ICES Journal of
Marine Science

ICES Journal of Marine Science (2018), 75(6), 2016–2024. doi:10.1093/icesjms/fsy126

Original Article

Comparison of the performance of age-structured models with few survey indices

Hiroshi Okamura[1,*], Yuuho Yamashita[2], Momoko Ichinokawa[1], and Shota Nishijima[1]

[1]*Research Center for Fisheries Management, National Research Institute of Fisheries Science, Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan*
[2]*Fisheries Management Department, Hokkaido National Fisheries Research Institute, Fisheries Research and Education Agency, 116 Katsurakoi, Kushiro, Hokkaido 085-0802, Japan*

**Corresponding author: tel: + 81 45 788 7514; fax: + 81 45 788 5004; e-mail: okamura@fra.affrc.go.jp.*

Okamura, H., Yamashita, Y., Ichinokawa, M., and Nishijima, S. Comparison of the performance of age-structured models with few survey indices. – ICES Journal of Marine Science, 75: 2016–2024.

Received 24 April 2018; revised 9 August 2018; accepted 16 August 2018; advance access publication 12 September 2018.

Age-structured models have played an important role in fisheries stock assessment. Although virtual population analysis (VPA) was once the most widely used stock assessment model for when catch-at-age information is available, (hierarchical) statistical catch-at-age analysis (SCAA) is about to take that position. However, the estimation performance of different age-structured models has not been evaluated sufficiently, especially in cases where there are few available abundance indices. We examined the performance of VPA and SCAA using simulation data in which only the abundance indices of spawning stock biomass and recruitment were available. The simulation demonstrated that VPA with the ridge penalty selected by minimizing retrospective bias provided near-unbiased abundance estimates without catch-at-age error and moderately biased estimates with catch-at-age error, whereas SCAA with random-walk selectivity suffered from problems in estimating parameters and population states. Without sufficient information on abundance trends, naïvely using SCAA with many random effects should be done cautiously, and comparing results from various age-structured models via simulation tests will be informative in selecting an appropriate stock assessment model.

Keywords: ridge penalty, state-space model, statistical catch-at-age model, time-varying selectivity, virtual population analysis.

Introduction

Age-structured models have been widely used to estimate historical shifts in population sizes and fishing mortalities for assessment of fisheries stock (Quinn and Deriso, 1999). There are two traditional age-structured models used in fisheries science: virtual population analysis (VPA) and statistical catch-at-age analysis (SCAA). VPA estimates the abundance-at-age and the fishing mortality-at-age from catch-at-age data by backward solving the deterministic population dynamics equation and assuming no or negligible proportion-at-age error. In contrast to VPA, SCAA allows for errors in catch-at-age data and internally estimates the stock recruitment relationship by assuming a stochastic population dynamics equation. SCAA has become increasingly popular as computational ability has improved (Nielsen and Berg, 2014; Berg and Nielsen, 2016). However, there have been few studies comparing the performance of VPA with that of SCAA (Radomski *et al.*, 2005; Butterworth and Rademeyer, 2008).

Classical VPA usually requires strong assumptions about the population sizes and fishing mortalities for the terminal age and year to backward solve the equation from only catch-at-age data (Lassen and Medley, 2001). To limit dependence on these strong assumptions for the terminal year and to compensate for information deficiency from using only the catch-at-age data, a fitting process to a relative abundance index, such as catch-per-unit-effort (CPUE), is typically incorporated into VPA (Lassen and Medley, 2001). However, when age-specific abundance indices are missing or incomplete, the estimates of fishing mortalities for the terminal year in the VPA tuned by age-limited or age-aggregated abundance indices become unstable due to lack of information on selectivity-at-age. We need a relatively weak assumption on the selectivity patterns to overcome this (Punt et al., 2014). For example, the selectivity-at-age for the terminal year might be assumed to be equal to the selectivity-at-age averaged for the last few years, typically three or five (Ichinokawa and Okamura, 2014; Hashimoto et al., 2018). However, nobody has formally identified how sensitive stock assessment results are to such a weak assumption about selectivity-at-age. However, Okamura et al. (2017) have developed a new method, the ridge virtual population analysis (r-VPA), that does not require the assumption of constant selectivity-at-age over recent time, even in data-poor situations such as having few abundance indices. In r-VPA the fishing mortalities at the terminal year can be stably estimated using the ridge regression penalty selected by minimizing retrospective bias (Mohn, 1999). Ridge VPA was used for assessing some Japanese fishery stocks in 2017 (Kurota et al., 2018; Yamashita et al., 2018; Yukami et al., 2018).

SCAA commonly assumes that the selectivity pattern is constant over time under a separability assumption, meaning that the fishing mortality-at-age is divided into separate age and year components (Butterworth and Rademeyer, 2008). Some simulation studies have demonstrated that the assumption of a constant selectivity pattern in SCAA could cause serious bias in abundance estimates when the selectivity pattern has actually changed over time (Radomski et al., 2005; Ichinokawa et al., 2014). Nielsen and Berg (2014) presented a simple state-space assessment model (SAM) to estimate a time-varying selectivity pattern following a random walk with a multivariate normal distribution as well as the transition of population size. While the SCAA with selectivity modelled by the random walk, such as SAM, is generally expected to perform better than that without random walk selectivity (Punt et al., 2014), the performance of the state-space SCAA remains unknown when there is large uncertainty and limited information. Because the number of parameters to be estimated tends to be large when using age-structured models, good estimation performance may not be always guaranteed (Auger-Méthé et al., 2016).

The objective of this paper is to examine how to choose an appropriate age-structured model, especially when there are few available abundance indices and the annual selectivity-at-age pattern varies systematically (i.e. time-varying selectivity patterns). For this, the simulated catch-at-age data were generated to emulate a constant selectivity pattern and a non-constant selectivity pattern using the northern Japan Sea stock (JSS) of walleye pollock (*Gadus chalcogrammus*) in the Sea of Japan (Yamashita, 2017). We then examined the performance of VPA and SAM by applying the models to the simulated catch-at-age data and abundance indices. We discuss when and where we should use VPA or SAM as well as the potential future direction of age-structured stock assessment modelling in data-poor situations.

