NRC Research Press

ARTICLE

Improved estimation of age composition by accounting for spatiotemporal variability in somatic growth

Giancarlo M. Correa, Lorenzo Ciannelli, Lewis A.K. Barnett, Stan Kotwicki, and Claudio Fuentes

Abstract: Age composition is defined as the proportion of a fish population belonging to each age class and is an informative input to stock assessment models. Variations in somatic growth rates may lead to larger errors in age composition estimates. To reduce this source of error, we compared the performance of four methods for estimating age compositions of a simulated fish population: two methods based on age–length keys (ALK, pooled and annual) and two model-based approaches (generalized additive models (GAMs) and continuation ratio logits (CRLs)). CRL was the most robust and precise method, followed by annual ALKs, particularly when significant growth variability was present. We applied these methods to survey age subsample data for Pacific cod (*Gadus macrocephalus*) in the eastern Bering Sea, estimating age compositions that were then incorporated in its stock assessment model. The model that included age compositions estimated by CRL displayed the highest consistency with other data in the model. CRL approach has utility for estimating age compositions employed in stock assessment models, especially when substantial variation in somatic growth is present.

Résumé : La composition selon l'âge désigne la proportion d'une population de poissons appartenant à chaque classe d'âge et constitue un intrant informatif dans les modèles d'évaluation de stocks. Des variations des taux de croissance somatique peuvent accroître l'erreur associée aux estimations de la composition selon l'âge. Pour réduire cette source d'erreur, nous avons comparé la performance de quatre méthodes d'estimation de la composition selon l'âge d'une population de poissons simulée, dont deux méthodes basées sur des clés âge–longueur (CAL, regroupées et annuelles) et deux approches basées sur des modèles (modèles additifs généralisés, ou MAGs, et modèles logistiques des cotes de continuité, ou MLCCs). La méthode MLCC s'avère la plus robuste et précise, suivie des CAL annuelles, particulièrement quand une variabilité significative de la croissance est présente. Nous avons appliqué ces méthodes à des données d'évaluation sous-échantillonnées pour l'âge pour la morue du Pacifique (*Gadus macrocephalus*) dans la mer de Behring orientale, pour estimer des compositions selon l'âge qui ont ensuite été incorporées dans le modèle d'évaluation du stock de l'espèce. Le modèle qui comprend les compositions selon l'âge estimées par MLCC présente la plus grande cohérence avec d'autres données dans le modèle. L'approche MLCC est utile pour estimer les compositions selon l'âge employées dans les modèles d'évaluation de stocks, particulièrement en présence de variations considérables de la croissance somatique. [Traduit par la Rédaction]

Introduction

Marine fish populations inhabit a changing environment, where oceanographic, ecological, and fishery drivers affect vital rates, including somatic growth, sexual maturation, natural mortality, and spatial distribution (Gertseva et al. 2017; Thorson et al. 2015). Numerous studies have documented substantial variability in marine fish somatic growth rates related to environmental, community, and density-dependent factors (Baudron et al. 2014; Puerta et al. 2019*b*; Stawitz et al. 2015; Thorson and Minte-Vera 2016). For instance, fish growth rates are positively correlated with temperature (Gislason et al. 2010) and prey availability (Gale et al. 2013), but there are also top-down regulators such as size-selective predation mortality (Aikio et al. 2013; Sinclair et al. 2002). Variation in growth is an important driver of population fluctuations (Stawitz and Essington 2019) and has implications for the assessment and ecological studies of fish stocks (Helser and Brodziak 1998; Lee et al. 2018; Punt et al. 2015; Thorson et al. 2017, 2015). Growth variability can also affect multiple data types used in stock assessment models, such as age compositions, an informative input for recruitment, natural mortality, and population structure (Chen et al. 2003; Magnusson and Hilborn 2007; Ono et al. 2015).

Age composition is defined as the proportions of a population belonging to each age class, and it can be estimated from fishery-independent and -dependent sources, typically by a two-stage sampling. The first stage (length subsampling) collects length information on a random sample obtained from the total catch in a haul. In the second stage (age subsampling), individuals are subsampled from the length subsample for aging (Quinn and Deriso 1999). From this age subsample, ages are estimated through laboratory analysis, by counting growth increments deposited with regular periodicity in the otoliths or other hard parts. Using this information, fisheries scientists may construct a classic age–length key (ALK; Fridriksson 1934) matrix describing the probability of being a specific age at a given length and use this to assign age to individual fish in the length subsample (age assignment). Finally, using information in haul catches, length subsample, and an ALK, one may estimate abundance-at-age per haul and then the age composition of the population using a design-based procedure (Stewart and Hamel 2014; Thorson and Haltuch 2019).

Received 4 May 2020. Accepted 3 August 2020.

G.M. Correa and L. Ciannelli. College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97330, USA.
L.A.K. Barnett and S. Kotwicki. National Marine Fisheries Service, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, WA 98195, USA.
C. Fuentes. Department of Statistics, Oregon State University, Corvallis, OR 97331, USA.
Corresponding author: Giancarlo M. Correa (email: moroncog@oregonstate.edu).

Methods used to estimate abundances-at-age can have effects on estimates of age composition. In recent years, some researchers have opted to use alternative approaches such as state-of-art statistical models to standardize the abundance estimates of each age class. For example, Thorson and Haltuch (2019) developed a spatiotemporal statistical model to estimate abundance-at-age, resulting in improved compositional data that led to better-fitted stock assessment models. Likewise, Berg et al. (2014) applied a generalized additive model (GAM) to estimate an abundance index for each age for three species in the North Sea. These standardization techniques still require an age assignment method for individuals in the length subsample, and ALKs are normally preferred (Gulland and Rosenberg 1992). However, scientists should be cautious when constructing an ALK, since it may be influenced by spatial and temporal variation in vital rates (Ailloud and Hoenig 2019), including somatic growth. These effects, if ignored, can then lead to suboptimal age composition estimates (Aanes and Vølstad 2015; Gerritsen et al. 2006).

Recent studies applied model-based approaches for age assignment as an alternative to ALKs. Kvist et al. (2000) used continuation ratio logits (CRLs), a type of logistic model for an ordinal response (e.g., age; Agresti 2010), with generalized linear models (GLMs) to analyze variations of the age proportions of lesser sandeel (*Ammodytes marinus*) across space in the North Sea. Subsequent studies focused on estimating age compositions, smooth length distributions, and spatial differences in ALKs using CRL and GLM for estimation, incorporating spatial differences through a stratification procedure (Gerritsen et al. 2006; Rindorf and Lewy 2001; Stari et al. 2010). Berg and Kristensen (2012) tried to overcome the need for a spatial stratification using GAM for fitting CRL to model age proportions as a smooth function of length and geographical position. Finally, other approaches have modeled age as the response variable directly using GLM or GAM (Ochwada et al. 2008; Puerta et al. 2019*b*, 2019*a*). However, the performance of these various procedures for age assignment has rarely been explored.

The primary goal of this study is to evaluate the performance of classic ALKs and two model-based methods to estimate age compositions, focusing on the age assignment step of this procedure. We are especially interested in their performance when used for a fish population with substantial spatiotemporal variability in somatic growth. The two model-based methods tested are GAM (assuming age as the response variable) and CRL (using GAM for estimation). We implement a simulation experiment that assumes known (true) age composition and can, therefore, quantify the performance of each method with respect to the "true" values. Moreover, we assess their robustness in two population scenarios: one with and one without spatiotemporal variability in somatic growth. Our secondary goal is to perform a system-based test to assess the consistency of age compositions estimated by the evaluated methods with other data used in a stock assessment model. For this, we use survey data for Pacific cod (*Gadus macrocephalus*) in the eastern Bering Sea (EBS) as a case study because there is evidence that this stock has experienced spatiotemporal variation in somatic growth rates in recent years (Ciannelli et al. 2020; Puerta et al. 2019*a*, 2019*b*). We hypothesize that more precise age compositions may be more consistent with the model structure and therefore lead to better fits by the model.

This research contributes to the improvement of age assignment and age composition estimation of fish populations that can be included in stock assessment models or applied to other ecological studies in which age structure of a population is required, especially for cases with significant spatiotemporal variability in somatic growth.

Methods

Simulation experiment

We create an operating model to simulate the dynamics of an age-structured fish population with spatiotemporal variability in somatic growth. Although we base our simulation on the growth variability and dynamics of Pacific cod in the EBS, our operating model can be used to test the dynamics of other species in the future. A brief description of the EBS and the Pacific cod stock can be found in Appendix A.

Somatic growth variability

Somatic growth variability is simulated by incorporating two main sources of variation:

1. *Variance of size-at-age in a single year*. We assume two processes that impact this variance. First, there is a contribution of processes that generate different sizes for a given age (e.g., differences in date of birth, size-selective predation) with similar effects across space and that are present in all fish populations. Second, some fish stocks experience spatial differences in growth triggered by variable conditions (e.g., environment or prey composition) across space (Adams et al. 2018; Ciannelli et al. 2020; Gertseva et al. 2017). The latter source creates systematic spatial trends in growth rates and, consequently, increases the variance of size-at-age at the population scale (Fig. 1).
2. *Growth variability across multiple years*. We assume that growth rate (k, the growth coefficient of the von Bertalanffy growth function; see below) varies over time, resulting in temporal changes in the mean size-at-age. This has been observed for many stocks, where changing temperature or ecological regimes produced changes in growth rates (Ciannelli et al. 2020; Hernandez-Miranda and Ojeda 2006; Hunter et al. 2019).

