived stochastically from the effort-dynamics model:

$$\text{(12)} \qquad F_t = F_{t-1}\left(\frac{SB_{t-1}}{\gamma SB_0}\right)^{\lambda}$$

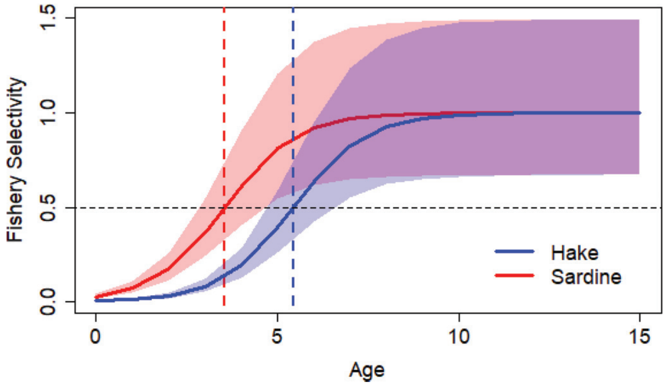
where we specify that acceleration rate $\lambda = 0.2$ and the ratio of equilibrium SB to unfished SB $\gamma = 0.35$ (Thorson et al. 2013). This effort-dynamics model is used to generate contrast in F while also ensuring that F is correlated with process errors affecting biomass (e.g., total biomass, exploitable biomass, or SB). Here we choose SB as the population attribute driving the dynamics of F but other choices give similar time-series characteristics to F. Under this fishing behavior, SB_t tends to decrease from the approximately unfished initial value (SB_1) towards the equilibrium level γSB_0 over time and the rate of the decrease in SB_t is controlled by λ (examples of SB_t trajectory can be found in Fig. 1 in Thorson and Cope 2015).

Two types of life history are investigated in the two simulation experiments. The first one represents a "slow" or "periodic" type of life history, roughly based upon Pacific hake (*Merluccius productus*), and the second one represents a "fast" or "opportunistic" type of life history, roughly based on Pacific sardine (*Sardinops sagax*). The definition and value of all of the life history parameters for the two fishes are shown in Table 1 and a detailed description of how the life history parameters are derived can be found in Thorson and Cope (2015). For each replicate, the population dynamics of the slow and fast species is simulated by the OM for 20 years (T = 20). The parametric fishery selectivity form and the associated selectivity variability for the two fishes are compared in Fig. 1.

Sampling model in simulation experiments

Age composition is sampled by the fishery every year from the simulated population, assuming a multinomial distribution with a constant sample size of n_{comp}:

Fig. 1. Comparison of the parametric fishery selectivity for the two types of life history as a function of age. The shaded areas show the ±1 standard deviation range of selectivity variation that is induced by the nonparametric deviation term. The vertical broken lines denote the age at 50% selection in the fishery. [Color online.]



$$A_t \sim \text{Multinomial}(C_{a,t}, n_{\text{comp}}) \tag{13}$$

The fishery also provides an index of abundance I_t for every year in the modeled period. It is drawn from a lognormal distribution with coefficient of variation of CV_{abund} (where the value of CV_{abund} varies among different simulation scenarios) and constant catchability of $q = 0.0001$:

$$\ln(I_t) \sim \text{N}\left(\log(qB_t), \ln\left(1 + CV^2_{\text{abund}}\right)\right) \tag{14}$$

where B_t represents the exploitable biomass in year t:

$$B_t = \sum_{a=0}^{A} N_{a,t} w_a S_{a,t} \tag{15}$$

Estimation model in simulation experiments

We compare the performance of five EMs under autocorrelated deviations in age- and time-varying fishery selectivity.

EM1 ("constant selectivity")

The first EM ignores the variation in fishery selectivity, which is a common practice in stock assessments. Selectivity in this EM is a logistic function of age:

$$\hat{S}_a = \frac{1}{1 + e^{-\hat{S}_{\text{slope}}(a - \hat{S}_{50})}} \tag{16}$$

This function is the same as the parametric part of the true selectivity shown in eq. 6.

EM2 ("IID deviations")

The second EM assumes that the variation in fishery selectivity is independent of age and time. More specifically, fishery selectivity in this EM is assumed to be a product of the true parametric function for selectivity and a deviation term having a lognormal distribution with log mean of 0 and log standard deviation of $\hat{\sigma}_S$:

$$\hat{S}_{a,t} = \frac{1}{1 + e^{-\hat{S}_{\text{slope}}(a - \hat{S}_{50})}} \times e^{\hat{\varepsilon}_{a,t}} \tag{17}$$

$$\hat{\varepsilon}_{a,t} \sim \text{N}\left(0, \hat{\sigma}^2_S\right) \tag{18}$$

where $\hat{\sigma}_S$ is estimated via the iterative approach proposed by Methot and Taylor (2011). Particularly, $\hat{\sigma}_S$ in this EM is iteratively tuned to match the relationship within an accuracy of 0.01:

$$\hat{\sigma}^2_S = \text{SD}(\hat{\boldsymbol{\varepsilon}})^2 + \frac{1}{6T}\sum_{a=2}^{7}\sum_{t=1}^{T}\text{SE}(\hat{\varepsilon}_{a,t})^2 \tag{19}$$

where $\text{SD}(\hat{\boldsymbol{\varepsilon}})$ is the standard deviation of $\hat{\boldsymbol{\varepsilon}}$ and $\text{SE}(\hat{\varepsilon}_{a,t})$ is the standard error of $\hat{\varepsilon}_{a,t}$ that is estimated from the inverse Hessian. We use

$$b = 1 - \frac{\frac{1}{6T}\sum_{a=2}^{7}\sum_{t=1}^{T}\text{SE}(\hat{\varepsilon}_{a,t})^2}{\hat{\sigma}^2_S} \tag{20}$$

as a measure of how informative each simulation replicate is regarding estimating $\hat{\boldsymbol{\varepsilon}}$.

EM3 ("2D AR deviations")

The third EM also has the true parametric selectivity function for selectivity but specifies that the deviations in fishery selectivity are autocorrelated among both ages and years and thereby follows a multivariate normal distribution (see eqs. 6 and 7). In this EM, $\hat{\sigma}_S$ is fixed at the value that EM2 provides and ρ_a and ρ_t are fixed at the values externally estimated from samples. In detail, we extract $\hat{\boldsymbol{\varepsilon}}$ estimates from EM2 and then estimated ρ_a and ρ_t by fitting a stand-alone model to the extracted $\hat{\boldsymbol{\varepsilon}}$ samples. In the stand-alone model, the extracted samples are assumed to follow the multivariate normal distribution described in eq. 7. Based on that assumption, the only two estimable parameters (ρ_a and ρ_t) in the stand-alone model are estimated simultaneously via the maximum likelihood approach. This method is similar to the "external" method that Johnson et al. (2016) investigated to estimate recruitment autocorrelation in integrated assessment models. Johnson et al. (2016) found that the "external" method can provide an adequate estimate of recruitment autocorrelation when more than 40 years of recruitment estimates are available, although it has not previously been tested for use when estimating two autocorrelation parameters (as we do here).

EM4 ("REML estimation")

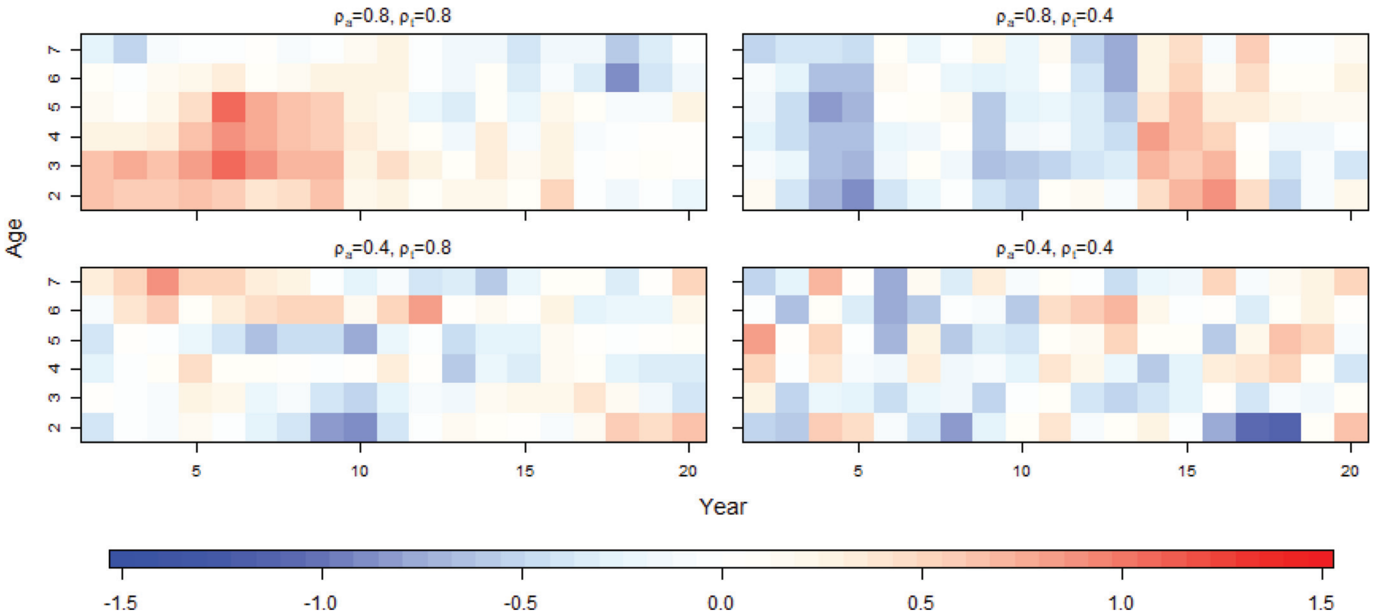
The fourth EM is the same as EM3 except that the three hyper-parameters for age- and time-varying selectivity are fixed at the values internally estimated using the restricted maximum likelihood (REML) (Harville 1974). This EM includes two steps: first, the three hyper-parameters are estimated by treating all other estimated parameters as random effects (while specifying an improper, uniform prior on all fixed effects) and second, the EM is rerun to estimate all parameters other than the three hyper-parameters by fixing the three hyper-parameters at the values estimated from the previous step.

EM5 ("perfect information")

The last EM is the same as the previous two EMs except that the three hyper-parameters ($\hat{\sigma}_S$, ρ_a, and ρ_t) for semi-parametric age- and time-varying selectivity are fixed at the true values that are used to generate the autocorrelated age- and time-varying selectivity in the OM. This EM cannot be implemented in practice (because we never know the true value of these parameters except in a simulation study) and is included as a reference to demonstrate the ideal performance of EM3 and EM4 when the three externally or internally estimated hyper-parameters are the same as the truth.

