Fisheries Research 276 (2024) 107039

Contents lists available at ScienceDirect

Fisheries Research

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

# Performance gains from incorporating dynamic numbers by length-within-age in fishery assessment models

Richard McGarvey [a,*], Richard D. Methot [b], André E. Punt [c], Janet M. Matthews [a], Ian G. Taylor [b], John E. Feenstra [a], Kathryn Doering [d,e]

[a] South Australian Research and Development Institute (Aquatic Sciences), Henley Beach, South Australia, Australia
[b] NOAA Fisheries, Northwest Fisheries Science Centre, Seattle, WA, USA
[c] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA
[d] Caelum Research Corporation Under Contract to NOAA Fisheries, Northwest Fisheries Science Centre, Seattle, WA, USA
[e] NOAA Fisheries, Office of Science and Technology, Silver Spring, MD, USA

## ARTICLE INFO

Handled by: Jie Cao

*Keywords:*
Dynamic length-within-age
Next-generation model
Platoons
Rosa Lee phenomenon
Slice partition

## ABSTRACT

Simulation testing is fundamental for evaluating which features to include in a next-generation stock assessment model. One feature, contemplated but rarely adopted, is to extend the standard age-only population array to estimate population numbers dynamically by both age and length-within-age. This allows the model to capture changes over time in the shape of the length-at-age distribution when mortality is asymmetric with respect to fish body length. Here we use simulated data from an individual-based fishery model to evaluate three assessment models. Two are versions of Stock Synthesis (SS3), one employing a standard formulation and a second using 'platoons' to account for dynamic numbers by length-within-age. The third model, using a 'slice partition' to model numbers by length-within-age, was designed for assessing fisheries with a knife-edge truncation of the catch at a regulated legal minimum length. Platoons produced substantially better estimates of mean length-at-age, selectivity, and biomass than SS3-age for the simulated fishery scenarios with a higher level of individual growth variation. SS3 with platoons also produced better estimates than the slice model for these high growth variation scenarios, primarily when selectivity was estimated and was not knife-edged. The three assessment models performed well, and similarly, for lower individual growth variability, except that the slice model performed modestly better under knife-edge selectivity for which it was designed. Incorporating dynamic population numbers by length-within-age can be ranked as a medium priority for a next generation assessment model. Incorporating platoons in SS3 showed a clear performance gain, with moderate improvement evident, when there is high individual variation in growth. If length-within-age is incorporated, platoons are preferable to slices for implementation flexibility: by specifying the use of only one platoon, the model reduces to a standard age-based form, leaving all other model features unchanged.

## 1. Introduction

Since the 1980s, fishery assessment models have generally represented change in fish population number by age and cohort, but not also by fish length within each age class. The distribution of fish lengths about the mean is usually assumed not to change over time, in particular under differing levels of fishing mortality. However, this ignores that faster growing fish incur higher rates of fishing mortality sooner than slower growing fish because faster growing fish reach legal or gear-selected size at a younger age. This more rapid reduction in fish numbers in the upper range of each cohort's length-at-age distribution changes its shape. This length-asymmetric mortality reduces the population's mean length and its standard deviation, leading to a bias called the Rosa Lee phenomenon in age-only models (Lee, 1920; Kraak et al., 2019.)

The Rosa Lee phenomenon can only be addressed within assessment models that dynamically represent both age and length. This can be done by representing the spread of fish by length in each cohort as a continuous distribution or partitioning each cohort into discrete length bins. The models of Deriso and Parma (1988) and Parma and Deriso

---

* Corresponding author.
*E-mail address:* rickmcg2222@gmail.com (R. McGarvey).

https://doi.org/10.1016/j.fishres.2024.107039
Received 21 September 2023; Received in revised form 1 April 2024; Accepted 25 April 2024

(1990), and those, both steady-state (Smith and Botsford, 1998) and dynamic (Banks et al., 1991; Kristensen et al., 2006), that analytically solve the von Foerster equation, model a continuous distribution of lengths at age. In practice, the von Foerster equation can be solved only for a limited range of special cases. Discrete length bins are used in assessment models that apply growth-transition matrices to each cohort (e.g., Frøysa et al., 2002; Allen Akselrud et al., 2017), which is computationally expensive. While advancements with the growth transition approach continue to be made (Zhang and Cadigan, 2022), 'growth groups' (see below) provide a more computationally efficient breakdown of a cohort by growth trajectory (Punt et al., 2002).

Punt et al. (2020) note that much of stock assessment science is focused on the use of 'integrated' methods of assessment. These methods separate the development of the hypothesized model of the population dynamics (time-dependent equations) from the likelihood sub-model that relates model predictions to the data (see Maunder and Punt, 2013 and Punt et al., 2013 for reviews of integrated model frameworks based on age-structured and size-structured population dynamics models). Punt et al. (2020) highlight several areas that need further development as part of the 'next generation' of stock assessment packages, including the need to capture age and size/stage dynamics simultaneously yet computationally efficiently. Of the major stock assessment packages used worldwide, GADGET [Globally applicable Area Disaggregated General Ecosystem Toolbox] (Begley, 2014) integrates age-length dynamics using a computationally intensive approach. Stock Synthesis (SS3; Methot and Wetzel, 2013) and Casal2 (Doonan et al., 2016) include age-length dynamics approximately using 'platoons' (SS3) or 'growth paths' (Casal2). The latter two approaches are computationally efficient (for SS3, see Taylor and Methot, 2013), but their performance has not been evaluated in detail. In this paper we evaluate two methods to model dynamic length-within-age: platoons and slice partitions.

The slice method for partitioning the continuous length-at-age distribution (McGarvey et al., 2007) was developed for fisheries whose catches by length are truncated under a regulated legal minimum length (McGarvey and Fowler, 2002). The cohort lengths-at-age are partitioned into bins (slices) that vary in width and grow with the cohort. With each time step a single new slice is defined as that portion of the length-at-age distribution that grows across the legal minimum length into harvestable size. The number of fish in each slice is accounted for as the cohort grows and animals die. The more rapid removal of faster growing fish is modelled by the right-hand (larger fish) slices reaching harvestable size sooner and incurring higher (fishing) mortality over more model time steps. Rosa Lee bias is thereby addressed because the model accounts for the more rapid removal of the faster growing fish in the population (and the length selectivity of the catch) when fitting to catch length-at-age samples. Since the early 2000s, slice partition models have been used to assess three fish stocks off South Australia, all regulated by legal minimum length, King George whiting *Sillaginodes punctata* (Fowler and McGarvey, 1999), southern garfish *Hyporhamphus melanochir* (McGarvey et al., 2007), and snapper *Pagrus auratus* (McGarvey and Feenstra, 2004).

SS3 models size-specific survivorship by representing the overall distribution of length-at-age as the sum of several overlapping length-at-age normal distributions called growth platoons (Taylor and Methot, 2013). Platoon-specific selectivity-at-age is derived from selectivity-at-length then used when applying fishing mortality, such that the larger platoons experience higher *F* at an earlier age and the Rosa Lee effect is addressed. Details of the slice and SS3 assessment models are provided in these cited references, and in Sections 2.4 and 2.5 below.

The overall goal of this study is to evaluate the relative gains of including length-within-age in next generation stock assessments. The study had two objectives: (1) to compare the estimation performance of the standard version of SS3 (here termed SS3-age) to that of an SS3 configuration that incorporates platoons, and (2) to compare the estimation performance of the SS3 platoon approach to that of the slice approach. The first objective quantifies the benefit of accounting for length-within-age for assessment purposes, and the second compares two methods of doing so.

## 2. Methods

We used an individual-based model (IBM), originally developed to assess the value of data sets that added catch in number to the usual data of catch in weight and effort (McGarvey et al., 2005), to create simulated data sets based roughly on a South Australian snapper stock. The quantities of this simulated fished population serve as the true values that the assessment models are intended to estimate. The sets of true values are fourfold: (1) the suite of input parameters to the IBM, for recruitment, growth, mortality, and length selectivity; (2) the changing numbers in the simulated population by cohort year, age and length for each monthly time step of the IBM; (3) the output data sets generated by the IBM to which the assessment models are fitted or conditioned on, catch total by weight, catch rate, and yearly random samples of the catch by age and length; and (4) management quantities of recruitment, population number, and biomass. Data sets (3) are summed over IBM months to give semester-time step resolution used by the semi-yearly assessment models. The assessment models' yearly estimates of (4) are compared to IBM true as the primary basis for evaluating assessment model performance. Fishing mortality, growth variability, and selectivity are modeler-specified to generate each of the eight scenarios used here for performance comparisons among the three assessment models.

This simulation model is not based on any of the three assessment models and so offers an independent way to evaluate biases in estimates of management quantities from these models and to compare them. It thus extends, and improves on, previous evaluations of the effects of individual variation in growth such as Punt et al. (2017) by modelling individuals rather than platoons (which would match one of the estimation methods).

Eight scenarios were created by changing the values of the parameters of the IBM, then 100 simulated data sets were created from each scenario. Each data set was analyzed by three assessment models: the standard version of SS3 using length selectivity but only one platoon (labelled as "SS3-age"), an SS3 version that incorporates five platoons ("SS3-platoons"), and an assessment model using slices ("slice model"). Performance was evaluated by calculating relative errors from the IBM "truth" for four yearly management quantities: recruitment, total fishable population numbers, fishable biomass, and fishing mortality, in addition to estimated selectivity by length and mean length-at-age.

### 2.1. Individual-based data simulator

The IBM had a monthly time-step, and covered 30 years (nominally "1990" to "2019"), spanning a plausible range of time when age-length data might be available for modern fisheries. In each monthly time step, all fish grow and may die due to either natural causes or fishing. The probability of being caught depends on logistic length selectivity and user-specified yearly instantaneous maximum *F*. There is no maximum age imposed. The population is initialized at the start of 1990 by assuming that all animals grow according to the same growth curve and mortality is constant during a 'burn-in' period of 30 or 48 years (see Section 2.2). The IBM produces simulated catch and effort totals, and samples of the catch by age and length. Annual true values for biomass and population number were taken as those at the start of the first month of each year. The specified 12 IBM values for *F* were held constant through each 12-month year. The following sections summarize key features of the IBM.