Material and methods

Virtual population analysis

We used two tuned VPA procedures. One has a constant selectivity-at-age assumption (cs-VPA; Ichinokawa and Okamura, 2014), in which the latest selectivity-at-age was assumed to be equal to the selectivity-at-age averaged for the past 5 years, and then only the fishing mortality at the terminal age and year was estimated. The other is r-VPA, which has a ridge penalty for fishing mortalities-at-age in the terminal year (Y) instead of estimating them independently (r-VPA; Okamura et al., 2017). In r-VPA, the ridge penalty $\lambda \sum_{a=2}^{A-1} F_{a,Y}^2$ was added to the objective function of tuned VPA. The ridge penalty λ was selected by minimizing the retrospective bias of total biomass where Mohn's rho (Mohn, 1999) was used as an indicator of retrospective bias (Okamura et al., 2017). The "peeling" in Mohn's rho was set to 5. For both VPAs (cs-VPA and r-VPA), the population dynamics were modelled by Pope's approximation equation (Pope, 1972) and the oldest age group was a "plus group" containing individuals of a specified age or older. The fishing mortalities for the oldest and second-oldest age groups were assumed to be equal. The details of the VPA specification are provided in Supplementary Material A.

Statistical catch-at-age model

We used the state-space assessment model (SAM) with the logarithms of number-at-age and fishing mortality-at-age following a random walk (Nielsen and Berg, 2014), because SAM is a statistical catch-at-age assessment model that has recently been used for ICES stock assessments (Berg and Nielsen, 2016) and it is versatile, flexible, and easily applied to the same datasets as VPA. Specifically, the fishing mortality-at-age was assumed to have a multivariate normal distribution with a common and constant correlation coefficient ρ:

$$\log F_y = \log F_{y-1} + \xi_y,\ \xi_y \sim \mathrm{MVN}(0, \Sigma), \tag{1}$$

where $F_y = (F_{1,y}, F_{2,y}, \ldots,\ F_{A+,y})$, and Σ is the variance–covariance matrix with diagonal element σ_a and off-diagonal element $\rho\sigma_a\sigma_{a'}$ ($a \neq a'$). As in the VPAs, the oldest age group was a plus group and the fishing mortalities for the oldest and second-oldest age groups were assumed to be equal. Observation equations with a log-normal error structure were constructed for abundance indices and catch-at-age data, in contrast with the observation equation in VPA, which is applied to abundance indices alone (i.e. it assumes that the catch-at-age error is small enough to be ignored). The details of the SAM specification are provided in Supplementary Material B. A summary of the characteristics of VPA and SAM used in this paper is provided in Table 1.

Specification of the simulation test

We generated the simulation data from the 2016 stock assessment report for JSS walleye pollock in the Sea of Japan because the selectivity pattern of the JSS walleye pollock is thought to have recently changed and the abundance indices do not have full age-specific information (Yamashita, 2017). The stock recruitment relationship for the JSS walleye pollock is unclear because of a drastic linear decline (Ichinokawa et al., 2017; Okamura et al., 2017). We therefore fitted a linear regression model to the logarithm of the recruitments (the numbers at age 2) estimated in the 2016 stock assessment:

$$\log(R_y) \sim N\left(\mu_y, \sigma^2\right), \tag{2}$$

where R_y ($=N_{2,y}$; $y = 1980, \ldots, 2015$) is the recruitment in year y, $\mu_y = \omega_0 + \omega_1 \times y$ ($\hat{\omega}_0 = 14.2$ and $\hat{\omega}_1 = -0.086$), and σ is the standard deviation of the log-recruitments ($\hat{\sigma} = 0.66$). The recruitment in year y was then generated using

Table 1. Summary of the characteristics of VPA and SAM.

	VPA		
Solving method	**cs-VPA Backwad**	**r-VPA Backwad**	**SAM Forward**
Catch formula	Pope's approximation	Pope's approximation	Baranov's equation
Abundance	Fixed effects	Fixed effects	Random effects
Fishing mortality	Fixed effects	Fixed effects	Random effects
Catch-at-age	Exact	Exact	Including errors
Selectivity	The recent selectivities-at-age are constant	No assumption	Random walk
Stock–recruitment relastionship	No assumption	No assumption	Random walk
Observation model	Ordinary likelihood for abundance indices	Penalized likelihood for abundance indices	Ordinary likelihood for abundance indices and catch-at-age
Process model	Deterministic model	Deterministic model	Stochastic random walk models for abundance and fishing mortality

$$\log\left(\tilde{R}_y\right) \sim N\left(\hat{\mu}_y, \hat{\sigma}^2\right), \tag{3}$$

where $\hat{\mu}_y$ and $\hat{\sigma}^2$ are least-squares estimates.

The sums of fishing mortalities in year y ($\tilde{F}_y$) were assumed to be equal to those estimated in the 2016 stock assessment (Yamashita, 2017). The values for selectivity-at-age a in year y were probabilistically generated by

$$S_{a,y} = \exp\left[-b(a - m_y)^2 + \varepsilon_{a,y}\right], \tag{4}$$

where b is the scale parameter, m_y is the age corresponding to the highest selectivity, and $\varepsilon_{a,y}$ is normally distributed noise from $N(0, \tau^2)$. The selectivity-at-age values were normalized so that $\sum_{a=2}^{10+} \tilde{S}_{a,y} = 1$. The fishing mortality-at-age values were then calculated by $F_{a,y} = \tilde{F}_y \times \tilde{S}_{a,y}$. The number-at-age ($a = 3, \ldots, 10+$) in year y ($y = 1980, \ldots, 2015$) was then estimated by

$$N_{a,y} = N_{a-1,y-1}\exp(-M_{a-1} - F_{a-1,y-1}), \tag{5}$$

where M_a is the age-specific natural mortality coefficient (0.3 for $a = 2$ and 0.25 for $a \geq 3$; Yamashita, 2017). The catch-at-age values were generated by the Baranov catch equation (Quinn and Deriso, 1999). The spawning stock biomass (SSB) was calculated by multiplying the number-at-age by the weight-at-age and the maturity proportion-at-age and summing across ages.

