Mixture of Time-Dependent Growth Models with an Application to Blue Swimmer Crab Length-Frequency Data

Luke R. Lloyd-Jones,[1,*] **Hien D. Nguyen,**[2,**] **Geoffrey J. McLachlan,**[2,***] **Wayne Sumpton,**[3,****] **and You-Gan Wang**[4,*****]

[1]Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia
[2]School of Mathematics and Physics, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia
[3]Queensland Department of Agriculture and Fisheries, Ecosciences Precinct, Joe Baker St, Dutton Park, Brisbane, Queensland 4102, Australia
[4]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland 4000, Australia
*email: luke.lloydjones@uqconnect.edu.au
**email: h.nguyen7@uq.edu.au,
***email: g.mclachlan@uq.edu.au,
****email: wayne.sumpton@daf.qld.gov.au
*****email: yougan.wang@qut.edu.au

SUMMARY. Understanding how aquatic species grow is fundamental in fisheries because stock assessment often relies on growth dependent statistical models. Length-frequency-based methods become important when more applicable data for growth model estimation are either not available or very expensive. In this article, we develop a new framework for growth estimation from length-frequency data using a generalized von Bertalanffy growth model (VBGM) framework that allows for time-dependent covariates to be incorporated. A finite mixture of normal distributions is used to model the length-frequency cohorts of each month with the means constrained to follow a VBGM. The variances of the finite mixture components are constrained to be a function of mean length, reducing the number of parameters and allowing for an estimate of the variance at any length. To optimize the likelihood, we use a minorization–maximization (MM) algorithm with a Nelder–Mead sub-step. This work was motivated by the decline in catches of the blue swimmer crab (BSC) (*Portunus armatus*) off the east coast of Queensland, Australia. We test the method with a simulation study and then apply it to the BSC fishery data.

KEY WORDS: Blue swimmer crab; Growth model estimation; Length-frequency data; Minorization–maximization algorithm; Mixture modeling.

1. Introduction

Understanding how aquatic species grow is fundamental in fisheries because stock assessment often relies on growth dependent statistical models. The statistical model used to make inference about growth depends on the data available; three common data types used for growth model estimation are (i) individual length and time-at-liberty data from tag-recapture experiments; (ii) length-frequency data gathered from fishers and/or research surveys; and (iii) length data that are directly aged via otolith readings, scales, and or calcified structures such as spines. It is common to use tag-recapture data to estimate the parameters of a potential growth model. However, tagging studies can be costly and ineffective for some species, especially crustaceans (McPherson, 2002; González-Vicente et al., 2012). Length-frequency data are recordings of individual lengths at a time point and can be interpreted as a mixture distribution; this interpretation allows for the identification of separate modes, which may be attributed to different age cohorts. Following the modes of these cohorts through time provides a natural way to model growth of an aquatic species (Pauly and Morgan, 1987; Montgomery et al., 2011).

The von Bertalanffy growth model (VBGM) is commonly used in fisheries research. The VBGM has a biological basis, with its derivation stemming from a differential equation that models growth rate as the difference between the rate of anabolism versus the rate of catabolism. Since its inception (Von Bertalanffy, 1938), the VBGM has remained one of the models of choice for representing growth in aquatic species largely due to its biological interpretability and success in outperforming competing multi-parameter functions (Pauly, 1979; Chen et al., 1992; Essington et al., 2001). In its length form, the solution to von Bertalanffy's differential equation with relevant assumptions is

$$L(t) = L_\infty\{1 - \exp[-K(t - T_0)]\},$$

where L_∞ represents the asymptotic length, K controls the curvature, and T_0 is defined to be the time when an individual would have had length 0 if its post-juvenile growth

stage is projected backwards. In this article, we consider L_∞ to be a random variable with expectation μ_∞ and variance σ_∞^2, although we do not explicitly model this variability. It is debatable whether both L_∞ and K should be considered random, because there are problems with over parameterization. In this article, we treat K as a common population parameter denoted by k; Wang et al. (1995) and Eveson et al. (2007) detail the principal reasons for doing so.

This work was motivated by the Department of Agriculture and Fisheries, Queensland, Australia, who were concerned about declining catches of the blue swimmer crab (BSC) (*Portunus armatus*). This research was part of a larger study to assess the reasons for declining catch rates including: evidence associated with overfishing, sperm limitation, disease, and flood events. Growth estimation was a primary part of this assessment and, thus, data were required to study the growth process. As blue swimmer crabs cannot be directly aged, tag-recapture data and length-frequency data are commonly used. However, tagging data are difficult to gather for crustaceans because exoskeleton molting (the process of shedding of a crab's exoskeleton before expansion and rehardening) leads to loss of tags (Potter et al., 1991). Additionally, tagged crabs are often recaught quickly by the fishery before molting can occur. This leads to measurement error because the new measurement is being taken close to the old measurement resulting in tag-recapture data that are often unreliable. Length-frequency data are simpler to obtain, but fewer models have been developed for growth estimation that can incorporate individual variability and time-dependent covariates such as seasonality; the BSC has a known seasonal growth pattern with growth hibernation in the winter (Kumar et al., 2003). To provide improved estimates of growth for inclusion in the larger stock assessment, we set out to use length-frequency data to make inference about individual variability and seasonality of BSC growth with a fast and robust estimation procedure.

Approaches that incorporate the estimation of a parametric growth curve into the length-frequency analysis have been developed (Macdonald and Pitcher, 1979; Wang and Somers, 1996; Fournier et al., 1998; Leigh and Hearn, 2000). In their seminal work, Macdonald and Pitcher (1979) showed how mixture modeling could be applied to fisheries length-frequency data. Fournier et al. (1998) presented a likelihood-based method that used a mixture of normal distributions with one component distribution for each age class. In other methods, Laslett et al. (2004) used a two step approach by first fitting a finite mixture of normal distributions to the length-frequency data and then modeled the mean estimates with a growth curve. Eveson et al. (2004), and more recently Dortel et al. (2015), incorporated a likelihood component for tag-recapture data, length-frequency data, and direct age-length data simultaneously. Growth model estimation using length-frequency data is especially challenging in short lived species (like BSC) because the number of cohorts is often small. This is compounded with the difficulty that the data collection time frame is usually only 1 or 2 years. The existing methods rely on the following assumptions: (i) recruitment is continuous and constant throughout the year or is a pulse function as in Hoenig (1987); (ii) each individual follows the same VBGM; (iii) the total instantaneous mortality is independent of time, i.e., both natural and fishing mortality rates are constant; and (iv) the natural mortality is known and fixed at a predefined value (using prior knowledge of the species). Using our methodology we consider (i) the recruitment pattern is *arbitrary* and *unknown*; (ii) *any* time-dependent explanatory variables affecting growth (e.g., seasonal changes in growth rate); and (iii) individual variability, which allows for different sized individuals at the same age.

In this article, we develop a new framework for estimating growth from length-frequency data that is generalizable across species and number of cohorts. We model growth using a generalized von Bertalanffy growth framework (Wang, 1999), which incorporates seasonal changes in growth over a yearly time scale. The seasonal model is further constrained such that a hibernation period can be modeled. We use the minorization–maximization (MM) algorithm paradigm (Hunter and Lange, 2004) with a Nelder–Mead (Nelder and Mead, 1965) step to optimize the likelihood and estimate the parameters of the mean constrained mixture of normal distributions for each of the length-frequency sub-samples. The variances are constrained to be a function of the mean length for each cohort, which reduces the number of parameters. Given this model, we show that the MM algorithm is guaranteed to monotonically increase the likelihood and that there is a consistent local maximizer of the log-likelihood function (taken to be the maximum likelihood estimate). The method was coded in the R programming language (R Core Team, 2015) and is available at the *Biometrics* website on Wiley Online Library. We test the method with a simulation study and then apply it to the BSC fishery data.