#### 2.1.1. Recruitment

A single recruitment time series with no trend or spawner-recruitment relationship was used for all simulated datasets, with a mean of approximately 10,000 age-0 individuals born each year. A single recruitment time series was used (Fig. S1) to avoid an unworkably large number of scenarios. Recruitment had four peak years, 1990, 1998, 2005, and 2009 (Fig. S1). The length at birth was sampled from a normal distribution with a mean of 1 cm and a standard deviation of 0.1. Recruitment during the pre-1990 burn-in period was fixed at 10,000 fish per year.

#### 2.1.2. *Growth*

Growth was modeled using a von Bertalanffy monthly growth increment formula, i.e.:

$$L_{t+\Delta t} = L_t + (L_\infty - L_t)\cdot\left(1 - e^{-K/12}\right) \tag{1}$$

Individual variation in growth was modeled by randomly assigning values for $K$ and $L_\infty$ to each animal at birth. These individual $K$ and $L_\infty$ values were sampled from a bivariate normal distribution (Williams and Shertzer, 2005) with mean ($\mu_K$, $\mu_{L_\infty}$) and covariance matrix defined by ($\sigma_K^2$, $\sigma_{L_\infty}^2$, $\rho$). The variances, $\sigma_K^2$ and $\sigma_{L_\infty}^2$, were computed from scenario-specific CV's, i.e., $\sigma_K = \mu_K \cdot CV_K$ and $\sigma_{L_\infty} = \mu_{L_\infty} \cdot CV_{L_\infty}$, with the covariance between $K$ and $L_\infty$ calculated as $\sigma_K \cdot \sigma_{L_\infty} \cdot \rho$. The input 'true' growth parameter values are given in Table 1.

#### 2.1.3. *Length selectivity and mortality*

Fishery selectivity was a logistic function of fish length parameterized in terms of lengths at 50% and 95% selectivity. Probabilities of death were calculated for each monthly time step given user-specified fully-selected fishing mortality $F$ and natural mortality $M$. The monthly probability of capture, $F_{\text{sel}}$, given each fish's length, was calculated as $\frac{F}{12}$ multiplied by logistic length selectivity, while the monthly probability of death by natural causes was $\frac{M}{12}$. A random number was drawn for each simulated fish to determine whether it dies, based on an overall mortality probability of $\left(\frac{F_{\text{sel}}}{12} + \frac{M}{12}\right)$ in a monthly time step. If death occurs, a second random number determined the cause of mortality with respective probabilities of $\frac{F_{sel}}{(F_{sel}+M)}$ and $\frac{M}{(F_{sel}+M)}$.

#### 2.1.4. *Data generation*

Age and length data were obtained by taking a random sample from all the captured fish in each year. The age, length, month, and year of capture were recorded from each sampled individual. The sampled fish were aggregated by half year to compute catch proportions by age and length (or for the slice model, length moment statistics) to produce data sets from each half-yearly time-step ("semester") used by the assessment models. Total catch in weight and a measure of total effort over each semester were summed from the IBM. Monthly effort was computed as catch weight divided by fishable biomass, and CPUE by semester as the total catch biomass divided by effort. No additional observation error was added beyond the demographic stochasticity in individual monthly death and growth processes because the focus of this study is on the dynamics of size-survivorship.

**Table 1**
The values for the IBM simulation parameters not varied in the scenarios.

| Parameter | Value |
|---|---|
| Recruitment | Time series in Fig. S1 |
| Natural mortality | 0.1 $\text{yr}^{-1}$ |
| Growth | |
| • Mean of $L_\infty$, $\mu_{L_\infty}$ | 100 cm |
| • Mean of $K$, $\mu_K$ | 0.1 $\text{yr}^{-1}$ |
| • $\rho$ | -0.4 |
| Weight at length | |
| • Coefficient | 3.0E-8 |
| • Power exponent | 2.86 |

### 2.2. *Scenarios*

The eight IBM scenarios (1A to 4B; Table 2) comprise all combinations of (1) two assumed yearly time series of fishing mortality ($F$), (2) two logistic length-selectivity ogives, and (3) two CV values for individual growth variability. Unlike previous studies such as that of Punt et al. (2017), all eight scenarios include individual variation in growth as this is realistic for most actual fished populations.

The two $F$ time series were (1) a time-invariant $F = 0.4$ $\text{yr}^{-1}$, and (2) a "one-way trip" with a gradual rise in $F$ over the first 10 years (nominally 1990–1999) from 0.025 to 0.25 $\text{yr}^{-1}$ in increments of 0.025 $\text{yr}^{-1}$, with $F$ from 1999 onward set to 0.25 $\text{yr}^{-1}$. As 'burn-in', the IBM was run for 30 years prior to 1990 at $F = 0.4$ $\text{yr}^{-1}$ for the constant-$F$ option, and for 48 years prior to 1990 at $F = 0$ under the one-way trip option (Table 2). The longer duration for $F = 0$ provides more information on older fish.

All length selectivity scenarios used $L_{50} = 40$ cm, but logistic steepness was varied by the choice of $L_{95}$: (1) a 'knife-edge' scenario was run to simulate the effect of a fishery managed under a legal minimum length ($L_{95} = 41$ cm), with the (2) 'less steep' ($L_{95} = 45$ cm) logistic scenario modeling selectivity for a fishing gear. Fishable biomass and fishable population numbers were defined as fish of length greater than the $L_{50}$ of 40 cm for knife-edge selectivity and greater than 30 cm for "less steep" selectivity because the size range of fish in the harvestable stock varies for the two selectivity scenarios.

Two scenario options were considered for individual growth variability, which determines the variation of $K$ and $L_\infty$ for individual fish at birth. The two growth variability options were (1) CV $= 0.1$ (A scenarios) and (2) CV $= 0.2$ (B scenarios), where $\text{CV} = CV_K = CV_{L_\infty}$.

### 2.3. *Behavior of the IBM simulator*

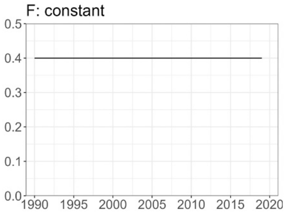
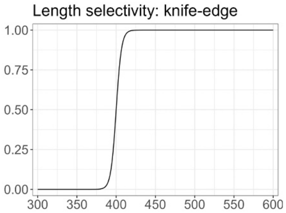
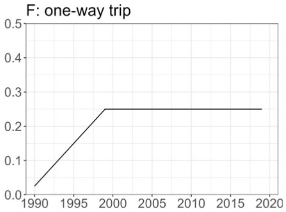
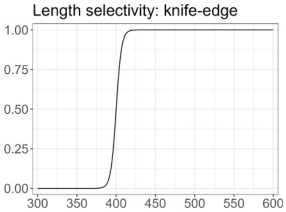
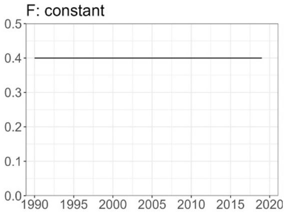
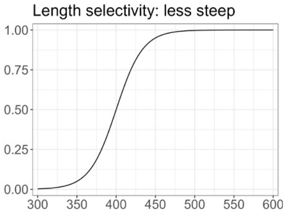
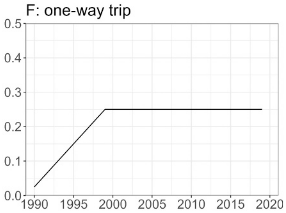
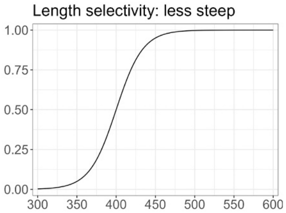
In order to better understand the data generated by IBM, we examined (1) how well effort represents trends in fishing mortality (where $F$ is assumed to vary in proportion to effort in the slice model), (2) how well CPUE represents trends in biomass (biomass is fitted to CPUE in both SS3 models), and (3) how strongly the Rosa Lee phenomenon affects the mean length-at-age and the length-at-age probability density functions (pdfs).

The IBM-computed fishing effort (catch weight divided by fishable biomass) agreed well with the specified 'true $F$' (see Fig. S2 showing one replicate for two scenarios). The monthly variation shown in Fig. S2 reflects the demographic stochasticity that arises from the individual probabilities for fishing and natural mortality. Similarly, and as expected given how effort was defined in the IBM, the IBM biomass matches the generated CPUE nearly identically (see Figs S3 and S4 for annual and semester values for two replicates and two scenarios). This low error is intended because the goal of this paper is to evaluate the impact of how growth and length-at-age are modelled.

The Rosa Lee effect reduces the apparent mean length-at-age of cohorts in the fishable-size population because fishing mortality selectively removes larger, faster growing fish. Thus, the (true) population mean length-at-age will be less than the hypothetically unexploited population mean length-at-age, the latter determined here by the von Bertalanffy growth of individual fish specified by sampled $K$ and $L_\infty$.