We examined three scenarios for the peak of the selectivity-at-age curve: (i) constant, (ii) age-raising (increase in fishery targeting older fish), and (iii) age-lowering (increase in fishery targeting younger fish) (Figure 1). For the constant scenario, m_y was set to 4 for all years. For the age-raising (resp., age-lowering) scenarios, m_y was set to change from 3 to 5 (resp., from 5 to 3) over the last 10 years. We set b and τ in the selectivity equation [Equation (4)] to 0.1.

Three abundance indices—one SSB and two recruitments (R)—for tuning the population dynamics models were generated with log-normal errors to emulate the stock assessment of the JSS walleye pollock (Yamashita, 2017). In this study, the proportionality constants (catchabilities for abundance) of all indices were fixed at 1 (this setting does not introduce any loss of generality). The standard deviations for the logarithms of the two recruitments were set to 0.4 and that for the logarithm of SSB was set to $0.4/\sqrt{10}$ because the ratios of precision between these logarithms were assumed to be 1:1:10 in the stock assessment (Yamashita, 2017). To examine the impact of full age-specific information being available, we examined the case with age-specific abundance indices having a standard deviation of the logarithms of indices equal to 0.2 (8 indices, from ages 2 to 9). The abundance indices were available for all years (1980–2015) in all cases of the simulation.

The age-structured models tested in the simulation were cs-VPA, r-VPA, and SAM. The performance of SAM was found to be sensitive to the assumption made about the standard deviation of log F_y in preliminary test analyses. In addition, it is difficult to make a model with very many parameters converge when only a few abundance indices are available. We therefore assumed different standard deviations for the age classes 2–4, 5–7, and 8–10+ because this setting generally provided small biases for data-rich cases with eight age-specific abundance indices (Supplementary Figure S1). This setting was used for all scenarios when we fitted SAM to the simulated data. The other standard deviations, for number-at-age and catch-at-age, were age-aggregated in SAM to avoid unstable convergence and due to an indication of robustness to these parameters in preliminary analyses. In addition, we did not assume any parametric form of the stock recruitment relationship and instead used a random walk for the recruitment in SAM because the stock recruitment relationship is unclear for JSS walleye pollock (Ichinokawa *et al.*, 2017). When age-specific abundance indices were used for tuning, only r-VPA and SAM were used because cs-VPA is not usually used for data-rich situations (Ichinokawa and Okamura, 2014).

The performance of the age-structured models was evaluated using the relative bias of the population and fisheries parameters [total number (N_y), age-specific fisheries mortalities ($F_{a,y}$), spawning stock biomass (SSB_y), recruitment (R_y), and recruitment per spawning stock biomass (RPS_y)]. For total number, spawning stock biomass, and recruitment, we calculated $(N_Y^{\mathrm{Est}} - N_Y^{\mathrm{True}})/N_Y^{\mathrm{True}}$, $(SSB_Y^{\mathrm{Est}} - SSB_Y^{\mathrm{True}})/SSB_Y^{\mathrm{True}}$, and $(R_Y^{\mathrm{Est}} - R_Y^{\mathrm{True}})/R_Y^{\mathrm{True}}$, respectively, where Y is the terminal year (2015),

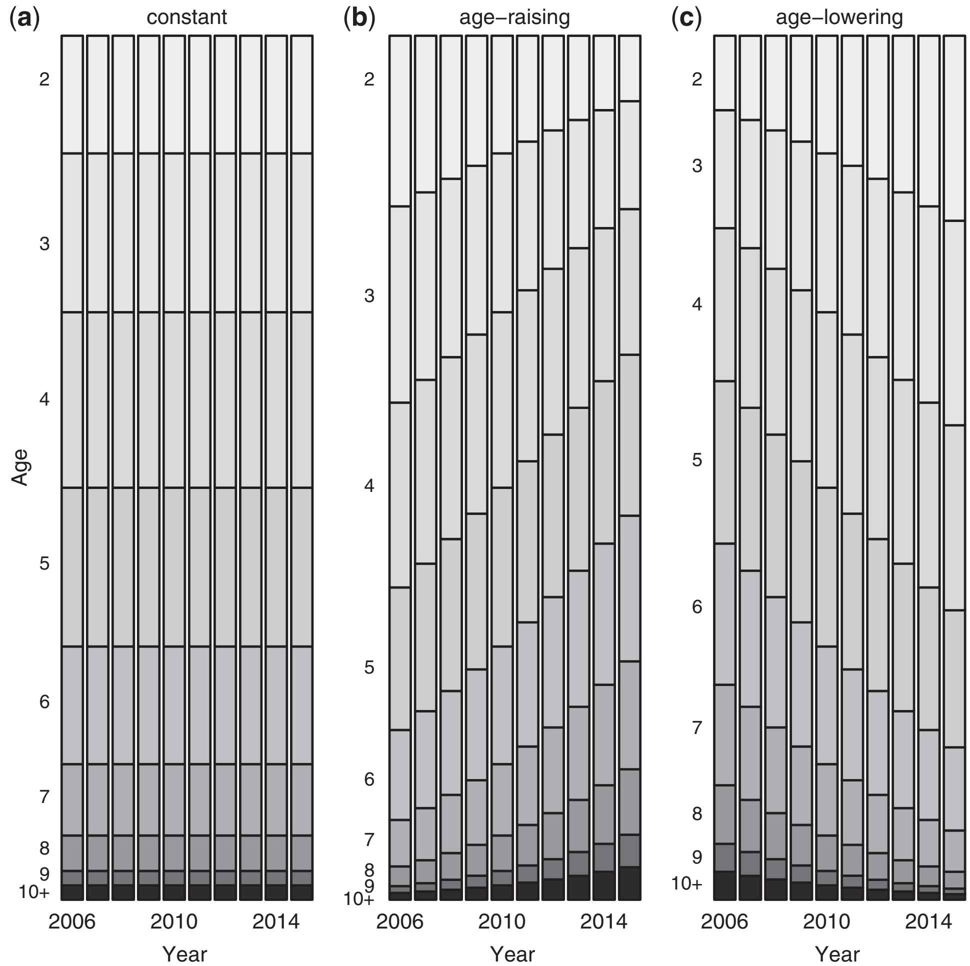

Figure 1. The selectivity-at-age pattern generated in the simulation: (a) constant, (b) age-raising, and (c) age-lowering. These plots have no random error, though we added a random error, $\varepsilon_{a,y} \sim N(0, 0.1^2)$, in the simulation.