We consider two spatiotemporal scenarios for somatic growth: (*i*) no spatial – no temporal (No S – No T) scenario, or control scenario, where only the first source of variance of size-at-age in a single year (i.e., processes with similar effects across space) is considered and there is no temporal variability; and (*ii*) spatial–temporal (S–T) scenario, where we also include systematic spatial trends and temporal growth variability.

Spatially explicit age-structured population model

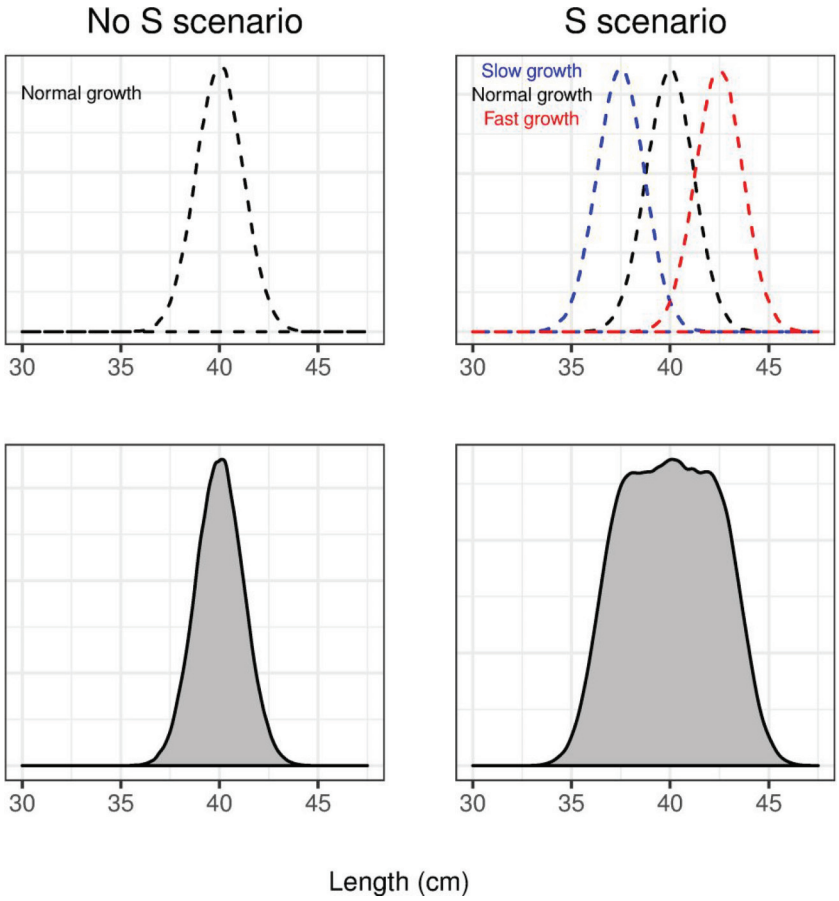
We simulate abundance at age a, location s, and year y ($N_{a,s,y}$) throughout an area composed of 9376 cells, each 11 km by 11 km:

$$N_{a,s,y} = \begin{cases} R_{s,y} & \text{if } a = 0 \\ N_{a-1,s,y-1}\exp(-S1_a F - M) & \text{if } a > 1 \end{cases}$$

where $a \in \{0, 1, \ldots, A\}$, M is the natural mortality rate, F is the fishing mortality rate, $S1_a$ is the fishery selectivity-at-age. Recruitment varies in space and time: $R_{s,y} = \exp(\beta_{N_y} + \omega_{N_s} + \varepsilon_{N_{s,y}})$, where ω_{N_s} and $\varepsilon_{N_{s,y}}$ are the spatial and spatiotemporal variation in log-density, respectively. Then, recruit density by year is $\beta_{N_y} = \log(R_y/\text{Area})$, where Area is the study area (km²) and annual recruitment is $R_y = R_0 \times \exp[\text{Normal}(-\sigma_R^2/2, \sigma_R^2)]$, where R_0 is the average recruitment for all simulated years, σ_R^2 the variance of recruitment, and we use conventional bias correction for the lognormal distribution.

The spatial and spatiotemporal terms, ω_{N_s} and $\varepsilon_{N_{s,y}}$, are simulated from a Gaussian random field using the R package "RandomFields" (Schlather et al. 2015). We use an isotropic Matern covariance function and specify a spatial scale of 85 km for all simulated spatial fields, where the spatial and spatiotemporal terms have a standard deviation of 0.5 and 0.2, respectively. We vary these parameter values to evaluate their effects on results.

Fig. 1. Simulated spatial variability scenarios exemplified by frequency distributions of lengths. (1) No S: for a given age, the variability of length-at-age (left panel) is composed of processes with similar effects across space. (2) S: for a given age, the variability of length-at-age (right panel) is, in addition to (1), composed by large-scale processes that produce areas that favor slow or fast growth rates. Lower panels (grey curves) show the length-at-age variability at the population scale. If σ_a increases, the amplitude of the curves also increases in the four panels. [Colour online.]



The mean length-at-age ($L_{a,s,y}$) for the initial population is calculated as

$$L_{a,s,y} = L_\infty\{1 - \exp[-k_{s,y}(a - t_0)]\}$$

where L_∞ is the asymptotic length, t_0 is a constant representing an intercept that adjusts the model along a time axis, and $k_{s,y}$ is the growth coefficient simulated as

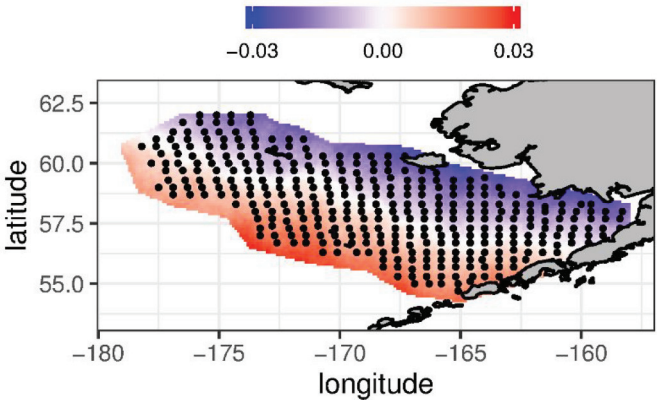
$$k_{s,y} = k + \omega_{k_s} + \vartheta_{k_y}$$

where k is the initial growth parameter, ω_{k_s} is related to the systematic spatial component of variance in size-at-age in a given year, and ϑ_{k_y} is related to the temporal variability. These two terms are equal to zero for the No S – No T scenario. For the S–T scenario, ω_{k_s} is a simulated spatial field using the R package "gstat" (Pebesma 2004), considering a Matern model and a linear increasing trend from the inner to the outer region (Fig. 2). ϑ_{k_y} follows a linear trend from the first to the last simulated year. Finally, ω_{k_s} and ϑ_{k_y} values are normalized between −0.03 and 0.03 to create the desired variability in growth.

Then, somatic growth in time is modeled as follows (Methot and Wetzel 2013):

$$\tilde{L}_{a,s,y} = L_{a,s,y} + (L_{a,s,y} - L_\infty)[\exp(-k_{s,y}\Delta y) - 1]$$

Fig. 2. Simulated survey in the eastern Bering Sea: 332 sampling stations (black dots) and simulated spatial field of variability for somatic growth, ω_{k_s} (colour scale). We used the R package "mapdata" (Becker et al. 2018) to include the map of Alaska. [Colour online.]



where Δy is the fraction of the year when the survey takes place, and $\tilde{L}_{a,s,y}$ is the mean length-at-age at the survey time.

The numbers-at-age are distributed across the defined length bins (the length bin width is 1 cm) following a normal distribution. The proportion in length bin l for age a ($\varphi_{l,a,s,y}$) is calculated as

Table 1. Names, symbols, and values of indices, variables, and parameters used in the simulation.

Name	Symbol	Value
Indices		
Age	a	
Year	y	
Length bin (cm)	l	
Grid	s	
Sampling station	i	
Replicate	n	
Individual sampled in the age subsample	j	
Variables		
Abundance-at-age	N_a	
Abundance-at-length-at-age	$N_{l,a}$	
Recruitment	R	
Spatial variation in recruitment density	ω_N	
Spatiotemporal variation in recruitment density	ε_N	
Recruitment density in log space	β_N	
Spatial variation in growth coefficient	ω_k	
Temporal variation in growth coefficient	ε_k	
Length-at-age	L_a	
Length-at-age at the time of a survey	$\tilde{L}_a$	
Proportion in length bin l for age a	$\varphi_{l,a}$	
Standard deviation in length for age a	σ_a	
Simulated encounter probability	$P_{l,a}$	
Simulated positive catch rates	$c_{l,a}$	
Parameters		
Maximum age (years)	A	20
Maximum length (cm)	L	120
Fishing mortality	F	0.46
Natural morality	M	0.34
Average recruitment	R_0	4.51e08
Standard deviation in recruitment	σ_R	0.66
Initial growth coefficient	k	0.1376
Asymptotic length (cm)	L_∞	118.6
Adjustment parameter in growth equation	t_0	–0.168
Lower limit of the smallest length bin (cm)	$L'_{\min}$	0.5
Lower limit of the largest length bin (cm)	$L'_{\max}$	119.5
Standard deviation in length for age 0	σ_0	0.8–3.5[a]
Standard deviation in length for age A	σ_A	3–9.5[a]
Length at 50% of retention of the survey selectivity curve (cm)	$S_{50\%}$	18
Slope of the survey age selectivity curve	S_{slope}	1
Fraction of a year when survey takes place	Δy	0.25
Area swept (km^2)	w	0.05
Dispersion for error term for positive catches rates	σ_c	0.7