The magnitude of the Rosa Lee effect is explored under varying levels and time patterns of fishing mortality and different choices for the steepness of logistic selectivity by tracking the IBM monthly lengths of all fish created in a cohort of 10,000 recruits, selectivity knife-edged, for three constant levels of fully-selected fishing mortality. Mean length-at-age was computed and plotted for every twelfth month of age. The dashed lines in Figs S5 and S6 are the means of length-at-age computed directly from the true mean values of the von Bertalanffy parameters ($\mu_K = 0.1$ $\text{yr}^{-1}$ and $\mu_{L_\infty} = 100$ cm) while the colored dots plot mean length-at-age of the IBM cohort for fishing mortality rates of 0, 0.25 and

**Table 2**
Parameter values or functions specifying the eight IBM scenarios (scenario 1A, scenario 1B, scenario 2A, etc.).

| Scenario | Variation in growth parameters | Fishing mortality | Length selectivity | $L_{50}$ (cm) | $L_{95}$ (cm) | Burn in time (years) |
|---|---|---|---|---|---|---|
| 1A<br>1B | CV = 0.1<br>CV = 0.2 |  |  | 40 | 41 | 30 ($F = 0.4$ yr$^{-1}$) |
| 2A<br>2B | CV = 0.1<br>CV = 0.2 |  |  | 40 | 41 | 48 ($F = 0$) |
| 3A<br>3B | CV = 0.1<br>CV = 0.2 |  |  | 40 | 45 | 30 ($F = 0.4$ yr$^{-1}$) |
| 4A<br>4B | CV = 0.1<br>CV = 0.2 |  |  | 40 | 45 | 48 ($F = 0$) |

0.5 yr$^{-1}$. Mean lengths-at-age for zero fishing mortality (red points in Fig. S5 and S6) exactly match the curve based on the mean values for $K$ and $L_\infty$ in Fig. S5 and nearly match in Fig. S6. This confirms that using discrete monthly growth increments and bivariate randomly sampled individual parameters to model growth in the IBM results in the same outcome as the von Bertalanffy formula for mean length-at-age. This removes the way growth is modelled in the IBM as a potential extraneous source of bias. The two increasing non-zero levels of fishing mortality lead to increasing differences between the expected mean length-at-age above approximately age 7, with the effect very small for $CV_K = CV_{L_\infty} = 0.1$ (Fig. S5), but substantial for $CV_K = CV_{L_\infty} = 0.2$ (Fig. S6). A Rosa Lee leftward shift under increasing fishing mortality in the pdf of length-at-age (Fig. S7; blue or green line for $F = 0.25$ and 0.4 yr$^{-1}$ compared with the red line for $F = 0$) is likewise evident from about age 7, and becomes extreme for older fish. However, relatively few fish survive to age 7 and older for $F = 0.4$ yr$^{-1}$ (Fig. S8).

### 2.4. Assessment models

Three models were compared using two assessment platforms; two versions of one platform, SS3, and the slice model.

#### 2.4.1. Stock synthesis

SS3 is an age- and length-structured population dynamics model fitted using ADMB (Fournier et al., 2012). It is capable of estimating fishery selectivity-at-length, and time series of recruitments and fishing mortality when supplied with time series of catch, fishery CPUE, and fishery length compositions. Addition of conditional age-at-length data (e.g., age data partitioned by each length bin) enables estimation of growth parameters when other model assumptions are correctly specified (Lee et al., 2019). All recruits are assigned to a single group of fish that all experience the same fishing mortality at age in an SS3 configuration without the platoons feature enabled (hereafter “SS3-age”). The mean length-at-age of this group follows a trajectory defined by the growth parameters. In each season, the standard deviation of length-at-age distributes these fish among the population’s 2-cm length bins. The product of this distribution and length selectivity produces the mean selectivity at each age for each morph, which is used to distribute fishing mortality across the ages. Similarly, body weight-at-age is built from numbers and body weight at length for each length bin.

The SS3-platoons is an extension of the typical configuration. The total recruitment is distributed among an odd numbered set of platoons (5 in this study), rather than a single platoon. Two additional controls need to be specified to use multiple platoons.

- An input vector specifies the distribution of recruits among the platoons. The vector is typically chosen to approximate a normal distribution. Here that is set to: 0.031, 0.237, 0.464, 0.237, and 0.031.
- The second SS3 control is the ratio of the standard deviation in length-at-age for each platoon to the standard deviation among the platoon means. A high ratio creates broad overlap between the platoons and a low ratio makes them more separated. The parameters of the growth model specify the mean length-at-age for the central of 5 platoons, with the mean lengths-at-age of the other platoons

calculated as offsets from those for the central platoon. The value of the overall standard deviation of length-at-age and the assigned ratio of within-to-between platoon standard deviation is used to calculate the offset between platoons (i.e., the between-platoon standard deviation) and the standard deviation of length-at-age within each platoon. For each platoon, the distribution of length-at-age is convolved with length selectivity to calculate that platoon's selectivity-at-age.

Early work on this project used a ratio of 1.0 and found little difference between the results for SS3-age and SS3-platoons. This is because high ratios create more overlap between the platoons. Subsequently, the ratio was set at 0.40 to provide more separation between the platoons (Fig. S9) and better fits to the simulated data. In the discussion we explain how this difference in performance led to changing the ratio factor into a parameter such that the SS3 model could estimate its value when suitable data are provided. That estimation capability was not available when we conducted the current study.

### 2.4.2. *Slice model*

In the slice model, mortality varies with fish length by acting separately on subintervals of the lengths-at-age of each cohort. This model keeps track of fish number within each subinterval (each 'slice') as an added dimension (length slice within age cohort) in the model population array. The distribution of fish by length prior to the effect of length-dependent mortality is assumed to be normal, and the mean length-at-age, $L(a)$, follows a von Bertalanffy growth curve:

$$L(a) = L_{\infty} \cdot (1 - \exp(-K \cdot a)), \tag{2}$$

where $L_{\infty}$ and $K$are estimated, and age $a$ is in units of years. The standard deviation of the cohort length-at-age pdf is assumed to increase with age and is modelled as a linear or power function of model-expected mean length, i.e.

$$\sigma_L(a) = \sigma_{L0} \cdot L(a)^{\sigma_{L1}}, \tag{3}$$

where $\sigma_{L0}$ and $\sigma_{L1}$ are estimated. The number of recruits of each cohort at its age of creation (two years of age for the slice model) is an estimated parameter.

The algorithm for assigning fish numbers by slice (Appendix A of McGarvey et al., 2007) takes advantage of two features of a normal pdf: (a) it is fully specified by its mean and standard deviation, and (b) numerical subroutines based on polynomial approximations (Abramowitz and Stegun, 1970) are available in ADMB to compute the z-scores along the length axis that define the upper and lower bounds of each slice, and hence the slice probabilities. The slice model can be implemented assuming other length-at-age distributions (e.g., gamma, lognormal) but this requires numerical methods to compute slice probabilities. Thus, the use of normal z-scores greatly improves computational efficiency, particularly when growth is estimated within the assessment as is the case here. Slices grow with the cohort because the length range of fish in each slice is defined by its position relative to the normal mean. In each

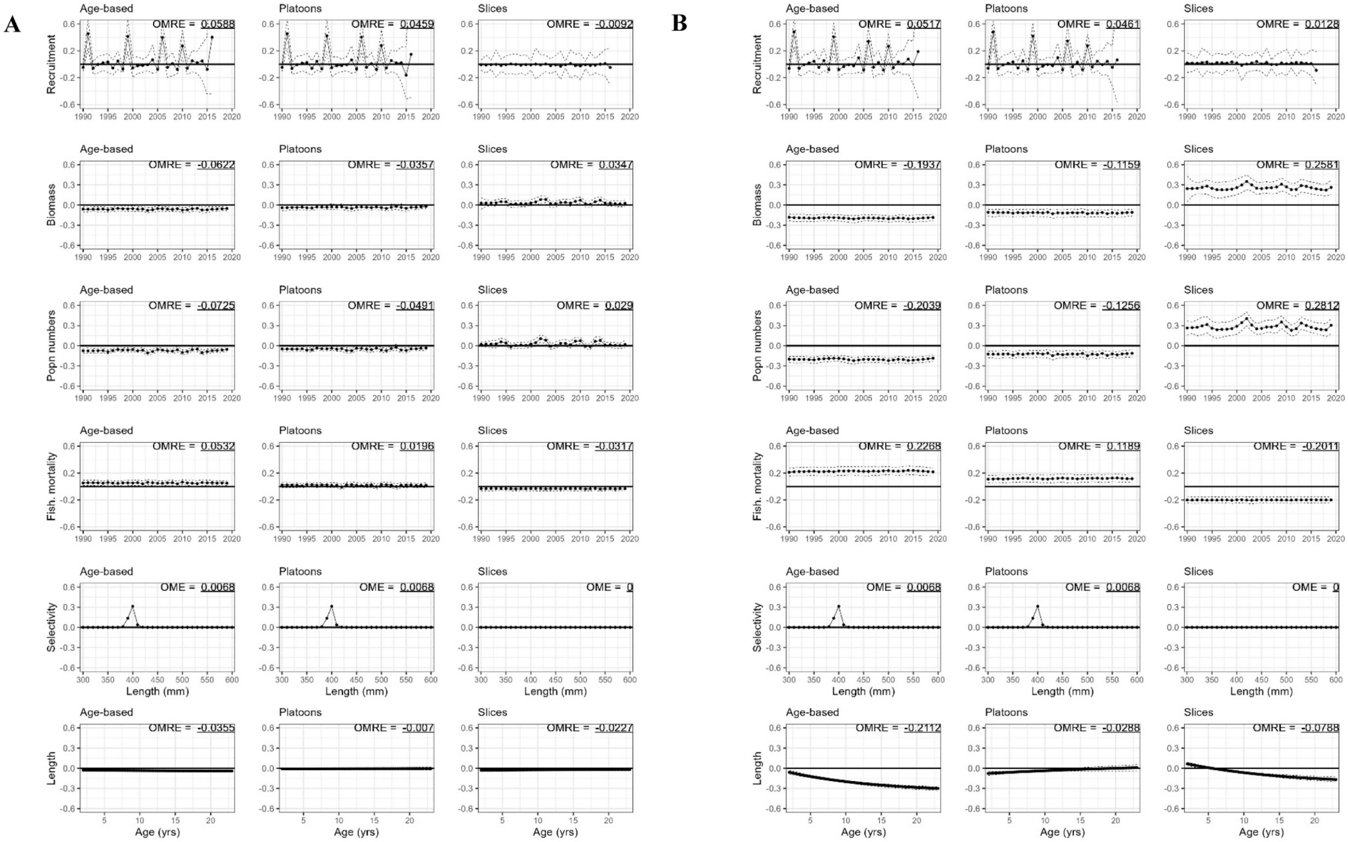


**Fig. 1.** Estimation performance for scenario 1: *F* constant through time and near-knife-edge length selectivity. Two choices for individual growth variability are presented: (A) CV = 0.1, and (B) CV = 0.2. Three assessment models are evaluated: SS3-age (Age-only, left column of panels), SS3-platoons (Platoons, middle column of panels), and slice model (Slices, right column of panels). Yearly values shown are fractional deviations from IBM-simulated true values summarized as mean relative error (points connected by solid lines) and the relative error 10% and 90% quantiles by year (dotted lines above and below the means) among 100 replicate runs. The 10% and 90% quantiles may not be visible on all plots because they are close to the means in some cases. The overall mean relative error (*OMRE*) is shown in the top right corner of each plot. For length selectivity, an actual rather than fractional deviation from IBM truth (*OEM*, overall error mean) is provided. The results for length selectivity and mean length-at-age (bottom two rows of graphs) are based on the evaluated logistic and von Bertalanffy functions respectively using IBM true parameter values and assessment model-estimated parameter values.

time step, one slice is created (and its z-score computed), as that proportion of the normal length-at-age pdf that newly grows across legal minimum length. A vector of z-scores is computed, defining the length partition into slices, as a cohort grows to each higher age, as illustrated in Figs. 1 and 2 of McGarvey et al. (2007).