2. Data and Methods

2.1. *Data*

The data available for the BSC fishery were scarce. Therefore, the final data set was a combined length-frequency data set made from separate data sets gathered using trawl and pot fishing gear over the years of 1984, 1985, and 1986. In our analysis both males and females were used even though the fishery only harvests males. The final data set used for analysis comprised 15,065 crabs.

The first data set was gathered using trawling methods described in Sumpton et al. (1994). The second data set was gathered via correspondence with commercial fishers using pots to target BSC. These data included length measurements from all female crabs captured but only male crabs greater than 155 mm because these males were targeted by fishers. These crabs were caught in offshore regions where larger crabs are known to migrate after full maturity. Therefore, the pot data consist of the full distribution of lengths for larger female crabs but a truncated distribution for male crabs.

2.2. *Growth Model Estimation for Length-Frequency Data*

Our objective was to use length-frequency data to estimate the parameters of the VBGM and a seasonal growth curve that models the change in growth through the year. We were also interested in estimating the dynamics of length-frequency cohorts for aquatic species (in particular BSC). It is important to note that the term cohort is defined to mean those individ-

uals that belong to separate modes or distributions. The term cohort can be thought of as a mixture component and will be used interchangeably in this sense throughout the following derivation. In fisheries, cohorts are hypothetical sets of individuals that come from the same pulse of reproduction; for many aquatic species, this reproductive event occurs annually. Therefore, we make the assumption that separate cohorts are assumed to be separated in age by 1 year. We also note that the length-frequency data are partitioned into monthly periods, with the length-frequency distribution for each month modeled with a finite mixture of normal distributions.

To model the length-frequency data, we used a finite mixture (of normal distributions) model approach, where the means of each of the length-frequency cohorts are constrained to follow a VBGM. The traditional VBGM assumes that the growth rate can be modeled by

$$\frac{dL}{dt} = k[L_\infty - L(t)],$$

where (k, L_∞) are growth parameters, L_∞ is the asymptotic length, and k determines the shape of the curve. We used the following generalized von Bertalanffy model proposed by Wang (1999) to model time-dependent covariates. The change in length with respect to time can be modeled as

$$\frac{dL}{dt} = [L_\infty - L(t)]g(\boldsymbol{\psi}, \mathbf{x}_t) + \sigma(t)\varepsilon(t), \qquad (1)$$

with $\varepsilon(t)$ being a zero mean error term representing the environmental perturbation and $\sigma(t)$ potentially accounting for the heteroscedasticity of the error process. Growth rate explanatory variables (such as seasonality) are incorporated into the model via a link function denoted $g(\boldsymbol{\psi}, \mathbf{x}_t)$, where $\boldsymbol{\psi}$ can be a vector or a scalar. The dynamic variable $\mathbf{x}_t$ represents the explanatory variables (deterministic) believed to be related to the growth rate. The solution to the differential equation (Wang, 1999) with initial conditions $L(b_0) = l_0$ (b_0 is time at birth) has the form

$$L(t) = l_0 + [L_\infty - l_0]\{1 - \exp[-z(b_0, t)]\} + w(b_0, t). \qquad (2)$$

In (2),

$$z(b_0, t) = \int_{b_0}^{t} g(\boldsymbol{\psi}, \mathbf{x}_u) \, du \text{ and} \qquad (3)$$

$$w(b_0, t) = \int_{b_0}^{t} \left\{ \exp\left[\int_{t}^{u} g(\boldsymbol{\psi}, \mathbf{x}_s)ds\right] \sigma(u)\varepsilon(u) \right\} \, du, \quad (4)$$

where $w(b_0, t)$ is a random variable with zero mean representing the aggregate effect of environmental perturbations (Wang, 1999). This model allows for time-dependent covariates to be incorporated into the VBGM (for the following derivation we assume $\mathbf{x}_t$ to be t). For many aquatic species (including BSC), the growth rate is known to be seasonal with a potential hibernation period during the winter months. Modeling the potential no-growth period was achieved by restraining the growth link function to be positive during the

integration step of the model derivation. Depending on the choice of link function, the integral in equation (3) is of varying difficulty. The difficulty rests on whether the roots of the trigonometric functions can be obtained. Seasonality over a year was modeled with the following link function

$$g(\boldsymbol{\psi}, t) = \max[k + \theta_1 \cos(2\pi t) + \theta_2 \sin(2\pi t), 0],$$

where $\boldsymbol{\psi} = (k, \theta_1, \theta_2)'$. Given the stated model, the roots for the trigonometric functions can be found and incorporated into the algorithm by letting the integral equal 0 between the two roots. An outline on how to incorporate this into the model is given in Lloyd-Jones et al. (2014).

The means of each cohort were hypothesized to follow a VBGM curve. Let t_i be the time caught for the ith individual $(i = 1, \ldots, N)$ in fraction of the year since January 1. Since the data are grouped by months for analysis, t_i is taken to be the midpoint of the month; i.e., for individuals caught in month m, $t_i = (m - 0.5)/12$ for months indexed $m = 1, \ldots, 12$. Let L_i (realized value l_i) be the length of individual i caught in year y ($y = 1, \ldots, Y$) at time t_i. Let μ_{jy} be cohort j's mean length in year y, where $j = 0, \ldots, J$ denotes the cohort and J the total number of cohorts. The mean length (under the VBGM) can be parameterized using the Fabens increment equation (Fabens, 1965) as

$$\mu_{jy}(t) = \mu_{0y} + (\mu_\infty - \mu_{0y})(1 - \exp[-z(t_0, t + j)]), \qquad (5)$$

where t_0 is the time from January 1 for the first group of juveniles in the first year modeled. For example, if the first cohort modeled were in February then t_0 is taken to be mid-February and $t_0 = 0.125$. For each year included in the analysis we reset the mean μ_{0y}, which represents the mean length of the first (or juvenile) cohort of the first modeled month. The mean μ_{0y} was reset so that growth was not dependent on the previous year. This allowed peak recruitment in the juveniles to have a different mean for each year.

At each month m, we let the density of individual i be given as

$$f(l_i) = \sum_{j=0}^{J} \pi_{jmy}\phi(l_i; \mu_{jmy}, \sigma_{jmy}^2), \qquad (6)$$

where π_{jmy} are the mixing proportions (with $\pi_{Jmy} := 1 - \sum_{j=0}^{J-1} \pi_{jmy}$) for each cohort and month, respectively, and $\phi(l_i; \mu, \sigma^2)$ is the normal probability density with mean μ and variance σ^2. In equation (6), μ_{jmy} is defined to equal equation (5) evaluated at t_i (the midpoint of month m in fraction of a year). The variances for each cohort in each month were constrained to be functions of the mean length of the cohort with the functional form

$$\sigma^2(\mu_{jmy}) := a\mu_{jmy}\exp(-b\mu_{jmy}), \qquad (7)$$

which for brevity we write as σ_{jmy}^2. This was done to reduce the parameter dimension and to allow for the variance to be known at any mean length. The variance terms (in the component mixtures) are an accumulation of variability in L_∞

and many other factors including size-dependent mortality, recruitment variability, and measurement errors. The variance function used was a Ricker function, which has precedence in the fisheries and ecological modeling literature (Ricker, 1954). This function has the desired properties of being positive if a is constrained to be positive, unimodal, and asymptotes toward 0.