Within the five EMs considered here, EM3–5 implement our new semi-parametric selectivity method. Unless otherwise noted (i.e., step 1 in EM4), R_0, S_{slope}, S_{50}, and F_t were estimated as fixed

Fig. 2. An example of the simulated selectivity deviation pattern ($\hat{\varepsilon}$) from each OM investigated in the second simulation experiment. ρ_a and ρ_t represent the among-age and among-year AR(1) coefficient for $\hat{\varepsilon}$, respectively. [Color online.]



effects and $\hat{\varepsilon}$ (in EM2–5) and R_t were estimated as random effects. The estimates of the three hyper-parameters ($\hat{\sigma}_S$, ρ_a, and ρ_t) for age- and time-varying selectivity vary among EMs. All other parameters are fixed at the true values, so the difference in EM performance is determined by how selectivity is parameterized and how well the three hyper-parameters can be estimated. In EM3 and EM4, ρ_a and ρ_t are constrained to be bounded by 0 and 1 through a logit transformation. In TMB, the marginal likelihood of fixed-effect parameters is calculated using the Laplace approximation to integrate across random effects (Kristensen et al. 2015), and fixed-effect parameters are then estimated via maximizing the marginal likelihood within the R computing environment (R Core Team 2015). The "nlminb" function is used in R to minimize the negative of the marginal log-likelihood, and after that, TMB predicts random effects using empirical Bayes (Kristensen et al. 2015). We implement a bias-correction algorithm to correct for differences between the median and mean of the lognormal recruitment as a function of the amount of the uncertainty within and variability among the estimated recruitment deviations based on the method introduced by Methot and Taylor (2011).

Simulation experiments

Two simulation experiments are conducted in this study. The first experiment is designed to evaluate the performance of the five EMs in accounting for autocorrelated deviations in age- and time-varying selectivity under three qualities of fishery data. Three cases corresponding to high-quality (n_{comp} = 200 and $\mathrm{CV}_{\mathrm{abund}}$ = 0.1), medium-quality (n_{comp} = 50 and $\mathrm{CV}_{\mathrm{abund}}$ = 0.2), and low-quality (n_{comp} = 15 and $\mathrm{CV}_{\mathrm{abund}}$ = 0.3) data from the fishery are investigated for each type of life history. Highly autocorrelated deviations in age- and time-varying selectivity are generated by the OM where both ρ_a and ρ_t are fixed at 0.8 and σ_S is fixed at 0.4. Two hundred simulation replicates with randomly generated process errors (in recruitment and selectivity) and observation errors (in age composition and index of abundance) are generated for each case, and each replicate then fits to the EM1–5. The five EMs are compared with respect to the interquartile range (IQR) of the relative error, RE = $(\hat{\theta}-\theta)/\theta$, in F and SB. Particularly, the estimates of F and SB in the terminal year of the assessment are important to stock status determination, so the five EMs are also compared using measures of mean relative error, MRE = mean(RE), and root mean square error, RMSE $= \sqrt{\mathrm{mean}(\mathrm{RE}^2)}$. MRE and RMSE measure the accuracy and precision of model estimates in the terminal year of the assessment, respectively.

The second simulation experiment is designed to evaluate the importance of the semi-parametric age- and time-varying selectivity method under various levels of selectivity autocorrelations (ρ_a and ρ_t). While $\hat{\sigma}_S$ is the same across OMs, ρ_a and ρ_t are fixed at either 0.8 or 0.4, generating four OMs from the 2 × 2 factorial combination of two levels of autocorrelation for age and year (Fig. 2). Two hundred simulation replicates with randomly generated process errors (in recruitment and selectivity) and observation errors (in age composition and index of abundance) are generated by each OM under the high-quality data case (n_{comp} = 200 and $\mathrm{CV}_{\mathrm{abund}}$ = 0.1), and each replicate then fits to the five EMs individually. Again, performance of the five EMs is compared based on the IQR of relative error in the estimates of F and SB, and the terminal year estimates of F and SB are also compared based on MRE and RMSE.

Case study application in Stock Synthesis

Finally, we implement the semi-parametric age- and time-varying selectivity method in Stock Synthesis (V3.30). The implementation is designed to be highly flexible in its interaction with existing Stock Synthesis features. In detail, users can specify (1) which parametric selectivity option to use (including any of the existing selectivity options in Stock Synthesis), which defines the selectivity in cases where deviations are estimated to be zero (i.e., $\hat{\varepsilon}_{a,t} = 0$), (2) the minimum and maximum age to use for selectivity deviations (the "age range"), and (3) the minimum and maximum year to use for selectivity deviations (the "year range").

We demonstrate this new Stock Synthesis feature using a real-world data set for North Sea herring. It should be noted that we did not use the same data and assumptions as in the official ICES stock assessment (ICES 2017). In this case study application, we apply the semi-parametric age- and time-varying selectivity option for the fishery in ages 1–8 and years 1947–2011 and specify that the parametric component follow a logistic selectivity at age curve (as parameterized in Stock Synthesis):

$$\hat{S}_{a,t} = \frac{1}{1 + e^{-\ln(19)(a-p_1)/p_2}} \times e^{\hat{\varepsilon}_{a,t}} \quad (21)$$

where p_1 and p_2 determine the age at inflection and width for 95% selection of the parametric component of selectivity, respectively. In this case study, we compare three Stock Synthesis configurations in which the assumptions for $\hat{\varepsilon}_{a,t}$ distribution corresponded to EM1–3. Other than that, the three runs, referred to as SS-EM1, SS-EM2, and SS-EM3 hereafter, have the same configuration. The configuration for REML estimation (EM4) is not investigated in this case study because estimating the three hyper-parameters using REML is not currently feasible using Stock Synthesis, which is written in ADMB (Fournier et al. 2012).

The data for this case study include (1) total catch from the fishery (CV = 0.05), (2) three indices of abundance from surveys (CV = 0.2), (3) age composition data from the fishery (input sample size = 65) and from one survey (input sample size = 15), and (4) empirical mass at age. The three surveys that provide the indices of abundance to the assessment model are the acoustic survey in the North Sea (HERAS), the international bottom trawl survey for young-of-the-year herring abundance index (IBTS_Age1), and a survey for the spawning component abundance index (SCAI). In addition to the index of abundance, HERAS also provides age composition information to the assessment model. The temporal range of each data set above are shown in Fig. A1 and more details about the data can be found in ICES (2017). We assume a Beverton–Holt curve for the stock–recruit relationship and *M* is known without error. The effective sample size for the fishery age composition data is estimated using the Dirichlet-multinomial weighting method (Trenkel et al. 2012; Thorson et al. 2017). In this case study, catchability, fishing mortality, the deviations in recruitment and fishery selectivity, the parameters for the stock–recruit function, and the parametric selectivity function for the fishery are the variables estimated by Stock Synthesis using maximum likelihood.

Results

Performance of the five EMs given autocorrelated deviations in selectivity

We first evaluated how informative the fishery data were in terms of estimating $\hat{\boldsymbol{\varepsilon}}$ for hake and sardine life histories. As expected, *b* (see eq. 20) was positively associated with the quality of fishery data (Fig. A2). While large variability existed in *b*, the median values for both hake and sardine were larger than 0.5 in the high-quality case, close to 0.3 in the medium-quality case, and close to 0.2 in the low-quality case.

We then examined how well the levels of selectivity variation ($\hat{\sigma}_S$) and autocorrelations (ρ_a and ρ_t) can be estimated by using the external method from 2D AR deviations (EM3) and the internal method from REML estimation (EM4). Estimates of the three hyper-parameters from EM4 were relatively accurate, as there was a small difference between each median estimate and the corresponding true value from the OM (Fig. 3). Not surprisingly, the higher the quality of fishery data, the more precise the three hyper-parameters from EM4. The lowest precision of the estimate of the three hyper-parameters from EM4 was found in the low-quality case for sardine. In general, EM3 was less accurate but more precise than EM4 in terms of estimating $\hat{\sigma}_S$, and it tended to underestimate the two AR(1) coefficients, ρ_t and especially ρ_a, in all three data quality cases.

No matter whether and how the EM accounted for autocorrelated deviations in age- and time-varying selectivity, the estimates of *F* and SB generally had greater imprecision than bias (Table 2). In the high-quality case, constant selectivity (EM1) corresponded to the least precise estimates of *F* and SB over the entire time assessment period for both hake and sardine (Fig. 4, left column). In terms of *SB*, while IID deviations (EM2) corresponded to more precise estimates than constant selectivity, it was still less precise than 2D AR deviations (EM3) and REML estimation (EM4) where the autocorrelations in age- and time-varying selectivity were estimated. REML estimation (EM4) and perfect information (EM5) differed minimally regarding estimation precision (i.e., IQR) because the three hyper-parameters were accurately estimated using REML (Fig. 3). The estimates of the three hyper-parameters from 2D AR deviations (EM3), however, were all biased towards zero, so EM3 was less precise (wider IQR) than EM4 for both hake and sardine. In terms of *F*, EM2–5 (time-varying selectivity) had very similar precisions but they were all pronouncedly more precise than EM1 (constant selectivity). It is worth noting that when data quality was high, EM was more precise (in terms of estimating terminal year SB) than an otherwise identical model without among-age autocorrelation in selectivity (Fig. A3). However, the improvement in model precision was marginal, since EM3 underestimated ρ_a to a large extent.

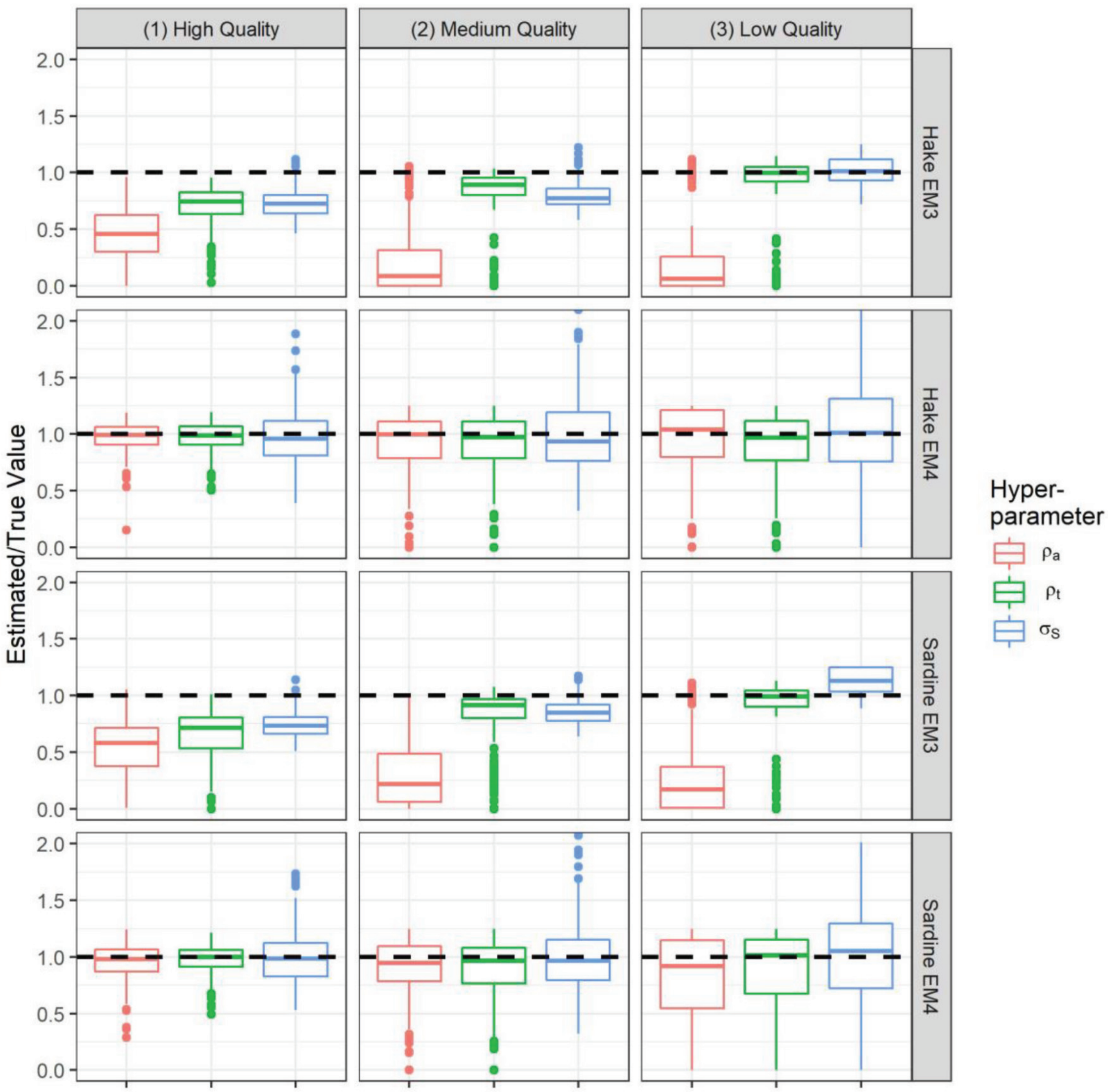
In the medium-quality case, performance of the five EMs in accounting for autocorrelated deviations in age- and time-varying selectivity ranked in the same order as in the high-quality case, but the difference among the five EMs was notably smaller than in the high-quality case (Fig. 4, middle column). 2D AR deviations (EM3) and REML estimation (EM4) were still more precise than IID deviations (EM2) and constant selectivity (EM1) in terms of SB, but the five EMs were already not differentiable with respect to the precision of *F* estimates. In the low-quality case, while EM1 was still least accurate and precise among the five EMs in term of SB, the other four EMs had similar bias and imprecision in terms of both *F* and SB (Fig. 4, right column). Therefore, both EM2 and EM3 were sufficient to account for autocorrelated time- and-age-varying selectivity in the low-quality case, even when ρ_a and ρ_t were both high (0.8).