The slice model was “effort-conditioned” with fully-selected fishing mortality $F$ related to monthly effort $E$ according to $F = qE$, where $q$ is an estimated catchability parameter. Total catch in weight by semester was fitted assuming a Gaussian likelihood with an assumed sigma (Table 3). Catch proportions-at-age were fitted assuming a multinomial sampling process. However, it is not convenient to fit slices to lengths-at-age data aggregated into fixed length bins (as in a multinomial likelihood) because slice bin partition lengths increase with time step as a cohort grows and also depend on the values for the growth parameters. Instead, slices are fitted to lengths-at-age following Fournier and Doonan (1987) who fitted to the absolute central moments of length-at-age. In the slice model, the ‘moment properties’ (mean, standard deviation, skewness and kurtosis) of the sampled length-at-age distribution are fitted for each age group based on a gaussian likelihood with the same variance parameter for all four moments. Moment properties are fitted in each time step for which there are sufficient numbers of animals sampled and aged for these statistics to be informative (1 aged fish for the mean, 2 for the standard deviation, 4 for skewness, 8 for kurtosis), and for which the number of model slices is also adequate (Eq. B8 of McGarvey et al., 2007). An advantage of fitting length-at-age moments rather than binned multinomial proportions is that it avoids binning error since the full precision of each length measurement is used to compute the data length moments. However, approximation is made in the model-predicted length moments by using slice midpoint lengths to compute model mean length-at-age, model standard deviation, etc.

**Table 3**
The dynamics, fitting method and estimable parameters of the two assessment model approaches.

| Parameter | SS3 | Slice |
|---|---|---|
| Yearly recruitment | Age-0; penalty on deviations from the mean | Age-2; estimated independently by year |
| Unfished recruitment ($R_0$) | Estimated | N/A |
| Growth | 5 parameters estimated: $L_0$, $L_\infty$, $K$, CV at $L_0$, and CV at $L_\infty$ | 4 parameters estimated: $L_\infty$, $K$, $\sigma_{L0}$ and $\sigma_{L1}$ |
| Fishing mortality | Calculated to match catch | Related to fishing effort |
| Selectivity | | |
| • Knife-edge | $L_{50}$ and $L_{95}$ known | $L_{50}$ and $L_{95}$ known |
| • Less steep | $L_{50}$ and $L_{95}$ estimated | $L_{50}$ and $L_{95}$ estimated |
| Conditioned on | Catch | Effort |
| Likelihood components | | |
| • Recruitment deviation penalty | Lognormal, fixed sigma = 0.7; reduced lambda | N/A |
| • CPUE | Lognormal | N/A |
| • Catch | N/A – $F$ matches Catch | Normal, fixed sigma = 500 |
| • Age-composition | N/A | Multinomial |
| • Length-composition | multinomial | N/A |
| • Length-at-age | N/A | Gaussian fit to length moments, mean, SD etc, single fixed sigma = 15 |
| • Conditional age-within-length | multinomial within 2-cm length bin | N/A |

## 2.5. *Features common to the assessment models*

The remaining features of the three assessment models were configured as similarly as possible. All models used a semester time step,

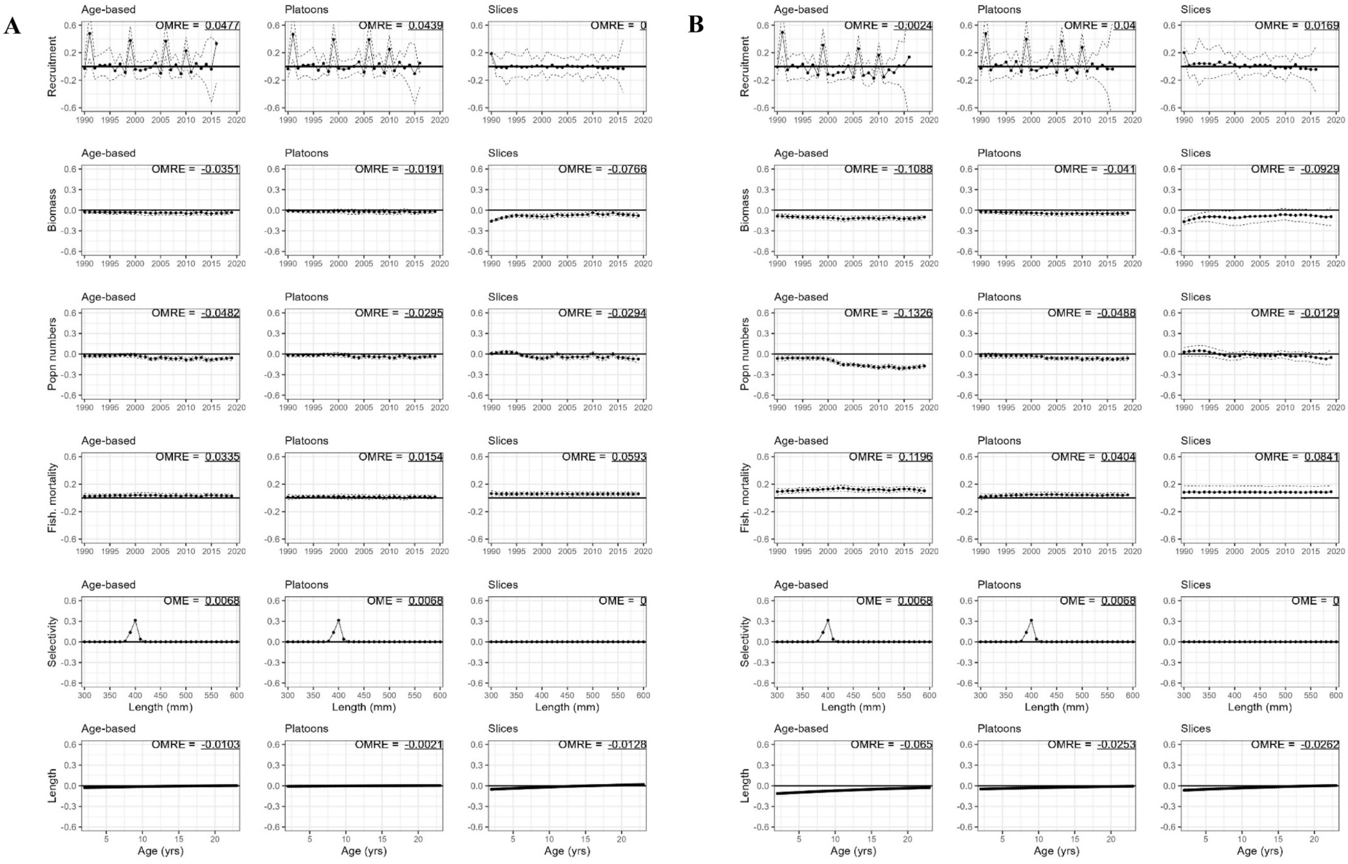


**Fig. 2.** Estimation performance for scenario 2: One-way trip $F$ and near knife-edge length selectivity. Figure description as for Fig. 1.

monthly simulated data were aggregated into semester, and $M$ was assumed to be 0.1 $\text{yr}^{-1}$. Selectivity was correctly assumed to be a logistic function of length. Growth was estimated within the models, governed by the correctly assumed von Bertalanffy equation with $t_0$ correctly known to be zero. Table 3 lists the estimable parameters of the SS3 and slice models, and some of the key differences, which are elaborated on below.

The three estimation models pre-specified the values of the selectivity parameters for the four scenarios employing knife-edge selectivity (1A–2B), which simulates a fishery where selectivity is regulated by a fixed (and therefore known) legal minimum length. The logistic selectivity parameters were estimated for the 'less-steep' scenarios (3A–4B). For the models with knife-edge selectivity, the slice model assumed the known true values for $L_{50} = 40$ cm and $L_{95} = 41$ cm. For the SS3 models, the values for the selectivity parameters were fixed to $L_{50} = 39.5$ cm and $L_{95} = 40.5$ cm as an approximation to knife-edged selectivity, although this resulted in a small offset relative to the slice estimation approach due to SS3 selectivity being evaluated for bins with width of 2 cm.

### 2.6. *Evaluating estimation performance*

The three assessment models were evaluated in terms of their ability to estimate yearly time-series of four management quantities: recruitment, total fishable population numbers, fishable biomass, and instantaneous fishing mortality, $F$. Numbers and biomass for the two SS3 models were taken from the start of the first semester, while for the slice model, an average across the first semester was taken assuming Baranov continuous decline. Annual $F$ was the average of the estimates for the first and second semesters. The yearly relative errors were calculated as $\left(\widehat{I}_y - I_y\right)/I_y$, where $I_y$ is the true (IBM-known) value and $\widehat{I}_y$ is its corresponding model-estimate. A yearly time series of each management quantity $\left\{\widehat{I}_y;\ y = 1990, \ldots, 2019\right\}$ was estimated for each combination of scenario, assessment model, and data set.

Bias by year for each management quantity was calculated as the mean of relative errors over the 100 replicate runs ($MRE_y$, plotted as the main solid line with dot markers by year in Fig.s 1–4. The $MRE_y$'s were time-averaged over all years to give the overall mean relative error ($OMRE$) to provide an overall measure of bias for each management quantity, assessment estimator and scenario. The yearly plotted $MRE_y$'s and the $OMRE$ statistics presented in Figs. 1–4 were the primary measures of model performance.

In addition to the main statistical performance measures in Figs. 1–4, we present model-estimated and IBM outputs for a single replicate in Fig. 5, the first of 100 data sets for scenario 4B. Multiple replicate comparisons are also presented in Supplementary Files S11-S13 for Scenario 1A.