[a]Values for the low-σ_a and high-σ_a cases, respectively.

$$\varphi_{l,a,s,y} = \begin{cases} \Phi\left(\dfrac{L'_{\min} - \tilde{L}_{a,s,y}}{\sigma_{a,s,y}}\right) & \text{for } l = 1 \\ \Phi\left(\dfrac{L'_{l+1} - \tilde{L}_{a,s,y}}{\sigma_{a,s,y}}\right) - \Phi\left(\dfrac{L'_{l} - \tilde{L}_{a,s,y}}{\sigma_{a,s,y}}\right) & \text{for } 1 < l < L \\ 1 - \Phi\left(\dfrac{L'_{\max} - L_{a,s,y}}{\sigma_{a,s,y}}\right) & \text{for } l = L \end{cases}$$

where Φ is the standard normal cumulative density function, L'_l is the lower limit of length l, $L'_{\min}$ is the lower limit of the smallest length bin, $L'_{\max}$ is the lower limit of the largest bin, L is the index of largest length bin, and $\sigma_{a,s,y}$ is the standard deviation of the length of a fish of age a and it is calculated as

$$\sigma_{a,s,y} = \sigma_0 + \left(\frac{\tilde{L}_{a,s,y} - L_0}{L_\infty - L_0}\right)(\sigma_A - \sigma_0)$$

where σ_0 and σ_A are the standard deviation for length at age 0 and A, respectively, and L_0 is the length at age 0. σ_a is related to the variance of size-at-age in a single year caused by processes with similar effects across space. Moreover, to evaluate the effects of σ_0 and σ_A on our results, we also compare two cases: low and high overlap in length distributions among ages ("low-σ_a" and "high-σ_a", respectively; refer to the online Supplementary material, Fig. S1[1]).

Using the spatiotemporal information of the simulated population, we simulate a survey similar to the Continental Shelf Bottom Trawl Survey of Groundfish and Invertebrate Resources performed in the EBS (Fig. 2; Conner and Lauth 2017) to obtain haul catches and length and age subsamples. A detailed description of this process can be found in Appendix B and a summary of the parameters used in this simulation experiment in Table 1.

Age composition estimation

Using the simulated survey data, we estimate age compositions following the next three steps, and, for simplicity, we keep the

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2020-0166.

haul or station subscript i but omit the year subscript y, but note that these calculations are performed for each year.

First expansion

To expand the number of individuals in the length subsample (c_i) to the total haul catch ($\tilde{c}_i$), we calculate the subsampling intensity: $\lambda_i = \tilde{c}_i/c_i$. Then, the abundance-at-length ($\hat{c}_{l,i}$) is

$$\hat{c}_{l,i} = \frac{\tilde{c}_{l,i}}{\lambda_i}$$

Estimation of abundance-at-age per haul

First, a brief introduction to ALKs is presented. An ALK is constructed from the age subsample where body length and aging structures have been collected. Suppose the ages to estimate are $a = J, \ldots, A^*$, where J and A^* represent the minimum and maximum estimable age, respectively, and the probability that a fish is age a given that it belongs to a length bin l is $q_{l,a}$, $l = 1, \ldots, L$ (Kimura 1977):

$$\sum_a q_{l,a} = 1$$

and where each $q_{l,a}$ is an element of the ALK matrix. Combining information from an ALK with abundance-at-length for station i ($\hat{c}_{l,i}$), we can obtain the abundance-at-age ($\hat{c}_{a,i}$):

$$\sum\nolimits_l \hat{c}_{l,i} q_{l,a} = \hat{c}_{a,i} \quad (1)$$

As we observe, an ALK distributes $\hat{c}_{l,i}$ among different ages to calculate $\hat{c}_{a,i}$. In this study, we evaluate alternative approaches to estimate $\hat{c}_{a,i}$ from $\hat{c}_{l,i}$, where age assignment plays a critical step. We use information from the age subsample to compare the following methods:

Pooled ALK

When the sample size is low, an ALK typically lacks age information for several length bins and a "pooled" ALK, composed of information from multiple years, is preferred (e.g., Carpi et al. 2015; Gulland and Rosenberg 1992). This method reproduces this approach in an extreme case, pooling data from all simulated years to construct a unique ALK that is used to calculate $\hat{c}_{a,i}$ for all years. To reproduce a low sample size scenario, we consider uniquely for this method that the number of age subsampling stations is 50% (166) of the standard number (332). Finally, $\hat{c}_{a,i}$ is estimated using eq. 1.

Annual ALK

For this method, $\hat{c}_{a,i}$ is estimated using a year-specific or annual ALK using eq. 1.

GAMs

For this method, we assume

$$g[\mathbb{E}(a_j)] = \alpha + s_1(l_j) + s_2(\text{lon}_j, \text{lat}_j) + \epsilon_j \quad (2)$$

where g is the log-link function, a_j is the age of a sampled individual j and follows a Tweedie distribution (but we also tested a Gaussian distribution as assumed in previous studies; Ochwada et al. 2008; Puerta et al. 2019*a*, 2019*b*), (lon, lat) represents the geographical coordinates where an individual was sampled, s_1 and s_2 represent the smooth functions (thin plate regression splines) for length and geographic location (both used 10 as the dimension of the basis to represent the smooth term), respectively, and ϵ_j is the error term.

Then, using the information in the length subsample, we predict ages to be assigned to individuals. Since these predicted ages are continuous, we use a simple rounding to obtain the final predicted ages ($\hat{a}$). Finally, $\hat{c}_{a,i} = \hat{c}_{l,i}$, where $a = \hat{a}$.

CRL models

CRLs are a type of model for ordered categorical responses (e.g., ages). We implement "A^* – J" CRL models using GAM for estimation. It follows that the conditional probability of a fish being age a given that it is at least that age is equal to

$$\pi_a = P(Y = a | Y \geq a) = \frac{p_a}{p_a + \ldots + p_{A^*}}, \qquad a = J \ldots A^* - 1$$

where π_a follows a binomial distribution and

$$g[\mathbb{E}(\pi_{aj})] = \alpha_a + \beta_a l_j + s_a(\text{lon}_j, \text{lat}_j) + \epsilon_{aj}$$

where g is the logit link function, and j is an individual sampled. Then, we calculate the estimated unconditional probabilities $\tilde{p}_a$ from the conditional probabilities $\hat{\pi}_a$ (Rindorf and Lewy 2001):

$$\begin{aligned} \tilde{p}_J &= \hat{\pi}_J \\ \tilde{p}_a &= \hat{\pi}_a\left(1 - \sum\nolimits_{j=J}^{a-1} \tilde{p}_j\right) = \hat{\pi}_a \prod\nolimits_{j=J}^{a-1} (1 - \hat{\pi}_j), \quad a > J \end{aligned} \quad (3)$$

For more details about CRL see Berg and Kristensen (2012) and Kvist et al. (2000). After the model implementation, we predict proportions by age class using the information in the length subsample. Finally, $\hat{c}_{l,i}$ is distributed among these predicted proportions to calculate $\hat{c}_{a,i}$. We used the mgcv package in R (Wood 2017) to implement these model-based approaches (GAM and CRL). We also evaluated the exclusion of location (lon_j, lat_j) for these model-based approaches for the No S – No T scenario, since it might be unnecessary.

See Fig. 3 and Fig. S2[1] for an example of how these four methods represent the transition from length to age (i.e., age assignment). Moreover, for comparison purposes, they assume ages 1 and 8 as the first and maximum estimable ages, respectively.

Second expansion

We use a design-based approach to estimate the age composition of the entire population ($\hat{p}_a$) (Stewart and Hamel 2014; Wakabayashi et al. 1985), expanding $\hat{c}_{a,i}$ to the total area sampled by the survey:

$$\hat{p}_a = \frac{\sum_i \hat{c}_{a,i}}{\sum_i \sum_a \hat{c}_{a,i}}$$

Performance metrics

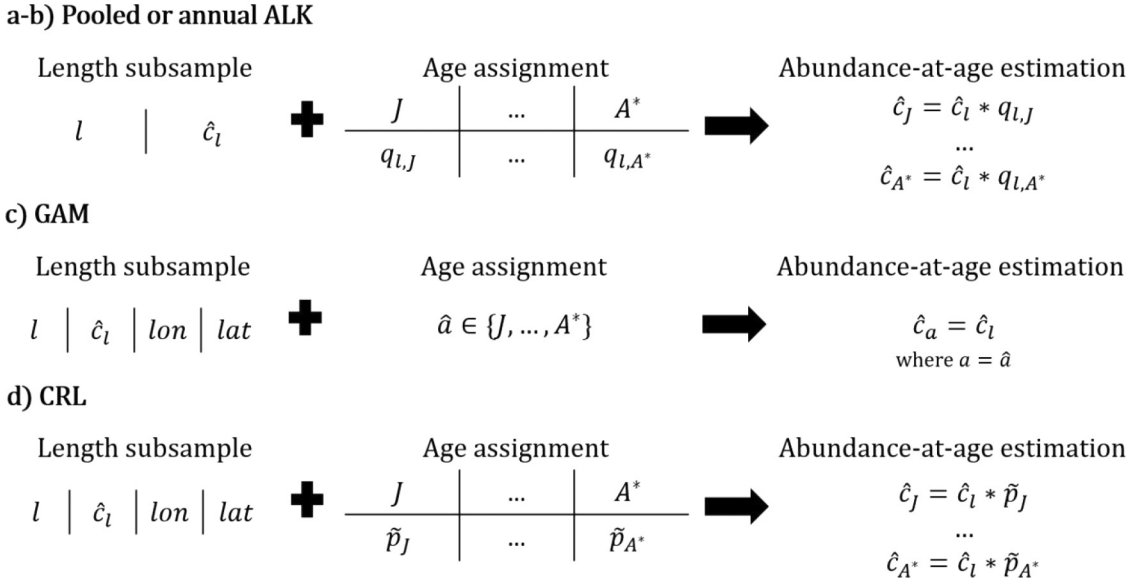
We perform 200 replicates (N_{rep}) and assess the performance of each method comparing the estimated ($\hat{p}_{a,y}$) using simulated survey data and true ($p_{a,y}$) age composition of the population for the two evaluated scenarios.

