

Including unsexed individuals in sex-specific growth models

| Journal: | Canadian Journal of Fisheries and Aquatic Sciences |
|--|---|
| Manuscript ID | cjfas-2016-0450.R1 |
| Manuscript Type: | Article |
| Date Submitted by the Author: | 17-Feb-2017 |
| Complete List of Authors: | Minto, Cóilín; Galway-Mayo Institute of Technology, Marine and Freshwater Research Centre Hinde, John; National University of Ireland - Galway, School of Mathematics, Statistics and Applied Mathematics Coelho, Rui; Centre of Marine Sciences (CCMAR). Univ. Algarve,; Portuguese Institute for the Ocean and Atmosphere (IPMA, I.P.), |
| Please Select from this Special Issues list if applicable: | |
| Keyword: | Dimorphism, EM algorithm, missing data, non-linear clustering, partial classification |
| | |

SCHOLARONE™ Manuscripts

Including unsexed individuals in sex-specific growth

2 models

³ Cóilín Minto¹, John Hinde² and Rui Coelho^{3,4}

- 4 ¹Corresponding author: Marine and Freshwater Research Centre (MFRC), Galway-Mayo In-
- 5 stitute of Technology (GMIT), Dublin Road, Galway, Ireland. E-mail: coilin.minto@gmit.ie.
- 6 Phone: +353 91 742 514. Fax: +353 91 758 412.
- 7 ²School of Mathematics, Statistics and Applied Mathematics, National University of Ireland,
- 8 Galway, Ireland. E-mail: john.hinde@nuigalway.ie
- ⁹ Portuguese Institute for the Ocean and Atmosphere (IPMA), Av. 5 de Outubro s/n, 8700-305
- 10 Olhão, Portugal. E-mail: rpcoelho@ualg.pt
- ⁴Centre of Marine Sciences (CCMar), Univ. Algarve, Campus de Gambelas FCT Ed. 7, 8005-
- 12 139 Faro, Portugal.

13

14 Abstract

Sexually dimorphic growth models are typically estimated by fitting growth curves to individuals of known sex. Yet, macrospically ascribing sex can be difficult, particularly for immature animals. As a result, sex-specific growth curves are often fit to known-sex individuals only, omitting unclassified immature individuals occupying an important region of the age-length space. We propose an alternative whereby the sex of the unclassified individuals is treated as a missing data problem to be estimated simultaneously with the sex-specific growth models. The mixture model we develop includes the biological processes of growth and sexual dimorphism. Simulations show that where the assumed growth model holds, the method improves precision and bias of all parameters relative to the data ommission. Ability to chose the correct combination of sex-specific and sex-generic parameters is also improved. Application of the method to two shark species -where sex can be ascribed from birth- indicates improvements in the fit but also highlights the importance of the assumed model forms. The proposed method avoids discarding unclassified observations thus improving our understanding of dimorphic growth.

- 30 Key words: Dimorphism; EM algorithm; missing data; non-linear clustering; partial classifi-
- 31 cation.

Introduction

Growth is a central process in the life history of an organsim (Beverton and Holt 1957; Werner 33 1986; Starck and Ricklefs 1998; Nentwig 2012) and describing growth is a regular task across 34 many taxa (Lacointe 2000; Krebs and Cowan 1962; Starck and Ricklefs 1998; Jones et al. 35 2011). Reflecting the importance of understanding growth, modelling growth has a long and 36 productive history in ecology (Robertson 1923; Brody 1926; von Bertalanffy 1938). A com-37 monly applied parametric functional form for the relationship between length and age of given 38 life history stages is the von Bertalanffy model (von Bertalanffy 1938; Chen et al. 1992). Early 39 methods to fit this non-linear model used the linear differenced relationship between mean 40 lengths separated by a unit of age (Ford 1933; Walford 1946). Maximum likelihood fitting under the assumption of normally distributed errors was introduced by Kimura (1980). A so-42 phisticated range of fitting methods now exists, including: non-normal error distributions, more 43 stable re-parameterizations (Gallucci and Quinn 1979; Schnute 1981; Francis 1988), hierarchi-44 cal methods (Helser and Lai 2004; Brunel and Dickey-Collas 2010; Weisberg et al. 2010), and error-in-variables approaches (Cope and Punt 2007). 46 Sexually dimorphic growth occurs in many taxa (Fairbairn et al. 2007). Differential selection 47 pressures, mortality schedules and the degree of reproductive investment affect the ultimate 48 differences in size (Roff 1982; Parker 2006). Where it occurs, sexually dimorphic growth has important implications for management. Natural mortality can be size-dependent (Pauly 1980). 50 Harvesting methods are also often size-based, such that the probability of removal changes with 51 size, e.g., small fish might escape through the mesh whereas larger fish are too big to escape 52 (Myers and Hoenig 1997). Where differences exist in the growth of the sexes, selective size-53 based removal will enact differential mortality between the sexes (Kendall and Quinn 2012). 54 To understand and mitigate for size-selective removals, it is essential that sex-specific growth 55 characteristics are well described. 56 On first inspection, fitting sex-specific growth models to length-at-age data requires that the sex 57 of an individual is known. For maturing and mature animals, sex can often be assigned visually 58 from primary or secondary sexual characteristics, but immature fish, amphibians, insects and 59 reptiles often require histological methods to distinguish between the sexes (e.g., Vitale et al.