and the superscripts "Est" and "True" denote the estimate from each age-structured model and the true value in the simulation. We calculated $\frac{1}{8}\sum_{a=2}^{9}(F_{a,Y}^{\mathrm{Est}} - F_{a,Y}^{\mathrm{True}})/F_{a,Y}^{\mathrm{True}}$ for age-specific fishing mortalities and $\frac{1}{5}\sum_{y=Y-4}^{Y}(RPS_y^{\mathrm{Est}} - RPS_y^{\mathrm{True}})/RPS_y^{\mathrm{True}}$ for recruitment per spawning stock biomass.

The parameters of the stock assessment models were estimated using a maximum likelihood approach. All simulation runs were conducted using the statistical software R (R Core Team, 2017) and the template model builder (TMB; Kristensen *et al.*, 2016).

In additional test scenarios, to avoid situations that are overly advantageous to VPA, we included the following three factors: (i) catch-at-age error, (ii) smoothed fishing mortalities, and (iii) a flat-topped selectivity curve instead of the above dome-shaped selectivity curve. The catch-at-age with errors was generated by extracting a number randomly from a multinomial distribution with a sample size of 100 and the proportion of the catch-at-age composition without error and enlarging it to the original sample size (Hashimoto *et al.*, 2018). Zero catches cause a problem in VPA (see Supplementary Material A). Thus, to avoid zero catch, we generated the catch-at-age including errors by combining 10% of the original catch-at-age and 90% of the catch-at-age newly generated from the multinomial distribution. Among the additional factors, the first factor (catch-at-age error) was the most influential while results were insensitive to the latter two. We will therefore discuss the catch-at-age error in detail and only briefly mention the other two factors in the "Results" section.

The simulation was repeated 200 times for each selectivity scenario combined with a sensitivity test scenario and a different number of abundance indices. The total number of scenarios was therefore 48: 3 (constant, age-raising, and age-lowering selectivity) × 2 (presence or absence of catch-at-age errors) × 2 (presence or absence of smoothing for fishing mortality) × 2 (dome-shaped or flat-topped selectivity pattern) × 2 [3 (SSB_y and $2R_y$s) or 8 ($N_{a,y}$ ($a = 2, \ldots, 9$)) abundance indices].

An outline of the generation of simulated data is given in Supplementary Material C.

Results

The simulation without catch-at-age error showed that using cs-VPA led to biased estimates of abundance-related parameters (N_y, SSB_y, R_y, and RPS_y) for the scenarios with temporally changing selectivity and provided nearly unbiased estimates for the scenario with constant selectivity (Figure 2). The r-VPA provided nearly unbiased results for abundance-related parameters when the annual selectivity-at-age varied or did not vary systematically