The likelihood function can be written as

$$\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{y=1}^{Y} \prod_{m=1}^{M} f(l_i)^{\mathbb{I}(m_i=m)\mathbb{I}(y_i=y)}, \tag{8}$$

where $\mathbb{I}(m_i = m)$ represents the indicator function, which takes value 1 if individual i's month m_i is equal to the month of interest m and zero otherwise (similarly for the year index) and m runs over all months being modeled for year y. To estimate $\boldsymbol{\theta} = (\boldsymbol{\psi}, \mu_\infty, \boldsymbol{\mu}_0', \boldsymbol{\pi}_0', \dots \boldsymbol{\pi}_{J-1}', a, b)'$, where $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0Y})'$, $\boldsymbol{\pi}_j = (\boldsymbol{\pi}_{j1}', \dots, \boldsymbol{\pi}_{jY}')'$, $\boldsymbol{\pi}_{jy} = (\pi_{j1y}, \dots, \pi_{jM_y})'$, we employed the MM algorithm paradigm (Hunter and Lange, 2004).

There are $M - R$ (where R represents the number of months that contain recruitment) parameters for the mixing proportions, and Y starting means. Using the MM algorithm allows for an optimization scheme to be devised without the constraints of the probabilistic setup of the Expectation–Maximization (EM) algorithm (McLachlan and Krishnan, 2008). The online Supplementary Materials section (Web Appendix A) provides the derivation and summary of the MM algorithm used, its convergence properties, and a proof that there exists a consistent local maximizer of the log-likelihood function. Model selection is performed with the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz et al., 1978). Once a model is chosen, the goodness of fit (GOF) is assessed via the Anderson–Darling (AD) test (Anderson and Darling, 1952) to evaluate the difference between the empirical cumulative distribution function (ECDF) and the estimated CDFs for each subpopulation and each month.

3. Simulation Studies

3.1. *Setup*

Four simulation scenarios were designed to test the method's robustness to departure of the data from the underlying modeling assumptions. Simulation 1 imitates the model with no deviations from its assumptions and provides a base case for comparison. The simulation parameters were chosen to conform to the expected parameters (from previous knowledge) for the BSC. Three cohorts were simulated and included a juvenile cohort (less than 1 year old), a 1 year old adult cohort, and a 2 year old cohort. The mean of each of the cohorts followed a VBGM with seasonal link function with parameter vector $\boldsymbol{\psi} = (1, 2, 2)$. This constrained the seasonal curve to have a no growth period between mid-June to mid-October and to have a peak in February. The parameters of the seasonal link function were chosen to test the method's ability to model a no growth period, which was expected but may be a deviation from the true parameters for the BSC. We set the mean asymptotic length $\mu_\infty = 190$,

initial length $\mu_{01} = 60$, and the variance function parameters $a = 10$, $b = 0.02$ in equation (7). Given the parameters, we simulated 1 year of data with 3000 individuals drawn for each month from $N(\mu_{jm1}, \sigma_{jm1}^2)$ over all $j = 0, 1, 2$ and for each $m = 1, \dots, 12$. The mixing proportions were generated such that juvenile proportions diminished over the year and were incorporated into the 1 year old cohort as the year proceeded (simulated mixing proportion parameters are presented in Web Appendix B). The proportion of individuals in the largest class remained approximately equal over the year. The total data set comprised 36,000 simulated individuals with each simulation scenario run 1000 times. This number of individuals and simulation runs were used for all simulations below.

Simulation 2 tested the methods robustness to the number of cohorts or mixture groups present in the data. It may be difficult to decide how many distributions best fit the data and thus this simulation was used to investigate the methods ability to estimate key model parameters given a misspecified number of mixtures. Simulated individuals were generated under the same parameters as Simulation 1 but with only two cohorts. When the method was implemented, the program was set to estimate three distributions despite the data only containing two simulated cohorts.

Simulation 3 examines the model's assumption of component normality. To challenge the method, three different distributions were used for each cohort. The juvenile cohort was drawn from a gamma distribution with shape and scale parameters determined by the monthly mean and variance calculated under the VBGM and variance function in Simulation 1. The gamma distribution was chosen for group one to simulate a potential longer tail for the juvenile class. The 1 year old cohort was simulated from a log-normal distribution (to deviate from normality) with location and scale parameters again calculated from the monthly mean and variance values. The 2 year old cohort was simulated under a truncated normal distribution with a left hand side truncation at 175 and means and variances as in simulation one. The truncated normal distribution was chosen to investigate the method's ability to deal with a truncation at a minimum size length that is often present in fisheries, which was the case for large male BSCs caught in pots.

Simulation 4 tests the model's assumption of cohort means constrained to follow a VBGM, by simulating under a generalized Gompertz growth model. The generalized Gompertz growth model of Lloyd-Jones et al. (2014) was used to constrain the means of the cohorts with all other parameters as in Simulation 1 used. The method was then fitted with the means assumed to follow a VBGM to estimate model parameters.

3.2. *Results*

The method was applied to each of the simulation scenarios. Estimates of key model parameters are summarized in Table 1 with more extensive summaries of mixing proportions, cohort means, and estimated fitted mixture distributions to simulated data (especially for Simulations 2, 3, and 4) available in online Supplementary Materials (Web Appendix B). The parameter estimates for Simulation 1, agreed very well with the true values with less than 1% error for all key model pa-

Table 1
Summary of key mean model parameter estimates, standard errors, percent error, and coefficient of variation (CV) for the four simulation studies over 1000 replicates

		k	θ_1	θ_2	μ_∞	μ_{01}	a	b
	True values	1.00	2.00	2.00	190	60	10	0.02
Simulation 1								
	Estimates	1.00	2.00	2.00	189.9	60.0	10.0	0.020
	Standard errors	0.030	0.039	0.036	0.29	0.25	0.34	0.0003
	% error	0.0	0.0	0.0	0.05	0.0	0.2	0.0
	CV (%)	3.0	2.0	1.8	0.2	0.4	3.4	1.5
Simulation 2								
	Estimates	1.02	2.00	2.00	189.5	60.0	10.3	0.020
	Standard errors	0.066	0.050	0.048	1.1	0.34	0.63	0.0005
	% error	2.0	0.0	0.0	0.3	0.0	3.0	0.0
	CV (%)	6.5	2.5	2.4	0.6	0.6	6.1	2.5
Simulation 3								
	Estimates	0.914	2.07	2.04	190.2	60.1	10.1	0.020
	Standard errors	0.039	0.050	0.047	0.27	0.33	0.49	0.0003
	% error	8.6	3.5	2.0	0.1	0.2	1.0	0.0
	CV (%)	4.3	2.4	2.3	0.1	0.5	4.9	1.5
Simulation 4								
	Estimates	0.823	1.12	1.16	199.2	57.6	10.5	0.020
	Standard errors	0.017	0.021	0.019	0.50	0.23	0.36	0.0002
	% error	17.7	44.0	42.0	4.8	4.0	5.0	0.0
	CV (%)	2.1	1.9	1.6	0.3	0.4	3.4	1.0

rameters, a maximum coefficient of variation (CV) of 3.4%, and a good fit of the estimated mixture distribution to the simulated data (Table 1 and Figure 1).

