Original Article

There is no best method for constructing size-transition matrices for size-structured stock assessments

Lee Cronin-Fine ⬤ [1]* and André E. Punt[2]

[1]Quantitative Ecology and Resource Management, University of Washington, Ocean Teaching Building, Suite 300, Seattle, WA, Box 357941 98195, USA
[2]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, Box 355020 98195, USA

*Corresponding author: tel: 917-880-9738; e-mail: lcroninfine@gmail.com.

Stock assessment methods for many invertebrate stocks, including crab stocks in the Bering Sea of Alaska, rely on size-structured population dynamics models. A key component of these models is the size-transition matrix, which specifies the probability of growing from one size-class to another after a certain period of time. Size-transition matrices can be defined using three parameters, the growth rate (k), asymptotic size (L_∞), and variability in the size increment. Most assessments use mark-recapture data to estimate these parameters and assume that all individuals follow the same growth curve, but this can lead to biased estimates of growth parameters. We compared three approaches: the traditional approach, the platoon method, and a numerical integration method that allows k, L_∞, or both to vary among individuals, under a variety of scenarios using simulated data based on golden king crabs (*Lithodes aequispinus*) in the Aleutian Islands region of Alaska. No estimation method performed best for all scenarios. The number of size-classes in the size-transition matrix and how the data are generated heavily dictate performance. However, we recommend the numerical integration method that allows L_∞ to vary among individuals and smaller size-class widths.

Keywords: numerical integration, platoon method, size-structured model, size-transition matrix

Introduction

Crustaceans are a highly valuable fisheries resource. For example, the crab fisheries in the Bering Sea and Aleutian Islands region of Alaska produced an estimated gross ex-vessel revenue of $259.3 million in 2016 (Garber-Yonts and Lee, 2018). Crab fisheries, including those in the Bering Sea, use stock assessment models to support management. These models estimate the size and trend of a population as well as reference points, status relative to reference points, and provide information needed to apply harvest control rules. Finfish assessments typically use some version of an age-structured population dynamics model, which rely on age-composition data. However, the moulting of crustacean exoskeletons on an irregular basis removes potential indicators of age, making it difficult to obtain such data.

Stock assessment of hard-to-age species can be based on size-structured population dynamics models. These models track groups within the population by size instead of age, which allow population dynamic processes, such as mortality, to vary with size. A wide variety of data types are used to estimation parameters (e.g. size-composition of the annual catches and tag-recapture data; Punt *et al.*, 2013). Many fisheries use size-structured stock assessment methods for assessment [e.g. crab stocks in the Bering Sea and Aleutian Islands region of Alaska, American lobster (*Homarus americanus*) in the Atlantic, rainbow abalone (*Haliotis iris*) off New Zealand, and tiger and endeavour prawns off Australia; ASMFC, 2015; Buckworth *et al.*, 2015; Szuwalski and Turnock, 2016; Marsh and Fu, 2017].

A central aspect of a size-structured population dynamics model is the size-transition matrix (X). This matrix specifies the probability of an individual growing from one size-class to another after a certain period of time. Equation (1) is an example of a size-transition matrix with four size-classes:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} & X_{1,4} \\ 0 & X_{2,2} & X_{2,3} & X_{2,4} \\ 0 & 0 & X_{3,3} & X_{3,4} \\ 0 & 0 & 0 & X_{4,4} \end{bmatrix}, \qquad (1)$$

where $X_{i,j}$ represents the probability of an individual transitioning from size-class i to size-class j after one time step. In Equation (1), the rows represent the starting size-class and the columns the ending size-class. The lower triangle portion of the matrix is all zero since we assume individuals cannot shrink. The modeller chooses the range of sizes defining each size-class. Each row of the size-transition matrix sums to one, since individuals must fall into one of the size-classes after growth.

The values within the size-transition matrix are determined using an underlying growth function. A variety of growth functions are available (e.g. Francis, 1988; Turnock and Rugolo, 2013; Punt *et al.*, 2016). Stock assessments commonly use the von Bertalanffy growth curve [Equation (2); Punt *et al.*, 2016]:

$$l_2 = l_1 + (L_\infty - l_1)(1 - e^{-kt}), \qquad (2)$$

where l_1 is the initial size, L_∞ is the asymptotic size, k is the growth rate, t is the number of time steps, and l_2 is the ending size after t time steps, i.e. expected size after growth is a linear function of current size. In this article, the von Bertalanffy growth curve is always the underlying growth curve, although the estimation methods described below extend naturally to other functional forms between current and eventual size. The following function determines the values within the size-transition matrix:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} F\big(l_{1i}, l_{2j}, L_\infty, k, \theta'\big) \mathrm{d}l_{2j} \mathrm{d}l_{1i}, \qquad (3)$$

where i represents the initial size-class (which ranges from l_i^- to l_i^+), l_{1i} is the initial size within size-class i, j represents the ending size-class (which ranges from l_j^- to l_j^+), l_{2j} is the resulting size within size-class j, F is the product of the probability distribution functions (pdfs) for growth variation and the distribution of individuals within the initial size-class, and θ' is a vector of parameters that define the pdfs. The parameters needed to construct the size-transition matrix are estimated using tag-recapture data, which consists of three components for each individual: the size-at-release (size when the individual is first tagged), the size-at-recapture, and the time-at-liberty (how long an individual was in the wild after being tagged).

A common issue when estimating size-transition matrices is how to account for individual variation in growth. The traditional method assumes that the underlying growth curve represents the average growth of the population, with deviation from the curve due to process error. Realistically, each individual in a population follows their own growth trajectory. Punt *et al.* (2016) and Sainsbury (1980) showed that assuming there is only a single growth curve when there is individual variation in growth could bias estimates of growth parameters.

A second ("platoon") method addresses the assumption that all individuals roughly follow the same growth curve by dividing the population into a predetermined number of groups (e.g. 3 or 5), where each group has its own size-transition matrix. Each size-transition matrix has its own growth curve by having either k

or L_∞ (with expected growth based on the von Bertalanffy equation) vary among groups. There is still process error around each growth curve to account for individual variation within each platoon. This method has been used to model finfish growth (e.g. Punt *et al.*, 2001, 2017; Methot and Wetzel, 2013). It allows for multiple growth curves within the population, potentially improving the ability to account for individual variation in growth. However, within each group, individuals still follow one growth curve.

A third method to account for individual variation in growth is to allow individuals to have their own growth curves by allowing either k, L_∞, or both to vary for each individual. This is accomplished by adding additional integrals to Equation (3) depending on how many parameters vary. For example, if k and L_∞ vary Equation (3) becomes:

$$X_{i,j} = \int_{l_i^-}^{l_i^+} \int_{l_j^-}^{l_j^+} \int_0^\infty \int_0^\infty F\big(l_{1i}, l_{2j}, L_\infty, k, \theta'\big) \mathrm{d}L_\infty \mathrm{d}k \mathrm{d}l_{2j} \mathrm{d}l_{1i}, \qquad (4)$$

where F is now the product of the pdfs for growth variation, the distribution of individuals within the initial size-class and the growth parameter distributions. Several methods exist that allow either k or L_∞, but not both to vary among individuals when constructing a size-transition matrix (e.g. Troynikov, 1998; Wang *et al.*, 1995). We developed a new method that allows k, L_∞, or both to vary among individuals by constructing the size-transition matrix using numerical integration.

For this article, we conducted a simulation experiment to evaluate the performances of the three methods for constructing size-transition matrices. Past work also used simulation to evaluate methods for constructing size-transition matrices (e.g. Wang *et al.*, 1995; Punt *et al.*, 2009). Data generation was based on golden king crabs (*Lithodes aequispinus*) in the Aleutian Islands region of Alaska. We evaluated the performance of each method by estimating size-transition matrices using these methods and comparing them to a "true" matrix using the differences between the "true" and estimated matrix values and equilibrium size-structure.