We also compared the estimates of *F* and SB in the terminal year of the assessment. The degree to which the precision of terminal year *F* and SB increased from EM1 and 2 to EM3 and 4 was largely affected by data quality rather than life history (Table 2). Given that the true ρ_a and ρ_t were both high, improving stock assessments by using 2D AR deviations (EM3) or REML estimation (EM4) was most likely to occur in cases with high-quality fishery data. Here we used IID deviations (EM2) as a reference for performance comparison across the investigated EMs. In the high-quality case for hake, the RMSEs of SB estimates from EM3 and EM4 were 31% and 46% smaller than hose from EM2, respectively, indicating that model performance was improved by accounting for the autocorrelations in selectivity deviations. The two percentages dropped to 19% and 28% in the medium-quality case and to two negligible values in the low-quality case, indicating that the improvement in model performance became increasingly small as the quality of fishery data decreased.

In terms of *F*, the RMSEs for EM3 and EM4 were similar to those for EM2 in all three data cases. A similar RMSE pattern was found for sardine (Table 2). In the high-quality case, the RMSEs of SB estimates from EM3 and EM4 were 30% and 48% smaller than those from EM2. In contrast, the two percentages dropped to 15% and 24% in the medium-quality case to two negligible values in the low-quality case, also suggesting that the improvement in model performance was positively associated with the quality of fishery data. With respect to the MREs of terminal year SB and *F*, including the autocorrelations in selectivity deviations (EM3–5) minimally improved model performance in comparison with the reference model (EM2), probably because the MREs for EM2 were already very small (generally <0.08).

Interestingly, whether higher data quality led to more precise estimates in the terminal year of the assessment was influenced by the method EM used to deal with autocorrelated age- and time-varying selectivity. Given that the deviations in age- and time-varying fishery selectivity were highly autocorrelated, EM1 and 2 (constant selectivity and IID deviations) failed to provide the most

Fig. 3. Boxplots for the estimates of the three selectivity hyper-parameters (ρ_a, among-age autocorrelation; ρ_t, among-year autocorrelation; σ_S, variance) that EM3 (2D AR deviations) and EM4 (REML) provided for Pacific hake (*Merluccius productus*) (top two rows) and Pacific sardine (*Sardinops sagax*) (bottom two rows) in the first simulation experiment. The lower and upper hinges mark the first and third quantiles and the two whiskers extend to the value no further than 1.5 interquartile range from the corresponding hinge. The three columns correspond to the three data quality cases. The *y*-axis shows the ratio of estimated to true hyper-parameter values, so horizontal broken lines represent unbiased estimation. [Color online.]



precise terminal year estimates for hake and sardine in the high-quality case (Table 2). In terms of SB, the RMSEs for EM1 (0.41) and EM2 (0.39) in the high-quality case were even slightly larger than those for EM1 (0.39) and EM2 (0.36) in the medium-quality case. Hence, the assessments assuming constant selectivity (EM1) or IID deviations (EM2) could not benefit at all (in terms of the precision of SB) from high-data quality when the true deviations in selectivity were highly autocorrelated in both dimensions. However, when 2D AR deviations (EM3) or REML estimations (EM4) were used in the assessment to account for autocorrelated deviations in age- and time-varying selectivity, higher data quality was always associated with more precise estimates of terminal year *F* and SB. This association was particularly strong for EM4 where the three hyper-parameters could be more accurately and precisely estimated using REML (Table 2).

Importance of the semi-parametric selectivity method under various levels of autocorrelation

In the first simulation experiment, we found that accounting for autocorrelated deviations in selectivity substantially improved EM performance in the high-quality case, given that ρ_a and ρ_t were both high (0.8). Here, we conducted another simulation experiment to evaluate the performance of the semi-parametric age- and time-varying selectivity method (in the high-quality case) under three other selectivity autocorrelation patterns. Particularly, the three selectivity autocorrelation patterns corresponded to the cases where the deviations in age- and time-varying selectivity were weakly (0.4) autocorrelated among ages, years, or both ages and years.

No matter the levels of among-age and among-year autocorrelations in selectivity deviations were weak or strong (0.4 or 0.8), the three hyper-parameters (σ_S, ρ_a, and ρ_t) for semi-parametric age- and time-varying selectivity were accurately and precisely estimated using REML (Fig. 5). Consequently, REML estimation (EM4) and perfect information (EM5) performed similarly in terms of both *F* and SB (Fig. 6). Regardless of the level of autocorrelation in selectivity deviations, the estimates of ρ_t and especially ρ_a from 2D AR deviations (EM3) were consistently biased towards zero (i.e.,

Table 2. Metrics mean relative error (MRE) and root mean square error (RMSE) of the terminal year estimates of attributes *F* and SB in the first simulation experiment.

	High data quality				Medium data quality				Low data quality			
	F		SB		*F*		SB		*F*		SB	
	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE
Pacific hake (*Merluccius productus*)												
EM1	0.02	0.18	0.07	0.41	0.02	0.23	0.09	0.39	0.03	0.27	0.11	0.43
EM2	0.02	0.13	0.06	0.39	0.03	0.22	0.06	0.36	0.07	0.27	0.02	0.37
EM3	0.02	0.12	0.02	0.27	0.03	0.20	0.02	0.29	0.07	0.26	0.00	0.34
EM4	0.01	0.11	0.00	0.21	0.03	0.19	0.00	0.26	0.06	0.25	0.02	0.36
EM5	0.01	0.11	0.00	0.20	0.02	0.19	−0.01	0.24	0.05	0.25	0.02	0.35
Pacific sardine (*Sardinops sagax*)												
EM1	0.00	0.20	0.11	0.50	0.01	0.22	0.09	0.45	0.00	0.27	0.21	0.51
EM2	0.02	0.13	0.07	0.44	0.03	0.19	0.04	0.41	0.07	0.32	0.05	0.41
EM3	0.01	0.11	0.04	0.31	0.02	0.19	0.03	0.35	0.06	0.30	0.04	0.38
EM4	0.02	0.11	0.00	0.23	0.03	0.19	0.02	0.31	0.06	0.30	0.05	0.45
EM5	0.02	0.11	0.00	0.22	0.02	0.17	0.02	0.30	0.04	0.27	0.06	0.38

Note: The three columns correspond to the three data quality cases and EM1–5 are the five estimation models compared in the first simulation experiment.

Fig. 4. The interquartile range of relative error in the estimates of *F* and SB for Pacific hake (*Merluccius productus*) (top two rows) and Pacific sardine (*Sardinops sagax*) (bottom two rows) in the first simulation experiment. The three columns correspond to the three data quality cases and EM1–5 are the five estimation models compared in the first simulation experiment. [Color online.]