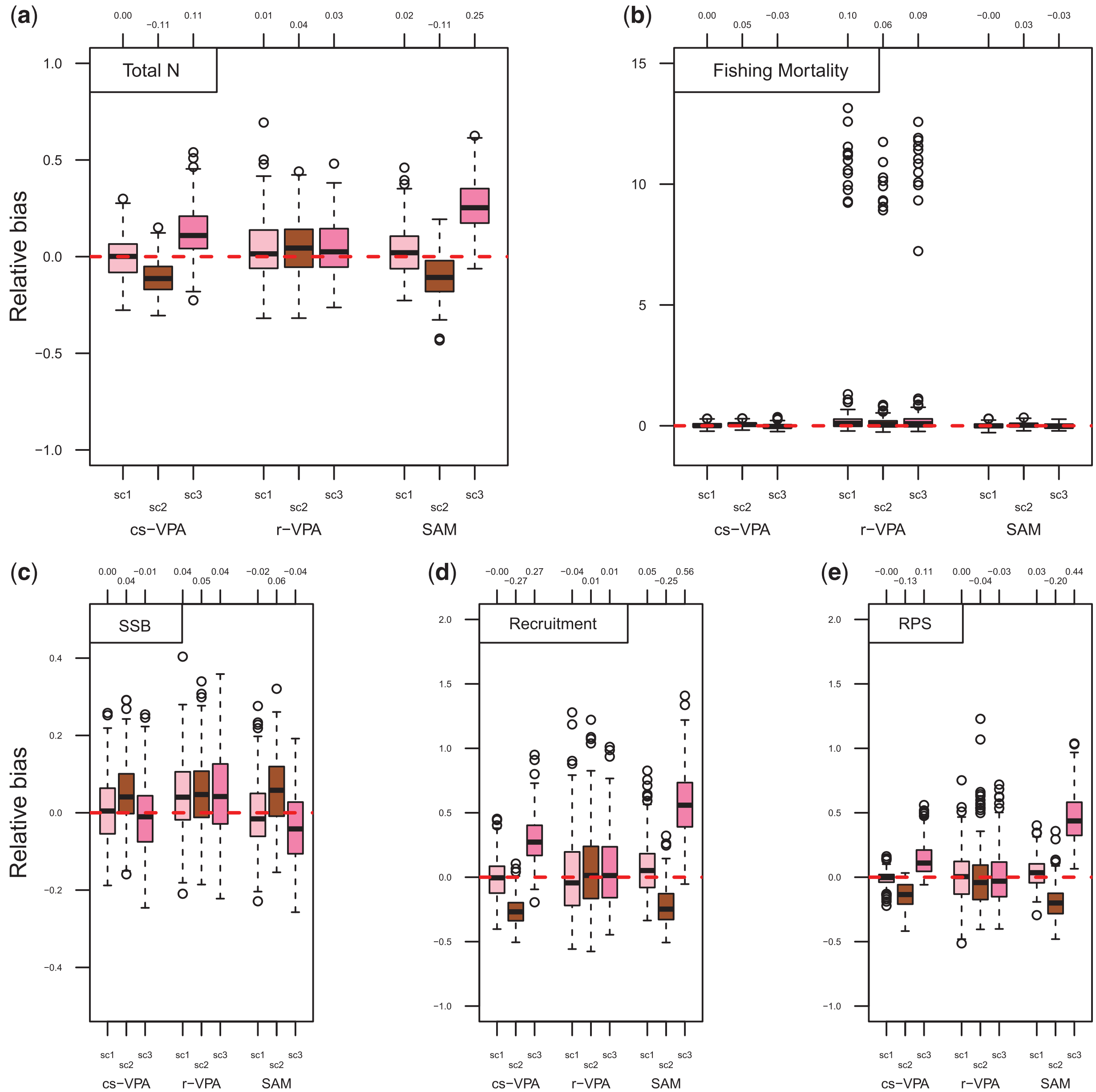

Figure 2. Relative biases of population and fisheries parameters when there is no catch-at-age error and few abundance indices are available: (a) total number, (b) fishing mortality, (c) spawning stock biomass, (d) recruitment, and (e) recruitment per spawning stock biomass. The labels sc1, sc2, and sc3 denote the scenarios for selectivity patterns: constant, age-raising, and age-lowering, respectively. The numbers on the plots indicate median values.

(Figure 2a, c, and d), although fisheries mortalities were overestimated for all scenarios (Figure 2b). Using SAM showed trends similar to the results from cs-VPA, although the results of SAM were slightly better than those of cs-VPA in terms of bias. In particular, when the selectivity-at-age had an age-lowering pattern (Figure 1c), the bias of cs-VPA and SAM for the latest recruitment was large (Figure 2d).

When errors were incorporated into catch-at-age data, we evaluated their magnitude using the median value of the square root of squared log deviances:

$$\text{median}\left[\sqrt{\left[\log\left(C_{a,y}^{\text{with error}}\right) - \log\left(C_{a,y}^{\text{without error}}\right)\right]^2}\right], \qquad (6)$$

where $C_{a,y}^{\text{with error}}$ and $C_{a,y}^{\text{without error}}$ are the components of catch-at-age matrix at age a in year y generated with and without errors, respectively, in the simulation. The values in Equation (6) approximately correspond to the coefficients of variation (CV) in catch-at-age because $\text{var}(\log(x)) \approx [\text{CV}(x)]^2$. The median values

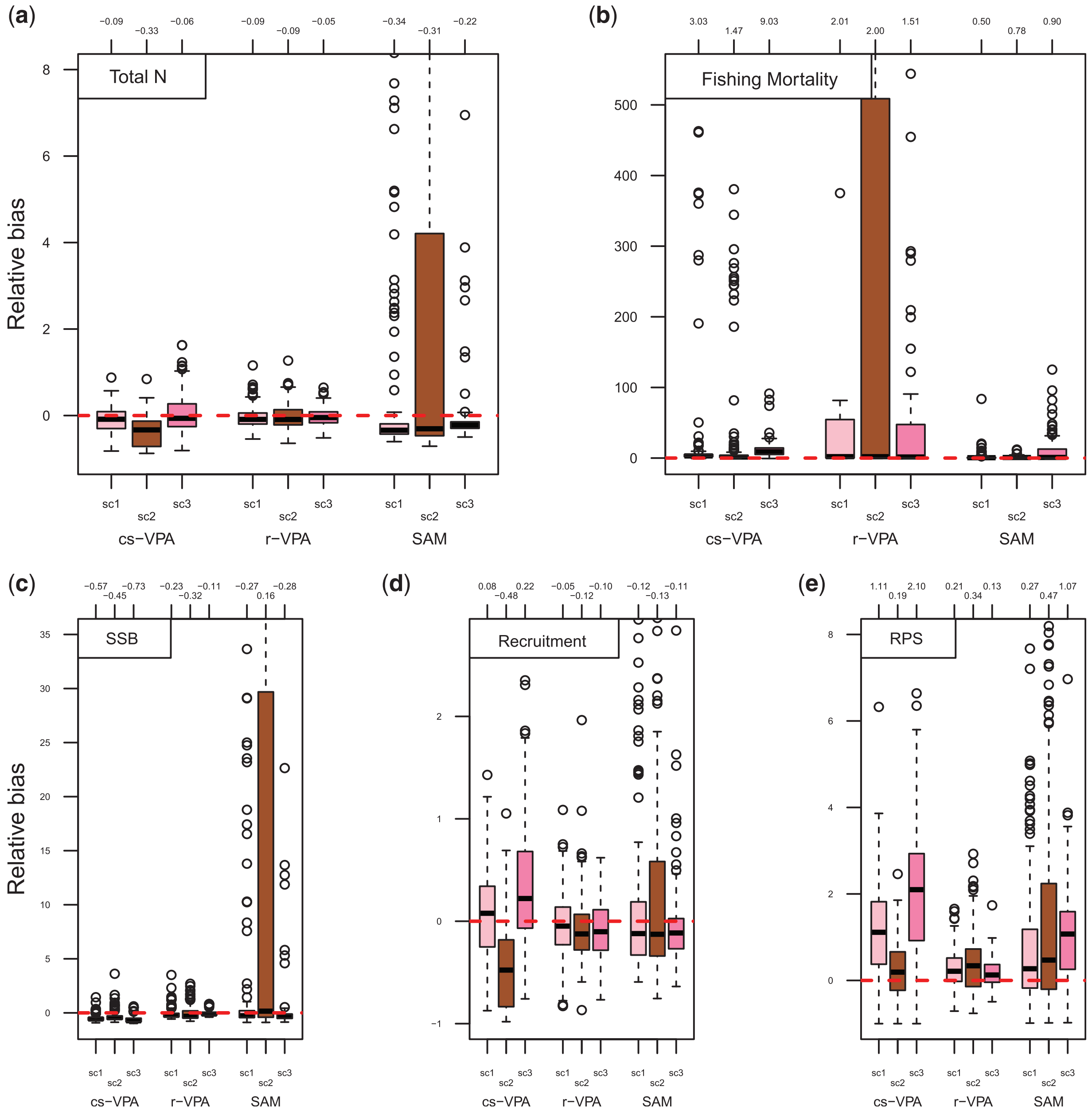

Figure 3. Relative biases of population and fisheries parameters when there is catch-at-age error and few abundance indices are available: (a) total number, (b) fishing mortality, (c) spawning stock biomass, (d) recruitment, and (e) recruitment per spawning stock biomass. The labels sc1, sc2, and sc3 denote the scenarios for selectivity patterns: constant, age-raising, and age-lowering, respectively. The numbers on the plots indicate median values.