Mean square error (MSE): used to give information on the accuracy of proportion estimates:

$$\text{MSE}_{a,y} = \frac{1}{N_{\text{rep}}} \sum_{n=1}^{N_{\text{rep}}} (\hat{p}_{a,y,n} - p_{a,y,n})^2$$

where n represents a replicate. We can also average over years (MSE_{year}), ages (MSE_{age}), and years and ages ($\text{MSE}_{\text{total}}$).

Fig. 3. Abundance-at-age estimation ($\hat{c}_a$) per haul by the four evaluated methods for a single length (l), year, and sampling station. Pooled and annual age–length key (ALK): uses length, abundance-at-length ($\hat{c}_l$), and an ALK as the age assignment method (see eq. 1). Generalized additive model (GAM): in addition to l and $\hat{c}_l$, uses longitude and latitude (lon and lat, respectively) information to implement the age assignment method (GAM model), where $\hat{a}$ is the predicted age (see eq. 2). Continuation ratio logit (CRL): the age assignment method is the CRL model, and $\tilde{p}_a$ are the predicted proportions (see eq. 3). J and A^* are the minimum and maximum estimable ages, respectively.



Mean relative error (MRE): used to give information on the bias of proportion estimates:

$$\mathrm{MRE}_{a,y} = \frac{1}{N_{\mathrm{rep}}} \sum_{n=1}^{N_{\mathrm{rep}}} \left(\frac{\hat{p}_{a,y,j} - p_{a,y,n}}{p_{a,y,j}} \right) \times 100\%$$

$\mathrm{MRE}_{a,y} < 0$ or $\mathrm{MRE}_{a,y} > 0$ values mean that the method underestimates or overestimates that proportion, respectively. As done for MSE, we can also average over years ($\mathrm{MRE}_{\mathrm{year}}$), ages ($\mathrm{MRE}_{\mathrm{age}}$), and years and ages ($\mathrm{MRE}_{\mathrm{total}}$).

The code to perform the simulation experiment and the CRL method can be found in https://www.github.com/gmoroncorrea/STageCompsEstimation.

Stock assessment model

We perform a system-based testing (Chang et al. 2017; Hinton and Maunder 2003) to investigate the consistency (compatibility) of age compositions estimated by the four aforementioned methods with regard to the other data sources (e.g., catch-per-unit effort (CPUE), length compositions) in the latest stock assessment model for the Pacific cod in the EBS. First, we estimate age compositions for this species by method and considering survey data from 1994 to 2016 (Conner and Lauth 2017). We compare these estimates and then include them in the stock assessment model, therefore having four models that differ exclusively in the age composition data (hereinafter referred to as SS pooled ALK, SS annual ALK, SS GAM, and SS CRL). We use the 2018 Pacific cod stock assessment model in the EBS configuration (Thompson et al. 2018) implemented in the Stock Synthesis (SS) modeling framework (Methot and Wetzel 2013). This model has been thoroughly evaluated, and therefore we assume that its structure is correctly specified. Data inputs used by this model are fishery landings, survey abundance, fishery length composition, and survey age composition data (Fig. S3[1]). We use the Francis data weighting method (Francis 2011) to tune the input sample size of compositional data. To evaluate the consistency of age compositions with the model structure and other data, we compare model fits using the calculated total and by component negative log-likelihood (NLL) by the stock assessment models. We also compare spawning biomass (SSB) and recruitment estimates to examine how different age compositions data sets influence these crucial time series.

We used the R software environment (R Core Team 2019) to perform the analyses of this study and the R package "ggplot2" (Wickham 2016) to generate the figures.

Results

The CRL method was the most robust (unbiased) method to estimate age composition in our simulation experiment across scenarios, and annual ALKs ranked second. Also, the stock assessment model that included age composition estimated by CRL displayed the highest consistency with other data.

Simulation experiment

The simulated population and survey successfully imitated the spatiotemporal variability in size-at-age observed for the Pacific cod in the EBS (Figs. S4–S9[1]; see also figures 2 and 5 in Ciannelli et al. 2020). Simulated haul catches ranged from 0 to ~300 individuals, lengths in the length subsamples from 12 to 110 cm, and ages in the age subsamples from 1 to 16. Different parameters to simulate the spatial and spatiotemporal terms (ω_{N_s} and $\varepsilon_{N_s,y}$) did not vary our results; therefore, we present results using the parameter values as specified above. However, different values of σ_0 and σ_A (low-σ_a and high-σ_a cases) generated substantial changes on the performance of some methods as explained below.

Age composition estimation

Data gaps (length bins without age information) were generally scarce for the pooled ALK; however, data gaps were common in the annual ALKs (Fig. S2[1]), especially for larger lengths. Regarding the GAM, diagnostic plots, which test the distribution assumption, constant variance, and distribution of residuals, confirmed that the Tweedie distribution outperformed a Gaussian distribution (Figs. S10–S11[1]); therefore, we retain results assuming the former. The exclusion of location (lon_j, lat_j) of the independent variables for GAM and CRL for the No S – No T scenario did not vary our results; therefore, we present results including them.

Table 2. Total mean square error (MSE_{total}, $\times 10^{-5}$) and mean relative error (MRE_{total}) averaged across replicates.

		No S – No T				S–T			
		Pooled ALK	Annual ALK	GAM	CRL	Pooled ALK	Annual ALK	GAM	CRL
Low-σ_a	MSE_{total}	2.83	2.71	3.13	**2.43**	127.6	3.77	2.63	**2.5**
	MRE_{total}	3.24	–4.71	–0.28	**0.12**	20.9	–4	1.75	**0.11**
High-σ_a	MSE_{total}	151.73	5.36	30.68	**5.07**	249.4	5.93	29.38	**5.12**
	MRE_{total}	25.42	–3.82	2.34	**0.2**	26.62	–3.3	5.13	**0.33**

Note: The best metric value per scenario and case is shown in bold.

CRL was the most robust method to estimate age compositions for both scenarios (No S – No T and S–T) and cases (low-σ_a and high-σ_a) (Table 2). When the overlap of length distributions across ages was low (low-σ_a case) and spatiotemporal variability in somatic growth was absent, the four methods led to small errors (MSE_{total}) with the CRL approach being the method that displayed the lowest value and therefore best performance. Differences in MRE_{total} values, a measure of bias, among methods were also small for this scenario (<±5%). Bias for pooled ALK was slightly positive (+3.24%), annual ALK bias was slightly negative (–4.71%), and GAM and CRL biases were close to 0%. In contrast, results changed drastically for the pooled ALK method when variability in somatic growth was present; however, the other methods were almost unaffected, reporting similar values in comparison with the No S – No T scenario. Similarly, MRE_{total} values supported CRL as the best method (~0%), the GAM model ranked second (~1.75%), and the pooled ALK was the worst (+20.9%).

Nevertheless, the performance of the pooled ALK and GAM got substantially worse when the overlap of length distributions across ages increased (high-σ_a case; Table 2). Specifically, errors (MSE_{total} values) of these two methods were quite larger, and there was also a small increase for annual ALK and CRL for both scenarios. Likewise, there was a large positive bias (MRE_{total}) for the pooled ALK, and this increased slightly for GAM in comparison with the low-σ_a case. The annual ALK and CRL were quite robust in comparison with the low-σ_a case for both scenarios.

We analyzed performance metrics by ages and year periods. For both scenarios and cases, errors (MSE_{age}) tended to be larger for younger ages and decreased for older ages (Fig. 4 and Fig. S12[1]). Except for the No S – No T and low-σ_a cases, CRL and pooled ALK exhibited the best and worst performance, respectively, for all ages. Regarding biases (MRE_{age}), those were always larger for older ages except for the CRL method, which displayed small biases for all ages. Moreover, the annual ALK always tended to underestimate proportions for older ages.

Examining these metrics by year period, the largest errors (MSE_{year}) were normally observed for the pooled ALK excepting the No S – No T and low-σ_a cases, which displayed comparable errors among methods and year periods (Fig. 5 and Fig. S13[1]). In terms of bias (MRE_{year}), the underestimation and overestimation observed for the annual ALK and GAM models, respectively, were present for all periods. CRL did not show substantial biases, while the magnitude of the bias for the pooled ALK increased for the last periods.

In summary, these results support the CRL model as the most robust method to estimate age compositions across scenarios, ages, and years and for different degrees of overlap of length distributions among ages. The annual ALK ranked second in terms of performance, while the GAM was substantially affected when σ_a increased. On the other hand, the pooled ALK showed the worst performance, producing quite large errors and biases.