2006). Histological methods are often not feasible for large-scale continuous sampling pro-61 grams. As a result, many sex-specific growth datasets contains the sex-designations: Female, 62 Male and Unclassified (FMU). 63 Extant approaches to fitting sex-specific growth curves to FMU data proceed by first removing 64 unsexed (immature or otherwise unclassified) individuals and then fitting sex-specific curves to the remaining known-sex individuals (e.g., using the methods of Kimura (1980)). Sex-specific 66 growth models of this type are thus fit to a subset of the observed age-length data; we term these FMU fits. FMU fits introduce an extrapolation of the model over an omitted space (typ-68 ically young immature animals) that may be very informative for certain parameters of the model, particularly the growth rate and intercept. As an extreme illustration, consider a species 70 with determinate growth (somatic growth ceases post-maturation) with all individuals under-71 going maturation at a given age (when the sex can be determined). Known-sex lengths will 72 be constant over age and therefore only informative for sex-specific asymptotic size. Without 73 additional assumptions, sex-specific growth rates and intercepts will be confounded. In real-74 ity, many species such as many molluscs, fish and reptiles have indeterminate growth (growth 75 continues post-maturation) but we contend that fitting sex-specific growth curves to FMU data 76 in the extant fashion amounts to extrapolation, the severity of which will depend on: ability 77 to determine sex; duration of the immature stage; degree of post-maturation growth; and how 78 protracted the maturation process is across individuals. 79 Instead of omitting unclassified individuals when fitting sex-specific growth models to FMU 80 data, we propose that the sex of the unclassified individuals be treated as a missing data prob-81 lem to be estimated simultaneously with the growth curves. We develop a mixture model 82 for this purpose where few restrictions are made on the form of the growth curves or on the 83 between- or within-individual variability but we do focus on the von Bertalanffy model, as a 84 commonly applied functional form (Pardo et al. 2013). 85 Our objectives are to: (1) develop a mixture model for including unclassified individuals in 86 sex-specific growth curves; (2) develop an accompanying expectation-maximisation (EM) fit-87 ting algorithm (Dempster et al. 1977); (3) test the performance of the method relative to the 88 extant FMU approach via simulation; and (4) fit to two species of shark where known sexes of 89

90 immature and mature individuals provide a useful real data test of performance.

91

2 Materials and Methods

- 93 We first motivate the unsexed individuals as a missing data problem, then propose a mixture
- 94 model solution with an accompanying EM-algorithm estimation routine; simulation tests and
- 95 real data applications follow.