of the square root of squared log deviances for all years and ages were 0.525 for Scenario 1 (constant selectivity-at-age), 0.748 for Scenario 2 (age-raising selectivity), and 0.380 for Scenario 3 (age-lowering selectivity). These values are comparable to the coefficients of variation for the JSS walleye pollock catch-at-age [at most 33% (year 2001) in Yamaguchi and Matsuishi (2007)] and those of other species [the ranges of CV of catch-at-age for rainbow trout in the Kenai River and Hauraki Gulf snapper are given as 7–40% and 6–100%, respectively, in Tables 8.2 and 8.4 in Quinn and Deriso (1999)]. For abundance-related parameters, all models provided biased estimates (Figure 3). The use of cs-VPA provided consistently large biases in all scenarios, underestimating fishing mortalities and overestimating abundance-related parameters. The use of r-VPA provided moderately large biases for abundance-related parameters. The use of SAM provided smaller biases than the use of cs-VPA and biases of a magnitude similar to those from r-VPA. For the age-raising scenario, although SAM showed relatively little bias for abundance-related parameters, the variations were large, indicating instability of estimation (Figure 3c and d).

When there was no catch-at-age error, the proportionality constants for SSB and the two Rs were estimated almost without

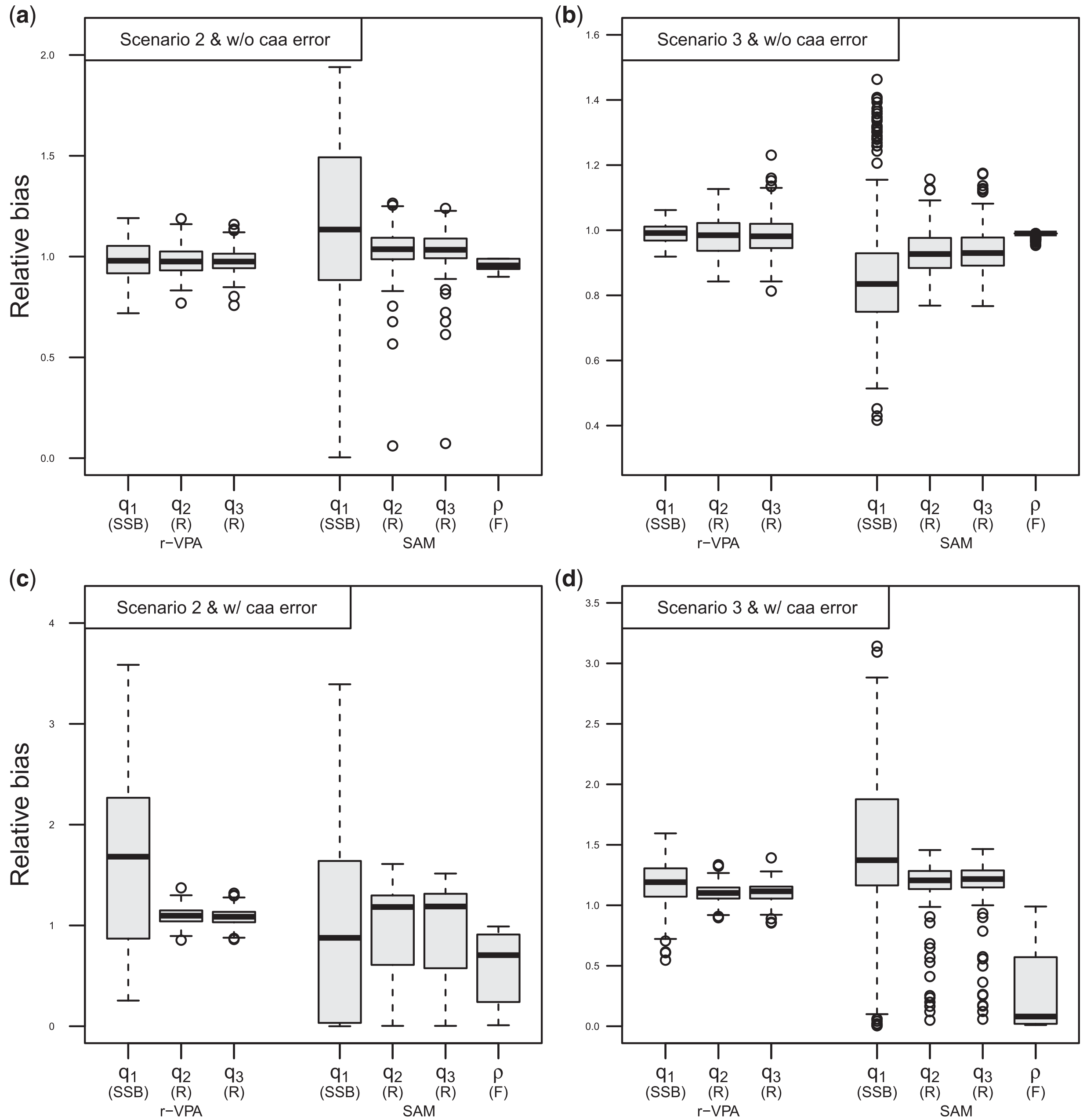

Figure 4. Estimates of proportionality constants and the correlation of fishing mortalities in Scenario 2 (selectivity with an age-raising pattern) and Scenario 3 (selectivity with an age-lowering pattern). Panels (a) and (b) correspond to the simulation without catch-at-age error and panels (c) and (d) correspond to the simulation with catch-at-age error. The proportionality constant q_1 is related to spawning stock biomass while q_2 and q_3 are related to recruitment.

bias by using r-VPA. In contrast, when using SAM, estimates were substantially biased, especially for SSB (Figure 4a and b). When there was catch-at-age error, r-VPA showed some small biases while SAM showed slightly larger biases in general with substantially larger variations in estimates (Figure 4c and d). The correlation coefficients between fishing mortalities-at-age in SAM were close to 1 when there was no catch-at-age error (Figure 4a and b), whereas they were smaller and had much larger variations when there was catch-at-age error (Figure 4c and d).