Stock assessment model

Survey haul catches for Pacific cod (1994–2016) in the EBS ranged from 0 to ~400 individuals, although there were a few outliers up to ~6000 individuals. Length subsamples ranged between 4 and 112 cm, and the age subsamples ranged between 1 and 17. Age compositions estimated by the evaluated methods were quite similar for younger ages; however, differences became larger for intermediate and older ages (Fig. S14[1]). Specifically, the pooled ALK displayed substantial differences in population structures for some years (e.g., 1995, 1996, 2002, 2004, 2011, 2015, 2016), when it did not seem to capture the variability across ages estimated by other methods.

Stock assessment models incorporating these age composition estimates reached convergence (positive defined Hessian and convergence criterion: 1e–04). These models used the same input sample size for survey age and fishery length composition data; therefore, observed differences are an exclusive consequence of different age composition data. The stock assessment model that included age compositions estimated using CRL (SS CRL) reported the lowest NLL (Table 3) and therefore the best consistency with other data. NLL by component showed that compositional data (length and age compositions) were better fitted by SS CRL as well; however, no substantial differences were found for indices of abundance. The SS pooled ALK model displayed the highest NLL, followed by the SS GAM model, results that agree with what we observed in our simulation experiment. SSB and recruitment estimates showed quite similar trends for the SS CRL and SS annual ALK. However, SSB values estimated by SS pooled ALK were generally larger, and recruitment estimates did not capture the variability observed for the other stock assessment models (Fig. 6).

Discussion

In this study, we implemented a simulation experiment to evaluate the performance of different methods applied in previous studies to estimate age compositions of a fish population. We showed, through two metrics, that the CRL approach is the most precise and accurate for estimating age compositions. Annual ALKs displayed quite good performance as well, both being methods quite robust to changes in the overlap of length distributions among ages and variations in somatic growth. Moreover, the incorporation of age compositions estimated by CRL in the Pacific cod stock assessment model resulted in a higher consistency with other data inputs.

Age composition estimation

Our results demonstrate some shortcomings of classic methods (ALKs) when estimating age compositions. When the sampling effort is low or insufficient within a year, data gaps are abundant in the age subsample. This missing information often leads to fisheries scientists taking subjective and uncertain decisions to fill these gaps (e.g., sharing data from adjacent years; Ailloud and Hoenig 2019). A pooled ALK might help to avoid these issues; however, it led to large errors and biases, especially for intermediate ages, as a consequence of sharing data from years with different somatic growth rates (S–T scenario). Moreover, these biases were aggravated when the variation of size-at-age was increased (high-σ_a case) regardless the variation in somatic growth. Aanes and Vølstad (2015) also found serious biases in age composition estimates when pooled ALKs were constructed for the Northeast Arctic cod (*Gadus morhua*); therefore, we consider that this strategy is not appropriate, especially when there is substantial variability in somatic growth or there is no information about it.

Annual or year-specific ALKs might be a reasonable solution to deal with pooled ALKs drawbacks, especially when the temporal

Fig. 4. Mean square error (MSE_{age}, a measure of errors) and mean relative error (MRE_{age}, a measure of bias) per age for both scenarios for the high-σ_a case. Metric values beyond the *y* axis limits are printed. [Colour online.]

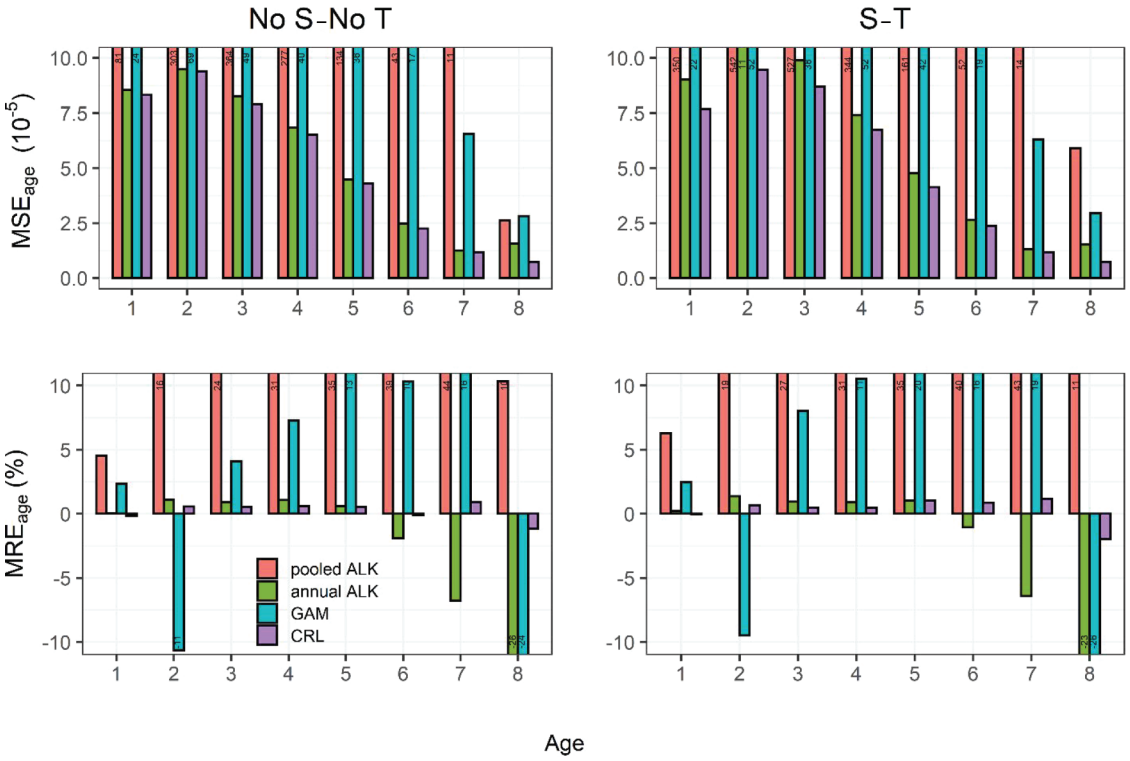
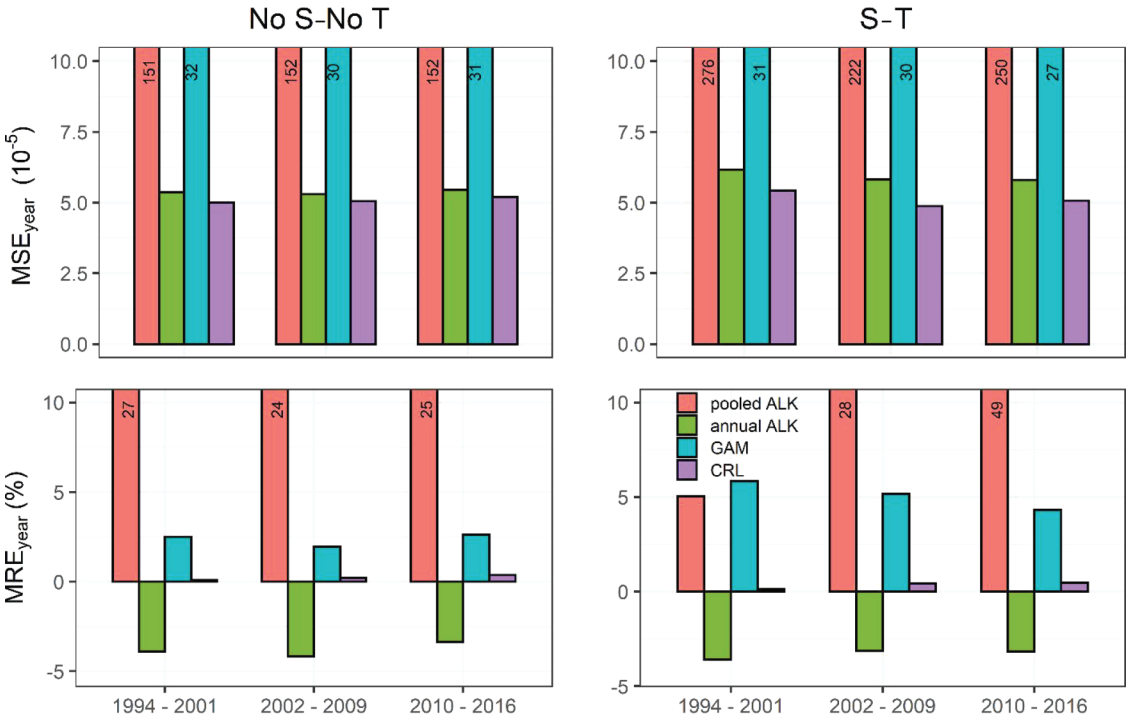


Fig. 5. Mean square error (MSE_{year}, a measure of errors) and mean relative error (MRE_{year}, a measure of bias) per period for both scenarios for the high-σ_a case. Metric values beyond the *y* axis limits are printed. [Colour online.]