Estimation methods
Method 1: Traditional
The first (traditional) method assumes that every individual in the population follows the same growth function and thus has the same growth parameter values. This assumption uses Equation (3) to determine the values within the size-transition matrix, but it can be difficult to solve. A common practice is to assume that, before growing, all individuals have a size equal to the midpoint of the initial size-class (Punt *et al.*, 1997). This modification changes Equation (3) to:

$$X_{i,j} = \int_{l_j^-}^{l_j^+} F\big(l_{2j}, \bar{l}_i, \theta'\big) \mathrm{d}l_{2j}, \qquad (5)$$

where $\bar{l}_i$ is the midpoint of the size-class i. The function F is the pdf for growth variation, with options such as the normal, lognormal, and the gamma distributions. This article assumes a gamma distribution for the pdf because it is used as the error structure for assessments of golden king crab (Siddeek *et al.*, 2008). Therefore, $F\big(l_{2j}, \bar{l}_i, \theta'\big)$ is:

$$F\left(l_{2j}, \bar{l}_i, \theta'\right) = \frac{(l_{2j} - \bar{l}_i)^{\alpha_i - 1} e^{-\beta_i(l_{2j} - \bar{l}_i)}}{\Gamma(\alpha_i)\beta_i^{-\alpha_i}}, \tag{6}$$

where α_i and β_i are the parameters that define the gamma distribution for the growth increment of individuals that were initially in size-class i. The values for the parameters α_i and β_i are computed from the expected growth increment of an individual initially of size $\bar{l}_i$ (μ_i), which is determined using Equation (2), and the coefficient of variance for the growth increment ($C.V.$):

$$\begin{aligned} \mu_i &= \frac{\alpha_i}{\beta_i} \\ (\mu_i \cdot C.V.)^2 &= \frac{\alpha_i}{\beta_i^2}. \end{aligned} \tag{7}$$

Therefore, there are only three estimable parameters: k, L_∞, and CV for this method.

The likelihood function is the sum over all recaptured individuals of the natural logarithm of the probability of an individual tagged in size-class i growing to the recapture size-class j given that it was at liberty for a time Δt (Moser *et al.*, 2002). These are the values within the size-transition matrix when accounting for time-at-liberty ($X^{\Delta t}$ using matrix multiplication). Raising the size-transition matrix to the time-at-liberty accounts for growth during the intermediate time steps. Individuals tagged but not recaptured are not included since they do not provide information on growth. The likelihood function is not based on Equation (2) since the goal is to create a size-transition matrix that fits the data. Selectivity is also included in the likelihood function to account for the probability of catching an individual of a particular size, i.e.:

$$\ln(L) = \sum_n \ln(S_{\tilde{l}_{j,n}} [X^{\Delta t_n}]_{\tilde{l}_{i,n}, \tilde{l}_{j,n}}), \tag{8}$$

where $S_{\tilde{l}_{j,n}}$ is the selectivity for the midpoint of the size-class into which individual n fell when it was recaptured, $\tilde{l}_{i,n}$ is the size-class of release for individual n, and $\tilde{l}_{j,n}$ is the size-class at recapture for individual n. The selectivity function needs to be pre-specified and is set, for this article, to that used to generate the simulated data sets (outlined below). The time-at-liberties are integers representing years for this study because the assessment of golden king crab focuses on annual growth (Siddeek *et al.*, 2008).

Method 2: Platoon
The "platoon" method assumes that each individual follows one of several (for this article three) growth curves. This framework is similar to finite mixture models in that the population is divided into sub-groups based on growth. For this article, each growth curve has the same growth rate parameter (k), but a different asymptotic size (L_∞; nominally "small", "medium", and "large"). There is a separate size-transition matrix for each growth curve, i.e.:

$$X_{p,i,j} = \int_{l_j^-}^{l_j^+} F_p\left(l_{2j}, \bar{l}_i, L_{\infty \cdot p}, k, \sigma_{within \cdot i}\right) dl_{2j}, \tag{9}$$

where $X_{p,i,j}$ is the probability of an individual in platoon p growing from size-class i to size-class j after one time step, $\sigma_{within \cdot i}$ is the variance about the platoon-specific growth increment when the initial size is $\bar{l}_i$, and $L_{\infty \cdot p}$ is the asymptotic size for platoon p.

The values for $L_{\infty \cdot p}$ are determined from the median L_∞ for the entire population ($\tilde{L}_\infty$) and the standard deviation for L_∞ (σ_L). The L_∞s for the "small" and "large" platoons are determined by adding or subtracting one standard deviation from $\tilde{L}_\infty$. The L_∞ for the "medium" platoon equals $\tilde{L}_\infty$. If there were five platoons, the additional two platoons would have L_∞'s that were $\pm$ two standard deviations from $\tilde{L}_\infty$. The pdf for growth variability within each platoon (F_p) is assumed to be a gamma distribution.

Taylor and Methot (2013) argued that the total variance in growth increment ($\sigma_{total \cdot i}^2$) can be broken down into two parts: the variance within platoons ($\sigma_{within \cdot i}^2$) and the variance among platoons ($\sigma_{between \cdot i}^2$). $\sigma_{within \cdot i}^2$ defines the variance in the growth increment from the underlying growth curve for each platoon while $\sigma_{between \cdot i}^2$ defines the variation in the expected growth increment among platoons. The values for the three variances are determined using ρ, which equals ${\sigma_{within \cdot i}}/{\sigma_{between \cdot i}}$. Details on how this is done can be found in Supplementary Appendix 1. The value of ρ controls the extent to which the platoons overlap. This parameter is pre-specified because it is difficult to estimate (*sensu* Taylor and Methot, 2013). We explored the sensitivity to the pre-specified value for ρ by setting it to 0.75, 1.5, 2, 2.5, or 4.

Three parameters are estimated when using the platoon method; k, $\tilde{L}_\infty$, and σ_L. Equation (8) cannot be used to estimate the parameters because the platoon structure is a theoretical concept. The likelihood function needs to account for the probability of an individual being in one of the three platoons. Therefore, the parameters are estimated using:

$$L = \sum_n \sum_p S_{\tilde{l}_{j,n}} \tau_p \left[X_p^{\Delta t_n}\right]_{\tilde{l}_{i,n}, \tilde{l}_{j,n}}, \tag{10}$$

where τ_p is the proportion of recruits that settle to platoon p and $X_p^{\Delta t_n}$ is the size-transition matrix for platoon p raised to the time-at-liberty. Multiplying each platoon's size-transition matrix by the proportion of new recruits to that platoon addresses the uncertainty related to which platoon a tagged individual belongs.

The values for τ_p are 16%, 68%, and 16% to the "small", "medium" and "larger" platoons since we assumed that new recruits are randomly assigned to each platoon such that the distribution of recruits is approximately normal and mirrors how L_∞ is divided between platoons. This ensures that the average of each individual's size-transition matrix across all platoons is approximately the average size-transition matrix across all individuals if there were no platoons.

Method 3: Numerical integration
This method assumes that individuals follow their own growth curve. A single size-transition matrix is created, but instead of focusing on variation around a single curve, this method focuses on the variation in the values for the growth parameters by allowing either k, L_∞, or both to vary for each individual. For our work, we assume $L_\infty \sim$ lognormal $(\overline{L_\infty}, \sigma_{L_\infty})$, $k \sim$ lognormal $(\overline{k}, \sigma_k)$ and that their values and distributions are independent. This is a simplifying assumption that ignores underlying ecological process that suggests the two growth parameters may be correlated (e.g. Pilling *et al.*, 2002), although the method can be extended to allow for such correlation. The values within the matrix are determined using Equation (4). However, the pdf for growth variation cannot be written down in closed form when

all individuals grow according to their own growth curves. The size of an individual after one-time step depends on its growth parameters and initial size. Thus, there is no potential variation in the ending size of an individual. To calculate the probability of growing into a particular size-class, we force the range over which one of the growth parameters is integrated in Equation (4) to depend on the other. Three variants of this method (Methods 3a, 3b, and 3c) differ in terms of whether k, L_∞, or both vary for each individual. The likelihood function used in Method 1 [Equation (8)] is used for Method 3. Supplementary Appendix 2 provides technical details on how the numerical integration method is implemented in this article.