When we used age-specific abundance indices and there was no catch-at-age error, the biases for r-VPA and SAM were greatly reduced, indicating that SAM could better estimate the abundance-related parameters (Supplementary Material D: Figure S1). When there was catch-at-age error, some biases were still detected, although they were considerably smaller than in cases with fewer abundance indices (Supplementary Material D: Figure S2). The use of SAM generally resulted in smaller biases than using r-VPA, especially when there was catch-at-age error. However, the variations in the relative biases of SAM were much greater than those of r-VPA, particularly for the abundance-related parameters (Supplementary Material D: Figure S2).

Other sensitivity tests for the smoothness of fishing mortality and the form of selectivity curve produced results that were not conspicuously different from the simulation results described above.

Discussion

Our simulation study demonstrated that incorrect assumptions about recent selectivity-at-age in VPA, even weak ones, could cause serious biases in the abundance-related parameters for situations lacking sufficient abundance indices. However, the newly developed r-VPA (Okamura *et al.*, 2017) was insensitive to systematically variable selectivity-at-age and provided almost unbiased estimates for abundance-related parameters when there was no catch-at-age error. When including substantially serious catch-at-age errors into the simulated data, r-VPA provided somewhat biased results, although such biases were not serious. However, r-VPA was not able to accurately reproduce selectivity-at-age patterns. There are two explanations for this. To begin with, shrinkage methods improve the model's predictability by reducing variance at the cost of increasing bias. We therefore should not expect an unbiased estimate from r-VPA. Second, as shown for 14 different resources in Butterworth *et al.* (2014), accurately estimating selectivities will typically be less important than accurately estimating other biological parameters when calculating catch limits, especially for medium- to long-lived species. In fact, researchers often take the average of recent selectivities when projecting a population model into the future, taking the large uncertainty of the latest selectivity estimates into account (Yamashita, 2017; Yamashita *et al.*, 2018), and such a procedure may substantially mitigate the impact of the uncertainty of selectivity estimates. Although the influence of bias on selectivities needs to be examined further, as a whole, r-VPA, which can robustly estimate abundance-related quantities even when there are few available abundance indices, is a promising stock assessment method for data-poor situations.

However, there may be cases in which estimating the latest selectivities accurately is more important, such as for a short-lived species (Butterworth *et al.*, 2014). For such cases, we should expect that an SCAA model that allows time-varying selectivity will work better than VPA (Nielsen and Berg, 2014; Punt *et al.*, 2014). However, our simulation demonstrated that SAM did not always work well when there were few available abundance indices. This may be because the abundance index of SSB did not provide information on the age composition of older fish and consequently the estimation model could not differentiate between uncertainties of abundance and age composition. The fact that the estimates of the proportionality constant for SSB from SAM were worse, whether overestimated or underestimated (Figure 4), corroborates this. Although the correlation coefficients (ρ) of age-specific fishing mortalities were close to 1 in our simulation when there were few abundance indices and no catch-at-age error (Figure 4), they were not close to 1 when there were many (age-specific) abundance indices (Supplementary Material D: Figure S3). A correlation coefficient close to 1 means that the selectivity patterns are similar among ages and the model could not estimate time-varying selectivities. This indicates that there is a parameter estimation problem in SAM when there are few abundance indices. Although SAM is a promising stock assessment model (Berg and Nielsen, 2016), its estimates did not always turn out to be accurate and precise in our simulation. One cause may be due to differences in the specifications of the simulation model and SAM. However, Auger-Méthé *et al.* (2016) showed that even simple linear state-space models could lead to biased results in parameter and state estimation even when there is no misspecification in the state-space model. We should therefore use state-space models such as SAM with caution, in particular when the number of available abundance indices is small. Using r-VPA and SAM simultaneously and then comparing their results will be important, especially when there are few abundance indices. In addition, using age-specific abundance indices that are as accurate as possible will be effective for SAM (Figures 2 and 3 and Supplementary Material D: Figures S1 and S2). A refined and advanced standardization for fisheries-dependent CPUE data will play a central role in obtaining accurate age-specific abundance indices (Okamura *et al.*, 2018).

When the catch-at-age included substantial errors, both VPA and SAM underestimated total numbers and SSB (Figure 3). Because the recent proportion of older fish in the simulated data was low, the impact of catch-at-age errors on older fish would tend to be larger than that on younger fish. This may lead to underestimation of older fish abundances in VPA when there are catch-at-age errors because misallocation of older fish to younger age groups can occur more easily. Similarly, the underestimation of abundance in SAM was caused by the underestimation of abundances of older fish. This may be also because of the larger impact of catch-at-age errors on older fish as a result of the difference of the multinomial distribution used in the simulation and the age-blocked log-normal distribution used in SAM (the multinomial distribution assumes different variance–covariances for different ages).

We did not examine the impact of systematically changing unknown biological parameters, such as natural mortality, in this study. Such changes will cause serious bias in both VPA and SAM estimates (Hurtado-Ferro *et al.*, 2015; Okamura *et al.*, 2017). Assessing the effects of changing these parameters is a necessary investigation for the future. In addition, we did not examine the estimation performance of biological reference points, such as maximum sustainable yield (MSY), because the stock recruitment relationship of JSS walleye pollock was unclear and estimating it was difficult. VPA estimates MSY externally after estimating SSB and recruitment (Ichinokawa *et al.*, 2017), whereas SAM estimates MSY internally at the same time as estimating abundance (Nielsen and Berg, 2014). Although internal and simultaneous estimation would be desirable in a statistical sense, which of VPA and SAM is better in terms of bias and variance in MSY estimation remains to be examined when there are few available abundance indices, given that our simulation showed that r-VPA generally outperformed SAM in terms of abundance estimation.