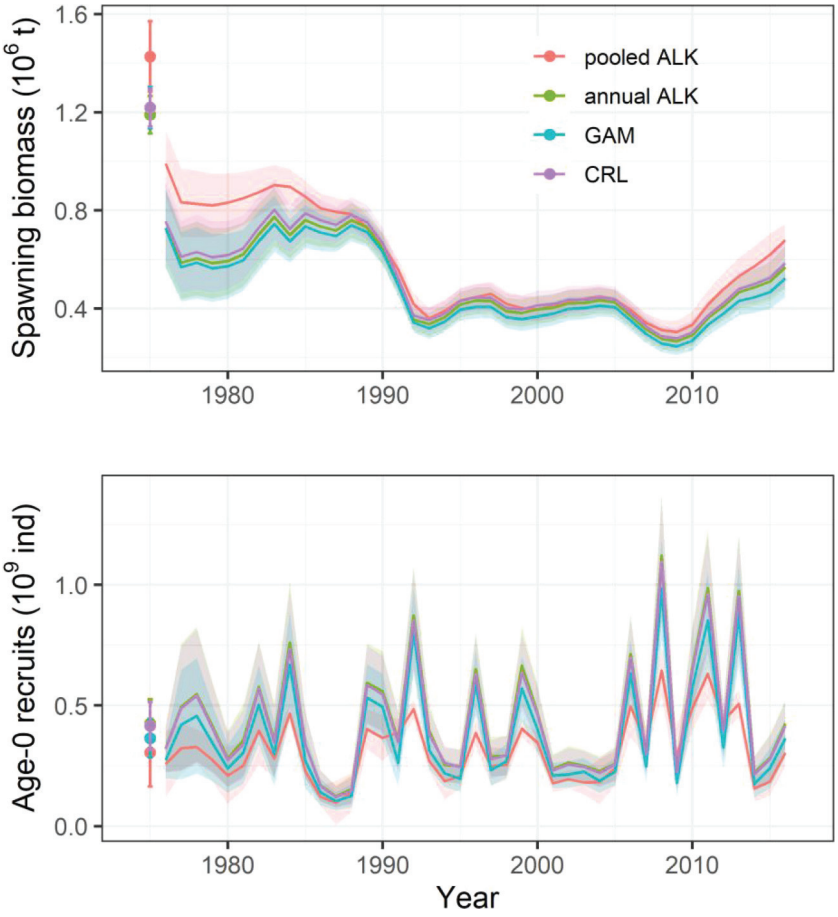


variability is substantial. However, although annual ALKs account for changes in growth rates in time, their performance might still be affected by the spatial variability in growth. Moreover, another drawback of this method is that it introduces a considerable number of data gaps for larger lengths, possibly influenced by a proportional age sampling design assumed in this study (see Appendix B), which collects more data for lengths with higher abundances in the length subsample (Quinn and Deriso 1999). These data gaps are the main cause of the observed underestimation in age proportions for older ages in both scenarios. Some

Table 3. Negative log-likelihood (NLL) by component of stock assessment models (SS, Stock Synthesiss) that incorporated age compositions estimated by the four evaluated methods.

Component	SS pooled ALK	SS annual ALK	SS GAM	SS CRL
Total	92.53	75.18	88.96	72.93
Catch	5.9e–13	0.0105e–13	0.06e–13	0.35e–13
Equilibrium catch	11.4e–05	6.3e–05	9.6e–05	6.8e–05
Survey	–29.7	–40.7	–40.05	–40.15
Length composition	74.27	72.1	72.66	71.77
Age composition	76.32	61.83	71.26	59.43
Recruitment	–29.23	–19.03	–16.08	–19.09

Fig. 6. Spawning biomass and recruitment time series estimated by stock assessment models (continuous lines) that incorporated age compositions estimated by evaluated methods using Pacific cod data. Dots show the estimated value at a virgin state, and shaded areas are 95% confidence intervals estimated by the stock assessment model. [Colour online.]



authors have proposed strategies to deal with this problem (Ailloud and Hoenig 2019; Hoenig et al. 2002; Isermann and Knight 2005; Kimura and Chikuni 1987), for example, implementing a combined classic-inverse ALK (an inverse ALK describes the probability of length given age) into one likelihood function (Ailloud et al. 2019). However, simpler strategies are normally preferred, such as "borrowing" data from previous years exclusively for these lengths; however, they can also lead to similar consequences as we observed for the pooled ALK.

Although the GAM model reduced total bias over annual ALKs in the low-σ_a case, it still displayed poor performance for age 5 and older. We suspect that this was caused by predicting a single age from length information and treating the response as continuous when it is discrete, similar to previous studies (Puerta et al. 2019*a*, 2019*b*). Large errors for older ages appeared because there is a higher overlap between length distributions at those ages. This was confirmed when the overlap was increased (high-σ_a case), where GAM showed a substantially worse performance for all ages. Similar results were also found for the age slicing method, which assigns a single age from length information using the von Bertalanffy growth curve (Ailloud et al. 2015; Kell and Kell 2011). We can conclude that methods that assign a unique age from length information, such as the GAM in this study or the age slicing method, may be inappropriate when a fish population has a high overlap of length distributions among ages.

CRL was the most robust method in the simulation experiment. This method accounts for spatial and temporal variability in the model, and the response variable is proportions-at-age, so rounding decisions become unnecessary. CRL has been previously used for different purposes and has shown promising results (Berg and Kristensen 2012; Gerritsen et al. 2006; Kvist et al. 2000; Stari et al. 2010). CRL procedures outperformed ALKs when using fishery or survey data (Berg and Kristensen 2012) and also have the benefit of smoothing data gaps when age data are scarce for some lengths

(Stari et al. 2010); therefore, data gaps are absent for large lengths. However, it also has some issues related to data availability. For instance, model convergence issues were found when an age older than 8 was designated as the maximum estimable age. This was due to age and length information being commonly scarce for older ages, particularly if they are dominated by low-abundance cohorts or if there is a low sampling effort. CRL cannot deal with limited amounts of data and led to issues when estimating the parameters of the GAM model. Similar problems might arise for younger ages, which might be underrepresented in haul catches due to selectivity effects. Therefore, we recommend using CRL for data-rich scenarios.

Stock assessment model

One potential reason for the similar age composition estimated by the four evaluated methods using Pacific cod survey data is that the abundance-at-age was estimated per depth stratum defined by NOAA scientists in the EBS (Conner and Lauth 2017) and then aggregated to maintain consistency with its practical implementation. This strategy is expected to produce more precise estimates of abundances given the spatial distribution of fish abundance geographically and with depth (Wakabayashi et al. 1985), which can improve estimates of age compositions regardless the age assignment procedure. A second potential reason is that the Pacific cod age subsamples were collected using a fixed strategy (i.e., collecting otoliths for aging from a constant number of individuals that is in some sense independent of the abundance in the catch; Quinn and Deriso 1999) for most of the time series, which might lead to smaller differences in age composition estimates across methods. Nevertheless, these small differences may lead to improvements in the performance of stock assessment models as demonstrated by previous studies (Berg et al. 2014; Thorson and Haltuch 2019).

A simple approach such as system-based testing performed in this study supported age composition estimated by CRL according to the fit of the assessment model to that data as well as its consistency with other data inputs in the Pacific cod stock assessment model. A system-based testing examines the consistency among different data inputs in a stock assessment model, and it has principally been applied for CPUE data and involves evaluating how different CPUE standardized time series fit with the predicted values by the model and how consistent they are with ancillary data (Chang et al. 2017; Hinton and Maunder 2003). However, since likelihood functions are determined by observed or input data, using NLL to directly compare and rank stock assessment models that are based on different input data might not be completely appropriate, and other alternatives should be contrasted. For instance, the ratio between input and effective sample size of compositional data has been proposed to be a good indicator to evaluate different compositional data sets (Thorson and Haltuch 2019). Area of confidence ellipses, which compare uncertainty of pairs of parameters conjunctly, also appears as an alternative and has been applied for abundance-at-age indices (Berg et al. 2014).

The quality of a stock assessment is related to the quality and quantity of the input data (Mace et al. 2001), which means that more consistent data result in more certain stock assessment estimates (Berg et al. 2014). Age composition data are especially informative for growth, recruitment, selectivity, and natural mortality (Magnusson and Hilborn 2007; Maunder and Piner 2015), and many stock assessment models are driven primarily by this source of data (Francis 2011); therefore, there is a need to obtain better estimates. Some efforts have been made to obtain more precise age composition estimates using model-based approaches (Berg et al. 2014; Maunder et al. 2020; Thorson and Haltuch 2019). These novel approaches can be used conjunctly with CRL in future studies, improving the entire estimation process. But there is also a need to evaluate other effects ignored in this study, such as aging imprecision, produced when scientists read otoliths or other hard parts (Candy et al. 2012), or how the age subsample is sampled: fixed or proportional (Quinn and Deriso 1999). The evaluation of these interacting factors might lead to a better understanding and improvement of the estimation of age composition and therefore of stock assessment outputs.

Conclusions and future research

The CRL approach appears as a robust alternative to classic approaches (ALKs) to estimate age compositions. While we focused on estimation in the face of spatiotemporal variability in somatic growth, future studies could broaden the scope of our inferences by evaluating the performance of these methods given different life-history parameters, sampling strategies, spatial and temporal variation in natural morality, and the strength of density dependence. It is also worth evaluating how fishery size-specific selectivity might affect our results, since it may increase mortality for larger individuals of an age class, producing a reduction of the mean length-at-age of that age class in time (Lee 1912). Pacific cod age compositions estimated by the evaluated methods did not substantially differ for younger ages, possibly influenced by other factors related to the sampling strategy or the age composition estimation process. The use of age compositions estimated by CRL in the stock assessment model resulted in better fits and was more consistent with other data inputs. However, we recommend using other approaches to evaluate consistency of different input data sets for a stock assessment model. This study demonstrates the feasibility of using model-based approaches as CRL to estimate age composition and its application in stock assessment models.

Acknowledgements

We are grateful to the Habitat Information for Stock Assessment program (ID: NA16OAR4320152) for providing funding for this research for Giancarlo M. Correa. Grant Thompson provided valuable comments on the stock assessment model and James Thorson on the implementation of the simulation experiment. We are also very grateful to Caitlin Allen Akselrud, Carey McGilliard, and three anonymous reviewers for their valuable comments and to the scientists of the NOAA Alaska Fisheries Science Center and vessel crew that have contributed to the eastern Bering Sea groundfish survey.