Simulation evaluation

The methods described above, each with different assumptions about individual variation in growth, result in nine candidate methods (Table 1, with accompanied abbreviations that will be used henceforth). The formulae for the variation in the growth increment for each method are listed in Supplementary Table S3. We evaluated these methods using pseudo tag-recapture data generated using an individual-based operating model with biological characteristics mirroring golden king crab in the Aleutian Islands (100 simulated data sets for each scenario). The assessment of golden king crab is based on a linear relationship between initial and ending size, which matches the assumptions of von Bertalanffy growth curve. Previous analyses used similar techniques for simulating tag-recapture data to evaluate methods for estimating size-transition matrices (e.g. Kanaiwa *et al.*, 2008; Punt *et al.*, 2009). Two operating models are used. In the first operating model (OP1), each individual follows the same underlying growth curve, with deviations in growth increment due to process error. In contrast, individuals in the second operating model (OP2) have their own k and L_∞ values. The number of data points could impact the performance of each estimation method. The golden king crab fishery has approximately 1500 tag-recapture data points (Siddeek *et al.*, 2016). Thus, we considered pseudo tag-recapture data sets with 500, 1500, 3000, and 4000 data points to explore the impact of sample size.

The width of the size-classes represents a trade-off between computational burden (quicker to have larger size-classes) and biological precision (smaller size-classes allow more detail to be captured). Szuwalski *et al.* (2014) showed that increasing size-class width could increase estimates of abundance. Therefore, the nine estimation methods constructed size-transition matrices with size-class widths of 5, 10, or 15 mm for each data set. Changing the size-class width results in changes to the number of

size-classes. A size-class width of 5 mm has 18 size-classes, 10 mm has 9, and 15 mm has 6.

We simulated tag-recapture data by representing aspects of golden king crab life history, such as natural mortality and removal by the fishery, as a series of Bernoulli trials (Supplementary Figure S1). The first step is to obtain a value for the size-at-release for an individual. That individual is projected forward until it is recaptured, dies in the wild, or it is at liberty for more than the maximum time-at-liberty for golden king crab (7 years). For every yearly time step, the individual must first survive the year. The most recent assessment of golden king crab assumed that natural mortality (M) is the same for all individuals regardless of size (i.e. 0.18 year^{-1}; Pengilly, 2016). If the individual survives, it grows according to Equation (2), where t equals 1. In OP1, all individuals follow the same growth curve, with variation in the growth increment due to process error, which is gamma distributed, i.e.:

$$\overline{GI} = (L_\infty - l_1)(1 - e^{-kt})$$
$$GI \sim \Gamma(\alpha_1, \beta_1) \qquad , \qquad (11)$$
$$l_2 = l_1 + GI$$

where $\overline{GI}$ is the mean growth increment for an individual with initial size l_1, GI is the growth increment with process error and α_1 and β_1 define the gamma distribution whose values are determined using Equation (7) with $\overline{GI}$ replacing μ_i. In OP2, each animal follows its own growth curve, with k and L_∞ randomly chosen from lognormal distributions.

After growing, the individual encounters the fishery. The probability of an individual being caught is dictated by the selectivity curve and fishing effort (q). An individual is considered caught by the fishery (and hence included in the tag-recapture data set) if a random draw from $U[0, 1]$ is less than the selectivity corresponding to the size of the individual multiplied by fishing effort. The assessment for the golden king crab assumes logistic selectivity:

$$S = \left[1 + \exp\left(-\ln(19)\frac{l_2 - \theta_{50}}{\theta_{95} - \theta_{50}}\right)\right]^{-1}, \qquad (12)$$

where θ_{50} is the size-at-50%-selectivity and θ_{95} is the size-at-95%-selectivity. We assumed that the values for the selectivity parameters are known when applying the estimation methods.

Four variants of the operating models are considered depending on how the initial sizes for individuals are selected (using actual data or a uniform distribution) and the maximum time-at-liberty (1 or 7 years). Tag-7 is a base-case as it is the closest

Table 1. Methods for constructing a size-transition matrix and associated estimable parameters.

| Method	Method abbreviation	Description	Parameters estimated
Method 1	Fixed	Individuals follow a common curve	$k, \underset{\sim}{L}_\infty, \sigma^2$
Method 2a	Plat 0.75	Platoon with $\rho = 0.75$	$k, \underset{\sim}{L}_\infty, \sigma_L$
Method 2b	Plat 1.5	Platoon with $\rho = 1.5$	$k, \underset{\sim}{L}_\infty, \sigma_L$
Method 2c	Plat 2	Platoon with $\rho = 2.0$	$k, \underset{\sim}{L}_\infty, \sigma_L$
Method 2	Plat 2.5	Platoon with $\rho = 2.5$	$k, \underset{\sim}{L}_\infty, \sigma_L$
Method 2c	Plat 4	Platoon with $\rho = 4.0$	k, L_∞, σ_L
Method 3a	Vary k	Variation in k	$L_\infty, \overline{k}, \sigma_k$
Method 3b	Vary Linf	Variation in L_∞	$k, \overline{L}_\infty, \sigma_{L_\infty}$
Method 3c	Both	Variation in L_∞ and k	$\overline{L}_\infty, \sigma_{L_\infty}, \overline{k}, \sigma_k$
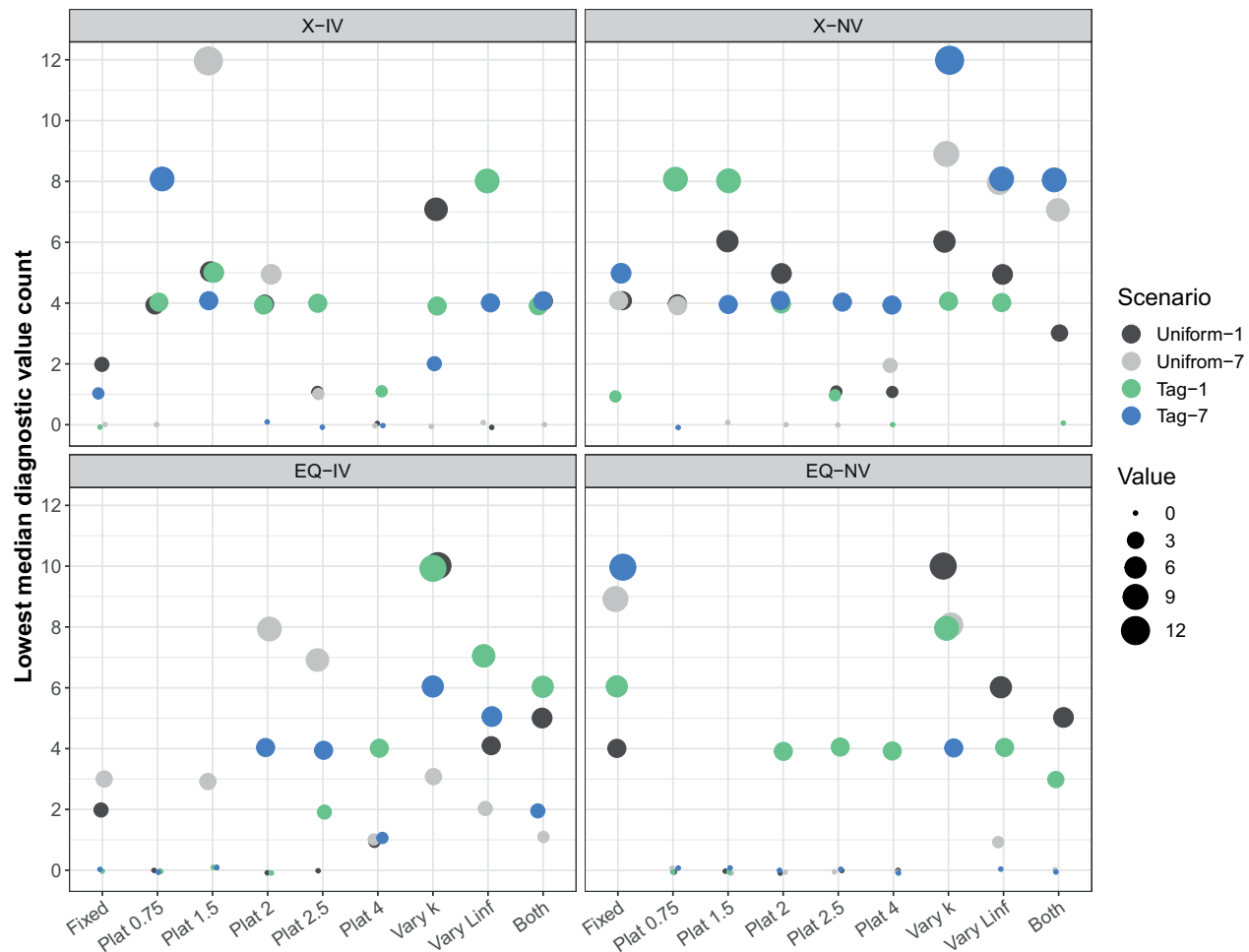
Figure 1. Counts for the number of times the median value, over 100 runs, of a particular diagnostic for an estimation method was among the lowest (meaning performed best) for a particular scenario. Methods were considered not appreciably different from the lowest value for the matrix diagnostic if the difference in median values was less than 0.1 from that of the lowest. For the equilibrium diagnostic, the discrepancy threshold was 0.01. The colours represent how the initial size was generated ("Uniform"—uniform distribution, "Tag"—actual tag-recapture data) and the maximum time-at-liberty (1—one year or 7—seven years). The counts arise from the options for sample size (four options) and size-class width (three options) with a maximum possible value of 12 for each colour for each estimation method. The titles indicate the diagnostic (X—matrix diagnostic, EQ—equilibrium diagnostic) and whether there is individual variation in growth in the operating model (IV—Individual Variation; OP2, NV—No Individual Variation; OP1).