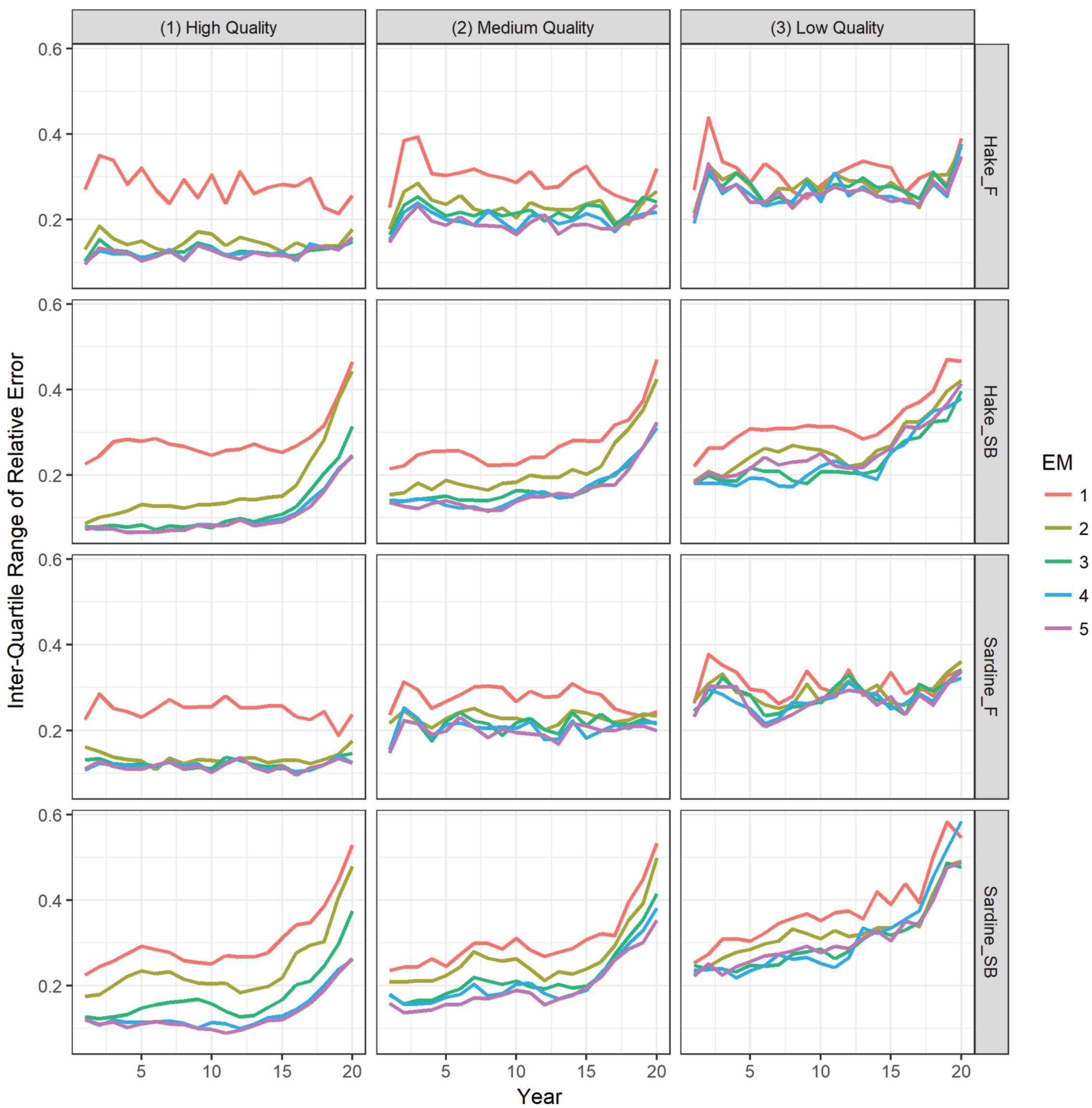
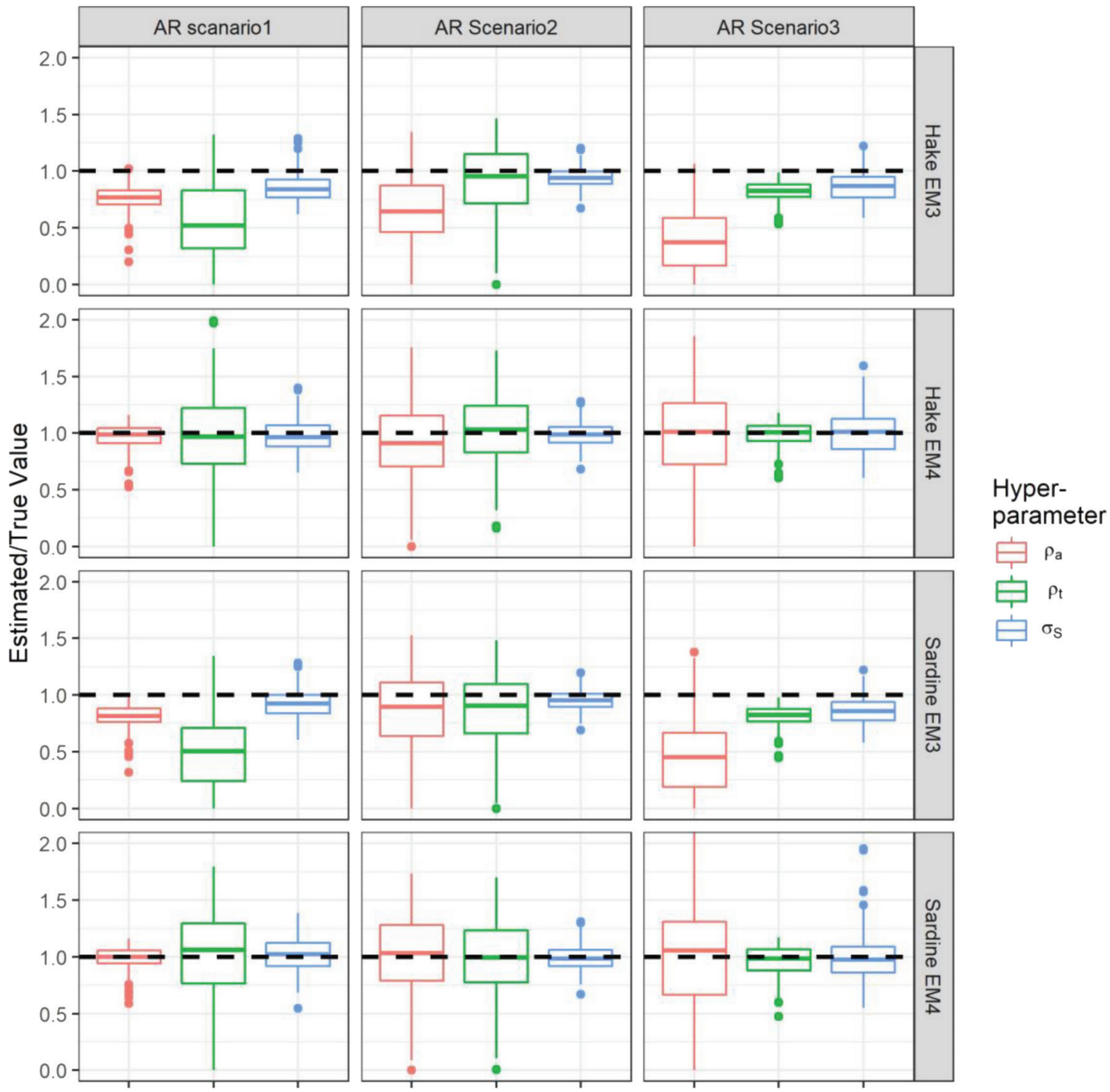


Fig. 5. Boxplots for the estimates of the three hyper-parameters (ρ_a, among-age autocorrelation; ρ_t, among-year autocorrelation; σ_S, variance) that EM3 and 4 provided for Pacific hake (*Merluccius productus*) (top two rows) and Pacific sardine (*Sardinops sagax*) (bottom two rows) in the second simulation experiment. The lower and upper hinges mark the first and third quantiles and the two whiskers extend to the value no further than 1.5 interquartile range from the corresponding hinge. The three columns correspond to the three selectivity autocorrelation scenarios (from left to right: ρ_a = 0.8 and ρ_t = 0.4, ρ_a = 0.4 and ρ_t = 0.4, ρ_a = 0.4 and ρ_t = 0.8). The *y*-axis shows the ratio of estimated to true hyper-parameter values, so the horizontal broken lines represent unbiased estimation. [Color online.]



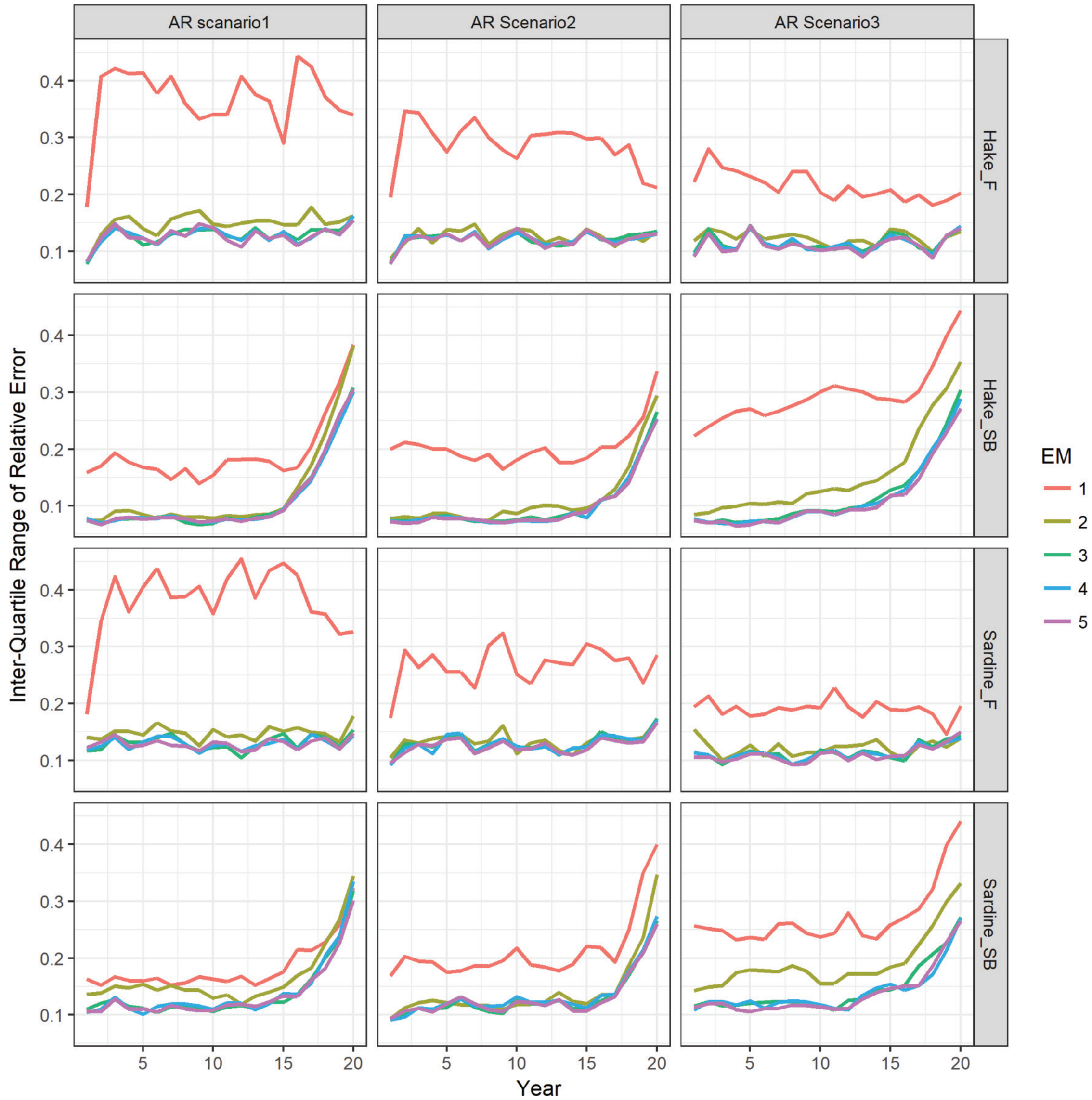
underestimated) (Fig. 5), whereas EM4 only slightly outperformed EM3 regarding estimating *F* and SB (Fig. 6). Generally speaking, constant selectivity (EM1) was by far the worst-performing model and accounting for autocorrelated selectivity deviations using EM3 or EM4 resulted in more precise SB estimates, especially when ρ_t was high (Fig. 6). However, IID deviations (EM2) seemed to perform well enough in terms of the precision of *F* estimates, and the improvement in precision (in terms of estimating *F*) from EM2 to EM1 was positively related to the value of ρ_a (Fig. 6).

Similar to the first simulation experiment, we also evaluated the performance of the semi-parametric age- and time-varying selectivity method based upon the estimates of *F* and SB in the terminal year of the assessment. The estimates of *F* and SB had much greater imprecision than bias (Table 3), so we chose RMSE as the primary metric assessing EM performance. 2D AR selectivity (EM3) and REML estimation (EM4) had very similar performance and both outperformed IID deviations (EM2) and especially constant selectivity (EM1) in the cases where the deviations in age- and time-varying selectivity were weakly autocorrelated in at least one dimension (Table 3). Understandably, the importance of accounting for autocorrelated age- and time-varying selectivity (by using EM3 or EM4) was more pronounced when the true ρ_a or/and ρ_t was/were high (0.8).