References

Aanes, S., and Vølstad, J.H. 2015. Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Can. J. Fish. Aquat. Sci. **72**: 938–953. doi:10.1139/cjfas-2014-0408.

Adams, G.D., Leaf, R.T., Ballenger, J.C., Arnott, S.A., and McDonough, C.J. 2018. Spatial variability in the growth of Sheepshead (*Archosargus probatocephalus*) in the Southeast US: Implications for assessment and management. Fish. Res. **206**: 35–43. doi:10.1016/j.fishres.2018.04.023.

Agresti, A. 2010. Analysis of Ordinal Categorical Data, Second. ed. John Wiley & Sons, New Jersey.

Aikio, S., Herczeg, G., Kuparinen, A., and Merilä, J. 2013. Optimal growth strategies under divergent predation pressure. J. Fish Biol. **82**: 318–331. doi:10.1111/jfb.12006. PMID:23331153.

Ailloud, L.E., and Hoenig, J.M. 2019. A general theory of age–length keys: combining the forward and inverse keys to estimate age composition from incomplete data. ICES J. Mar. Sci. **76**(6): 1515–1523. doi:10.1093/icesjms/fsz072.

Ailloud, L.E., Smith, M.W., Then, A.Y., Omori, K.L., Ralph, G.M., and Hoenig, J.M. 2015. Properties of age compositions and mortality estimates derived from cohort slicing of length data. ICES J. Mar. Sci. **72**: 44–53. doi:10.1093/icesjms/fsu088.

Ailloud, L.E., Lauretta, M.V., Walter, J.F., and Hoenig, J.M. 2019. Estimating age composition for multiple years when there are gaps in the ageing data: the case of western Atlantic bluefin tuna. ICES J. Mar. Sci. **76**(6): 1690–1701. doi:10.1093/icesjms/fsz069.

Baudron, A.R., Needle, C.L., Rijnsdorp, A.D., and Tara Marshall, C. 2014. Warming temperatures and smaller body sizes: synchronous changes in growth of North Sea fishes. Glob. Chang. Biol. **20**: 1023–1031. doi:10.1111/gcb.12514. PMID:24375891.

Becker, R., Wilks, A., and Brownrigg, R. 2018. mapdata: Extra Map Databases. R package version 2.3.0.

Berg, C.W., and Kristensen, K. 2012. Spatial age–length key modelling using

continuation ratio logits. Fish. Res. **129–130**: 119–126. doi:10.1016/j.fishres.2012.06.016.

Berg, C.W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fish. Res. **151**: 91–99. doi:10.1016/j.fishres.2013.10.005.

Candy, S.G., Nowara, G.B., Welsford, D.C., and McKinlay, J.P. 2012. Estimating an ageing error matrix for Patagonian toothfish (*Dissostichus eleginoides*) otoliths using between-reader integer errors, readability scores, and continuation ratio models. Fish. Res. **115–116**: 14–23. doi:10.1016/j.fishres.2011.11.005.

Carpi, P., Santojanni, A., Donato, F., Colella, S., Čikeš Keč, V., Zorica, B., et al. 2015. A joint stock assessment for the anchovy stock of the northern and central Adriatic Sea: comparison of two catch-at-age models. Sci. Mar. **79**: 57–70. doi:10.3989/scimar.03903.29A.

Chang, S.-K., Liu, H.-I., Fukuda, H., and Maunder, M.N. 2017. Data reconstruction can improve abundance index estimation: An example using Taiwanese longline data for Pacific bluefin tuna. PLoS ONE, **12**: e0185784. doi:10.1371/journal.pone.0185784. PMID:28968434.

Chen, Y., Chen, L., and Stergiou, K.I. 2003. Impacts of data quantity on fisheries stock assessment. Aquat. Sci. **65**: 92–98. doi:10.1007/s000270300008.

Ciannelli, L., Tolkova, I., Lauth, R., Puerta, P., Helser, T., Gitelman, A., and Thompson, G. 2020. Spatial, interannual, and generational sources of trait variability in a marine population. Ecology **101**(1): e02907. doi:10.1002/ecy.2907. PMID:31587266.

Coachman, L.K. 1986. Circulation, water masses, and fluxes on the southeastern Bering Sea shelf. Cont. Shelf Res. **5**(1–2): 23–108. doi:10.1016/0278-4343(86)90011-7.

Conner, J., and Lauth, R.R. 2017. Results of the 2016 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources (Technical Memorandum). NOAA, Seattle, Wash.

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

Fridriksson, A. 1934. On the calculation of age distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. Rapp. Proces-Verbaux Reun. Cons. Int. Pour Explor. Mer. **86**: 1–14.

Gale, B.H., Johnson, J.B., Bruce Schaalje, G., and Belk, M.C. 2013. Effects of predation environment and food availability on somatic growth in the livebearing fish *Brachyrhaphis rhabdophora* (Pisces: Poeciliidae). Ecol. Evol. **3**: 326–333. doi:10.1002/ece3.459. PMID:23467582.

Gerritsen, H.D., McGrath, D., and Lordan, C. 2006. A simple method for comparing age–length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). ICES J. Mar. Sci. **63**: 1096–1100. doi:10.1016/j.icesjms.2006.04.008.

Gertseva, V., Matson, S.E., and Cope, J. 2017. Spatial growth variability in marine fish: example from Northeast Pacific groundfish. ICES J. Mar. Sci. **74**: 1602–1613. doi:10.1093/icesjms/fsx016.

Gislason, H., Daan, N., Rice, J.C., and Pope, J.G. 2010. Size, growth, temperature and the natural mortality of marine fish. Fish Fish. **11**: 149–158. doi:10.1111/j.1467-2979.2009.00350.x.

Gulland, J.A., and Rosenberg, A.A. 1992. A review of length-based approaches to assessing fish stocks (No. 323). FAO, Rome.

Helser, T.E., and Brodziak, J.K.T. 1998. Impacts of density-dependent growth and maturation on assessment advice to rebuild depleted U.S. silver hake (*Merluccius bilinearis*) stocks. Can. J. Fish. Aquat. Sic. **55**(4): 882–892. doi:10.1139/f97-290.

Hernandez-Miranda, E., and Ojeda, F.P. 2006. Inter-annual variability in somatic growth rates and mortality of coastal fishes off central Chile: an ENSO driven process? Mar. Biol. **149**: 925–936. doi:10.1007/s00227-006-0249-9.

Hinton, M.G., and Maunder, M.N. 2003. Methods for standardizing CPUE and how to select among them (No. MWG-7). Inter-American Tropical Tuna Commission.

Hoenig, J.M., Hanumara, R.C., and Heisey, D.M. 2002. Generalizing Double and Triple Sampling for Repeated Surveys and Partial Verification. Biom. J. **44**(5): 603–618. doi:10.1002/1521-4036(200207)44:5<603::AID-BIMJ603>3.0.CO;2-4.

Hunter, A., Speirs, D.C., and Heath, M.R. 2019. Population density and temperature correlate with long-term trends in somatic growth rates and maturation schedules of herring and sprat. PLoS ONE, **14**: e0212176. doi:10.1371/journal.pone.0212176. PMID:30840654.

Hurst, T., Miller, J., Ferm, N., Heintz, R., and Farley, E. 2018. Spatial variation in potential and realized growth of juvenile Pacific cod in the southeastern Bering Sea. Mar. Ecol. Prog. Ser. **590**: 171–185. doi:10.3354/meps12494.

Isermann, D.A., and Knight, C.T. 2005. A Computer Program for Age–Length Keys Incorporating Age Assignment to Individual Fish. North Am. J. Fish. Manag. **25**: 1153–1160. doi:10.1577/M04-130.1.

Kell, L.T., and Kell, A. 2011. A comparison of age slicing and statistical age estimation for mediterranean swordfish (*Xiphias gladius*). Collect. Vol. Sci. Pap. ICCAT, **66**(4): 1522–1534.

Kimura, D.K. 1977. Statistical Assessment of the Age–Length Key. J. Fish. Res. Board Can. **34**: 317–324. doi:10.1139/f77-052.

Kimura, D.K., and Chikuni, S. 1987. Mixtures of Empirical Distributions: An Iterative Application of the Age–Length Key. Biometrics, **43**: 23. doi:10.2307/2531945.

Kvist, T., Gislason, H., and Thyregod, P. 2000. Using continuation-ratio logits to analyze the variation of the age composition of fish catches. J. Appl. Stat. **27**: 303–319. doi:10.1080/02664760021628.

Lee, Q., Thorson, J.T., Gertseva, V.V., and Punt, A.E. 2018. The benefits and risks of incorporating climate-driven growth variation into stock assessment models, with application to Splitnose Rockfish (*Sebastes diploproa*). ICES J. Mar. Sci. **75**: 245–256. doi:10.1093/icesjms/fsx147.

Lee, R.M. 1912. An investigation into the methods of growth determination in fishes by 763 means of scales. ICES J. Mar. Sci. **1**(63): 3–34. doi:10.1093/icesjms/s1.63.3.

Mace, P.M., Bartoo, N.W., Hollowed, A.B., Kleiber, P., Methot, R.D., Murawski, S.A., et al. 2001. Marine Fisheries Stock Assessment Improvement Plan. US Dep. Comm. NOAA Tech. Memo. NMFS-F/SPO-56.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish Fish. **8**: 337–358. doi:10.1111/j.1467-2979.2007.00258.x.

Maunder, M.N., and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. **72**: 7–18. doi:10.1093/icesjms/fsu015.