Simulation 2 showed close agreement between estimated parameters and the true parameters with a maximum of 2% error and 6.5% CV across all estimated model parameters. Simulation 2 challenged the model in the number of mixture groups, with this most evident in the estimates of the mixing proportions (Web Table 2). Mixing proportions for the third component had estimates that were close to zero.

The results for Simulation 3 showed larger deviances between model parameters and true values. The percent error was largest for k, with 8.6% error. However, the % CV for each of the parameters remained small. In this model scenario the deviation of component normality resulted in an underestimate of k.

Simulation 4 showed the largest deviances between model parameters and true values. Significant deviations from the true parameters of the seasonal curve were evident, especially for θ_1 (44% error). The misspecification of the growth curve resulted in a large underestimation of k and overestimation of μ_∞, although the percent error for μ_∞ was much smaller than that of k. Large deviations in parameter estimates were seen in this simulation but again the % CV remained small for all model parameters (Table 1).

4. Application to BSC

4.1. Setup
The presented model is general in that it can be used to model any number of cohorts over any number of years. The follow-

ing description outlines the constraints that were placed on the model based on the data and prior knowledge of the BSC fishery. Further descriptions and summaries of the initial data investigation are available in online Supplementary Materials (Web Appendix C).

Recruitment was defined to be the entering of juveniles to the fishery, that is, when juveniles reach a length susceptible to being caught by the fishery (approximately 50 mm). Months that contain new recruitment were deemed ineligible for use in modeling because the change in the mode is not driven by growth but by smaller individuals entering the fishery. Therefore, differences between these months will not follow a VBGM. The decision rule for when recruitment has stopped was taken to be when the scaled density height of the juvenile population peaks, i.e., the next month's scaled peak is smaller than the previous month's peak. This is quite a subjective notion and will need to be defined by the modeler based on experience and prior knowledge of a species' population dynamics. The month when the density peaked was used as the first month to model for that year.

As a preliminary investigation, the BSC length-frequency data were visualized via frequency histograms for all available months in years 1984, 1985, and 1986 (see Web Appendix C). This allowed for a visual diagnosis of the months to include and exclude from the model depending on whether they contained significant recruitment. The raw data included length-frequency measurements from October, 1984 to June, 1986. However, only the months of February–August, 1985, and February–May, 1986 were included in the analysis. All other months were deemed ineligible due to the recruitment pattern observed in the length-frequency histograms; in total, 11 months were modeled over 2 years.
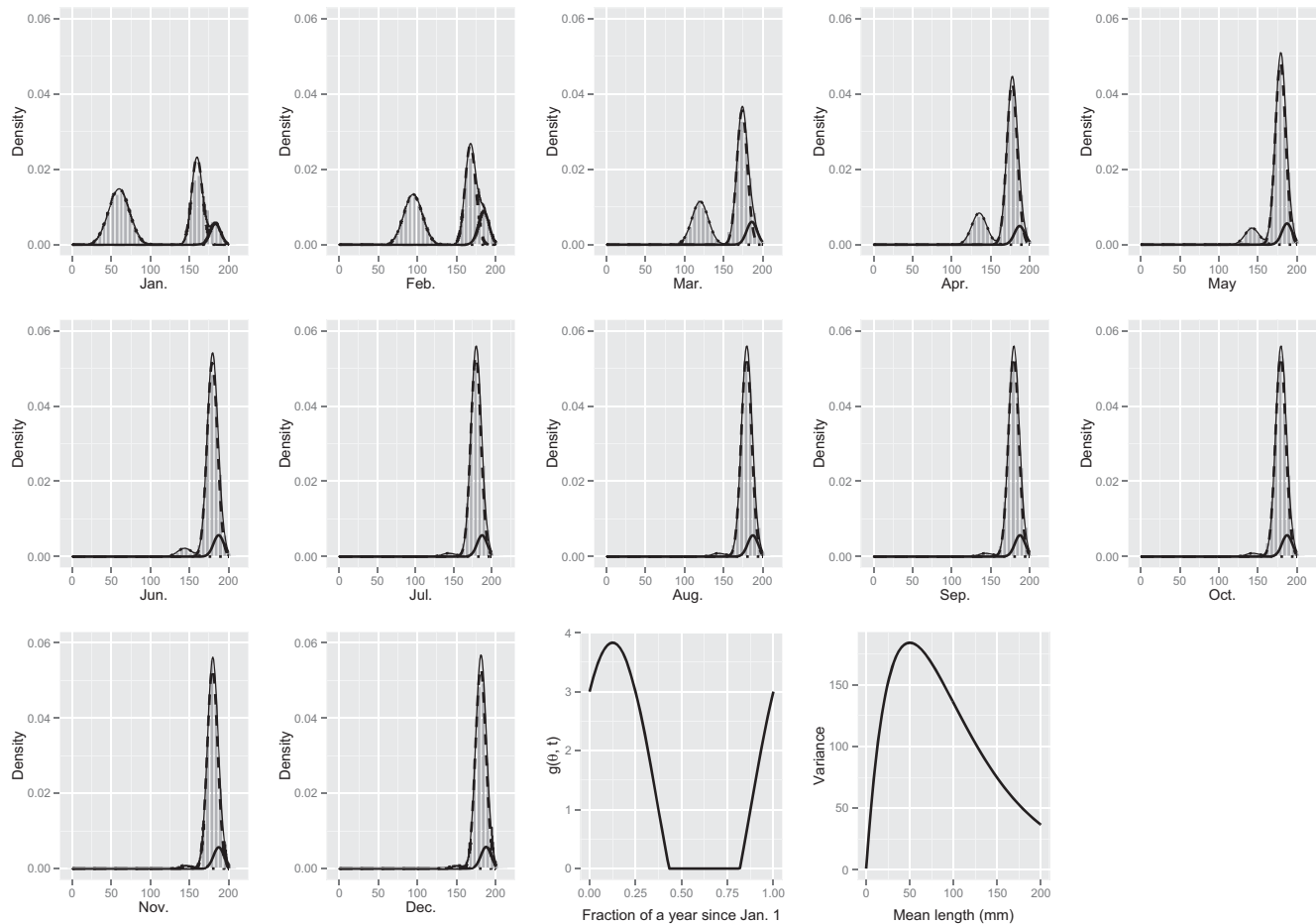
Figure 1. Simulation 1—windows 1–12 represent the mixture fits to the simulated length-frequency data from January to December. The cohort curves are represented with dashed lines (of same width) ordered by length with order juvenile cohort, 1 year cohort, and 2-year-old cohort. The thin solid line represents the sum of the mixture distributions. Window 13 contains the estimated seasonal curve as a function of fraction of a year. Window 14 contains the variance as a function of the mean length.

Additional initial analyses included the fitting of an unconstrained (in means and variances) mixture of two and three normal distributions to the length-frequency data for all available months in years 1985 and 1986 (see Web Appendix C). The BSC is known to be a short lived species, reaching its maximal length within 2–3 years (Potter et al., 1983; Johnston et al., 2011). We concluded that it was likely that the BSC length-frequency data were best modeled with a mixture of three normal distributions (Web Appendix C outlines further evidence for this choice). We hypothesized that these three distributions represented the juvenile, adult, and "close to asymptotic length" cohorts. It was further assumed that the adult cohort (second mixture) was 1 year older than the juvenile cohort and that the "asymptotic cohort" was 2 years older than the juvenile cohort; prior knowledge of the biology of the BSC contributed to this decision. Although a mixture of three normal distributions was hypothesized to be the best model for the data, a mixture of two normal distributions was also fitted and the AIC and BIC calculated to infer the best model.