representation of the actual gold king crab tag-recapture data. Uniform-7 explores the impact of the tagged animals representing a larger range of sizes compared to the distribution in Tag-7 (Supplementary Figure S2). Tag-1 explores reducing the conflict between discrete (size-transition matrix) and continuous (von Bertalanffy) growth since the number of time steps has been reduced to one. Uniform-1 combines the factors that constitute Uniform-7 and Tag-1, and would be expected to lead to the best performance because it provides the most informative data (broad range of sizes-at-release) and avoids dealing with the distinction between continuous and discrete growth. We used 864 scenarios to explore the impact of sample size, size-class width, initial size distribution, and maximum time-at-liberty (Supplementary Table S4).

The parameters of the operating model, their associated values, and the information on which those values are based are given in Supplementary Table S5. Fishing effort was chosen so that the distribution of the time-at-liberty resembles the actual

distribution of the time-at-liberty for golden king crab (Supplementary Figure S3).

Evaluation

Three methods are used to evaluate the performance of the estimation methods. The first determines how well each method mimicked the simulated data. This involved first creating a distribution for the simulated sizes-at-recapture. The model-estimate of this distribution was computed by: (i) grouping the simulated data by size-class-at-release and time-at-liberty, (ii) multiplying the number of individuals in each of these groups by the row of $\mathbf{X}^{\mathrm{T}}$ (where T is the time-at-liberty) representing their predicted size-classes-at-recapture, and (iii) summing the predicted size-at-recapture distribution over all recaptures. We visually compared these values and the number of size-at-recaptures for each size-class from the tag-recapture data produced through simulation when estimating the size-transition matrix.
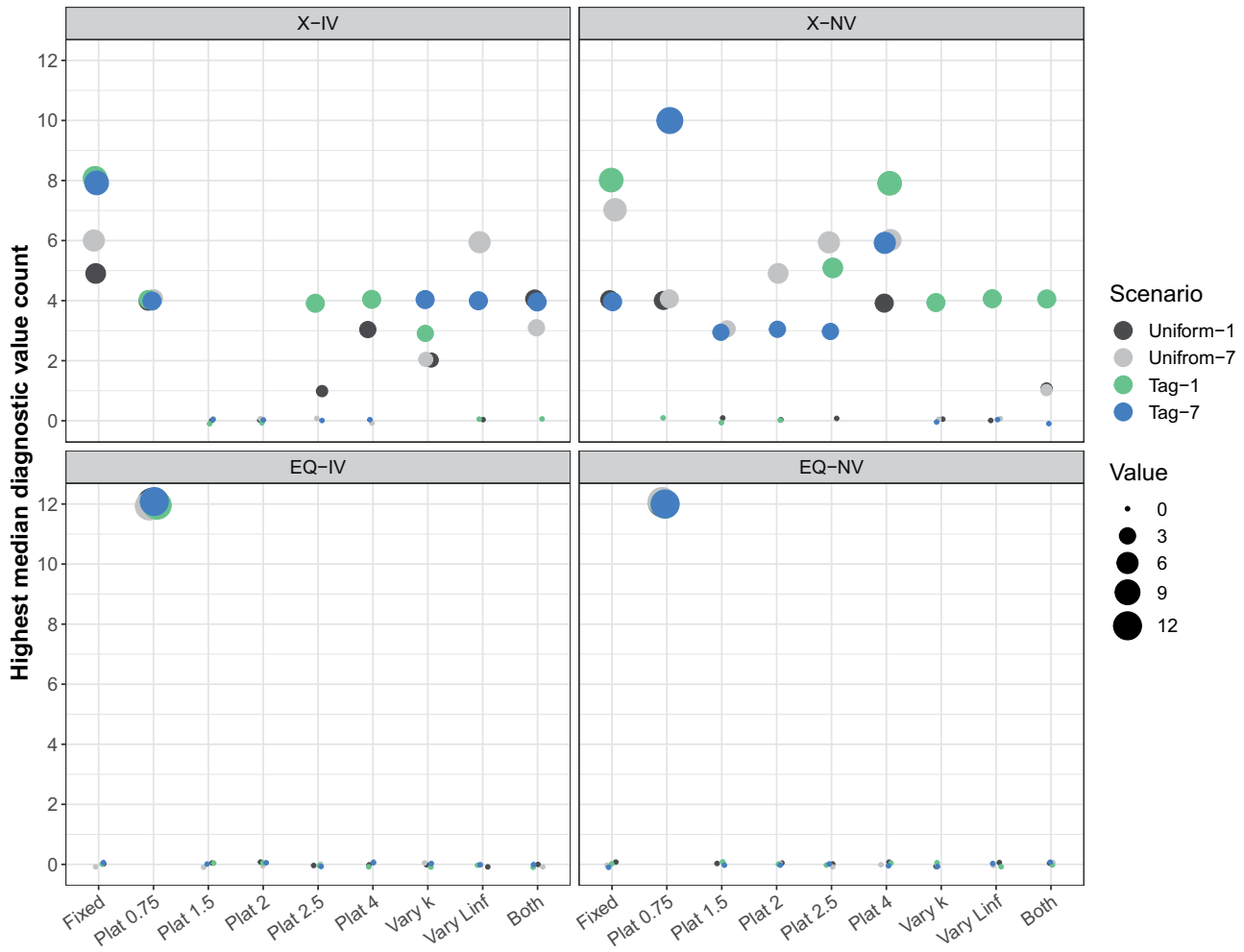
Figure 2. As for Figure 1 except the counts are of estimation methods that are among the highest (meaning performing poorly) median value.

The second performance metric, the matrix diagnostic, compares the "true" and estimated values for the size-transition matrix. The "true" matrix for both operating models is constructed using simulation. For OP1, values for l_{1i} are randomly drawn ($U[l_i^-, l_i^+]$) and inserted into the first line of Equation (11) (with $t = 1$) to determine the expected growth increment using the L_∞ and k values from OP1. This value, and the CV from OP1 are used to determine a gamma distribution from which a random value is drawn, added to l_{1i} (representing the ending size) and recorded. This process is repeated 30 000 000 times. The proportion of ending sizes that fall within each size-class determines the "true" size-transition matrix. In contrast, the creation of the "true" matrix for OP2 draws new values for L_∞ (lognormal distribution [$\overline{L}_\infty, \sigma_{L_\infty}$]) and k (lognormal distribution [$\overline{k}, \sigma_k$]) for each l_{1i}. These values are inserted into Equation (2) (with $t = 1$) to determine the ending sizes. The remaining steps match those used for constructing OP1's "true" size-transition matrix after obtaining the ending sizes.