The assessment model population mean length-at-age (prior to any Rosa Lee effect), $L(a)$, was calculated for each replicate using Eq. 2. Eq. 2 was also used to calculate the true population mean length-at-age, using the true mean values for $L_\infty$ and $K$ (i.e. $L_\infty = \mu_{L_\infty}$, $K = \mu_K$). Similarly, the estimated selectivity ogive for each replicate was calculated using the estimated values for $L_{50}$ and $L_{95}$. Selectivity varies between 0 and 1, so differences between estimated and true values were plotted as actual rather than relative differences. Performance in terms of estimating the value for $L_\infty$ is also presented because this parameter was expected to be particularly sensitive to size-selectivity.

## 3. Results

Overall, all three assessment models had very little bias for the scenarios with low growth variability ("A") - $OMRE$ values were less than 0.08 in absolute magnitude (Figs. 1–4). However, several models had $OMRE$ values $> 0.1$ for the high growth variability "B" scenarios. SS3-platoons, with a peak absolute value for $OMRE$ of 0.13, was

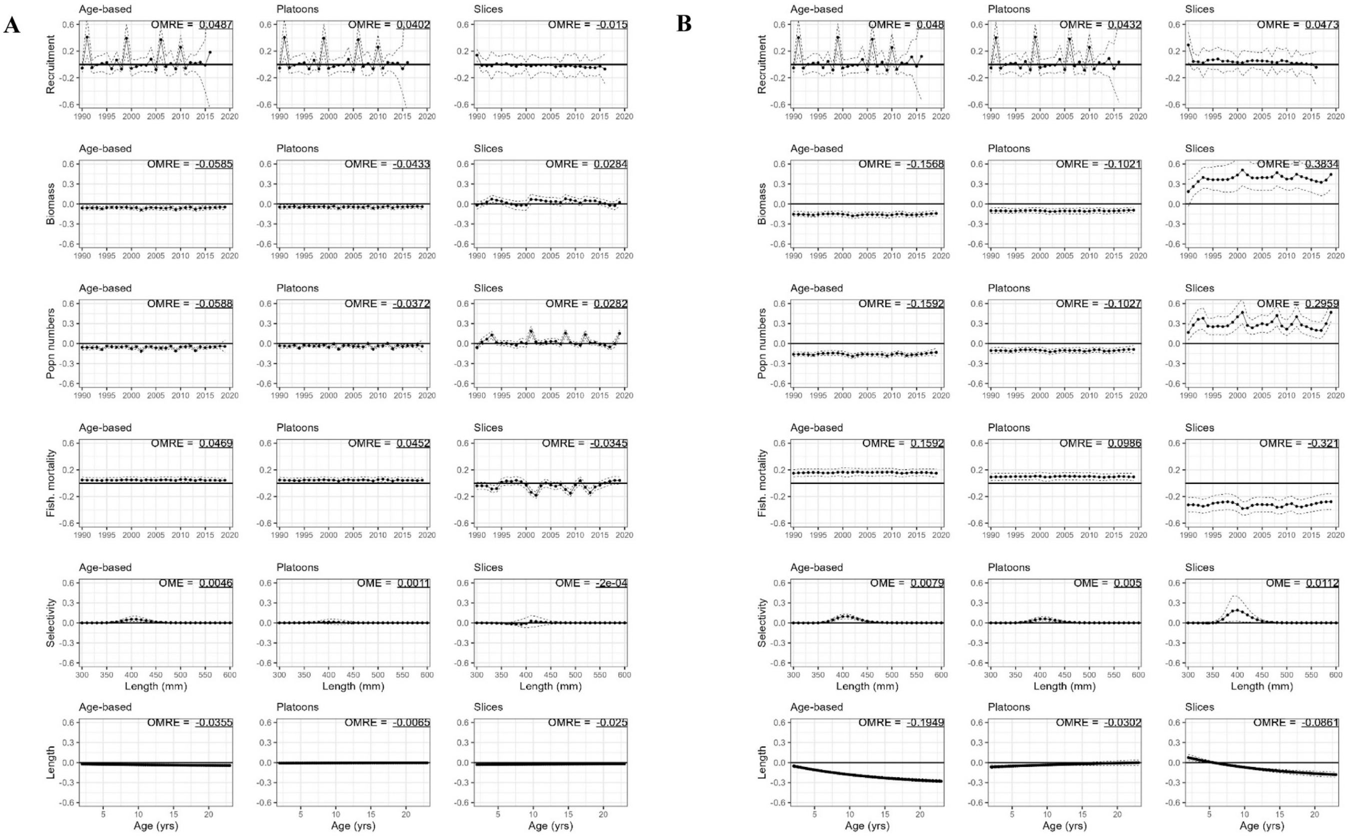


**Fig. 3.** Estimation performance for scenario 3: constant $F$ and less-steep logistic selectivity. Figure description as for Fig. 1.

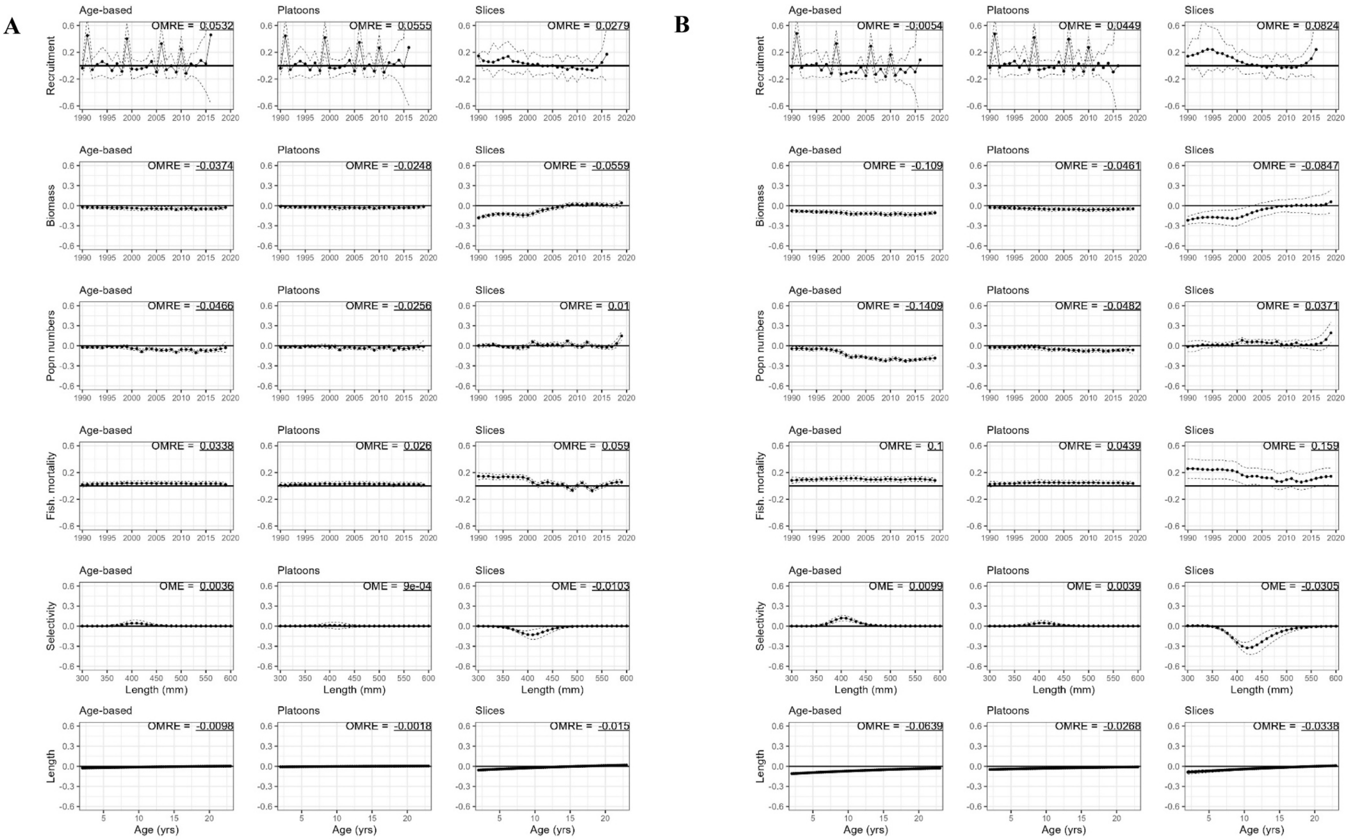


**Fig. 4.** Estimation performance for scenario 4: one-way trip for yearly *F* and less-steep logistic selectivity. Figure description as for Fig. 1.

consistently less biased at estimating biomass, numbers and *F* than SS3-age or the slice model, which had *OMRE* values as high as 0.38. The bias for SS3-age and the smaller bias for SS3-platoons was opposite in sign relative to the bias for slice for scenarios 1 and 3 (constant *F*), and had the same sign for scenarios 2 and 4 (one-way-trip). As expected, when biomass and age were estimated low, *F* was estimated high, especially in the constant *F* scenarios (1 and 3). Among high-growth variability B scenarios, SS3-platoons was more biased for the constant *F* scenarios (1 and 3) with *OMRE* $> 0.1$ for some indicators, while the one-way-trip scenarios (2 and 4) had *OMRE* $< 0.05$. The SS3-platoons bias was of the same sign as that of SS3-age and opposite in sign to that for the slice model for the constant F scenarios. Thus, the yearly pattern of *F* had a larger effect on estimation performance than the steepness of the logistic selectivity ogive.

The SS3 models had a noticeable time-dependent bias in recruitment estimates for all scenarios (Fig. 1A-4B) - recruitment numbers were consistently overestimated by SS3 for four years with low recruitment (1991, 1999, 2006, 2010) (Fig. 1A-4B - years in which low IBM recruitment followed a high year (Fig. 5 and S1). These contributed to higher *OMRE* for SS3 than the slice model for low growth variability. However, the *OMRE* for the slice model was as high or higher than that for SS3 due to low frequency patterns in deviations for scenarios with high growth variability and sloped selectivity (3B and 4B). Overall, the slice model performed well with knife-edge selectivity, notably for these scenarios with CV = 0.1 (1A and 2A). This is unsurprising because this method was specifically designed for fisheries regulated by a legal minimum length. The slice model did much less well for higher growth variation (CV = 0.2), and SS3-platoons was best, although all three models had some bias.