The method was applied to the combined male and female data to provide estimates for model parameters from a larger number of samples and to provide a baseline comparison for the individual gender runs. The BSC is known to have a sexual dimorphism with respect to growth, therefore, growth parameters were also estimated for males and females separately. BSCs were hypothesized to show growth dormancy during the winter period and thus a seasonal growth curve was estimated. Standard errors for key model parameter estimates were obtained by jackknife resampling (Miller, 1974).

4.2. *Results*

The growth estimation method was applied to the combined, male, and female BSC length-frequency data set. Further descriptions and summaries of the BSC results including mixing proportions and cohort means for all modeled months are available in online Supplementary Materials (Web Appendix D). A GOF analysis for each of the chosen models is presented in the online Supplementary Materials (Web Appendix E).

Table 2

Summary of von Bertalanffy growth model parameter estimates and jackknife standard errors for combined (male and female), male, and female BSC length-frequency data. Estimates are reported for the fitting of a two component mixture of normal distributions (2G) and a three component mixture of normal distributions (3G)

	k	θ_1	θ_2	μ_∞	μ_{01}	μ_{02}	a	b	AIC	BIC
Combined (2G)	2.92	−0.402	0.491	156.2	71.8	84.7	90.6	0.0300	135765	135910
Combined (3G)	0.715	0.315	0.016	179.4	65.3	80.8	29.5	0.0247	135086	135315
Standard errors	0.031	0.182	0.109	2.01	1.18	0.929	4.42	0.001		
Male (2G)	3.45	1.74	−0.542	159.5	80.9	87.8	736	0.0466	76740	76875
Male (3G)	0.565	0.308	−0.473	196.2	73.8	77.3	56.3	0.0305	76392	76601
Standard errors	0.375	0.485	0.450	37.8	7.53	3.98	59.2	0.007		
Female (2G)	2.00	−1.49	1.33	149.0	61.2	81.5	14.6	0.0164	55905	56033
Female (3G)	0.787	0.0574	0.435	161.6	59.6	80.8	11.1	0.0155	55715	55917
Standard errors	0.099	0.630	0.280	5.37	0.921	0.970	1.39	0.001		

For the combined data, the mixture of three normal distributions showed lower AIC and BIC when compared to the two component mixture and thus jackknife standard errors were calculated for the parameter estimates from the three component model (Table 2). The combined results showed that $\mu_\infty \approx 180$ mm and $k = 0.715$. The estimated seasonal curve peaked in January and was lowest in June and does not show a period of no growth (Table 2 and Figure 2). The seasonal link function parameters θ_1 and θ_2 showed the largest jackknife standard errors.

For the male run, the mixture of three normal distributions model showed smaller AIC and BIC than the two component mixture. We saw a mean asymptotic length $\mu_\infty \approx 196$ mm and $k = 0.565$ (Table 2) when the three component mixture was used. The predicted mixture distribution showed poor fit for the first month of February, 1985 with the distribution of the juvenile class for this month appearing to be right skewed, which is particularly evident in the GOF analysis (Web Table 17 and Web Figure 12). The seasonal curve did not contain a period of no growth, but became very close to zero in the period between the months of April–June. The male estimate of k is smaller than that of the combined run with the seasonal curve showing greater amplitude than the combined run and predicts that growth peaks in October for male BSC. Males BSC showed a much smaller variance for the asymptotic length when compared to the combined run. Jackknife standard error estimates for the male parameters were much larger for all parameters when compared to the combined and female runs.

Females also exhibited a smaller AIC and BIC for the three component mixture of normal distributions than the two component mixture. Females were estimated to have a much smaller asymptotic length than males with $\mu_\infty \approx 162$ mm with a higher $k = 0.787$ estimate (Table 2). The seasonal curve had no period of zero growth for females and less amplitude when compared with the males (Figure 4). The seasonal curve peaked in March and was lowest in September. The predicted mixture distribution fitted well to the data, even for the initial juvenile months, with the GOF analysis showing good fit for all but two months (and only marginal p-values for these two

months). The method predicted that juveniles were included in the asymptotic adult group by August, 1985. Again, jackknife standard error estimates were largest for the seasonal link function parameters θ_1 and θ_2.

5. Discussion

Motivated by a decline in BSC stock and data scarcity, we sought to provide a method that could obtain VBGM growth parameters from available length-frequency data and provide extra information regarding the dynamics of the BSC fishery. Understanding seasonality in growth rate was a primary concern because the BSC is expected to show growth dormancy during winter months; this change in growth rate has implications for the larger demographic model used to make management recommendations to the fishery (De Lestang et al., 2003). The seasonal model is further constrained such that a hibernation period can be modeled, which is not present in many other growth model methods. The MM algorithm paradigm was used as the basis for the optimization routine, and a Nelder–Mead step optimized the nonlinear mean components of the likelihood. The variances were also constrained to be a function of the mean length for each cohort, which allows for an estimate of the variance at any mean length. Throughout the modeling process, collaboration with fisheries scientists guided method construction and modeling decisions such as how old the largest cohort should be, model starting parameters, number of groups to model, and months to exclude based on the known recruitment patterns of the fishery.