The matrix diagnostic is the sum of the absolute differences between the estimated and "true" values of the size-transition matrix divided by the number of positions in the upper triangular portion of the size-transition matrix times 100 (equation shown in Supplementary Appendix 3). This diagnostic is

not based on relative errors because some of the entries of the size-transition matrix can be very small. Relative errors can make the differences between small values seem inappropriately important. When comparing the platoon method to the "true" matrix, the values within each platoon matrix are multiplied by the proportions of recruits added to that platoon and summed up to get a single estimated value. Differences in the matrix diagnostic below the second decimal place, which represents the fourth decimal place for the values in the size-transition matrix, were not considered "appreciable" since the "true" matrix values were only repeatable to the third decimal place.

The third performance metric, the equilibrium diagnostic, compares the equilibrium size distribution between two theoretical population, each with a mortality rate of 0.18 year^{-1} (assumed rate of natural mortality for Aleutian Islands golden king crab) that used either the "true" or estimated size-transition matrix. We assumed that all recruits in these theoretical populations enter the first size-class. The equilibrium size-structure is determined by taking the right eigenvector of the product of the size-transition matrix and a matrix with the exponentiated mortality rate on the diagonals. The estimated and "true" equilibrium size-structures are compared using the average of the sum of the
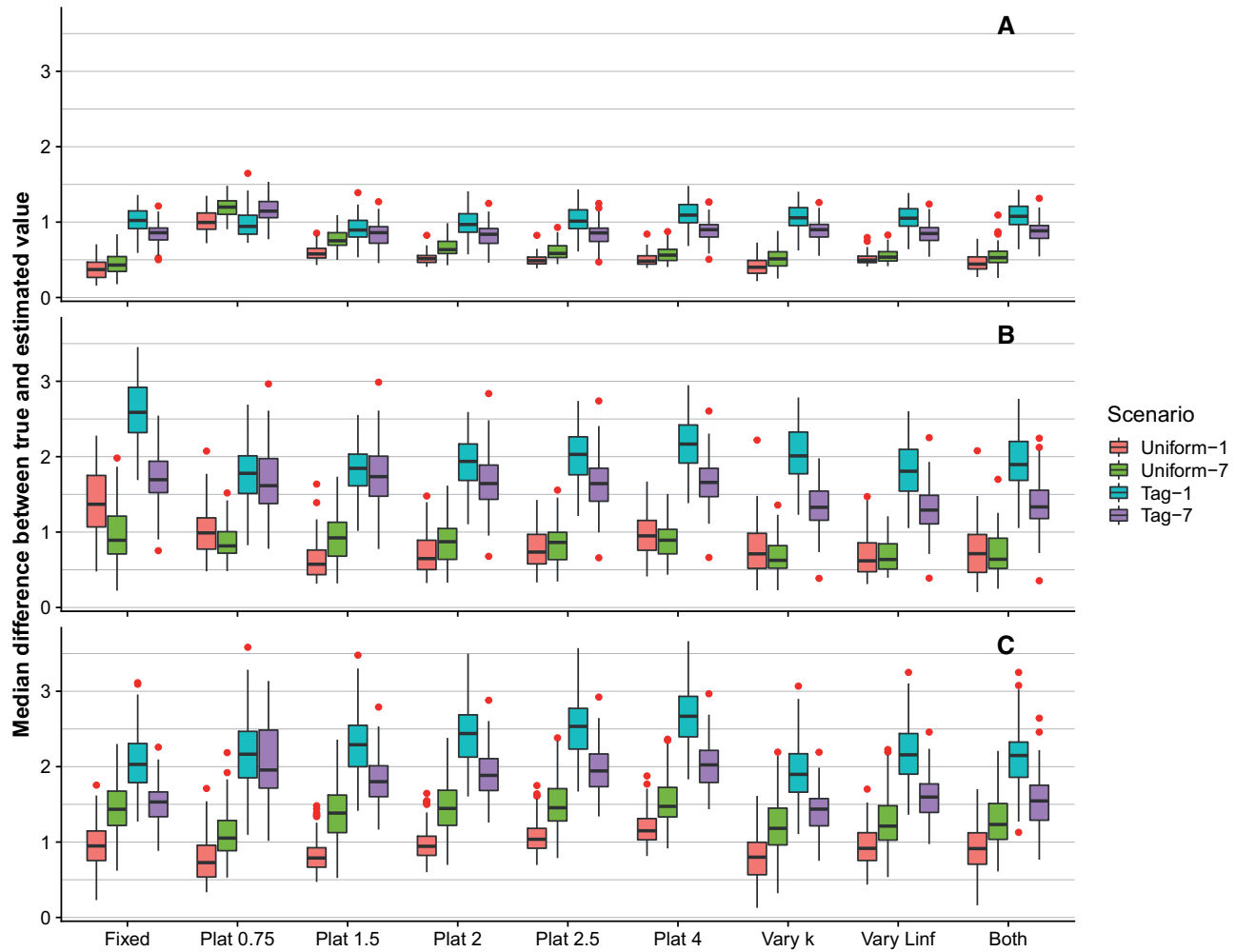
Figure 3. Boxplots of the matrix diagnostic for OP1. The colours indicate how the initial size was generated (Uniform—uniform distribution or Tag—actual tag-recapture data) and the maximum time-at-liberty (1 or 7 years). The letters in the upper right-hand corner indicate the number of size-classes (a = 18, b = 9, and c = 6).

absolute error times 100 for each simulation (equations shown in Supplementary Appendix 3).

Results

The estimation methods were able to mimic the simulated data sets well (Supplementary Figure S4). Fits improved with a smaller number of size-classes. Increasing the sample size also improved fits (Supplementary Figure S5). The simulated distribution of sizes-at-recapture depends on whether the initial sizes were generated from the actual tagging data versus a uniform distribution. However, this did not affect the estimation methods' ability to mimic the data (Supplementary Figure S6).

No estimation method was consistently among the best or worst performing for all scenarios for either the matrix or the equilibrium diagnostic (Figures 1 and 2; see Supplementary Table S6 for a tabular summary) except for Plat 0.75, which was always the worst for the equilibrium diagnostic (Figure 2). Estimation method performance was not consistent between diagnostics. For example, Plat 0.75 and Plat 1.5 performed well for the matrix diagnostic (top panels of Figure 1), especially when there was individual variation in growth (OP2; upper left panel of Figure 1) but

they almost were never among the best methods for the equilibrium diagnostic (bottom panels of Figure 1).

Boxplots of the matrix and equilibrium diagnostics were used to explore the differences in performance among estimation methods (medians of the boxplot distributions listed in Supplementary Tables S7–S10). In general, increasing sample size reduced the among-simulation variation in the diagnostics while differences in performance between estimation methods remained relatively consistent. Consequently, our boxplots only show results for a sample size of 1500 (this is approximately the tag-recapture sample size for golden king crab; see Supplementary Figures S7–S10 for the full set of boxplots showing all sample sizes).