96 A partially classified two-component mixture model

97 The observed data consist of a sample of $i = 1, \dots n$ individuals with recorded variables: length

98 l_i , age a_i , and observed sex $s_{obs,i}$ where $s_{obs,i} \in \{F, M, U\}$ (Female, Male, Unclassified). For

99 the overall population, following the notation of McLachlan and Peel (2000), we define a two-

100 component (female F and male M) mixture model with a length probability density function

101 of:

$$f(\boldsymbol{l}|\boldsymbol{a},\boldsymbol{\Psi}) = \pi_F f_F(\boldsymbol{l}|\boldsymbol{a},\boldsymbol{\theta_F}) + \pi_M f_M(\boldsymbol{l}|\boldsymbol{a},\boldsymbol{\theta_M}), \tag{1}$$

where: $\Psi = \{\pi_F, \theta_F, \theta_M\}$ is a vector of all free parameters; π_F is the mixing proportion,

which is the overall probability that the sex is female, $\pi_F = P(S = F)$, and the complement

104 $\pi_M=1-\pi_F$ is the overall probability that the sex is male; f_S is the sex-specific (female

or male) component density, e.g., lognormal for a single observation (presented without bias

106 correction):

$$f_S(l_i|a_i, \boldsymbol{\theta_S}) = \frac{1}{l_i \sigma_S \sqrt{2\pi}} \exp\left(-\frac{(\ln(l_i) - \ln(g(a_i, \boldsymbol{\Omega_S})))^2}{2\sigma_S^2}\right), \tag{2}$$

where $\theta_S = \{\sigma_S, \Omega_S\}$ is the sex-specific density parameter vector; $g(a_i, \Omega_S)$ is the sex-specific growth function, e.g., von Bertalanffy

$$g(a_i, \Omega_S) = L_{\infty,S} \left(1 - e^{-K_S(a_i - t_{0,S})} \right),$$
 (3)

- where $\Omega_S = \{L_{\infty,S}, K_S, t_{0,S}\}$ is the sex-specific growth curve parameter vector consisting here
- of the mean asymptotic length, Brody growth rate (rate at which asymptote is approached) and
- age at length zero, respectively. Alternative forms for g include other asymptotic, segmented,
- 112 and additive curves.
- 113 To estimate the parameters of the mixture model, we introduce the partially observed indicator
- variable Z that denotes the true sex of observation i

$$z_{i} = \begin{cases} 1, & \text{if observation } i \text{ is female, } s_{i} = F, \\ 0, & \text{if observation } i \text{ is male, } s_{i} = M. \end{cases}$$

$$(4)$$

- Letting i = 1, ..., m index the classified individuals for which the sex is known, we propose
- that for the remaining $j=m+1,\ldots,n$ individuals, z_j can be treated as missing data to be
- imputed from the posterior probability of an unclassified observation being female. Next, we
- introduce how classified and unclassified individuals are treated within the estimation.
- 119 Classified individuals
- For each i, the true value of $z_i \in \{0, 1\}$ and is known. The conditional density of the classified
- 121 observation can therefore be written

$$f(l_i|a_i, z_i, \mathbf{\Psi}) = f_F(l_i|a_i, \boldsymbol{\theta_F})^{z_i} f_M(l_i|a_i, \boldsymbol{\theta_M})^{1-z_i},$$
(5)

and the density of the group (female or male) membership

$$f(z_i|\mathbf{\Psi}) = \pi_F^{z_i} \pi_M^{1-z_i}. \tag{6}$$

- 123 Assuming independence, the joint density of the classified observation and the group member-
- ship can thus be written

$$f(l_i, z_i | a_i, \mathbf{\Psi}) = [\pi_F f_F(l_i | a_i, \boldsymbol{\theta_F})]^{z_i} [\pi_M f_M(l_i | a_i, \boldsymbol{\theta_M})]^{1-z_i}$$
(7)

125 The joint density of all the classified observations is

$$f(\boldsymbol{l}, \boldsymbol{z} | \boldsymbol{a}, \boldsymbol{\Psi}) = \prod_{i=1}^{m} [\pi_F f_F(l_i | a_i, \boldsymbol{\theta_F})]^{z_i} [\pi_M f_M(l_i | a_i, \boldsymbol{\theta_M})]^{1-z_i},$$
(8)

- the right-hand side of which is the likelihood of the classified data for a given set of growth model parameters.
- 128 Unclassified individuals: Expectation step
- 129 Unclassified individuals are typically immature but could also consist of a sample of unsexed
- 130 mature animals. The density of the unclassified observations is a mixture and written, as in
- 131 Equation (1), as:

$$f(l_i|a_i, \mathbf{\Psi}) = \pi_F f_F(l_i|a_i, \boldsymbol{\theta_F}) + \pi_M f_M(l_i|a_i, \boldsymbol{\theta_M}). \tag{9}$$