Although many of the modeling decisions were guided by prior knowledge of the fishery, the model rests on subjective choices such as which months to exclude due to recruitment and how many cohorts to model. The use of the AIC and BIC makes the choice of the number of cohorts to model more defensible, with an initial fitting of an unconstrained mixture of normal distributions recommended. The data used in the current analysis, from the 1980s, are unlikely to represent the current state of the fishery. However, these data were the only data available, which motivated a better modeling solution. The truncation of the length-frequency distribution of male
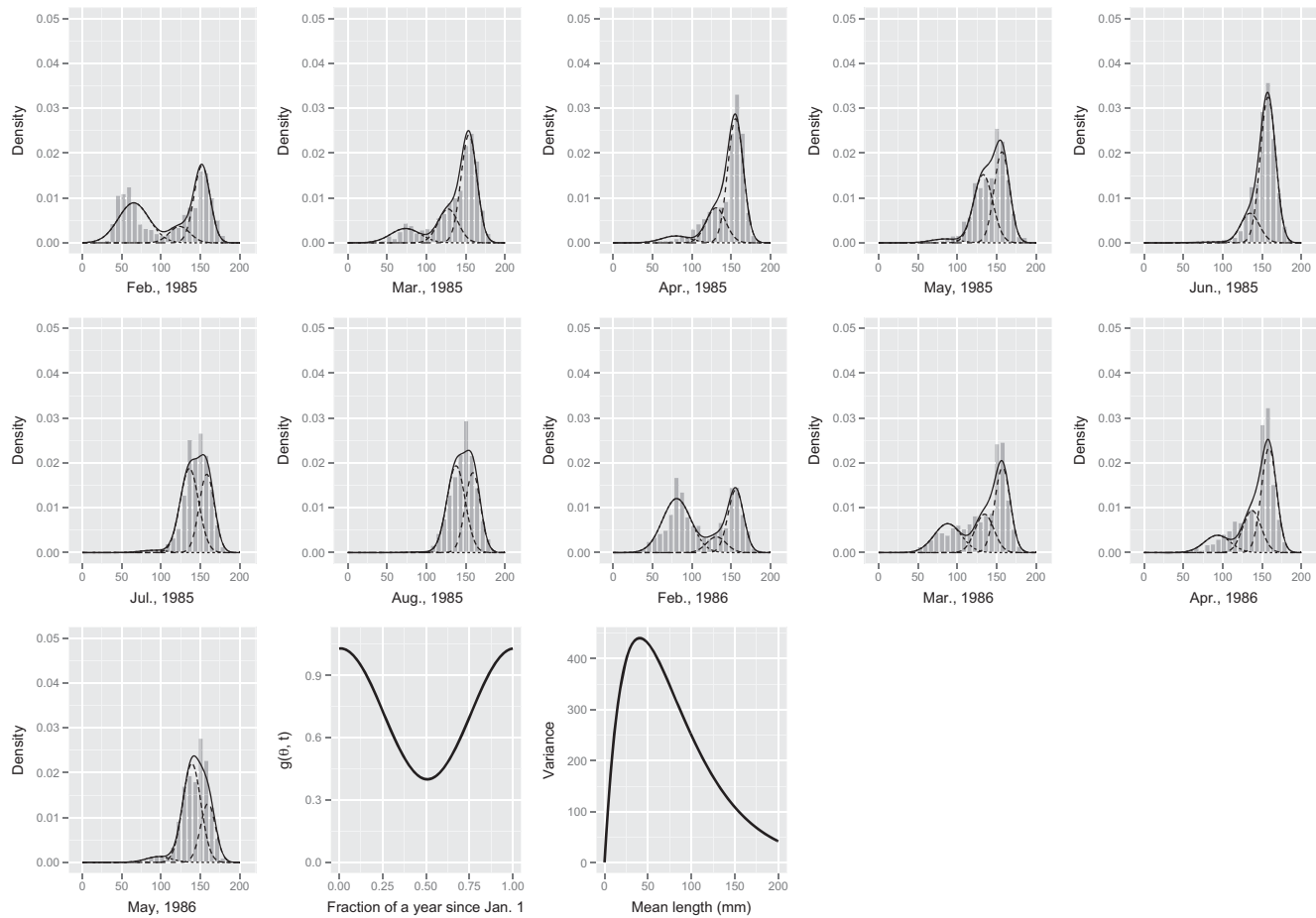
Figure 2. Windows 1–11 represent the mixture fits to the combined (male and female) length-frequency data (mm) with a subset of months excluding recruitment from 1985 and 1986. The cohort curves are represented with dashed lines (of same width) ordered by length with order juvenile cohort, 1 year cohort, and 2-year-old cohort. The thin solid line represents the sum of the mixture distributions. Window 12 represents the seasonal curve as a function of fraction of a year. Window 13 displays the variance as a function of the mean curve.

BSCs (captured in pots) breaks the key modeling assumption of normality. However, the simulations showed that the method is robust to departures from component normality. This truncation is a likely explanation for the poor GOF for the combined and male BSC analyses. Another key assumption is that each of the cohorts is separated in age by 1 year, which is unlikely to be the case given the potential for recruitment over many months for the BSC. The estimates for μ_∞ between the combined and individual runs are in line with the expectation of the fishery. We expected that the male estimates would be greater than the combined run and that male μ_∞ would be closer to 190 mm. The male seasonal is not in line with fluctuations in seasonal ocean temperature. This could be explained by fishing, as the months of February–April are the heaviest fishing months for BSC and, thus, larger crabs in these months are being removed from the length-frequency distribution. This may also be a result of adaptation, given male BSCs show this seasonal relationship but females are in line with expectation and are not fished. The uncertainty in model parameter estimates for the seasonal growth curve pa-

rameters must be considered when interpreting the seasonal curves.

Constraining the means of the cohorts to follow a parametric curve has the advantage of allowing for a mixture distribution to be fitted and growth parameters estimated concurrently. However, there are disadvantages in imposing such constraints such as the a priori choice of a parametric curve for model derivation. This is also reflected in the GOF comparison between the unrestricted mixture of three normal distributions, which shows superior GOF when compared to the constrained model at the cost of many more parameters and no inference on growth. Methods such as that presented in Laslett et al. (2004) allow for a parametric curve to be assumed postestimation of the means of the cohorts, which was explored in the online Supplementary Material. This allowed for visual reassurance that the VBGM was a reasonable choice for these data and we encourage this check before using this method for growth model estimation. The implementation of this method for other growth models, such as the Gompertz growth model, would be useful future work.
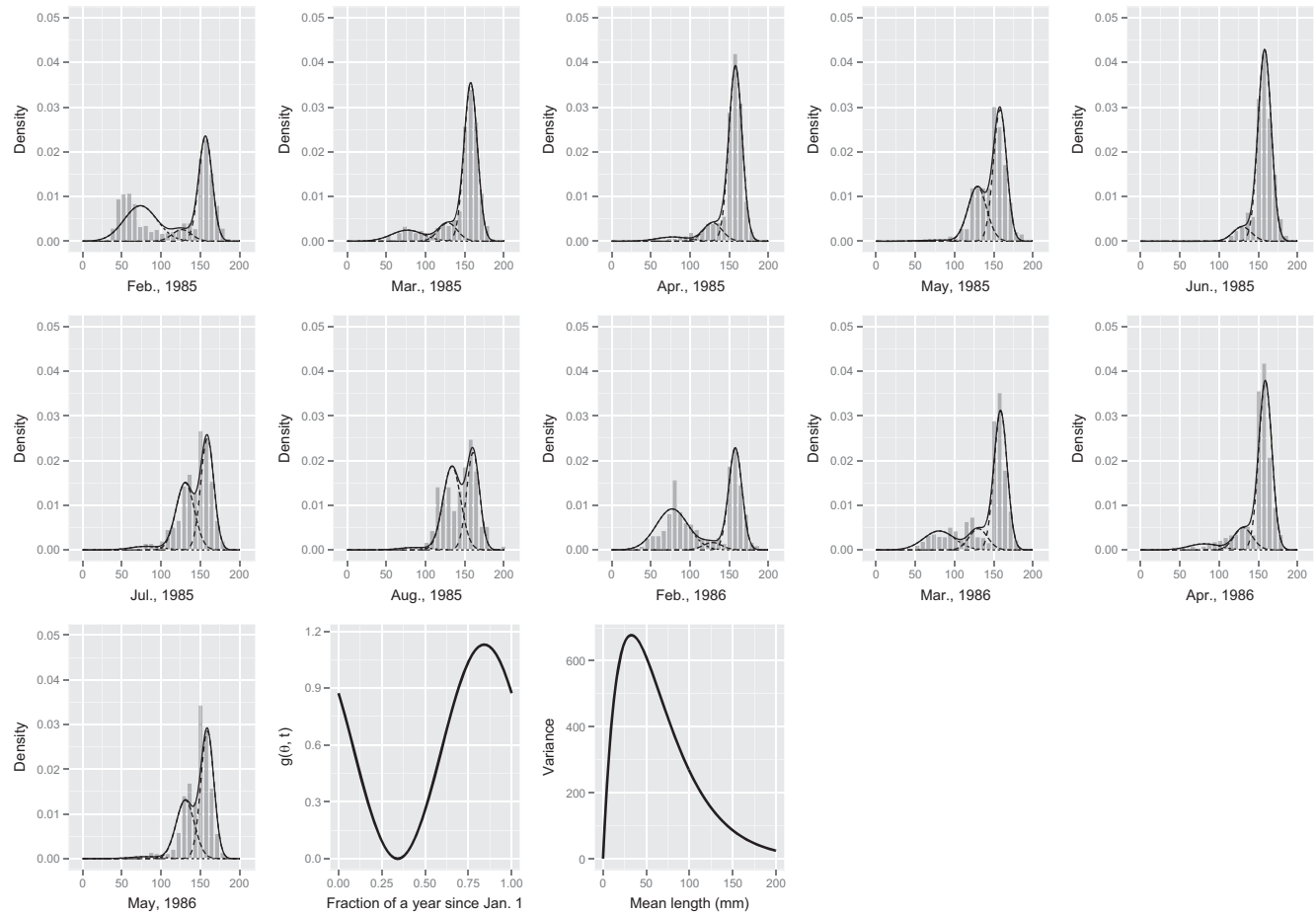
Figure 3. Windows 1–11 represent the mixture fits to the male length-frequency data (mm) with a subset of months excluding recruitment from 1985 and 1986. The cohort curves are represented with dashed lines (of same width) ordered by length with order juvenile cohort, 1 year cohort, and 2-year-old cohort. The thin solid line represents the sum of the mixture distributions. Window 12 represents the seasonal curve as a function of fraction of a year. Window 13 displays the variance as a function of the mean curve.