Case study application in Stock Synthesis

We further evaluated the performance of our semi-parametric selectivity approach in Stock Synthesis using a real-world data set for North Sea herring, finding that SS-EM3 (2D AR deviations) outperformed the other two SS configurations due to the autocorrelated selectivity of the North Sea herring fishery. Selectivity deviations from SS-EM2 had a strong 2D pattern, suggesting that they were likely to be autocorrelated in both dimensions and selectivity was misspecified in SS-EM2 (Fig. 7*a*). The deviations in selectivity were generally positive before 1977 and negative between 1977 and 1997. While the deviations in selectivity were still predominantly negative after 1997, they were much closer to zero

Fig. 6. The interquartile range of relative error in the estimates of *F* and SB for Pacific hake (*Merluccius productus*) (top two rows) and Pacific sardine (*Sardinops sagax*) (bottom two rows) in the second simulation experiment. The three columns correspond to the three selectivity autocorrelation scenarios (from left to right: ρ_a = 0.8 and ρ_t = 0.4, ρ_a = 0.4 and ρ_t = 0.4, ρ_a = 0.4 and ρ_t = 0.8). EM1–5 are the five estimation models compared in the first simulation experiment. [Color online.]



than between 1977 and 1997. Using the tuning method that we developed (eq. 19), $\hat{\sigma}_S$ was iteratively tuned to be 1.04 in SS-EM2. A very high value of b (0.9) was found for the North Sea herring fishery, indicating that the age composition data that the fishery provided were very informative regarding estimating semi-parametric age- and time-varying selectivity. We then fixed $\hat{\sigma}_S$ at 1.04 and externally estimated ρ_a (0.51) and ρ_t (0.79) using the selectivity deviation samples from SS-EM2.

Estimates of the three hyper-parameters suggested that selectivity of the fishery was highly variable (large $\hat{\sigma}_S$) and the deviations were highly autocorrelated among both ages and years (large ρ_a and ρ_t). With all three hyper-parameters being fixed at the estimated values above, SS-EM3 (2D AR deviations) provided smoother estimates of selectivity deviation on the age–year surface (Fig. 7*b*). While $\hat{\sigma}_S$ was fixed at the same value in the two EMs, the estimates of selectivity deviations from SS-EM3 were obviously larger (in absolute) than those from SS-EM2. It was because the penalty term for selectivity deviations in SS-EM2 served to pull the deviations towards zero but in SS-EM3 served to pull the selectivity deviations towards the values at adjacent ages and years when both ρ_a and ρ_t were positive and large. As the reference value for this penalty term changed from zero (SS-EM2) to those at adjacent grid points (SS-EM3), the peaks and valleys of the selectivity deviation surface from SS-EM3 can be farther away from zero.

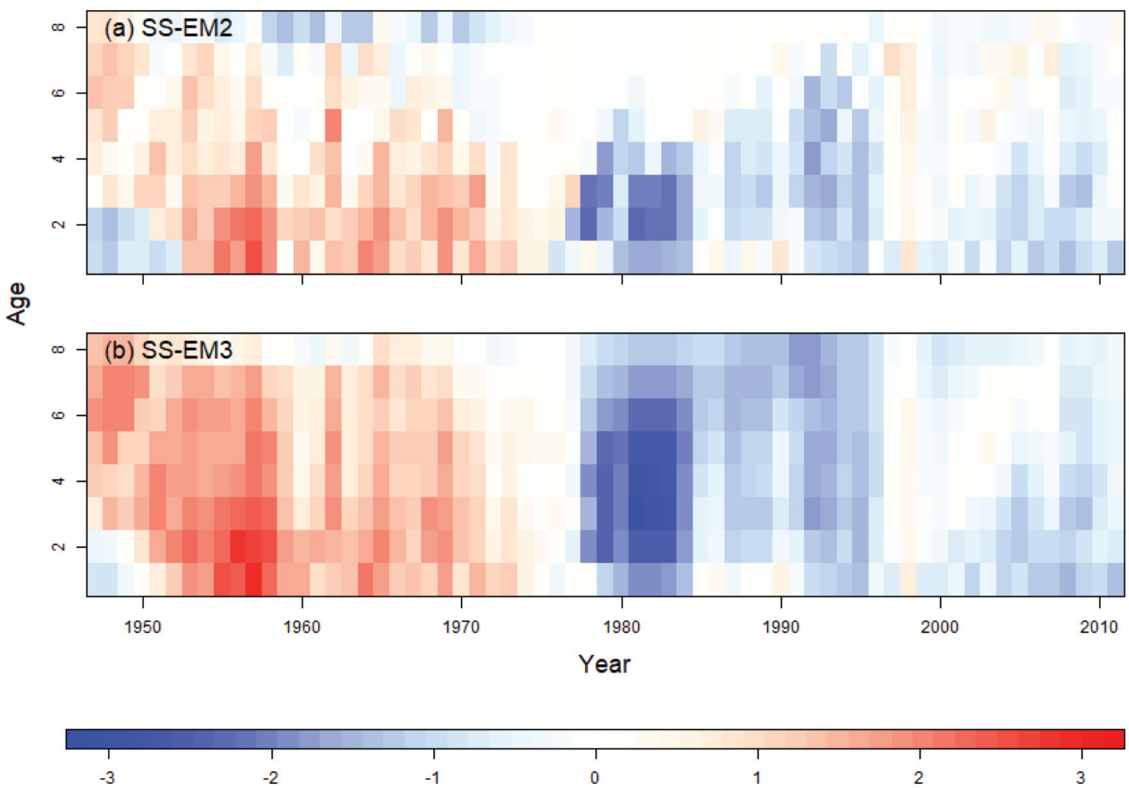
Inspecting residual patterns is another basic way of evaluating model performance. Large residuals are indicative of a lack of fit and temporal trends in residuals are indicative of model misspecification. The Pearson residuals in fit to the fishery age composition under IID selectivity deviations (SS-EM2) were smaller (maximum 1.23) than those under constant selectivity (SS-EM1, maximum 4.45), but they still had groups of positive or negative estimates across both age and year (Figs. A4 and A5). This systematic pattern in residuals implied that the assumption of independent selectivity deviations was violated. In contrast, the Pearson residuals in fit to the fishery age composition under 2D AR deviations (SS-EM3) were even smaller (maximum 0.98) and moreover distributed more randomly because of the among-age and among-year autocorrelations in selectivity deviations being

Table 3. Metrics mean relative error (MRE) and root mean square error (RMSE) of the terminal year estimates of attributes *F* and SB in the second simulation experiment.

	$\rho_a = 0.4$ and $\rho_t = 0.4$				$\rho_a = 0.4$ and $\rho_t = 0.8$				$\rho_a = 0.8$ and $\rho_t = 0.4$			
	F		SB		*F*		SB		*F*		SB	
	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE	MRE	RMSE
Pacific hake (*Merluccius productus*)												
EM1	0.00	0.17	0.04	0.28	–0.02	0.16	0.08	0.34	0.02	0.24	0.05	0.32
EM2	0.01	0.10	0.02	0.23	0.00	0.11	0.05	0.30	0.02	0.13	0.04	0.31
EM3	0.01	0.10	0.00	0.20	0.00	0.11	0.02	0.24	0.01	0.11	0.01	0.25
EM4	0.01	0.10	–0.01	0.20	0.00	0.11	0.01	0.23	0.01	0.11	0.00	0.23
EM5	0.01	0.10	0.00	0.19	0.00	0.11	0.01	0.22	0.01	0.11	0.00	0.22
Pacific sardine (*Sardinops sagax*)												
EM1	0.01	0.20	0.08	0.30	–0.01	0.17	0.07	0.32	0.02	0.24	0.02	0.29
EM2	0.01	0.12	0.03	0.25	0.02	0.11	0.02	0.27	0.03	0.14	0.00	0.30
EM3	0.00	0.11	0.03	0.23	0.02	0.11	0.00	0.22	0.03	0.13	0.00	0.26
EM4	0.00	0.11	0.03	0.23	0.02	0.11	–0.01	0.22	0.02	0.12	0.00	0.25
EM5	0.00	0.11	0.02	0.22	0.02	0.11	–0.01	0.21	0.02	0.12	0.00	0.25

Note: The three columns correspond to various levels of among-age (ρ_a) and among-year (ρ_t) autocorrelations in selectivity deviations and EM1–5 are the five estimation models compared in the second simulation experiment.

Fig. 7. Estimated selectivity deviation pattern ($\hat{\varepsilon}$) for the North Sea herring (*Clupea harengus*) fishery from (*a*) SS-EM2 and (*b*) SS-EM3 in the case study. [Color online.]

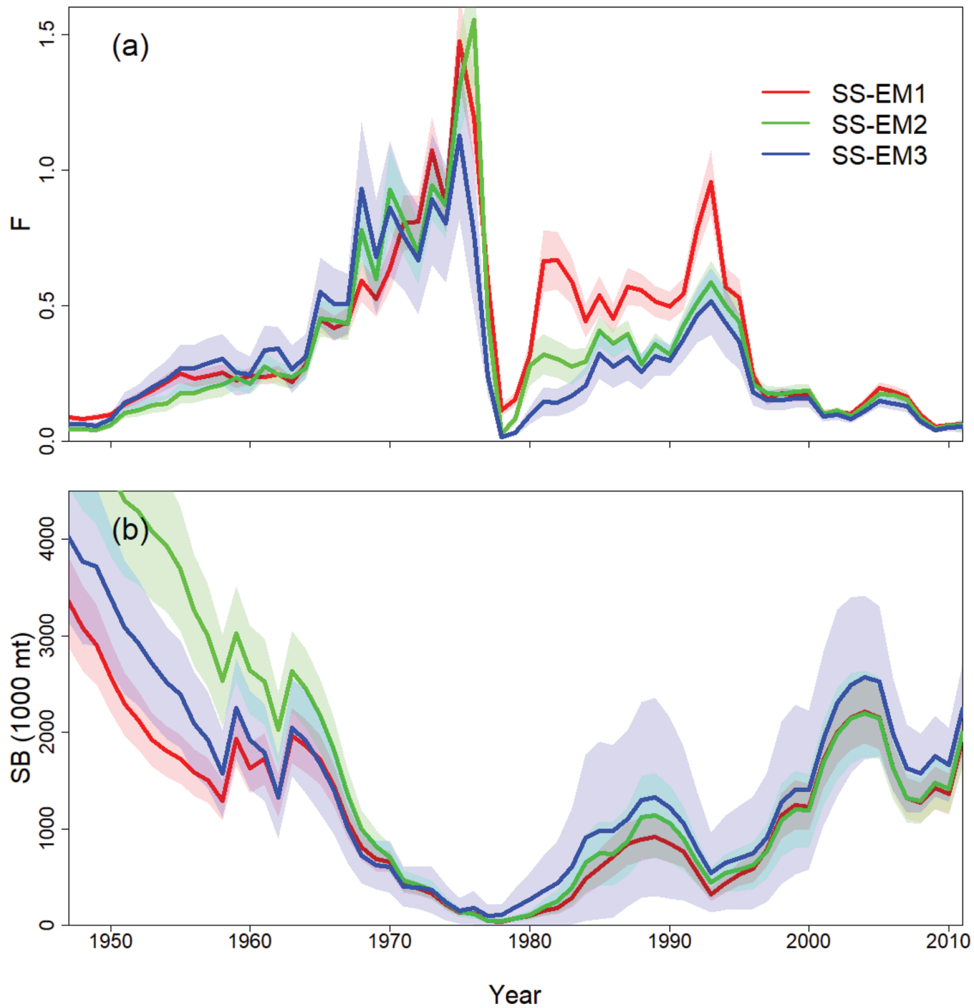


accounted for (Fig. A6). As such, applying the semi-parametric age- and time-varying selectivity method for the fishery of North Sea herring resulted in both improved fit and reduced model misspecification.