- For each j, the true value of $z_j \in \{0, 1\}$ but is now unknown. Similar to Dean et al. (2006) and
- 133 McLachlan and Krishnan (2008), a key quantity is the expected value of the group membership
- 134 indicator for unclassified individuals given the observed data and a given set of parameters
- 135 (denoted by an asterisk)

$$E(Z_i|a_j, l_j, \boldsymbol{\Psi}^*) = \Pr(Z_i = 1|a_j, l_j, \boldsymbol{\Psi}^*). \tag{10}$$

136 This pivotal probability is available via Bayes' theorem (Bayes 1764), as:

$$\Pr(Z_j = 1 | a_j, l_j, \mathbf{\Psi}^*) = \frac{\Pr(Z_j = 1) \Pr(l_j | z_j = 1, a_j, \mathbf{\Psi}^*)}{\Pr(l_j | a_j, \mathbf{\Psi}^*)},$$
(11)

$$= \frac{\pi_F^* f_F(l_j | a_j, \boldsymbol{\theta_F}^*)}{\pi_F^* f_F(l_j | a_j, \boldsymbol{\theta_F}^*) + \pi_M^* f_M(l_j | a_j, \boldsymbol{\theta_M}^*)}.$$
 (12)

The estimated component membership probabilities are denoted by z_j^* . Equation (12) thus provides a means for assigning the probability of group (i.e., female or male) membership to unclassified individuals for a given growth model and set of parameters. We can thus proceed, as for classified individuals, to an expected value for the log-likelihood, conditional on the

expected values of the group membership (termed the "complete data log-likelihood")

$$E[\ln L(\Psi|l_j, z_j^*, a_j)] = \sum_{j=m+1}^{n} z_j^* \ln(\pi_F f_F(l_j|a_j, \theta_F)) + (1 - z_j^*) \ln(\pi_M f_M(l_j|a_j, \theta_M)).$$
(13)

- 142 As an aside, the simplicity here is due to the log-likelihood based on a representation, as in
- Equation (8), that is linear in z_i^* . Combining the classified and unclassified individuals, the
- expected complete data log-likelihood of all observations is then given by

$$E[\ln L(\boldsymbol{\Psi}|\boldsymbol{l},\boldsymbol{z},\boldsymbol{a})] = \sum_{i=1}^{m} \left[z_{i} \ln(\pi_{F} f_{F}(l_{i}|a_{i},\boldsymbol{\theta_{F}})) + (1-z_{i}) \ln(\pi_{M} f_{M}(l_{i}|a_{i},\boldsymbol{\theta_{M}})) \right] + \sum_{j=m+1}^{n} \left[z_{j}^{*} \ln(\pi_{F} f_{F}(l_{j}|a_{j},\boldsymbol{\theta_{F}})) + (1-z_{j}^{*}) \ln(\pi_{M} f_{M}(l_{j}|a_{j},\boldsymbol{\theta_{M}})) \right]$$

$$(14)$$

- 145 EM algorithm
- 146 Estimation of the expected log-likelihood (Equation 14) is carried out by the following steps:
- 147 1. *E-step*
- Starting with a given set of parameters $\Psi^{(0)} = \{ \boldsymbol{\theta}^{(0)}, \pi_F^{(0)} \}$, calculate the expected value
- of the log-likelihood (Equation 14), which amounts to finding z_i^* in Equation (12).
- 150 2. *M-step*
- a. Maximise the expected value of the complete data log-likelihood with respect to the growth parameters

$$\boldsymbol{\theta}^{(1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, \mathrm{E}[\ln L(\boldsymbol{\theta}|\boldsymbol{l}, \boldsymbol{z}^{(0)}, \boldsymbol{a}, \pi_F^{(0)})]. \tag{15}$$