The nonlinearity of the mean function (with respect to the parameters) made it difficult to establish global convergence properties for our proposed algorithm. One could consider an alternative algorithm based on maximum-incremental steps (in the sense of Razaviyayn et al. (2013) Sec. 5) in order to establish such a result. However, this would lead to an algorithm that may be more computationally intensive; our simulations have demonstrated that the algorithm tended to converge to the required and anticipated limit points. MM algorithms are not applicable when there are no simple minorizers for the objective function under consideration but in this case we could find such minorizers. The MM algorithm relies heavily on local-convexity (or concavity, in the case of maximization) arguments, and derives its power from substituting a difficult optimization problem for a simpler one. This is often at the price of iteration or a slower rate of convergence. We further note that we could not establish the asymptotic normality of our ML estimator, due to the lack of differentiability in the form of the mean function, within the component densities. Unfortunately, the asymptotic normality of ML estimators for

mixture distributions with such component densities is still an open problem and we cannot anticipate its resolution in the immediate future.

The estimates arrived at from this method were motivated by concerns of diminished catches of the blue swimmer crab fishery and the requirement of seasonal parameters for the larger demographic model that would contribute to recommendations to the future management of the BSC fishery in Queensland, Australia. Sudden declines in BSC catches have been observed in other Australian fisheries with the decline attributed to biological factors, the distribution of the species, environmental changes, and fishery management decisions (Johnston et al., 2011). One key factor influencing the BSC is that recruitment success is known to be correlated with water temperature (De Lestang et al., 2003). Thus, the understanding of seasonal changes in growth for BSC was an important component in management decisions for the recovery of the species. It is challenging to assess the growth rate of crabs given the stepwise pattern of crustacean growth caused by molting. Further modeling of the more realistic step pattern
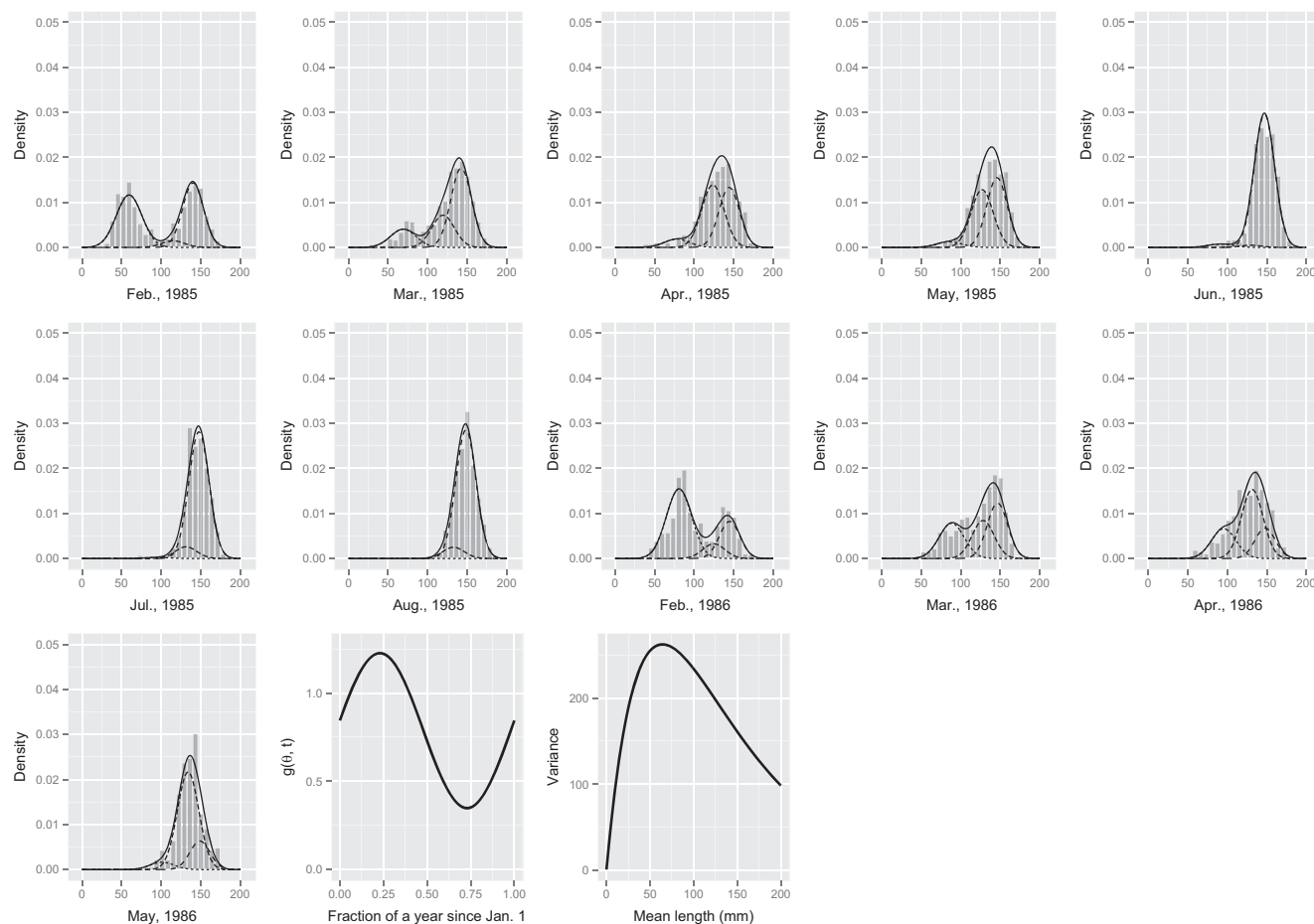
Figure 4. Windows 1–11 represent the mixture fits to the female length-frequency data (mm) with a subset of months excluding recruitment from 1985 and 1986. The cohort curves are represented with dashed lines (of same width) ordered by length with order juvenile cohort, 1 year cohort, and 2-year-old cohort. The thin solid line represents the sum of the mixture distributions. Window 12 represents the seasonal curve as a function of fraction of a year. Window 13 displays the variance as a function of the mean curve.

of crustacean growth is required to estimate BSC growth with higher certainty. Additionally, more recent length-frequency data or tag-recapture data for the BSC would be desirable. We intend on making the model work on a daily time scale rather than restricting growth to monthly periods. However, monthly progression modeling is common in fisheries research (Fournier et al., 1998; Lloyd-Jones et al., 2014; Punt et al., 2014). The availability of length-frequency data in fisheries makes this modeling paradigm desirable and the incorporation of seasonal components as well as a random L_∞ allows for extra information to be drawn from this methodology. Exploring finite mixtures of skew normal or t-distributions (Lee and McLachlan, 2013) to model the cohort distributions would be interesting future work. The method devised contributed to the management outcomes of the BSC fishery and we hope it can be extended and used for other species.