Lastly, we compared the estimates of *F* and SB that the three runs provided for North Sea herring (Fig. 8). Before 1975 and after 1995, the three runs provided similar *F* estimates but the associated uncertainties in SS-EM2 and SS-EM3 were notably larger than in SS-EM1 (Fig. 8*a*). During 1975–1995 when the deviations in selectivity abruptly changed from positive to negative (Fig. 7), however, SS-EM2 and especially SS-EM3 provided profoundly lower *F* estimates than SS-EM1. SB estimates from the three runs also differed notably but the largest differences occurred in the initial and terminal years of the assessment instead (Fig. 8*b*). More specifically, SS-EM3 provided 27% lower and 20% higher initial year SB estimates than SS-EM2 and SS-EM1, respectively. For the terminal year, SS-EM1 and SS-EM2 provided negligibly different (within 3%) SB estimates, which, however, were 12% lower than SS-EM3 provided. We also found that the estimates of *F* and SB from SS-EM2 and SS-EM3 had a notably larger level of uncertainty than those from SS-EM1. In summary, accounting for autocorrelated selectivity (SS-EM3) improved model fit, reduced the 2D pattern in the Pearson residuals for the fishery, and had a noticeable effect on the estimates and the associated uncertainty of critical population attributes.

We acknowledged that at least two obvious differences existed between the configuration of simulation experiments and that of the case study. First, the effective sample size was unknown and

Fig. 8. The estimates of North Sea herring (*Clupea harengus*) *F* and SB from the three Stock Synthesis runs in the case study. The shaded areas show the 95% confidence intervals of the estimates. [Color online.]



estimated in the case study but was assumed known without error in the simulation experiments. Second, the level of selectivity variation (1.04) in the case study was estimated to be much larger than that (0.4) assumed in the simulation experiments. Estimating the effective sample size in the case study induced additional uncertainty and bias in model estimates, and a large increase in the level of selectivity variation underlined the importance of accounting for age- and time-varying selectivity. To evaluate whether the rank of model performance in the two stimulation experiments was also robust for the case study, we conducted two sensitivity tests regarding data weighting and level of selectivity variation under high data quality (n_{comp} = 200 and CV_{abund} = 0.1) and high autocorrelations (ρ_a = 0.8 and ρ_t = 0.8) in selectivity deviations. As expected, the precision of SB estimates decreased when the effective sample size was estimated inside the model in comparison to being fixed at the true value (Fig. A7, middle column). However, 2D AR deviations (EM3) still had pronouncedly higher precision than either constant selectivity (EM1) or IID deviations (EM2). Moreover, this pattern was found to be more dramatic when the level of variation in selectivity was higher (Fig. A7, middle column). As such, accounting for the highly autocorrelated (ρ_a = 0.51 and ρ_t = 0.79) and variable ($\hat{\sigma}_S$ = 1.04) selectivity of the North Sea herring fishery (via our new semi-parametric selectivity method) was expected to be critical.

Discussion

This paper provides a new semi-parametric method to account for autocorrelated age- and time-varying (termed "2D autocorrelated") selectivity in age-structured assessment models. This method includes a parametric selectivity form and nonparametric process errors that can be autocorrelated among ages and years. We conducted a simulation experiment to evaluate the performance of this new method and found that it resulted in an increased precision of *F* and SB estimates for both hake and sardine, given that the deviations in selectivity were highly autocorrelated. Moreover, the degree to which the precision increases was positively related to the quality of fishery data and the accuracy of estimates of the three hyper-parameters ($\hat{\sigma}_S$, ρ_a, and ρ_t) for semi-parametric age- and time-varying selectivity. We conducted another simulation experiment to evaluate the importance of the semi-parametric selectivity method under various levels of autocorrelation in selectivity deviations. Given that the quality of fishery data is high, the EMs that used the new semi-parametric method to account for autocorrelated deviations in selectivity (2D AR deviations, REML estimation, and perfect information) outperformed the other EMs (constant selectivity and IID deviations) in terms of the precision of SB estimates, especially when the deviations in selectivity were highly autocorrelated. Regardless of data quality and selectivity autocorrelation, REML estimation (EM4) was more accurate than sample-based estimation (EM3) in estimating the three hyper-parameters. Consequently, REML estimation (EM4) was more precise than sample-based estimation (EM3) with respect to estimating both *F* and SB. However, the REML method involves estimating random effects, so it cannot be used in ADMB-based assessment packages (e.g., Stock Synthesis).

Not surprisingly, the 2D autocorrelated selectivity method was more important to the assessments that have high-quality age composition data. A larger number of effective samples per year corresponds to more informative age composition data and thereby more accurate estimates of the two autocorrelation coefficients (ρ_a and ρ_t). It also corresponds to a larger weight for age composition data in the objective function. When the likelihood term for age composition data makes a more significant contribution to the objective function, model estimates are expected to be more sensitive to how selectivity is parameterized in the EM. Thus, accounting for autocorrelated deviations in age- and time-varying selectivity was found to be most crucial to the high-quality (large *b*) data case.

For the high-quality data case, we conducted another simulation experiment to evaluate the importance of the 2D autocorrelated selectivity method under three other levels of autocorrelation in selectivity deviations. Results suggested that the semi-parametric selectivity method was more important in terms of the precision of *F* estimates when the true deviations in selectivity were highly autocorrelated among ages and was more important in terms of the precision of SB estimates when the true deviations in selectivity were highly autocorrelated among years. In an additional simulation that we introduced, we also found that the importance of accounting for autocorrelated selectivity deviations (in terms of the precision of SB estimates) was positively related to the level of variation in the corresponding selectivity ($\hat{\sigma}_S$).

Maunder (2011) found that estimating the effective sample size may lead to more accurate recruitment estimates when the EM ignores the large variation in selectivity. Real-world age composition data are likely correlated rather than independent due to, for example, age-specific aggregation or schooling (McAllister and Ianelli 1997). The positive among-age correlation in residuals for composition data results in overdispersion, i.e., the effective sample size is smaller than the actual sample size (Maunder 2011; Francis 2014; Thorson et al. 2017). The simulation experiment conducted by Maunder (2011) suggested that estimating the effective sample size could improve estimation performance when the effective sample size is less than one fifth of the actual sample size. We therefore recommend that future research examines (1) how the ratio of effective to actual sample size is affected by the level of autocorrelation in age- and time-varying selectivity and (2) the consequences of ignoring the overdispersion in composition data under various levels of autocorrelation in selectivity deviations. We believe that both topics will be feasible to explore using the 2D autocorrelated selectivity function to generate data.

We also implemented the 2D autocorrelated selectivity method in Stock Synthesis to evaluate the performance of the new Stock Synthesis feature using real data set for North Sea herring. We found that the age composition data from the fishery was very informative (b = 0.9) regarding estimating selectivity deviations. Moreover, selectivity of the fishery was highly variable ($\hat{\sigma}_S = 1.04$) and the deviations were highly autocorrelated among ages (ρ_a = 0.51) and years (ρ_t = 0.79). Thus, as expected, SS-EM3 (which implemented 2D autocorrelated selectivity) fitted the data better and provided more randomly distributed Pearson residuals of the fishery catch than the other two runs (which ignored selectivity variation or assumed independent selectivity deviations). It is important to note that the age- and time-varying selectivity feature increased not only model fit but also computation time. On a laptop with a four-core Intel processer, turning on the new selectivity feature in the North Sea herring case study increased the computation time from less than 1 min (SS-EM1) to more than 3 min (SS-EM2 and SS-EM3) due to the estimation of additional 520 (65 years × 8 ages) selectivity deviations for the fishery.

As an example of a case that might benefit from this new approach, a recent assessment of Pacific hake (Berger et al. 2017) found that the parameterization of time-varying selectivity induced profound uncertainty to terminal year estimates of SB and a better way to parameterize selectivity of the fishery was highly recommended by both the assessment authors and the reviewers. The parameterization in Stock Synthesis used in this assessment treats fishery selectivity as a time-varying process by adding annual deviations to the time-invariant selectivity parameters, each of which represents a change in selectivity from one age to the next (up to age 6, beyond which selectivity is assumed to be constant). The resulting selectivity, including annual deviations, is rescaled to sum to 1 in each year. Although the deviations are treated as independent of age and time, the parameterization introduces negative correlations among the deviations within each year, as the combination of the offset setup and the rescaling will cause a positive deviation in the parameter for any one age to reduce the selectivity for all other ages unless it is offset by a negative deviation for some other age. The hake assessment relies on MCMC sampling of the parameter space, so unnecessary parameter correlations reduce the sampling efficiency and lead to long run times (Berger et al. 2017). The 2D autocorrelated selectivity may overcome some of the challenges faced in the hake assessment by representing the variation over time as independent of the time-invariant selectivity (rather than rescaling the combination of the two factors) and by explicitly modeling the autocorrelation in age and time, rather than assuming independent deviation.

Before estimating the two autocorrelation coefficients (ρ_a and ρ_t) for selectivity deviations, we recommend checking the estimate of *b* from SS-EM2 to estimate the quality of age composition data regarding estimating $\hat{\boldsymbol{\varepsilon}}$ as well as the two autocorrelation coefficients. If the value of *b* is large (close to 1.0) and the Pearson residuals for the fishery have an obvious pattern across ages or years, then analysts could explore using the 2D autocorrelated selectivity method in Stock Synthesis. The choice of the age and year range over which selectivity is assumed to be time varying should depend upon the distribution of the quality of age compositions. We recommend focusing on only the data-rich age and time period when exploring the 2D autocorrelated selectivity method in Stock Synthesis. If the age and year range chosen for the semi-parametric selectivity method is too large (e.g., including the poor-sampled initial years and oldest age groups), the data may not be informative enough to estimate $\hat{\boldsymbol{\varepsilon}}$ and the two autocorrelation coefficients with reasonable accuracy and precision. Moreover, care should be taken when using *b* as a measure of the composition data quality because *b* is conditional on how properly the corresponding age composition data are weighted. In the simulation experiments, *b* was iteratively tuned based on the assumption that the effective sample size of the age composition data is known without error. If the effective sample size in a real-world assessment is over- or underestimated to a large extent, the calculation of *b* could be biased and uninformative.

Different from previous simulation studies where the standard deviation of a random effect was typically assumed known without error (Haltuch and Punt 2011; Johnson et al. 2016), we instead specified EM2–4, which estimate $\hat{\sigma}_S$, so that the simulation corresponds more closely to a real-world assessment process. The choice of the constraint imposed upon the magnitude of selectivity variation ($\hat{\sigma}_S$ in this case) could affect the estimates of important parameters and derived quantities (Francis 2011). Decisions regarding the value of $\hat{\sigma}_S$ were often subjective in past studies (Butterworth et al. 2003; Maunder et al. 2014; Punt et al. 2014), but we have evaluated both simple (penalized likelihood, EM2) and advanced (REML, EM4) approaches to uniquely estimate its value. The fact that $\hat{\sigma}_S$ was accurately and precisely estimated by EM2 in all three data quality cases provides additional credibility for implementing and exploring the semi-parametric selectivity method in Stock Synthesis for real-world stock assessments, in which the truth $\hat{\sigma}_S$ of any selectivity is not known. While σ_S cannot be estimated as a fixed-effect parameter in Stock Synthesis while treating $\hat{\boldsymbol{\varepsilon}}$ as a random effect, the Methot–Taylor tuning method serves as a feasible and reliable alternative to provide objective σ_S estimation.