- Non-linear growth models require numerical optimisation for this step.
- b. Update the unconditional probability of being female (sex ratio), if not assumed fixed, via

$$\pi_F^1 = \frac{\sum_{i=1}^m z_i + \sum_{j=m+1}^n z_j^{*(0)}}{n},\tag{16}$$

- which assumes the same overall sex ratio for the unclassified and classified; but this can be relaxed to classified or unclassified only, fixed at a given value, or modelled with covariates where data allow.
- 3. Replace $\Psi^{(0)}$ with $\Psi^{(1)} = \{\theta^{(1)}, \pi_F^{(1)}\}$ and repeat steps 1 and 2 until convergence, as assessed by a stopping criteria based on a relative change in the observed log-likelihood. The observed log-likelihood is given by

$$\ln L(\boldsymbol{\Psi}^*|\boldsymbol{l},\boldsymbol{z},\boldsymbol{a}) = \sum_{i=1}^{m} \left(z_i \ln(\pi_F^* f_F(l_i|a_i,\boldsymbol{\theta}_F^*)) + (1-z_i) \ln(\pi_M^* f_M(l_i|a_i,\boldsymbol{\theta}_M^*)) \right) + \sum_{j=m+1}^{n} \ln\left(\pi_F^* f_F(l_j|a_j,\boldsymbol{\theta}_F^*) + \pi_M^* f_M(l_j|a_j,\boldsymbol{\theta}_M^*)\right).$$
(17)

- Note that it may be possible to maximise this partially classified observed likelihood directly without the EM algorithm, but direct maximisation does not in general perform well, except in the vicinity of the maximum or via a grid search.
- 165 Parameter inference and standard errors
 - In the same manner as Kimura (1980), parameters may be bound or free across the sexes. We fit all combinations of bound and free parameters and choose the best fitting model by Bayesian Information Criterion (BIC using the: observed data log-likelihood (Equation 17); log of the number of observations; and number of parameters of the component distribution and the mixing proportion). A preliminary investigation of the performance of the Akaike Information Criterion showed it to select overly complex models compared to BIC. Various options are available to approximate the standard errors of the parameter estimates from an EM algorithm (McLachlan and Peel 2000). A key difficulty is obtaining the observed information matrix. The method we implement is to fit the model to convergence via the EM algorithm as above and then optimise the partially-classified mixture log-likelihood (Equation 17) directly, including the mixing proportion. The EM algorithm finds the maximum likelihood which would be difficult to do via direct optimisation. Once at the maximum, direct optimisation is relatively straightforward and can be used to estimate the curvature of the observed log-likelihood and

179 hence provide standard errors.

Method testing

181 Simulations

180

Our simulation framework study focusses on some of the primary factors influencing the es-182 timation of sex-specific von Bertalanffy growth models for non-hermaphroditic species. Two 183 overall functional types were tested: asymptotic ($L_{\infty,F} = 40 \text{cm}$; $k_F = 0.3 \text{year}^{-1} t_{0,F} =$ 184 -1 years; and $\sigma_F = 0.1$ lognormal error standard deviation); and a more 'linear' form ($L_{\infty,F} =$ 185 40cm; $k_F=0.1 {
m year}^{-1}~t_{0,F}=-1 {
m years})$ (Figure 1). The distribution of ages (50 female and 186 50 male per simulation) was drawn from a negative binomial distribution with a mean age of 187 5 and a shape parameter of one giving an age distribution with 5^{th} and 95^{th} quantiles at 0 and 188 16 years, respectively. Dimorphic differences (20%) for males were included by parameter 189 settings with: $L_{\infty,M}=0.8L_{\infty,F}$; $k_M=0.8k_F$ and $t_{0,M}=0.8t_{0,M}$. Early and late maturation 190 were included by setting the age at which 50% of the population are mature to ages two and 191 five, respectively; a maturation range between 25% and 75% mature was set to one year. To 192 emulate the scenario where immature animals cannot be sexed, we set all immature animals to 193 unclassified sex (see Discussion). These values provided a realistic but pragmatically limited 194 set of sexually-dimorphic differences (Figure 1) to illustrate the main aspects of performance 195 without over-loading with simulation output. All combinations of sex-specific and sex-generic 196 197 parameters (16 models in total) were fit to each simulation dataset and the best fit for each method chosen on the basis of the lowest BIC. Simulation fits were started at the true value 198 199 for all methods. Each simulation scenario was run 1,000 times with different random draws to provide stable sampling distributions. 200

[Figure 1 about here.]