Selectivity estimates exhibited a distinct pattern for the SS3 models with knife-edge selectivity (scenarios 1 and 2) because of the way the values for selectivity parameters were set for this selectivity pattern (see Table 3). SS3-platoons was least biased for less-steep selectivity and the slice model most biased when there was high growth variability (3B and 4B).

Length-at-age, which is calculated from the estimated growth parameters, was most biased for SS3-age for scenarios 1B and 3B with *OMRE* near 0.21. SS3-platoons was least biased with *OMRE* near 0.03 and was unbiased for the older ages. The slice model was intermediate, with *OMRE* near 0.08 but had larger bias for the older ages with constant *F*. SS3-age and the slice model had lower bias and were nearly unbiased for the older ages for the one-way trip scenarios 2B and 4B. The difference between SS3-platoons and the slice model is consistent with SS3-platoons performing best at estimating $L_\infty$ (Fig. S10). The model estimate of $L_\infty$ from SS3-platoons was within 20 mm of the true value of 1000 mm whereas SS3-age estimates were closer to 700 mm with no overlap with 1000 mm.

While the summary statistics of Figs. 1–4 reveal moderate bias for some combinations of scenario, assessment model and indicator, the plotted yearly estimates and true IBM values shown in Fig. 5 and S11-S13 indicate that the assessment models capture the general trends and absolute levels over time and versus length or age.

### 3.1. Scenarios 1A and 1B: constant F and near-knife-edge selectivity

The outcomes for lower and higher individual growth CVs differ substantially, with generally much greater bias evident for higher individual variation in growth (Figs. 1A, 1B). There is good agreement with true of mean length-at-age for the lower individual growth CVs (Fig. 1A).

Recruitment outcomes were similar for the higher CV data sets (Fig. 1B), although the SS3 estimates at the end of the time series are less precise because SS3 defines recruitment at age 0 so is estimating two more recruitments than the slice model, which defines recruits at age 2.

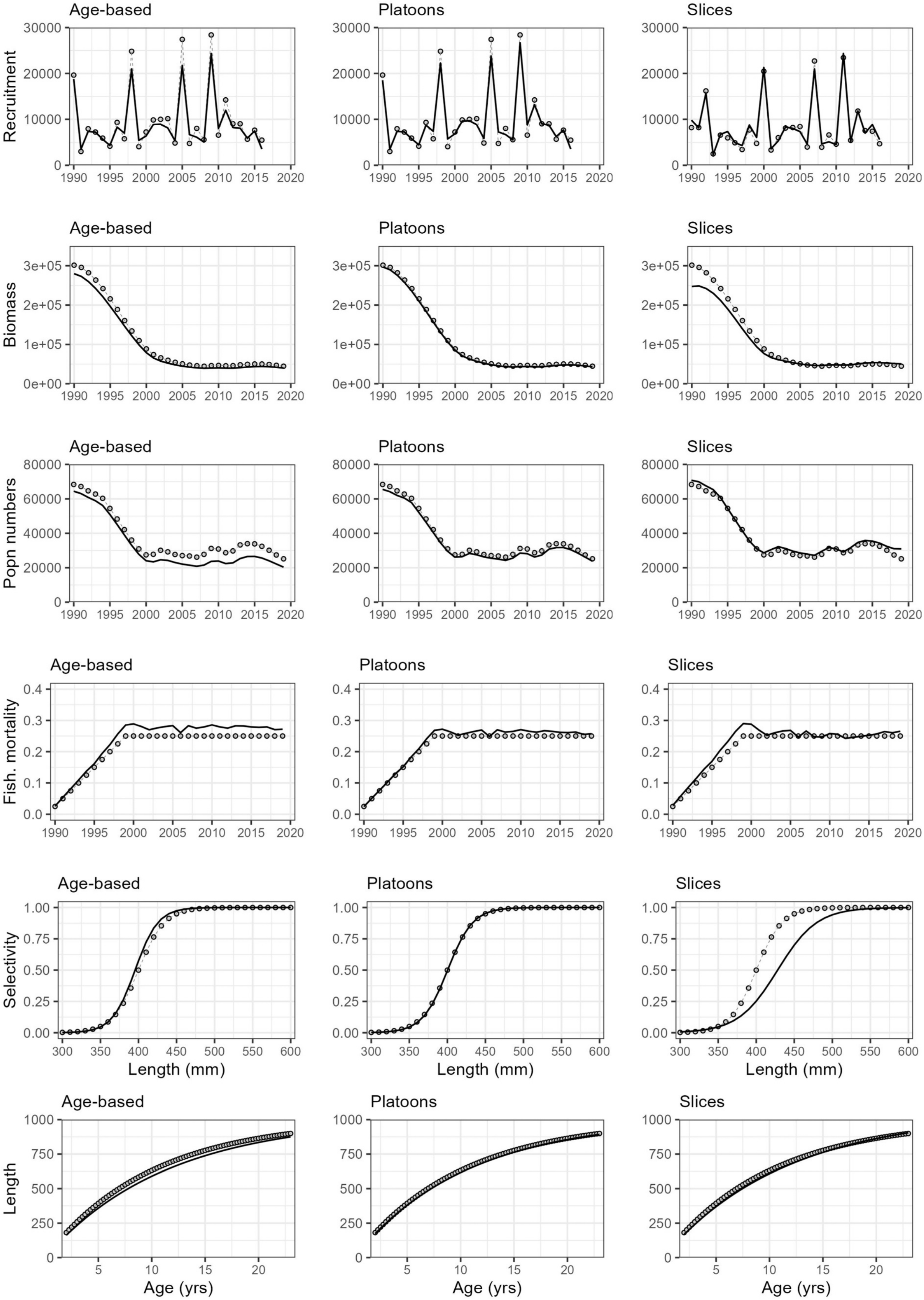


**Fig. 5.** Assessment results for a single replicate comparing IBM true (circle markers and dashed line) and estimated (solid line) values. This figure shows the first of 100 replicate data sets for scenario 4B, with *F* following a one-way trip, and the ‘less steep’ logistic selectivity. Note that slice model considers recruitment at age 2 and SS defines recruitment at age 0.

The three models provide biased estimates of biomass for the high-CV scenario (Fig. 1B), with the two SS3 models under-estimating (Fig. 1A) and the slice model leading to more substantial over-estimation (Fig. 1B). Mean length-at-age is less well estimated by the three models for CV = 0.2, although SS3-platoons produced nearly unbiased values, substantially outperforming SS3-age. This may also explain the notable improvement in the ability to estimate biomass and population numbers by SS3-platoons.

### 3.2. *Scenarios 2A and 2B: one-way trip and knife-edge selectivity*

Unlike scenario 1, a large reduction in performance was not evident for higher growth variability (Fig. 2B). The three models performed well for the low CV (Fig. 2A) scenario. For scenario 2B, SS3-platoons led to substantially less biased *F* estimates than SS3-age (Fig. 2B), while the slice model had a slight positive bias for *F*. SS3-platoons also improved estimation performance for biomass and population numbers (Fig. 2B). The slice model under high CV (Fig. 2B) showed modest under-estimation of biomass and over-estimation of *F*.

### 3.3. *Scenarios 3A and 3B: constant F and less-steep selectivity*

Overall, the results for scenario 3 were similar to those for scenario 1. For the low growth CV case (Fig. 3A), the slice model again performed better overall, primarily showing more accurate recruitment, as in scenario 1. Despite yearly variability in the deviations from true of SS3 recruitment, biomass and population numbers were close to unbiased for all three models, with the SS3 models showing less variation in relative error over time. Likewise, with low growth variability, all three models provided excellent estimates of selectivity and mean length-at-age, especially SS3-platoons. For the high growth CV scenario (Fig. 3B), all three models had appreciable bias in biomass, population numbers and fishing mortality. Moderate reduction in these biases was achieved by SS3-platoons relative to SS3-age.

### 3.4. *Scenarios 4A and 4B: one-way trip and less-steep selectivity*

The relative errors for the two SS3 models are similar in patterns to scenario 2, especially for biomass, population numbers, and fishing mortality. Biomass and population numbers from SS3 are nearly unbiased for the low-CV scenario (Fig. 4A), while the slice model was somewhat biased for recruitment, biomass and fishing mortality, notably during the early years when *F* was increasing.

## 4. Discussion

### 4.1. *Overall patterns in results – ability to estimate management quantities*

The bias in estimates of management quantities differed among scenarios and assessment models. The general trend was for the two SS3 models to overestimate *F* and underestimate biomass and population number, while the reverse trend was evident for the slice model.

The most general result is that higher individual variation in growth leads to substantially greater amounts of bias in estimates of biomass and fishing mortality. This reflects, at least in part, the focus of this simulation study on individual growth variation and the distributions of length-at-age, where other sources of variation and error were omitted, with the true value for natural mortality known, CPUE tracking biomass (effort tracking *F*) closely, and the functions for growth and selectivity known *a priori*. While the random variation among fish in their individual $K$ and $L_\infty$ values was substantial, Figs S7 and S8 show that the IBM length-at-age pdfs and histograms are approximately normal, suggesting the IBM is not producing highly unrealistic growth and lengths-at-age. Thus, even with only length-dependent processes varying in our simulated fishery and estimated by the assessment models, meaningful biases worth addressing were observed. This supports the case for more fully (dynamically) accounting for length-dependent mortality and growth processes in assessment modelling.

The second general result, observed only for the high-CV scenarios, is that SS3-platoons improved estimation performance substantially compared to SS3-age and the slice model. Biomass and population numbers were more accurately estimated when the assessment model included platoons. This is due at least in part to improved estimates of mean length-at-age under high individual growth variation. It is not clear why the biases for biomass and population numbers are negative for the two SS3 models while those for the slice model are positive.

The slice model performed better under the fishery management regime for which the slice model was created, of a regulated legal minimum length (scenarios 1 and 2), simulated here by setting selectivity to near knife-edge with *a priori* known parameters. However, these benefits were modest under low growth CV, with all three models performing relatively well. With high individual growth variability, slices showed greater bias when selectivity was estimated (scenarios 3B and 4B).