Ability to match the true size-transition matrix (matrix diagnostic)

Figures 3 and 4 show the distributions for the matrix diagnostic for OP1 and OP2, respectively. Low values imply better performance. In general, more size-classes resulted in lower values with a few exceptions such as Uniform-1, OP2, Plat 0.75 when comparing 6 vs. 9 size-classes. The best performing method varied
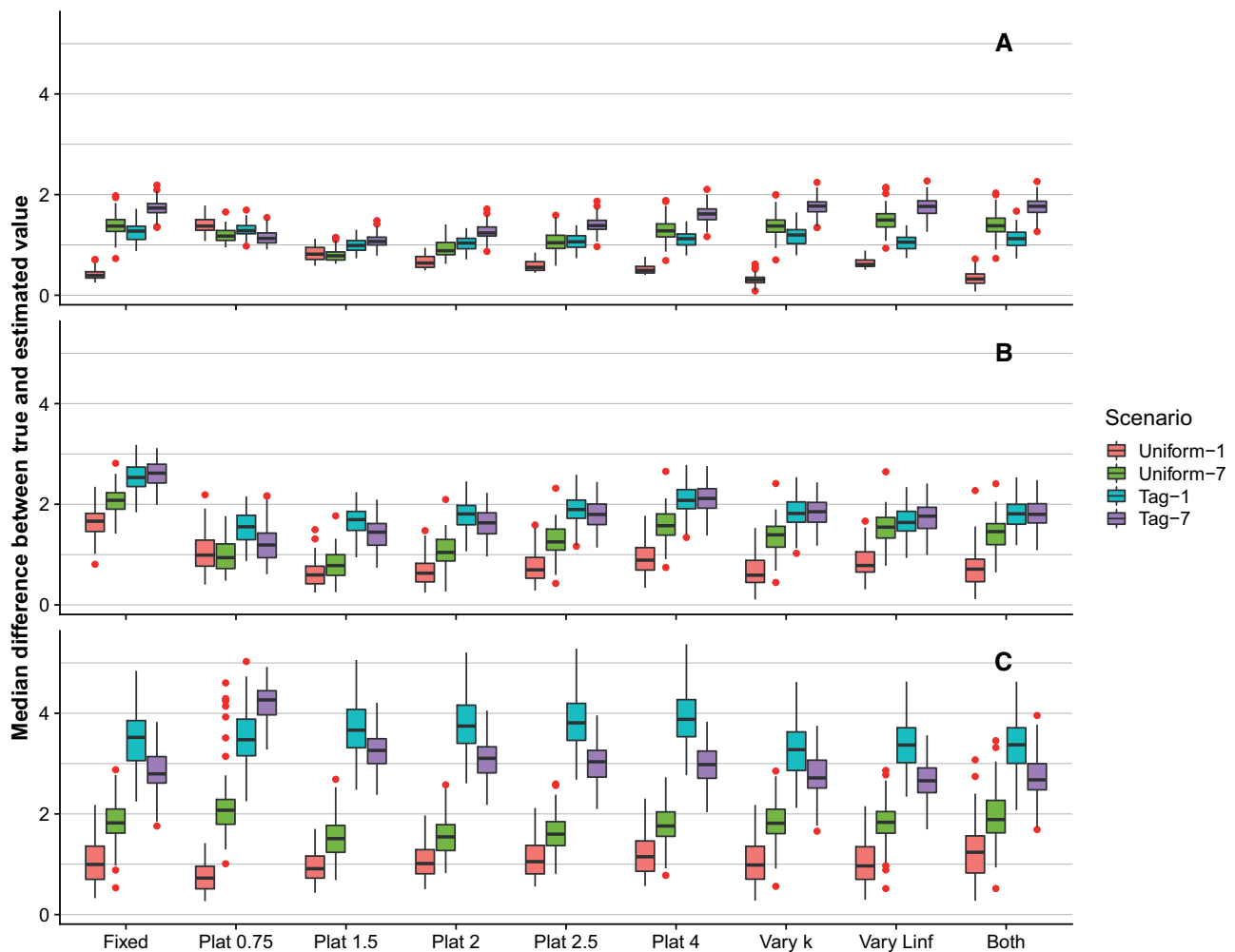
Figure 4. As for Figure 3, except that the operating model is OP2.

based on scenario and number of size-classes. Estimation methods performed best when the tag-recapture data were generated from a uniform distribution with a few exceptions such as when comparing Uniform-7 to Tag-1 for the numerical integration methods using 18 size-classes for OP2. The impact of the maximum time-at-liberty varied based on the initial size distribution. With a Uniform distribution, a 1 year time-at-liberty generally performed better than 7 years at liberty. However, the difference in performance between the maximum time-at-liberties was very small for 9-size-classes and OP1. The 7 year maximum time-at-liberty performed better for most scenarios using actual tagging data for OP1. However, for OP2, the difference between time-at-liberties depended on the number of size-classes with 7 years doing better with fewer size-classes and 1 year better with more size-classes.

Calculation of the equilibrium size-structure (equilibrium diagnostic)

Figures 5 and 6 show the distributions for the equilibrium diagnostic for OP1 and OP2, respectively. The lowest value (i.e. best) occurred with more size-classes, although the best performing method again depended on the number of size-classes and scenario. A uniform initial size distribution generally resulted in a

better performance except when comparing Tag-1 and Uniform-7 with OP2 and 18 size-classes. The influence of the maximum time-at-liberty varied based on the initial size distribution. A 1 year time-at-liberty performed best on average with a uniform distribution, with the exception of OP1 with 6 size-classes where there was barely any difference between the maximum time-at-liberties. When the tagging data were based on the actual tagging data, a 7 year maximum time-liberty performed better with fewer size-classes but the difference between the maximum time-at-liberties decreased as the number of size-classes increased. This trend resulted in 1 year time-at-liberty performing best when there were 18 size-classes, with OP2.

The poor performance of the plat 0.75 method

The poor performance of Plat 0.75 for the equilibrium diagnostic (Figure 2) was unexpected and relates to ρ. Small ρ values imply a large $\sigma_{between_i}$ and small σ_{within_i}, leading to less overlap in growth curves between platoons. Determining the equilibrium size-structure for the platoon method involves calculating the equilibrium size-structure for each platoon then combining them using the proportion of recruits to each platoon. The forced separation of growth trajectories with small ρ values is amplified when the equilibrium size-structure is calculated, leading to larger
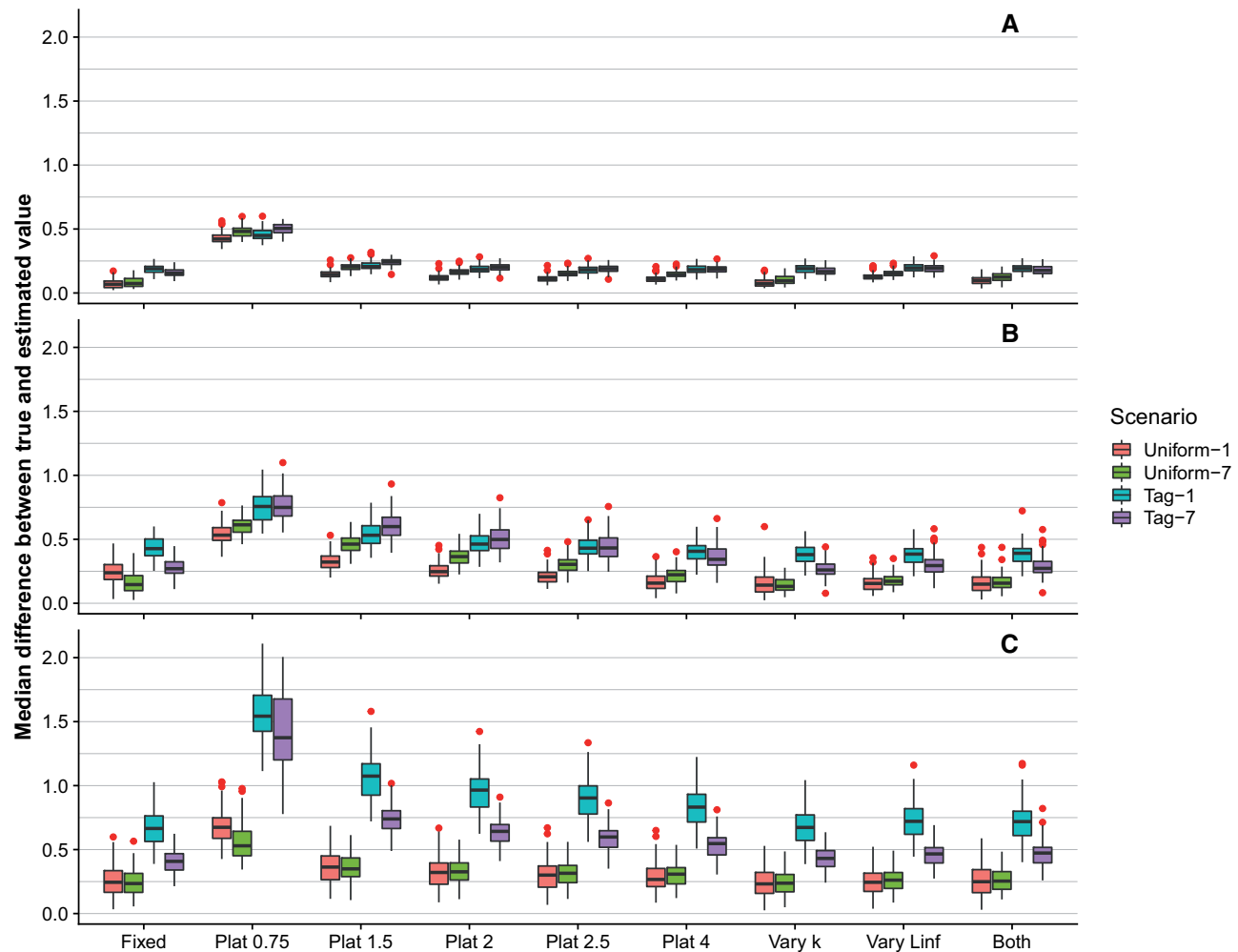
Figure 5. As for Figure 3, except that the metric is the equilibrium diagnostic.