The performance of the methods on the simulated data was appraised using: 1) parameter rootmedian-square error (RMSE); 2) classification rates (based on maximum a posteriori / binary classification); 3) ability to chose the correct model form via BIC; and 4) visually via boxplots of the relative error ($(\hat{\theta} - \theta_{true})/\theta_{true}$) of the parameter estimates.

206 Real data

207 Elasmobranchs (sharks and rays), represent a useful taxon to test the performance of the method given that the sex can be distinguished from birth, owing to two extensions of the pelvic fins 208 (claspers) on the males. We use two datasets on deepwater lantern shark species: velvet belly, 209 Etmopterus spinax, and smooth lanternshark, Etmopterus pusillus, which are found off the 210 south and southwestern coasts of Portugal (Coelho et al. 2005). The datasets consist of 733 (E. 211 spinax) and 518 (E. spinax) age-length readings. The data are cross-sectional in that there is a 212 213 single data point per individual. Females represented approximately 60% (E. spinax) and 43% (E. pusillus) of the samples, respectively. High proportions of the individuals were classified 214 215 as immature (E. spinax: 60%; E. spinax: 77%) based on macroscopic investigation of the reproductive organs (Coelho and Erzini 2007, 2008). 216 To test the performance of the method, we compare sex-specific von Bertalanffy fits from: (1) 217 all individuals, using mature and immature sex determinations; (2) only the mature animals -218 to mimic the FMU situation where sex cannot be distinguished prior to the onset of maturation; 219 and (3) all individuals with the sex of the immature animals treated as an unknown classification 220 problem and fit via the mixture model developed here. A grid of plausible starting values was 221 used to initialise fitting for each method. 222

223 Software

We developed the R package *lhmixr* (*l*ife *h*istory *mix*ture models in *R*: available on CRAN or at https://github.com/mintoc/lhmixr/) to fit sex-specific growth models with missing classifications. von Bertalanffy models are implemented with analytical gradients for all parameters in the complete-data log-likelihood and a generic binding formulation to allow parameters to be bound equal or free between the sexes. Normal and log-normal error distribution assumptions are included. Both real datasets are included and documented in the *lhmixr* package.

Results

230

233

234

235

Simulation performance 231

Example simulation fits show that, where fish can be classified early (Figure 2(A)), the FMU 232 approach can perform well in recovering the full data curve, though with decreased precision (Table 1, Figure 3). In contrast, for late maturing animals the example FMU fit performed poorly for males and females (Figure 2(B)) and in general had increased bias and decreased precision for the growth rate and intercept terms, particularly for strongly asymptotic growth 236 curves (Table 1, Figure 3 (C, E)).

238

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

237

[Table 1 about here.] 239

Overall, as judged by RMSE, the MIX model (mixture model developed here) outperforms the FMU approach for all parameters and both curve typologies (Table 1). The differences in RMSE are most pronounced in the asymptotic fits where the MIX-estimated intercept t_0 often displays an order of magnititude difference in RMSE over the FMU fits (Table 1). For early maturing animals with linear growth, the RMSE is reduced by half or more using the MIX method relative to the FMU approach (Table 1). The MIX method consistently had RMSE slightly larger than the full knowledge fit, reflecting the additional uncertainty when classification is needed (Table 1).

[Figure 2 about here.]

All methods had similar ability to detect sex-specific difference in the asymptotic size, irrespective of curve typology or maturation timing (Table 2, $L_{\infty,s}$ scenarios). Ability to detect differences in the growth rate is considerably reduced in the FMU fits compared to the full data fits, particularly for late maturing animals with more asymptotic growth where the correct model was identified in only 14% of simulations (Table 2, asymptotic growth form and late maturing FMU K_s scenario). This was improved to 46% chosen correct using the MIX model. The performance is not as poor for the more linear growth form (33% FMU chosen correctly), though the MIX model still improved the performance (69% chosen correctly). When t_0 differs by sex and the growth curve typology is asymptotic, the FMU approach performed poorly

267

268

269

270

271

272

275

278

in choosing the correct model (7% and 2% for early and late maturing animals, respectively) 258 (Table 2, FMU $t_{0,s}$ scenarios); this is somewhat improved for early maturing animals using the 259 MIX model (36%) but is low for the late maturing animals (24%) (Table 2, MIX $t_{0,s}$ scenar-260 ios). With asymptotic growth and where t_0 or K differ by sex and the FMU approach is used, 261 the chosen model is often that of no difference between the sexes (Table 2, FMU None row). 262 Ommitted models from Table (2) often included the correct model but with residual standard 263 deviation differing by sex (e.g., where only L_{∞} differs by sex, a model with $L_{\infty,s}$ and σ_s is the 264 265 most commonly chosen model of those not shown).