Finally, there is relatively little inter-replicate variation (Figs. 1–4; S11-S13), which can be attributed to the relatively low demographic variability, the lack of environmental stochasticity other than recruitment variability, and the fact that the simulated data from the IBM have relatively low sampling variability.

### 4.2. *Overall patterns in results – ability to estimate population parameters*

#### 4.2.1. *Length selectivity*

Length-selectivity was not estimated for the knife-edged scenarios (1 and 2), with the selectivity parameters set to their true values (for the slice model) or nearly true values (for the SS3 models) to represent fisheries where selectivity is regulated by a legal minimum length. The SS3 models outperformed the slice model in estimating length selectivity for the less-steep selectivity scenarios (3B and 4B). This could be due to higher length resolution achieved by the thinner 2-cm bin width of the SS3 models. The bin width of each slice in the slice model is inversely related to the number of annual time steps, with the result that these bins were relatively wide given a semester time step, and so presented a relatively coarse resolution for numbers versus length. It may also be that the moment method for fitting to lengths-at-age used by the slice model is not as accurate in resolving changes in the observed catch length samples as the cohort passes through the $L_{50}$ when selectivity is acting strongly.

#### 4.2.2. *Population mean length-at-age*

The ability to estimate fish growth, quantified by population mean length-at-age is important given the question of whether to include dynamic length-within-age in a next generation assessment package.

The slice model generally performed well at estimating population mean length-at-age, except for scenarios 1 and 3 under high variation in growth where the biases were similar in direction but smaller than for the SS3-age model. Greater individual variation in growth led to less reliable estimates of population mean length-at-age.

SS3 was able to nearly alleviate the negative bias in estimates of mean length-at-age using platoons. Improved growth estimation with platoons, achieved by accounting for dynamic population numbers by length within each age cohort, was expected insofar as this addresses the Rosa Lee phenomenon. This improvement with platoons was most noticeable for the scenarios with higher growth variability. This is also expected because a wider spread of lengths means more time steps over which fishing mortality acts more strongly on faster growing fish.

We conjecture that there is a strong interaction between error in estimating selectivity and the estimates of mean length-at-age. For example, the growth estimates will mistakenly be adjusted if $L_{50}$ is underestimated, such that there are more fish of given age in the catch lengths between estimated $L_{50}$ and the higher true $L_{50}$. This interaction is most noticeable in the results for scenario 4B for the slice model where the biases in management quantities are greater for the slice model showing relatively large errors in estimated selectivity, specifically for the high growth variation scenarios (Figs. 3B and 4B). Future work might seek to determine the *F* scenarios that lead to better separation of selectivity from recruitment patterns.

The ages at which underestimation of true mean length occurred differed between constant-*F* and one-way trip scenarios. This bias was greatest for older fish for the constant-*F* scenarios (Figs. 1B and 3B, mainly evident for SS3-age and the slice model) while bias was primarily evident at younger ages for the one-way trip *F* scenario (Figs. 2B and 4B). We hypothesize that data from the one-way-trip scenarios 2 and 4 contain more large fish early in the time series when *F* starts off near zero, and more older fish over all years due to this one-way trip leveling at a lower long-term level of *F* (0.25, versus 0.4 for Scenarios 1 and 3), so yielding more accurate estimates of the mean length of older fish. Testing this hypothesis could be explored in future work.

Overall, platoons showed better performance, at least in part owing to more accurate mean length-at-age, which provides correction for Rosa Lee bias.

#### 4.2.3. *Effects of the choice of simulator*

The IBM simulator was not based on any of three assessment models being evaluated because one goal of the paper was to compare alternative assessment models. Simulated data are sometimes generated from one of the models being evaluated in studies of assessment model performance. For example, SS3 has been extensively tested using simulated data when the operating model is SS3 (e.g., Anderson et al., 2014), as well as with independent simulators (e.g., Li et al., 2021). Using the same model framework to generate simulated data may tend to indicate better estimation performance than when the simulator and assessment model are structurally different.

Some of the results in this study may be caused by the way individual variation in growth was simulated in the IBM (i.e., bivariate normal for $K$ and $L_\infty$, with a correlation between $K$ and $L_\infty$ of $-0.4$). Other ways to represent growth variation may lead to different estimation outcomes.

### 4.3. *Which approach to modelling length-within-age?*

The slice model generally outperformed SS3 in terms of estimating yearly recruitment, with SS3 over-estimating recruitment in four years (Figs. 1–4). This result was initially thought to be due to the penalty imposed on the SS3 deviations in recruitment about the expected value, which tends to reduce the estimates of large recruitments and *vice versa*. In contrast to the slice model, SS3 implemented a recruitment-deviation penalty, which Methot and Taylor (2011) showed reduced excessive variation in estimated recruitment for years prior to commencement of age sampling. However, trial estimations with the penalty effectively eliminated did not alleviate the recruitment estimation bias.

It also is possible that the difference is due to the slice model's use of moments for length-at-age as a primary data source whereas SS3 fits to overall length distribution and the conditional age-at-length data. A preliminary comparison between the distribution of length-at-age calculated using SS3 with that produced by the IBM suggested a broader distribution in SS3 such that the length distribution of a large recruitment would "bleed" into the length range of adjacent small recruitments. This would tend to make SS3 balance underestimation of large recruitments against overestimation of the adjacent weak recruitments whereas the moment approach in the slice model was less susceptible to the effect. This difference warrants future investigation.

SS3-platoons provided more accurate estimates of length-at-age, which interacted with estimated selectivity to produce more accurate time series of biomass, numbers and *F*. The differences were small for the low growth variability scenarios, but the SS3-platoons outperformed the others for high growth variability.

### 4.4. *Caveats and future work*

The bias due to the Rosa Lee phenomenon was relatively small, with estimated mean lengths-at-age differing little between the two SS3 models, suggesting that standard models may be able to provide sufficient accuracy for assigning probable ages given a fish's measured length. However, the impact on estimated $L_\infty$ was very strong, with the platoon approach leading to much more robust estimation of this parameter (Fig. S10). This is important because it impacts the calculated length-at-age in an unfished population which is used for reference points. This should be explored further, particularly for stocks for which data on age-composition are generally lacking (e.g., tunas).

The method of fitting to lengths-at-age can potentially be as important to the accuracy of an assessment model as adding dynamic length-within-age. The models tested here used two distinct approaches: the moment property method used in the slice model, and a fixed-bin multinomial method in the SS3 models. The moment property method is closer in approach to a minimum statistic estimator than a fixed-bin multinomial, but is less flexible, and cannot work, for example, if the length-at-age distribution of individual cohorts is, for example, bimodal. Nevertheless, maximum likelihood estimators lead to highest precision when a sufficient statistic can be identified and used as the basis of estimation, as Chapman (1961) showed for tag time-at-large mortality estimation (reviewed by McGarvey, 2009), invoking the statistical theory of complete minimum sufficient statistics (Lehmann and Scheffé, 1950; Lehmann, 1983, p. 80). This suggests that fitting to fewer, but more directly informative, statistics can yield more precise estimates when fitting to raw lengths-at-age. Most of the length-at-age distribution information is conveyed by the mean and standard deviation, with skewness and kurtosis quantifying deviation from a normal shape. If lengths-at-age by cohort are exactly normal, then estimating that distribution's parameters using the mean and standard deviation provides a minimum variance estimator. If cohort lengths-at-age are roughly normal, the moment approach might still afford a more efficient estimator compared to the typically larger number of statistics fitted using a multinomial, namely the number of relative proportions by length bin. This could be investigated in future work. How use of a moment estimator might address another ongoing limitation of multinomial fits to lengths-at-age, namely estimating the sampling error variance (which has led some authors to adopt a Dirichlet distribution), also remains an additional topic of investigation. We also speculate that this fundamental difference in estimation approach may be the source of SS3's inaccurate estimation of weak recruitment following a strong recruitment, but investigation was beyond the scope of this paper.

Extensions of these eight fishery scenarios could be of interest in future simulation study. The IBM simulator here assumed standard forms for selectivity and growth, with fixed natural mortality. But natural mortality has been observed to decrease inversely with body length (Lorenzen et al., 2022) and so decline also with age (Lorenzen, 2022). Domed-shaped selectivity was investigated when platoons were first proposed (Taylor and Methot, 2013). Selectivity can also vary seasonally, which could be investigated by allowing IBM parameters of selectivity to vary by semester time step. These realistic alterations of the simulated fishery can be implemented to evaluate their impact on incorporating dynamic length-within-age on assessment performance. Another challenging scenario for estimating mean length-at-age occurs with fisheries where $L_{50}$ is not much below $L_\infty$ since most cohort growth would have occurred prior to reaching harvestable size and thus prior to the ages when most age-length data is collected in catch samples. Our assumption of a negative correlation ($\rho = -0.4$) between $K$ and $L_\infty$ could likewise be changed to a zero or positive correlation, where Vincenzi et al. (2014) observed a positive correlation among individual trout in streams.

The conclusion of this paper that the use of SS3-platoons improves estimation performance substantially compared to SS3-age when there is considerable variability in growth matches that of Punt et al. (2017).

Punt et al. (2017) noted that assessments based on size-structured or age-size-structured population dynamics models may perform even better than SS3-platoons when there is individual variation in growth, and showed this based on the simulated data sets they used. Future work should therefore explore whether size- or age-size-structured assessment methods outperform SS3-platoons for the scenarios developed for this paper.

### 4.5. *Sensitivity to model estimation settings*

In building the assessment models, we found that performances were quite sensitive to relatively subtle changes in the control parameters and other fixed inputs.

For the slice model, several modifications to the simulated fishery and slice model were made to identify potential causes of the observed biases: (1) the slice model was fitted to catch-in-number as well as catch-in-weight because together these data sets provide information on the mean weight of the catch, but no meaningful reduction in bias was observed. Catch-in-number total, when divided through the catch-in-weight total, gives the exact mean weight of a fish in the catch. This result may imply that age samples contain sufficiently more information about population structure than a single measure of mean body weight, that adding the additional information of catch-in-number is unnecessary. (2) The slice estimator time step width was reduced to quarterly and monthly, but no appreciable gain in performance was observed. In the future, further testing could be undertaken to search for the source of these biases. If identified, the changes needed to rectify the biases can be applied.