Although the semi-parametric age- and time-varying selectivity approach allows us to estimate the level of variation in selectivity ($\hat{\sigma}_S$) iteratively, care should be taken when applying this approach to real-world data sets. Acknowledging that this iteration algorism could be less efficient or even problematic in assessments when more than one $\hat{\sigma}_S$ needs to be tuned at the same time, we therefore recommend restricting the application of the semi-parametric age- and time-varying selectivity method to the most important fishery fleet(s). As for weighting composition data for the fleet to which the semi-parametric age- and time-varying selectivity approach is applied, we recommend using the Dirichlet-multinomial method (Thorson et al. 2017) rather than the more widely used iteration-based methods (McAllister and Ianelli 1997; Francis 2011). In comparison with those methods, the Dirichlet-multinomial method weights composition data inside the model based on maximum likelihood and thus gets rid of the interaction between the iteration for $\hat{\sigma}_S$ and that for effective sample size.

We are aware a caveat regarding the 2D AR(1) approach to account for age and time autocorrelated deviations in selectivity. This approach is treated as a first-order approximation of the complex 2D selectivity variation, which arises often due to a combination of several factors such as cohort strength, fishing gear, and behavior as well as the spatiotemporal distribution of the fish population. While some factors like fishing gear and behavior tend to be autocorrelated across time, some other factors like cohort strength are not necessarily being autocorrelated across time. It is important to note that the 2D AR(1) approach is not designed to deal with cohort-specific selectivity, but that is a topic that could be considered in future research. Ageing imprecision could be another important contributor to observed autocorrelation in age composition data. However, it is still unclear whether the inclusion of autocorrelated selectivity in the assessment model can reduce the ageing-caused biases in assessment outcomes.

Another caveat regarding the 2D AR(1) approach is that only positive correlation is allowed in selectivity deviations from two age bins. That constraint could be problematic, as selectivity deviations at two distant ages could be negatively correlated (e.g., Fig. 1A in Francis 2017). We recommend future research developing advanced AR structures to deal with more types of among-age correlation pattern for selectivity deviations. We acknowledge that the Pearson residuals for catch at age data from our approach are larger than those from the state-space approach that ICES used in the North Sea herring stock assessment (ICES 2017). The difference in fit between the two approaches is possibly due to the fact that the state-space model is more flexible than Stock Synthesis, as it also allows a random effect on survivorship.

Acknowledgements

We thank Tim Essington, Jim Ianelli, Michelle Mcclure, the Associate Editor, and two anonymous reviewers for providing valuable inputs to an earlier draft. We also thank Kasper Kristensen for developing Template Model Builder. Haikun Xu was partially funded by NOAA Award NA15OAR4320063 through the University of Washington Joint Institute for the Study of the Atmosphere and Ocean.

References

Berger, A.M., Grandin, C.J., Taylor, I.G., Edwards, A.M., and Cox, S. 2017. Status of the Pacific Hake (whiting) stock in U.S. and Canadian waters in 2017. Prepared by the Joint Technical Committee of the U.S. and Canada Pacific Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada.

Butterworth, D., Ianelli, J., and Hilborn, R. 2003. A statistical model for stock assessment of southern bluefin tuna with temporal changes in selectivity. Afr. J. Mar. Sci. **25**: 331–361. doi:10.2989/18142320309504021.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Opt. Methods Softw. **27**(2): 233–249. doi:10.1080/10556788.2011.597854.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. **68**(6): 1124–1138. doi:10.1139/f2011-025.

Francis, R.I.C.C. 2014. Replacing the multinomial in stock assessment models: a first step. Fish. Res. **151**: 70–84. doi:10.1016/j.fishres.2013.12.015.

Francis, R.I.C.C. 2017. Revisiting data weighting in fisheries stock assessment models. Fish. Res. **192**: 5–15. doi:10.1016/j.fishres.2016.06.006.

Gelman, A. 2005. Analysis of variance — why it is more important than ever. Ann. Stat. **33**(1): 1–53. doi:10.1214/009053604000001048.

Haltuch, M.A., and Punt, A.E. 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. **68**(5): 912–926. doi:10.1139/f2011-030.

Harville, D.A. 1974. Bayesian inference for variance components using only error contrasts. Biometrika, **61**(2): 383–385. doi:10.1093/biomet/61.2.383.

ICES. 2017. Herring Assessment Working Group for the Area South of 62 deg N (HAWG). ICES Document CM 2017/ACOM:07: 142.

Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., and Punt, A.E. 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fish. Res. **183**: 222–232. doi:10.1016/j.fishres.2016.06.004.

Kristensen, K., Nielsen, A., Berg, C., and Skaug, H. 2015. Template model builder TMB. J. Stat. Softw. **70**: 1–21.

Martell, S., and Stewart, I. 2014. Towards defining good practices for modeling time-varying selectivity. Fish. Res. **158**: 84–95. doi:10.1016/j.fishres.2013.11.001.

Maunder, M.N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. **109**(2): 311–319. doi:10.1016/j.fishres.2011.02.018.

Maunder, M.N., and Piner, K.R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fish. Res. **192**: 16–27. doi:10.1016/j.fishres.2016.04.022.

Maunder, M., Crone, P., Valero, J., and Semmens, B. 2014. Selectivity: theory, estimation, and application in fishery stock assessment models. Fish. Res. **158**: 1–4. doi:10.1016/j.fishres.2014.03.017.

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. **54**(2): 284–300. doi:10.1139/f96-285.

Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. **68**(10): 1744–1760. doi:10.1139/f2011-092.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

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

Punt, A.E., Hurtado-Ferro, F., and Whitten, A.R. 2014. Model selection for selectivity in fisheries stock assessments. Fish. Res. **158**: 124–134. doi:10.1016/j.fishres.2013.06.003.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Sampson, D.B., and Scott, R.D. 2012. An exploration of the shapes and stability of population–selection curves. Fish Fish. **13**(1): 89–104. doi:10.1111/j.1467-2979.2011.00417.x.

Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Can. J. Fish. Aquat. Sci. **71**(11): 1655–1666. doi:10.1139/cjfas-2013-0508.

Stewart, I.J., and Martell, S.J. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fish. Res. **158**: 40–49. doi:10.1016/j.fishres.2013.09.012.

Stewart, I.J., and Monnahan, C.C. 2017. Implications of process error in selectivity for approaches to weighting compositional data in fisheries stock assessments. Fish. Res. **192**: 126–134. doi:10.1016/j.fishres.2016.06.018.

Thorson, J.T., and Cope, J.M. 2015. Catch curve stock-reduction analysis: an alternative solution to the catch equations. Fish. Res. **171**: 33–41. doi:10.1016/j.fishres.2014.03.024.

Thorson, J.T., and 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. doi:10.1016/j.fishres.2013.10.002.

Thorson, J.T., Minto, C., Minte-Vera, C.V., Kleisner, K.M., and Longo, C. 2013. A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. Can. J. Fish. Aquat. Sci. **70**(12): 1829–1844. doi:10.1139/cjfas-2013-0280.
Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. **192**: 84–93. doi:10.1016/j.fishres.2016.06.005.
Trenkel, V.M., Bravington, M.V., and Lorance, P. 2012. A random effects population dynamics model based on proportions-at-age and removal data for estimating total mortality. Can. J. Fish. Aquat. Sci. **69**(11): 1881–1893. doi:10.1139/f2012-103.

Appendix A

Appendix Figs. A1–A7 appear on the following pages.

Fig. A1. The temporal coverage of the data included in the North Sea herring (*Clupea harengus*) case study. [Color online.]

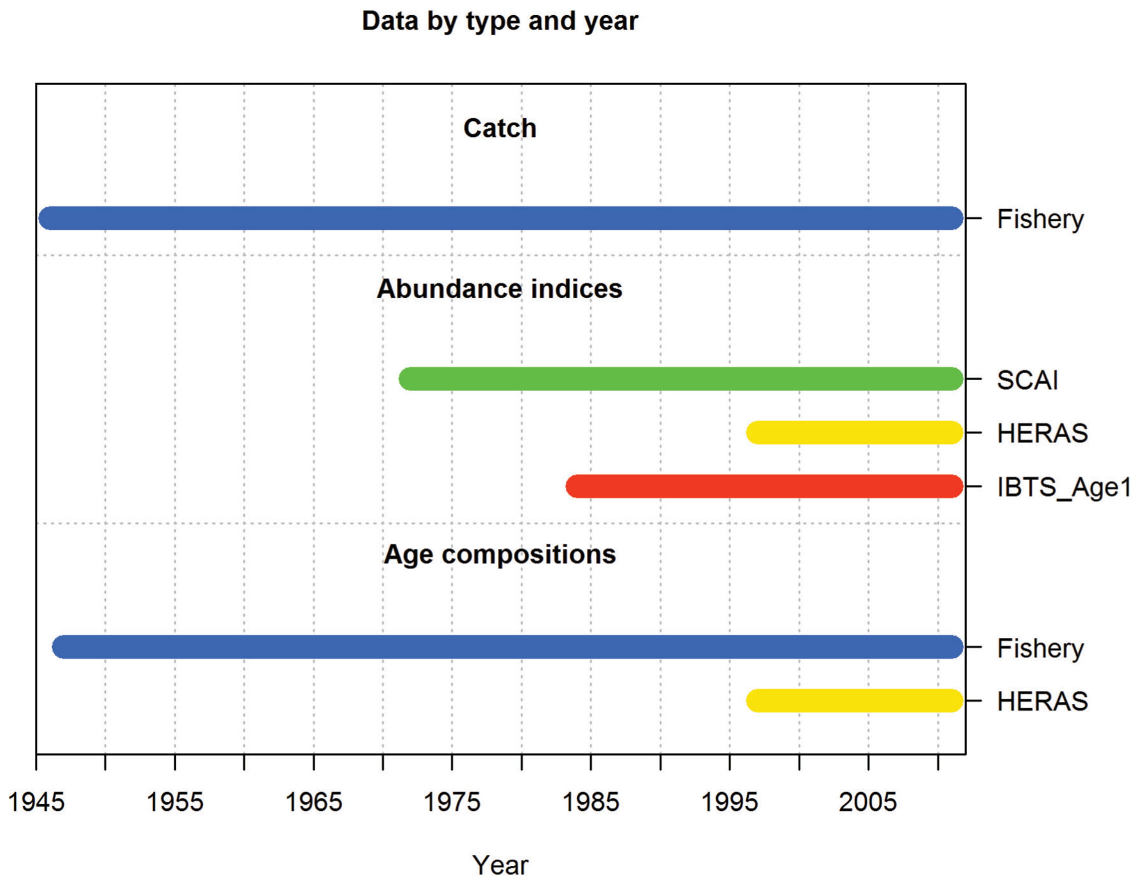


Fig. A2. Boxplots for b (see eq. 20) from the three data cases. b is a scaler (between 0 and 1) quantifying the richness of the fishery composition data in terms of estimating semi-parametric age- and time-varying selectivity. The lower and upper hinges mark the first and third quantiles and the two whiskers extend to the value no further than 1.5 interquartile range from the corresponding hinge. The three columns correspond to the three data quality cases. [Color online.]