differences between the "true" and estimated size-structures. Reversing the order of operation by combining each platoon's size-transition matrix first improves the ability to estimate the equilibrium size-structure. However, this ignores the platoon growth structure. Increasing ρ produces a similar effect because there is more overlap between the size-transition matrices (Supplementary Figure S11). This explains why platoon methods with higher ρ values occasionally perform best in terms of the equilibrium diagnostic (Figure 1). Unfortunately, estimating ρ is not an option because it is strongly negatively correlated with σ_L (results not shown).

Discussion

No estimation method consistently performed best across all factors considered in the analyses. The number of size-classes, maximum time-at-liberty, and how size-at-release was generated influenced performance. In addition, the matrix and equilibrium diagnostic metrics occasionally produced contradictory conclusions in terms of which estimation method performs best. Although no method performed best, the goal of this work was to determine which estimation method was most robust, and best for general use.

A surprising result is the conflicting outcomes between the matrix and equilibrium diagnostics. Specifically, Plat 0.75 and Plat 1.5 were amongst the best methods for the matrix diagnostic yet were almost never the best in terms of the equilibrium diagnostic. In fact, Plat 0.75 had the poorest performance for the equilibrium diagnostic for all scenarios. The difference in performance between the two metrics occurs in part because different size-transition matrices can produce the same value for the matrix diagnostic, but not for the equilibrium diagnostic. This is due in part because the diagonals of the size-transition matrix heavily influence the equilibrium diagnostic.

What influences performance?

A goal of this work was to determine which estimation method is the most robust; unfortunately, there is no simple answer to this question. In general, the lowest value for the matrix and equilibrium diagnostic occurred with Uniform-1, with the largest number of size-classes having the smallest values overall (top boxplot of Figures 3–6). When the "true" matrix had individuals following a common growth curve, the Fixed method performed best (Supplementary Tables S7 and S9) and when individuals followed their own growth curves, Vary k performed best (Supplementary Tables S8 and S10). Decreasing the number of size-classes resulted in other estimation
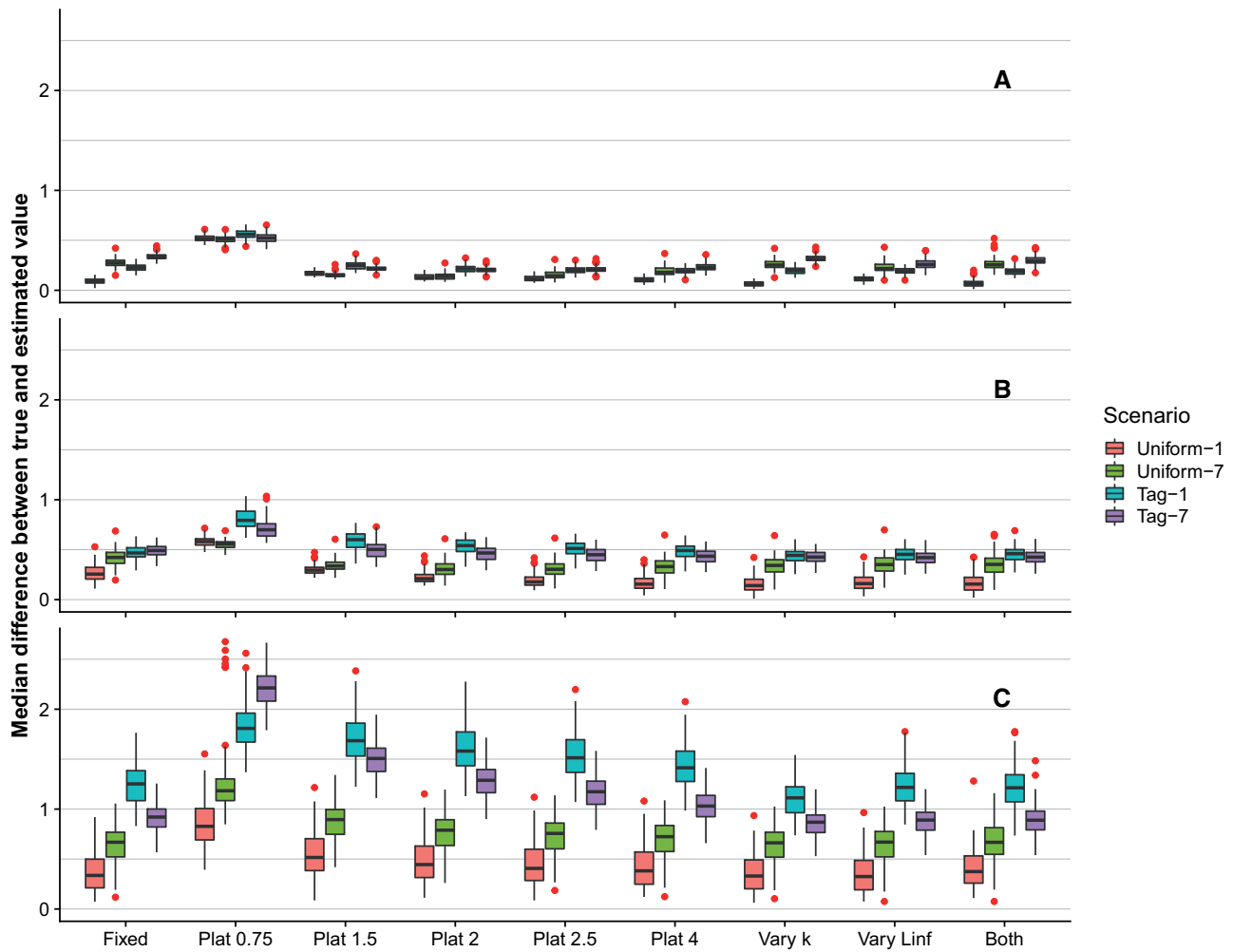
Figure 6. As for Figure 3, except that the metric is the equilibrium diagnostic and the operating model is OP2.

methods performing best. A possible reason for this is that fewer size-classes lead to an increase in size-class width causing the source of growth variability to be more difficult to detect. In addition, the assumption that individuals are uniformly distributed within a size-class is violated to a greater extent for wide size-classes.

The best estimation method also depends on the time-at-liberty. Size-transition matrices assume that growth is a discrete process. Thus, the distribution of size after two or more years for a given initial distribution of sizes is the product of the size-transition matrix two or more times. Likelihood Equations (8) and (10) rely on the assumption that the distribution of size within each size-class is uniform for all time steps, even post-growth. This is not the case, and the assumption becomes increasingly more violated each time the size-transition matrix is multiplied.