6. Supplementary Materials

Web appendices, referenced in Sections 2.2, 3.2, 4.1, and 4.2 are available with this article at the *Biometrics* website on Wiley Online Library. The blue swimmer crab data analyzed in Section 4 and an R program implementing the new method are available with this article at the *Biometrics* website on Wiley Online Library.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions* **19**, 716–723.
Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

Chen, Y., Jackson, D., and Harvey, H. (1992). A comparison of von Bertalanffy and polynomial functions in modelling fish growth data. *Canadian Journal of Fisheries and Aquatic Sciences.* **49**, 1228–1235.

De Lestang, S., Hall, N., and Potter, I. (2003). Reproductive biology of the blue swimmer crab (*Portunus pelagicus*, decapoda: Portunidae) in five bodies of water on the west coast of Australia. *Fishery Bulletin* **101**, 745–757.

Dortel, E., Sardenne, F., Bousquet, N., Rivot, E., Million, J., Le Croizier, G., and Chassot, E. (2015). An integrated Bayesian modeling approach for the growth of Indian Ocean yellowfin tuna. *Fisheries Research* **163**, 69–84.

Essington, T. E., Kitchell, J. F., and Walters, C. J. (2001). The von Bertalanffy growth function, bioenergetics, and the consumption rates of fish. *Canadian Journal of Fisheries and Aquatic Sciences.* **58**, 2129–2138.

Eveson, J. P., Laslett, G. M., and Polacheck, T. (2004). An integrated model for growth incorporating tag recapture, length frequency, and direct aging data. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 292–306.

Eveson, J. P., Polacheck, T., and Laslett, G. M. (2007). Consequences of assuming an incorrect error structure in von Bertalanffy growth models: A simulation study. *Canadian Journal of Fisheries and Aquatic Sciences.* **64**, 602–617.

Fabens, A. J. (1965). Properties and fitting of the von Bertalanffy growth curve. *Growth* **29**, 265–289.

Fournier, D. A., Hampton, J., and Sibert, J. R. (1998). MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to south pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 2105–2116.

González-Vicente, L., Díaz, D., Mallol, S., and Goñi, R. (2012). Tag loss in the lobster *Palinurus elephas* (Fabricius, 1787) and implications for population assessment with capture-mark-recapture methods. *Fisheries Research* **129**, 1–7.

Hoenig, J. (1987). Estimation of growth and mortality parameters for use in length-structured stock production models. In *Length-Based Methods in Fisheries Research. ICLARM Conference Proceedings*, volume 13, pages 121–128. Manila, Philippines: WorldFish.

Hunter, D. R. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statistician* **58**, 30–37.

Johnston, D., Harris, D., Caputi, N., and Thomson, A. (2011). Decline of a blue swimmer crab (*Portunus pelagicus*) fishery in Western Australia—History, contributing factors and future management strategy RefjtlFisheries Research **109**, 119–130.

Kumar, M. S., Xiao, Y., Venema, S., and Hooper, G. (2003). Reproductive cycle of the blue swimmer crab, *Portunus pelagicus*, off southern Australia. *Journal of the Marine Biological Association of the UK* **83**, 983–994.

Laslett, G. M., Eveson, J. P., and Polacheck, T. (2004). Fitting growth models to length frequency data. *ICES Journal of Marine Science: Journal du Conseil* **61**, 218–230.

Lee, S. X. and McLachlan, G. J. (2013). On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification* **7**, 241–266.

Leigh, G. M. and Hearn, W. S. (2000). Changes in growth of juvenile southern bluefin tuna *(Thunnus maccoyii)*: An analysis of length-frequency data from the Australian fishery. *Marine and Freshwater Research* **51**, 143–154.

Lloyd-Jones, L. R., Nguyen, H. D., Wang, Y.-G., ONeill, M. F., and Tierney, K. (2014). Improved estimation of size–transition matrices using tag–recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* **71**, 1385–1394.

Lloyd-Jones, L. R., Wang, Y.-G., and Nash, W. J. (2014). Generalised growth models for aquatic species with an appli-

cation to blacklip abalone (*Haliotis rubra*). *Ecological Modelling* **272**, 311–322.

Macdonald, P. and Pitcher, T. (1979). Age-groups from size-frequency data: A versatile and efficient method of analyzing distribution mixtures. *Journal of the Fisheries Board of Canada* **36**, 987–1001.

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, 2 edition. New York: Wiley.

McPherson, R. (2002). Assessment of T bar anchor tags for marking the Blue Swimmer Crab *Portunus pelagicus* (l.). *Fisheries Research* **54**, 209–216.

Miller, R. G. (1974). The jackknife—A review. *Biometrika* **61**, 1–15.

Montgomery, S. S., Walsh, C. T., Haddon, M., Kesby, C. L., and Johnson, D. D. (2011). Using length data in the Schnute Model to describe growth in a metapenaeid from waters off Australia. *Marine and Freshwater Research* **61**, 1435–1445.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal* **7**, 308–313.

Pauly, D. (1979). Gill size and temperature as governing factors in fish growth: A generalization of von Bertalanffy's growth formula. Technical report, Institut für Meereskunde.

Pauly, D. and Morgan, G. (1987). *Length-Based Methods in Fisheries Research*, volume 13. Manila, Philippines: WorldFish.

Potter, I., Chrystal, P., and Loneragan, N. (1983). The biology of the blue manna crab *Portunus pelagicus* in an Australian estuary. *Marine Biology* **78**, 75–85.

Potter, M., Sumpton, W., and Smith, G. (1991). Movement, fishery sector impact, and factors affecting the recapture rate of tagged sand crabs, *Portunus pelagicus* (l.), in Moreton Bay, Queensland. *Marine and Freshwater Research* **42**, 751–760.

Punt, A. E., Haddon, M., and McGarvey, R. (2014). Estimating growth within size-structured fishery stock assessments: What is the state of the art and what does the future look like? *Fisheries Research*.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Razaviyayn, M., Hong, M., and Luo, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization* **23**, 1126–1153.

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada* **11**, 559–623.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics* **6**, 461–464.

Sumpton, W., Potter, M., and Smith, G. (1994). Parasitism of the commercial sand crab *Portunus pelagicus* (l.) by the rhizocephalan *Sacculina granifera* boschma, 1973 in Moreton Bay, Queensland, Australia. *Marine and Freshwater Research* **45**, 169–175.

Von Bertalanffy, L. (1938). A quantitative theory of organic growth. Human biology. **10**, 181–213.

Wang, Y.-G. (1999). Estimating equations for parameters in stochastic growth models from tag–recapture data. *Biometrics.* **55**, 900–903.

Wang, Y.-G. and Somers, I. F. (1996). A simple method for estimating growth parameters from multiple length-frequency data in presence of continuous recruitment. *Fisheries Research* **28**, 45–56.

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.