Finally, r-VPA incorporates the concept of ridge and lasso regression (James *et al.*, 2013) into VPA through the predictive ability quantified by retrospective bias (Okamura *et al.*, 2017). This idea may also be effective for state-space modelling. Because TMB greatly reduces the computational burden even for complex hierarchical models, retrospective analysis for state-space age-structured models would not take an unreasonably long time. Developing a hybrid approach between r-VPA and SAM may be an attractive and interesting topic for future research.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

The authors thank Yuki Kanamori for her valuable comments.

Funding

This research was funded by JSPS KAKENHI grants (Numbers 17H07413 and 15K18737).

References

Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A. E., Lewis, M. A., Jonsen, I. D., and Flemming, J. M. 2016. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Scientific Reports, 6: 26677.

Berg, C. W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES Journal of Marine Science, 73: 1788–1797.

Butterworth, D. S., and Rademeyer, R. A. 2008. Statistical catch-at-age analysis vs. ADAPT-VPA: the case of Gulf of Maine cod. ICES Journal of Marine Science, 65: 1717–1732.

Butterworth, D. S., Rademeyer, R. A., Brandão, A., Geromont, H. F., and Johnston, S. J. 2014. Does selectivity matter? A fisheries management perspective. Fisheries Research, 158: 194–204.

Hashimoto, M., Okamura, H., Ichinokawa, M., Hiramatsu, K., and Yamakawa, T. 2018. Impacts of the nonlinear relationship between abundance and its index in a tuned virtual population analysis. Fisheries Science, 84: 335–347.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R. *et al.* 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72: 99–110.

Ichinokawa, M., and Okamura, H. 2014. Review of stock evaluation methods using VPA for fishery stocks in Japan: implementation with R. Bulletin of Japanese Society of Fisheries Oceanography, 78: 104–113 (in Japanese with English abstract).

Ichinokawa, M., Okamura, H., and 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. Fisheries Research, 158: 147–157.

Ichinokawa, M., Okamura, H., and Kurota, H. 2017. The status of Japanese fisheries relative to fisheries around the world. ICES Journal of Marine Science, 74: 1277–1287.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. An Introduction to Statistical Learning: with Applications in R. Springer, New York. 426 pp.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: automatic differentiation and Laplace approximation. Journal of Statistical Software, 70: 1–26.

Kurota, H., Yoda, M., Hayashi, A., Takegaki, S., Sasaki, Y., and Takahashi, M. 2018. Stock assessment and evaluation for the East China Sea stock of spotted mackerel (fiscal year 2017). *In* Marine Fisheries Stock Assessment and Evaluation for Japanese Waters (Fiscal Year 2017/2018), pp. 271–302. Fisheries Agency and Fisheries Research Agency of Japan. http://abchan.fra.go.jp/digests27/details/2710.pdf (last accessed 31 March 2018) (in Japanese).

Lassen, H., and Medley, P. 2001. Virtual Population Analysis—A Practical Manual for Stock Assessment. FAO Fisheries Technical Paper 400. Food and Agriculture Organization, Rome. 129 pp.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488.

Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96–101.

Okamura, H., Yamashita, Y., and Ichinokawa, M. 2017. Ridge virtual population analysis to reduce the instability of fishing mortality in the terminal year. ICES Journal of Marine Science, 74: 2427–2436.

Okamura, H., Morita, S. H., Funamoto, T., Ichinokawa, M., and Eguchi, S. 2018. Target-based catch-per-unit-effort standardization in multispecies fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 75: 452–463.

Punt, A. E., Hurtado-Ferro, F., and Whitten, A. R. 2014. Model selection for selectivity in fisheries stock assessments. Fisheries Research, 158: 124–134.

Quinn, T. J., and Deriso, R. B. 1999. Quantitative Fish Dynamics. Oxford University Press, Oxford. 560 pp.

Radomski, P., Bence, J. R., and Quinn, T. J. 2005. Comparison of virtual population analysis and statistical kill-at-age analysis for a recreational, kill-dominated fishery. Canadian Journal of Fisheries and Aquatic Sciences, 62: 436–452.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Yamaguchi, H., and Matsuishi, T. 2007. Effects of sampling errors on abundance estimates from virtual population analysis for walleye pollock in northern waters of Sea of Japan. Fisheries Science, 73: 1061–1069.

Yamashita, Y. 2017. Stock assessment and evaluation for the northern Sea of Japan stock of walleye pollock (fiscal year 2016). *In* Marine Fisheries Stock Assessment and Evaluation for Japanese Waters (Fiscal Year 2016/2017), pp. 337–393. Fisheries Agency and Fisheries Research Agency of Japan. http://abchan.fra.go.jp/digests27/details/2710.pdf (last accessed 31 March 2018) (in Japanese).

Yamashita, Y., Chimura, M., Ishino, M., Tanaka, H., Yamashita, N., and Hamatsu, Y. 2018. Stock assessment and evaluation for the northern Sea of Japan stock of walleye pollock (fiscal year 2017). *In* Marine Fisheries Stock Assessment and Evaluation for Japanese Waters (Fiscal Year 2017/2018), pp. 303–363. Fisheries Agency and Fisheries Research Agency of Japan. http://abchan.fra.go.jp/digests27/details/2710.pdf (last accessed 31 March 2018) (in Japanese).

Yukami, R., Nishijima, S., Isu, S., Watanabe, C., Kamimura, Y., and Hashimoto, M. 2018. Stock assessment and evaluation for the Pacific stock of chub mackerel (fiscal year 2017). *In* Marine Fisheries Stock Assessment and Evaluation for Japanese Waters (Fiscal Year 2017/2018), pp. 157–200. Fisheries Agency and Fisheries Research Agency of Japan. http://abchan.fra.go.jp/digests27/details/2710.pdf (last accessed 31 March 2018) (in Japanese).

Handling editor: Ernesto Jardim