The numerical integration method assumes that initial size is uniformly distributed, while the traditional and platoon methods assume the initial size for individuals is the midpoint of the starting size-class. Consequently, the numerical integration estimation methods are more consistent with the operating model when the initial sizes are generated from a uniform distribution (at least for a time-at-liberty of 1 year). The actual tagging data have a bell-shaped distribution over the entire size range (Supplementary

Figure S2), which could explain why having a greater number of size-classes performs better since the distribution of sizes within smaller size-class widths more accurately mimic a uniform distribution.

The bell-shaped distribution of tag-recapture data does not always adequately represent all size-classes. Some size-classes, typically the larger ones, can have fewer data points. This reduces the contrast in the data making it more difficult to determine the size-transition matrix values associated with the data poor size-classes.

Next steps

The next steps in this work relate to estimation methods and scenarios. Additional estimation methods could be considered such as allowing k rather than L_∞ to vary among platoons for the platoon method and allowing for other distributions of sizes within the initial size-class. The number of years of tagging data used to estimate the size-transition matrix had more impact than anticipated. Analyses examining the trade-off between using fewer tag-recaptures (perhaps only those tags at liberty for a year or two) and reducing bias vs. using more tags and increasing bias but reducing variance should be conducted. This trade-off will likely depend on total sample size. Our analyses assumed that

selectivity was known but there will always be some uncertainty in this regard. Future analyses should also examine sensitivity to errors in estimates of selectivity as well as assess how robust the results are to the parameters that determine growth and natural mortality. Many stock assessments that use size-structured models, including that for golden king crab, estimate the probability of a crab moulting. Tests should be conducted to examine how including moult probability influences the performance of the estimation methods.

Conclusions

This work has shown that numerical integration methods for estimating a size-transition matrix are viable. However, it is difficult to say which estimation method is the best due to confounding influences from the number of size-classes, maximum time-at-liberty, and the distribution of sizes at tagging. If we assume that crabs in the wild follow their own growth curve then Tag-7 and Tag-1 are the more realistic cases (non-uniform distribution of initial sizes, actual tagging data used). Therefore, we recommend Vary Linf as the default with smaller size-class widths and possibly restricting the tagging data to only 1 year. Vary Linf is always among the best-performing methods for Tag-1 for the matrix diagnostic (Figure 4; Supplementary Figure S12). For the equilibrium diagnostic, it is among the best performing when there are fewer size-classes (Figure 6; Supplementary Figure S13). Vary Linf was also never among the worst performing methods for Tag-1 (Figure 2). Plat 0.75 and Plat 1.5 dominate Tag-7, yet should be avoided as their performance, particularly in terms of estimating the equilibrium size-structure, is poor. However, other methods can be valid alternatives depending on which characteristics interest the analyst.

Supplementary data

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

References

Atlantic States Marine Fisheries Commission (ASMFC). 2015. American Lobster Benchmark Stock Assessment and Peer-review Report. Atlantic States Marine Fisheries Commission, Washington, DC. 493 pp.

Buckworth, R. C., Deng, R. A., Plagányi, E. E., Punt, A., Upston, J., Pascoe, S., Miller, M. T. *et al*. 2015. Northern Prawn Fishery RAG Assessments 2013–15. Final Report to the Australian Fisheries Management Authority, Research Project 2013/0005, June 2015. CSIRO, Brisbane. 177 p.

Francis, R. I. C. C. 1988. Maximum likelihood estimation of growth and growth variability from tagging data. New Zealand Journal of Marine and Freshwater Research, 22: 43–51.

Garber-Yonts, B., and Lee, J. 2018. Stock Assessment and Fishery Evaluation Report for King and Tanner Crab Fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands Area: Economic Status of the BSAI King and Tanner Crab Fisheries off Alaska, 2017.

Kanaiwa, M., Chen, Y., and Wilson, C. 2008. Evaluating a seasonal, sex-specific size-structured stock assessment model for the American lobster, *Homarus americanus*. Marine and Freshwater Research, 59: 41–56.

Marsh, C., and Fu, D. 2017. The 2016 stock assessment of paua (*Haliotis iris*) for PAU 5D. New Zealand Fisheries Assessment Report 2017/33.

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

Moser, S. M., Macintosh, D. J., Pripanapong, S., and Tongdee, N. 2002. Estimated growth of the mud crab *Sylla olivacea* in the Ranong mangrove ecosystem, Thailand, based on a tagging and recapture study. Marine and Freshwater Research, 53: 1083–1089.

Pengilly, D. 2016. 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. *In* Aleutian Islands Golden King Crab—2016 Tier 5 Stock Assessment. pp. 723–772. North Pacific Fishery Management Council, Anchorage, AK, USA.

Pilling, G. M., Kirkwood, G. P., and Walker, S. G. 2002. An improved method for estimating individual growth variability in fish, and the correlation between von Bertalanffy growth parameters. Canadian Journal of Fisheries and Aquatic Sciences, 59: 424–431.

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

Punt, A. E., Buckworth, R. C., Dichmont, C. M., and Ye, Y. 2009. Performance of methods for estimating size-transition matrices using tag-recapture data. Marine and Freshwater Research, 60: 168–182.

Punt, A. E., Campbell, R. A., and Smith, A. D. M. 2001. Evaluating empirical indicators and reference points for fisheries management: application to the broadbill sword fishery off eastern Australia. Marine Freshwater Research, 52: 819–832.

Punt, A. E., Haddon, M., and McGarvey, R. 2016. Estimating growth within size-structured fishery stock assessments: what is the state of the art and what does the future look like? Fisheries Research, 180: 147–160.

Punt, A. E., Huang, T.-C., and Maunder, M. N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusk species. ICES Journal of Marine Science, 70: 16–33.

Punt, A. E., Kennedy, R. B., and Frusher, S. 1997. Estimating the size-transition matrix for Tasmanian rock lobster, *Jasus edwardsii*. Marine and Freshwater Research, 48: 981–992.

Sainsbury, K. J. 1980. Effect of individual variability on the von Bertalanffy growth equation. Canadian Journal of Fisheries and Aquatic Sciences, 37: 241–247.

Siddeek, M. S. M., Watson, L. J., Barnard, D. R., and Gish, R. K. 2008. Aleutian Islands golden king crab (*Lithodes aequispinus*) stock assessment. *In* Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. pp. 443–495. North Pacific Fishery Management Council, Anchorage, AK.

Siddeek, M. S. M., Zheng, J., Punt, A. E., and Vanek, V. 2016. Estimation of size-transition matrices with and without probability for Alaska golden king crab using tag-recapture data. Fisheries Research, 180: 161–168.

Szuwalski, C., and Turnock, B. J. 2016. Stock assessment of eastern Bering Sea snow crab. *In* Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the

Bering Sea and Aleutian Islands Regions. pp. 167–251. North Pacific Fishery Management Council, Anchorage, AK.

Szuwalski, C. D., Foy, R. J., and Turnock, B. J. 2014. 2014 Stock assessment and fishery evaluation report for the Pribilof Island red king crab fishery of the Bering Sea and Aleutian Islands regions. *In* Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. pp. 546–605. North Pacific Fishery Management Council, Anchorage, AK.

Taylor, I. G., and Methot, R. D. 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fisheries Research, 142: 75–85.

Troynikov, V. 1998. Probability density functions useful for parametrization of heterogeneity in growth and allometry data. Bulletin of Mathematical Biology, 60: 1099–1122.

Turnock, B. J., and Rugolo, L. J. 2013. Stock assessment of eastern Bering Sea snow crab. *In* Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. pp. 39–167. North Pacific Fishery Management Council, Anchorage, AK.

Wang, Y.-G., Thomas, M. R., and Somers, I. F. 1995. A maximum likelihood approach for estimating growth from tag-recapture data. Canadian Journal of Fisheries and Aquatic Sciences, 52: 252–259.

Handling editor: Ken Andersen