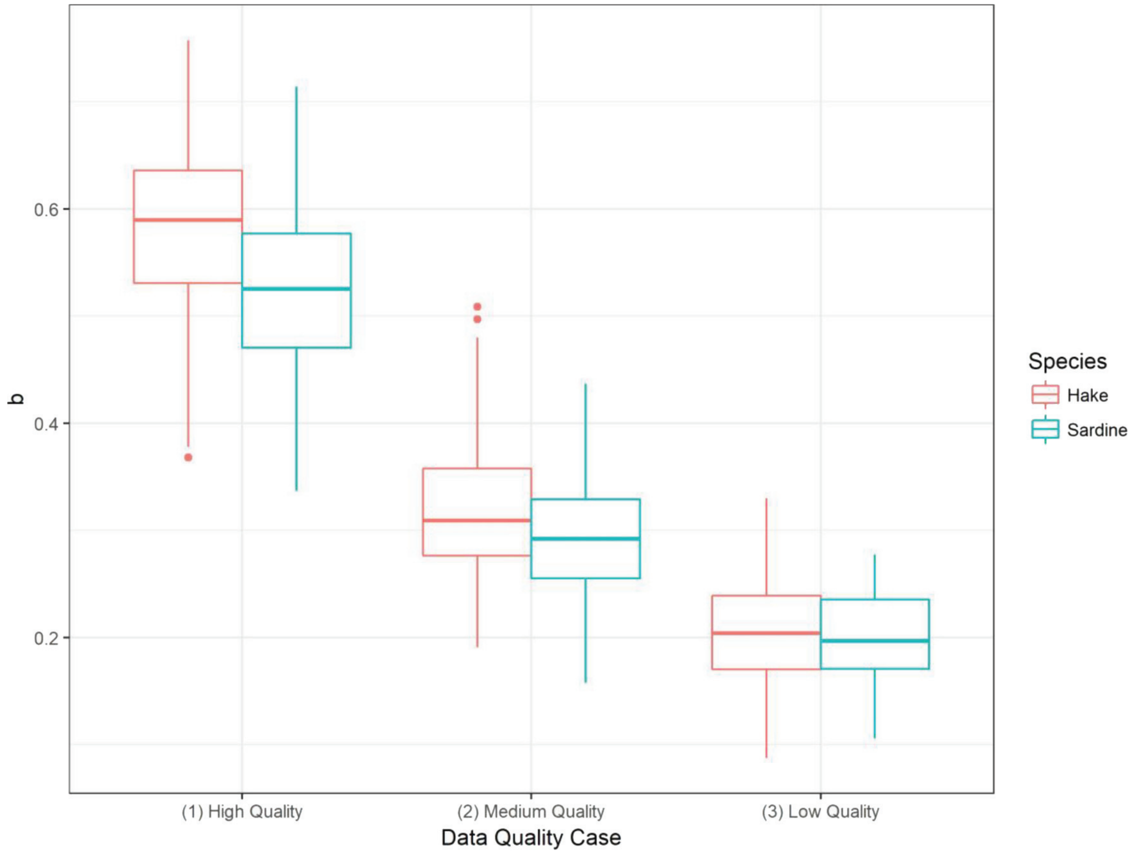


Fig. A3. The interquartile range of relative error in the estimates of F (top) and SB (bottom) for Pacific hake (*Merluccius productus*) in the additional simulation experiment (under the high data quality and selectivity autocorrelation case). EM6 is the same as EM3 except the among-age selectivity autocorrelation (ρ_a) was fixed at 0. The three columns correspond to the three OMs with different qualities of fishery data. [Color online.]

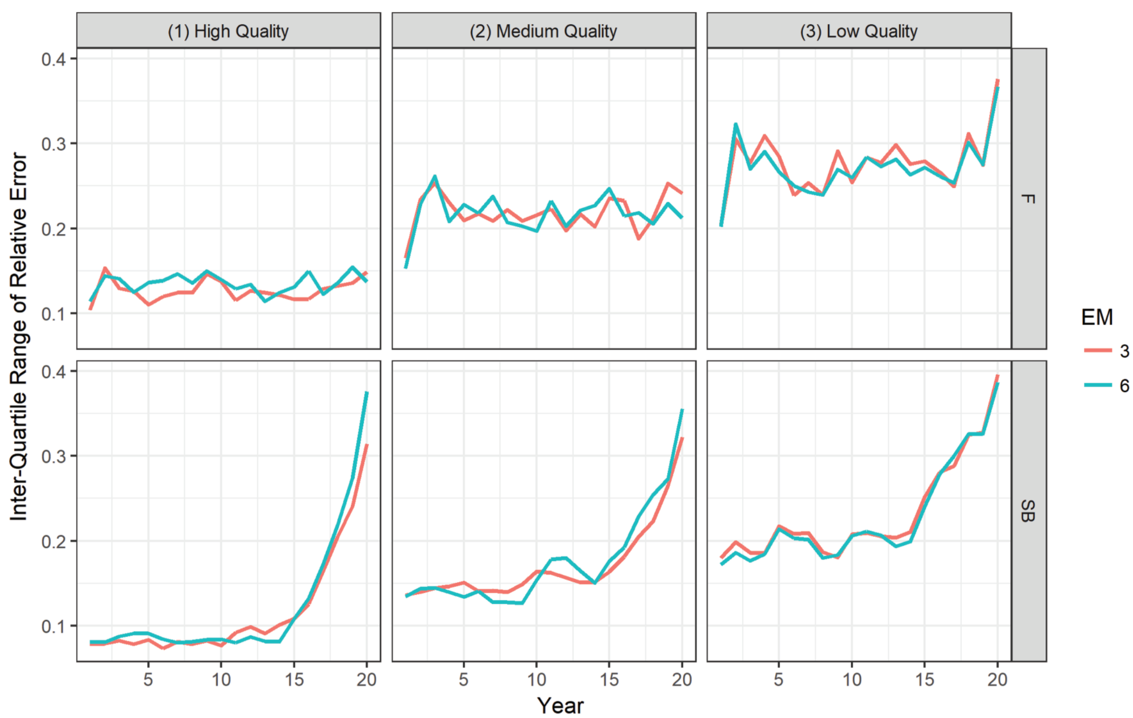


Fig. A4. The Pearson residuals in fit to the North Sea herring (*Clupea harengus*) fishery age composition from SS-EM1.

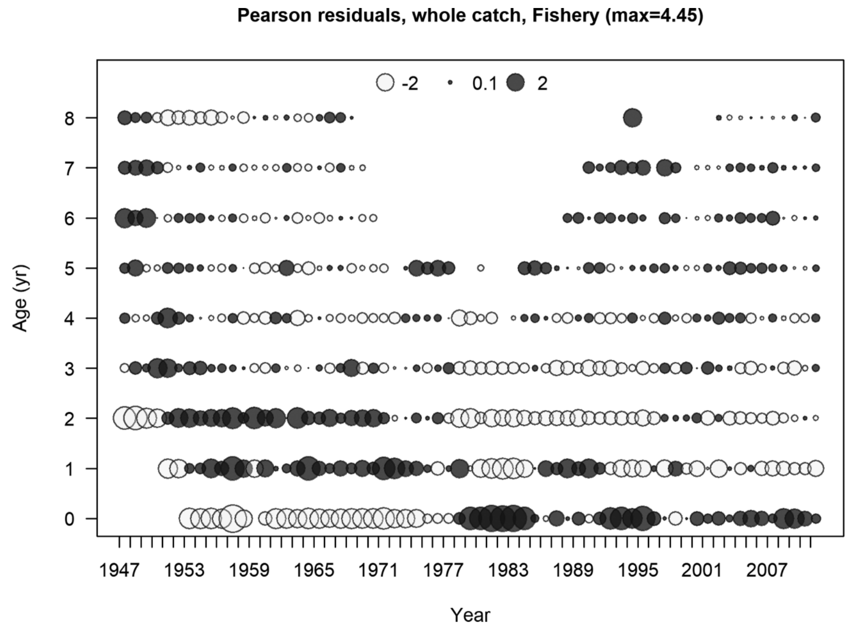


Fig. A5. The Pearson residuals in fit to the North Sea herring (*Clupea harengus*) fishery age composition from SS-EM2.

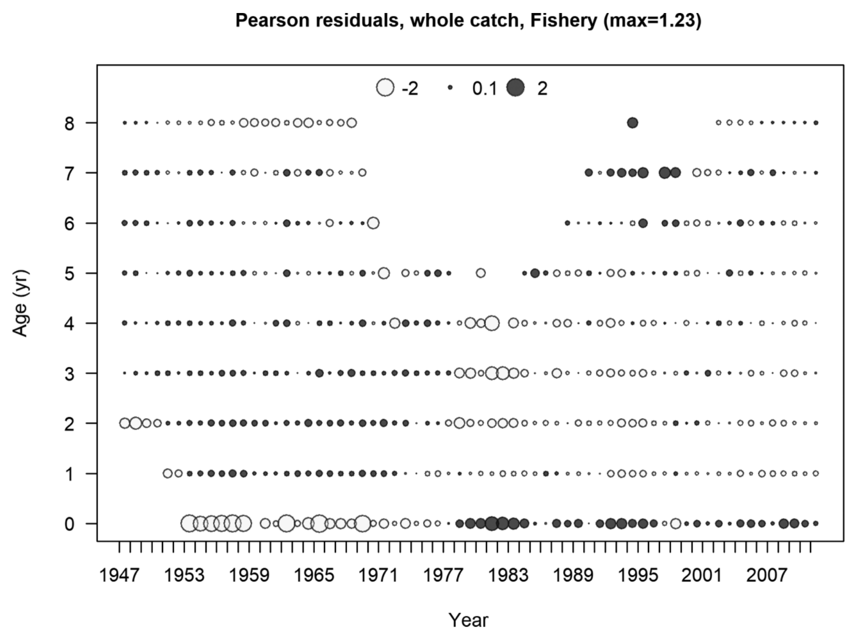


Fig. A6. The Pearson residuals in fit to the North Sea herring (*Clupea harengus*) fishery age composition from SS-EM3.

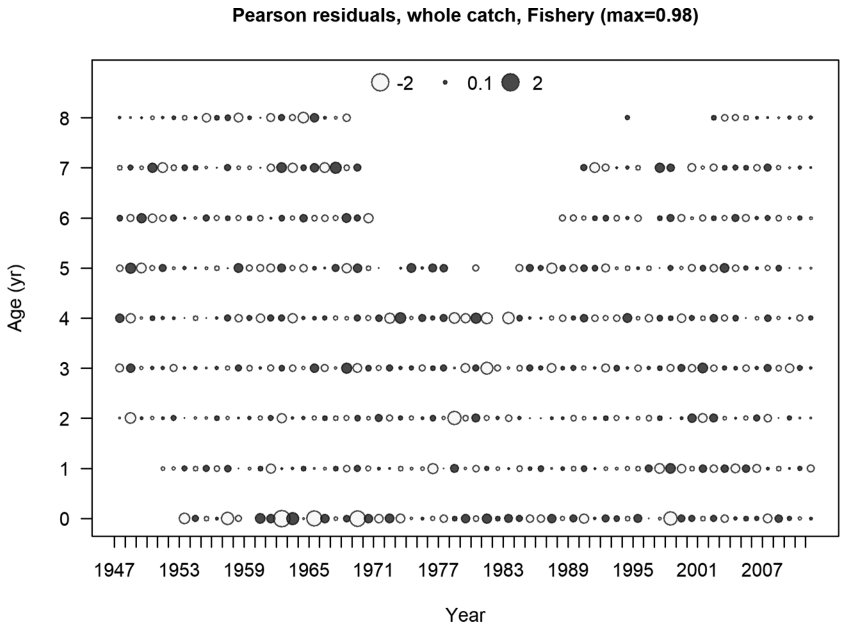


Fig. A7. The interquartile range of relative error in the estimates of *F* (top) and SB (bottom) for Pacific hake (*Merluccius productus*) in the sensitivity simulation test (under the high data quality and selectivity autocorrelation case). The three columns correspond to the base case in the simulation experiments (left), the effective sample size of the fishery age composition data are estimated (middle), and the true level of selectivity variation is doubled (σ_s = 0.8) compared with the base case (right). [Color online.]