Improvements in model performance for the SS3 models were achieved by several slow iterations of modifications to the model specifications for data binning, selectivity and recruitment variability. In hindsight, a general challenge with a simulation study having good separation between the data generation team and the estimation team is that the estimation team may not start with as thorough of a data investigation and sensitivity studies as might occur for a real assessment.

One challenge in applying platoons lies in selecting the ratio that controls the standard deviation of length within each of the platoons and their relative separation distances (Taylor and Methot, 2013). This ratio was initially set to 1.0 and resulted in little difference between the results for SS3-age and SS3-platoons. Analyses (not shown here) demonstrated that the ratio had a substantial effect on goodness-of-fit of the model to the data and guided us to use a value of 0.4 for the results presented here. The model constant representing the ratio has been changed by the SS3 team into a model parameter such that future investigations will have more flexibility of analysis.

### 4.6. *Conclusion*

None of the three assessment models led to unbiased estimates of all management quantities, and the estimation performance varied among scenarios, with the SS3 models generally estimating biomass most accurately and the slice model estimating recruitment most accurately. Including platoons substantially improved estimation performance for biomass and population numbers, and yielded much better estimates of population mean length-at-age, presumably by resolving the Rosa Lee bias. Even the platoon model still had overall bias in biomass estimation for the constant-$F$ scenario. This is perhaps not surprising given the low contrast a constant-$F$ (and biomass) scenario provides regarding the absolute size of the population. Overall, given the extent of improvement in assessment performance by the platoons SS3 model compared with the traditional age-based SS3, including dynamic length-within-age can be recommended as a feature worth including in the next-generation stock assessment package. This improvement was evident primarily for the high-growth-variability scenarios. Bias was reduced using platoons for biomass, population number and fishing mortality, which likely reflects, at least in part, the better (Scenarios 2B and 4B) or much better (Scenarios 1B and 3B) estimates of mean length-at-age using platoons. The improvements in $L_\infty$ estimation are striking.

We recommend platoons over slices if dynamic length-within-age were implemented in a next-generation stock assessment package. One disadvantage of slices is that the length resolution, specifically the average width of slice length bins, is bound to the number of time steps per year. A long time step means fewer and wider slices because a new slice is created with each model time step, which can result in excessively wide bins for a yearly model, or even, as here, with a half-yearly time step. A second practical advantage of using platoons is that it can be simply and reliably reduced to a standard SS3 assessment model. Within that model platform, platoons work even with other forms of model complexity like time-varying growth. This makes platoons the logical choice for application in a next-generation stock assessment package. With platoons, users can be given the choice of whether or not to adopt a dynamic length-within-age formulation. By simply choosing to use one platoon, the model reduces naturally to the standard formulation where the cohort length-at-age distributions are always normal.

## CRediT authorship contribution statement

**Richard McGarvey**: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Richard D. Methot**: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Andre E. Punt**: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Janet M. Matthews**: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Ian G. Taylor**: Software, Methodology, Investigation, Formal analysis. **John E. Feenstra**: Software, Methodology, Investigation, Formal analysis. **Kathryn Doering**: Investigation, Formal analysis.

## Declaration of Competing Interest

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

## Data Availability

Data will be made available on request.

## Acknowledgements

This publication was partially funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CICOES) under NOAA Cooperative Agreement NA20OAR4320271, Contribution No. 2024-1373.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fishres.2024.107039.

## References

Abramowitz, M., Stegun, I.A., 1970. Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables. National Bureau of Standards, Washington DC.

Allen Akselrud, C., Punt, A.E., Cronin-Fine, L., 2017. Exploring model structure uncertainty using a general stock assessment framework: the case of Pacific cod in the Eastern Bering Sea. Fish. Res. 193, 104–120.

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., 2014. ss3sim: an R package for fisheries stock assessment simulation with Stock Synthesis. PLoS One 9, e92725.

Banks, H.T., Botsford, L.W., Kappel, F., Wang, C., 1991. Estimation of growth and survival in size-structured cohort data: an application to larval striped bass (*Morone saxatilis*). J. Math. Biol. 30, 125–150.

Begley, J., 2014. Gadget User Guide. Available at ⟨http://www.hafro.is/gadget/userguide/userguide.html⟩.

Chapman, D.G., 1961. Statistical problems in dynamics of exploited fisheries populations. Proc. 4th Berkeley Symp. 4, 153–168.

Deriso, R.B., Parma, A.M., 1988. Dynamics of age and size for a stochastic population model. Can. J. Fish. Aquat. Sci. 45, 1054–1068.

Doonan, I., Large, K., Dunn, A., Rasmussen, S., Marsh, C., Mormede, S., 2016. Casal2: new Zealand's integrated population modelling tool. Fish. Res. 183, 498–505.

Fournier, D.A., Doonan, I.J., 1987. A length-based stock assessment method utilizing a generalized delay-difference model. Can. J. Fish. Aquat. Sci. 44, 422–437.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233–249.

Fowler, A.J., McGarvey, R., 1999. Development of an integrated fisheries management model for King George whiting (*Sillaginodes punctata*) in South Australia. Fisheries Research and Development Corporation of Australia Final Report No. 95/008. 232 pp.

Frøysa, K.G., Bogstad, B., Skagen, D.W., 2002. Fleksibest—an age-length structured stock assessment model. Fish. Res. 55, 87–101.

Kraak, S.B.M., Haase, S., Minto, C., Santos, J., 2019. The Rosa Lee phenomenon and its consequences for fisheries advice on changes in fishing mortality or gear selectivity. ICES J. Mar. Sci. 76, 2179–2192.

Kristensen, K., Lewy, P., Beyer, J.E., 2006. How to validate a length-based model of single species fish stock dynamics. Can. J. Fish. Aquat. Sci. 63, 2531–2542.

Lee, R.M., 1920. A review of methods of age and growth determination in fishes by mean of scales. Board Agric. Fish. Fish. Invest. II (4), 32.

Lee, H.-H., Piner, K.R., Taylor, I.G., Kitakado, T., 2019. On the use of conditional age at length data as a likelihood component in integrated population dynamics models. Fish. Res. 219, 204–211.

Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.

Lehmann, E.L., Scheffé, H., 1950. Completeness, similar regions, and unbiased estimation—Part 1. Sankhya 10, 305–340.

Li, B., Shertzer, K.W., Lynch, P.D., Ianelli, J.N., Legault, C.M., Williams, E.H., Methot, R.D., Brooks, E.N., Deroba, J.J., Berger, A.M., Sagarese, S.R., Brodziak, J.K.T., Taylor, I.G., Karp, M.A., Wetzel, C.R., Supernaw, M., 2021. A comparison of 4 primary age-structured stock assessment models used in the United States. Fish. Bull. 119, 149–167.

Lorenzen, K., 2022. Size-and age-dependent natural mortality in fish populations: Biology, models, implications, and a generalized length-inverse mortality paradigm. Fish. Res. 255, 106454.

Lorenzen, K., Camp, E.V., Garlock, T.M., 2022. Natural mortality and body size in fish populations. Fish. Res. 252, 106327.

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74.

McGarvey, R., 2009. Methods of estimating mortality and movement rates from single-tag recovery data that are unbiased by tag non-reporting. Rev. Fish. Sci. 17, 291–304.

McGarvey, R., Feenstra, J.E., Ye, Q., 2007. Modeling fish numbers dynamically by age and length: partitioning cohorts into 'slices. Can. J. Fish. Aquat. Sci. 64, 1157–1173.

McGarvey, R., Feenstra, J.E. 2004. Stock assessment models with graphical user interfaces for key South Australian marine finfish stocks. Final report for Australian Fisheries Research and Development Corporation Project No. 1999/145. 176 pp. ISBN: 0730853039.

McGarvey, R., Fowler, A.J., 2002. Seasonal growth of King George whiting (*Sillaginodes punctata*) from length-at-age samples truncated by legal minimum size. Fish. Bull. 100, 545–558.

McGarvey, R., Punt, A.E., Matthews, J.M., 2005. Assessing the information content of catch-in-numbers: a simulation comparison of catch and effort data sets. In: Kruse, G. H., Gallucci, V.F., Hay, D.E., Perry, R.I., Peterman, R.M., Shirley, T.C., Spencer, P.D., Wilson, B., Woodby, D. (Eds.), Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant College Program, University of Alaska, Fairbanks, pp. 635–653.

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68, 1744–1760.

Methot Jr, R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86–99.

Parma, A.M., Deriso, R.B., 1990. Dynamics of age and size composition in a population subject to size-selective mortality: effects of phenotypic variability in growth. Can. J. Fish. Aquat. Sci. 47, 274–289.

Punt, A.E., Allen Akselrud, C., Cronin-Fine, L., 2017. The effects of applying mis-specified age- and size-structured models. Fish. Res. 188, 58–73.

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

Punt, A.E., Huang, T.-C., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70, 16–33.

Punt, A.E., Smith, D.M., Cui, G., 2002. Evaluation of management tools for Australia's South East Fishery 1. Modelling the South East Fishery taking account of technical interactions. Mar. Freshw. Res. 53, 615–629.

Smith, B.D., Botsford, L.W., 1998. Interpretation of growth, mortality, and recruitment patterns in size-at-age, growth increment, and size frequency data. Can. Spec. Pub. Fish. Aquat. Sci. 125, 125–139.

Taylor, I.G., Methot Jr, R.D., 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. 142, 75–85.

Vincenzi, S., Mangel, M., Crivelli, A.J., Munch, S., Skaug, H.J., 2014. Determining individual variation in growth and its implication for life-history and population processes using the Empirical Bayes method. PLoS Comput. Biol. 10, e1003828.

Williams, E.H., Shertzer, K.W., 2005. Effects of fishing on growth traits: a simulation analysis. Fish. Bull. 103, 392–404.

Zhang, F., Cadigan, N.G., 2022. An age-and length-structured statistical catch-at-length model for hard-to-age fisheries stocks. Fish Fish. 23, 1121–1135.