Maunder, M.N., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., et al. 2020. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fish. Res. **229**: 105594. doi:10.1016/j.fishres.2020.105594.

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.

Ochwada, F.A., Scandol, J.P., and Gray, C.A. 2008. Predicting the age of fish using general and generalized linear models of biometric data: A case study of two estuarine finfish from New South Wales, Australia. Fish. Res. **90**: 187–197. doi:10.1016/j.fishres.2007.10.007.

Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., et al. 2015. The importance of length and age composition data in statistical age-structured models for marine species. ICES J. Mar. Sci. **72**: 31–43. doi:10.1093/icesjms/fsu007.

Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. **30**: 683–691. doi:10.1016/j.cageo.2004.03.012.

Puerta, P., Ciannelli, L., and Johnson, B. 2019*a*. A simulation framework for evaluating multi-stage sampling designs in populations with spatially structured traits. PeerJ, **7**: e6471. doi:10.7717/peerj.6471. PMID:30828489.

Puerta, P., Johnson, B., Ciannelli, L., Helser, T., and Lauth, R. 2019*b*. Subsampling populations with spatially structured traits: a field comparison of stratified and random strategies. Can. J. Fish. Aquat. Sci. **76**: 511–522. doi:10.1139/cjfas-2017-0248.

Punt, A.E., Haddon, M., and Tuck, G.N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? A case study based on pink ling in Australia. Fish. Res. **168**: 85–99. doi:10.1016/j.fishres.2015.04.002.

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

R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rindorf, A., and Lewy, P. 2001. Analyses of length and age distributions using continuation-ratio logits. Can. J. Fish. Aquat. Sci. **58**: 1141–1152. doi:10.1139/f01-062.

Schlather, M., Malinowski, A., Menck, P.J., Oesting, M., and Strokorb, K. 2015. Analysis, Simulation and Prediction of Multivariate Random Fields with Package Random Fields. J. Stat. Software, **63**(8): 1–25. doi:10.18637/jss.v063.i08.

Sinclair, A.F., Swain, D.P., and Hanson, J.M. 2002. Disentangling the effects of size-selective mortality, density, and temperature on length-at-age. Can. J. Fish. Aquat. Sci. **59**: 372–382. doi:10.1139/f02-014.

Stabeno, P.J., Bond, N.A., Kachel, N.B., Salo, S.A., and Schumacher, J.D. 2001. On the temporal variability of the physical environment over the south-eastern Bering Sea. Oceanogr. **10**(1): 81–98. doi:10.1046/j.1365-2419.2001.00157.x.

Stari, T., Preedy, K.F., McKenzie, E., Gurney, W.S.C., Heath, M.R., Kunzlik, P.A., and Speirs, D.C. 2010. Smooth age length keys: Observations and implications for data collection on North Sea haddock. Fish. Res. **105**: 2–12. doi:10.1016/j.fishres.2010.02.004.

Stawitz, C.C., and Essington, T.E. 2019. Somatic growth contributes to population variation in marine fishes. J. Anim. Ecol. **88**(2): 135–329. doi:10.1111/1365-2656.12921.

Stawitz, C.C., Essington, T.E., Branch, T.A., Haltuch, M.A., Hollowed, A.B., and Spencer, P.D. 2015. A state-space approach for detecting growth variation and application to North Pacific groundfish. Can. J. Fish. Aquat. Sci. **72**: 1316–1328. doi:10.1139/cjfas-2014-0558.

Stewart, I.J., and Hamel, O.S. 2014. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. **71**: 581–588. doi:10.1139/cjfas-2013-0289.

Thompson, G., Ianelli, J., and Lauth, R. 2018. Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands Area, Plan Team for the Groundfish Fisheries of the Bering Sea/Aleutian Islands. North Pacific Fishery Management Council, 605 W. 4th Avenue Suite 306, Anchorage, AK 99501.

Thorson, J.T., and Haltuch, M.A. 2019. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Can. J. Fish. Aquat. Sci. **76**: 401–414. doi:10.1139/cjfas-2018-0015.

Thorson, J.T., and Minte-Vera, C.V. 2016. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fish. Res. **180**: 45–53. doi:10.1016/j.fishres.2014.11.016.
Thorson, J.T., Monnahan, C.C., and Cope, J.M. 2015. The potential impact of time-variation in vital rates on fisheries management targets for marine fishes. Fish. Res. **169**: 8–17. doi:10.1016/j.fishres.2015.04.007.
Thorson, J.T., Ianelli, J.N., and Kotwicki, S. 2017. The relative influence of temperature and size-structure on fish distribution shifts: A case-study on Walleye pollock in the Bering Sea. Fish Fish. **18**: 1073–1084. doi:10.1111/faf.12225.
Wakabayashi, K., Bakkala, R., and Alton, M. 1985. Methods of the U.S.–Japan demersal trawl surveys. *In* Results of Cooperative U.S.–Japan Groundfish Investigations in the Bering Sea During May–August 1979. *Edited by* R.G. Bakkala and K. Wakabayashi. Int. North Pac. Fish. Comm. Bull.
Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Wood, S.N. 2017. Generalized additive models: an introduction with R. Second edition. Chapman & Hall/CRC texts in statistical science. CRC Press/Taylor & Francis Group, Boca Raton, Fla.

Appendix A. Case study system

The eastern Bering Sea (EBS) continental shelf extends more than 500 km eastward from the Alaskan coastline, with a steep shelf break at the western boundary. The shelf is divided into inner (<50 m depth), middle (50–100 m), and outer (100–200 m) regions, considering bathymetry and oceanographic characteristics (Coachman 1986; Stabeno et al. 2001). The cold pool is a water layer <2 °C formed from winter sea ice, which induces cooling and increases salinity and density of the surface water, particularly affecting the EBS middle shelf domain (Stabeno et al. 2001). The Pacific cod stock is an important groundfish species in the EBS food web and is distributed along the entire continental shelf. Researchers have observed spatial variation in size-at-age for this species along a gradient from the inner to outer shelf, with differences up to 5 cm, caused in large part by the cold pool (Ciannelli et al. 2020; Puerta et al. 2019*b*). Ciannelli et al. (2020) found that size-at-age was highest through age 5 when age 1 cohorts experience average temperature from 1.5 to 2.5 °C, which agrees with previous studies that conclude higher potential growth in age 0 cod for that range of temperatures (Hurst et al. 2018). Moreover, there is also a temporal variation in size-at-age. For example, individuals at a given age were ~6 cm smaller in 1994 than in 2016 (Ciannelli et al. 2020).

Appendix B. Sampling

We simulate a survey based on the EBS trawl survey (Conner and Lauth, 2017) at the end of the first quarter of each year. We model the sampling process separating encounter probability and positive catch rates. The encounter probability component defines the probability $d_{l,a,i}$ that catch for the sample station i (349 stations in a survey; Fig. 2) in the location s_i and year y and for the age a and length l is nonzero (Thorson et al. 2018):

$$P_{l,a} \sim \text{Bernoulli}(d_{l,a,i})$$

where

$$d_{l,a,i} = 1 - \exp(-w_i S2_a \mathbf{N}_{l,a,s_i,y})$$

where w_i is the area swept by sample i (0.05 km^2), $\mathbf{N}_{l,a,s,y}$ is a matrix representing the abundance-at-length-at-age obtained from $\varphi_{l,a,s_i,y}$ and $N_{a,s_i,y}$, and $S2_a$ is the logistic selectivity-at-age of the survey ($S2_a = 1$ for $a > 0$).

The catch rates $\mathbf{c}_{l,a,i}$, in numbers of individuals, are simulated as

$$\mathbf{c}_{l,a,i} = \begin{cases} 0 & \text{if } P_{l,a} = 0 \\ \text{Poisson}[r_{l,a,i}\exp(\psi_{l,a,i})] & \text{if } P_{l,a} = 1 \end{cases}$$

where $\psi_{l,a,i}$ accounts for variations in densities at fine spatial scales and $\psi_{l,a,i} \sim \text{Normal}(0, \sigma_c^2)$, and σ_c^2 is the variance of overdispersion. Moreover, the expected abundance when fish were encountered is

$$r_{l,a,i} = \frac{w_i S_a \mathbf{N}_{l,a,s_i,y}}{d_{l,a,i}}$$

Therefore, the total number of individual fish caught in a station s_i and year y is $c_i = \sum_{l,a} \mathbf{c}_{l,a,i}$. For the length subsampling, the number of individuals per length bin is $\tilde{c}_{l,i} = \sum_a \mathbf{c}_{l,a,i}$, and the total number is $\tilde{c}_i = \sum_{l,a} \mathbf{c}_{l,a,i}$.

For the age subsampling, we implement a proportional design to sample ages, where the number of individuals selected for aging for each length bin is proportional to its abundance in the length subsample (Kimura 1977; Quinn and Deriso 1999). We randomly select 95% of all stations (~332) to collect otoliths and assume a maximum number of individuals sampled in a station equal to 4. Since an individual with length bin l can have more than one true age in the matrix $\mathbf{c}_{l,a,i}$, we specify that the probability of a subsampled individual x in the length bin l being age a is proportional to its abundance in the catch as

$$\Pr(x_{l,i} = a) = \frac{\mathbf{c}_{l,a,i}}{\sum_{a=0}^{A} \mathbf{c}_{l,a,i}}$$

Finally, we are not assuming any source of aging error or bias made by laboratory analyses. This approach reproduces age structure collections in the real EBS survey, and the age composition is estimated using a design-based approach (Wakabayashi et al. 1985).