266 [Table 2 about here.]

Working with the best fitting model per simulation, the classification rates where typically lower when using the FMU model; whereas the MIX model recovered similar but consistently lower classification rates to the FULL data scenario (Table 3). Relative to the FMU approach, large gains in the classification rates are seen where growth is asymptotic (Table 3, third row). As expected, classification rates are poor across all methods where no true differences exist among the sexes (Table 3, "None" columns).

273 [Table 3 about here.]

Reflecting the RMSE results, differences in the relative error across methods are least for the 274 L_{∞} parameter, reflecting the predominant importance of mature animals for estimating that parameter (Figure 3). In many cases the relative error of the MIX model is comparable to that 276 277 of the full data scenario for these simulations (Figure 3).

[Figure 3 about here.]

Real data performance 279

The two elasmobranch species investigated spend a relatively long period of their observed 280 lifespan in the immature phase in addition to a relatively large size at maturation (Figure 4). 281 As a result the best fitting FMU t_0 estimates for female and male E. spinax and E. pusil-282 lus are considerably more negative than when the full data are used (Table 4, sixth column). 283

FMU-estimated intercepts also had considerably decreased precision with no evidence of a 284 sex-specific difference. These biases and lack of precision simply reflect that the FMU model 285 excludes those immature animals and thus only fits to the mature animals. In contrast, the MIX 286 approach results in more comparable estimates to the full data scenario. 287 288 Estimates of L_{∞} are similar across methods and both sexes for E. pusillus. The FMU and mixture models estimate a lower asymptotic size for female E. spinax (Figure 4(C,E)) most likely 289 because they are more influenced by the larger known sex females, which are fit poorly in the 290 291 full data scenario (Figure 4(A)). The E. spinax male asymptotic size is estimated as lower in the FMU fit compared to the full data fit; whereas the mixture model estimated a larger and 292 more uncertain male asymptotic size (Table 4, row 6). 293 The female E. pusillus mixture model estimated growth rate is faster than the full data scenario 294 and results in a higher curve (Figure 4(F)). As a result many of the smaller fish for a given age 295 are classified as male, whereas this is not true from the full data, where there is a mixture of 296 males and females. This may arise from the choice of functional form in that there appears to be 297 a broadly similar growth pattern for until older ages where the male growth slows considerably 298 for both spacies(see Discussion). As the FMU approach has smaller males and the immature 299 animals fall under both curves, the predicted classification is unity for males but very poor for 300 females (Table 4). The mixture model improves on this but goes to a solution that separates 301 the sexes more than is apparent in the true data (Figure 4). Overall the classification rates from 302 all models, including the full data scenario, are poor across both sexes owing to a lack of clear 303 304 separation.

[Figure 4 about here.]

[Table 4 about here.] 306

Discussion

305

307

308

309

Discarding unclassified data degrades the precision and accuracy of fitted sex-specific growth models. Here, we have proposed a redress that simultaneously attempts to classify unclassified individuals while estimating parameters of a growth curve of interest. To our knowledge this 310

is the first recognition of the potential to couple classification and life history model parameter estimation in this manner.

313 Performance

Where the fitting assumptions match the data-generating assumptions, as in our simulations, 314 the mixture model offers general improvements on the estimation of sex-specific growth curves 315 from partially-classified data (Table 1, Figure 3). Since a lognormal distribution and the von 316 317 Bertalanffy model are the most common assumptions for these data we think that the method should also improve fits to real datasets. 318 Where animals can be classified at early ages, the FMU approach may provide good estimates 319 of the parameters of the growth curve (Figure 2). This does have the cost of reduced precision 320 (Table 1, Figure 3), which will decrease the ability to correctly choose differences between 321 the sexes (Table 2). In reality, the impact of the ommission of unclassified individuals will 322 vary by population depending on how early immature animals can be classified. Even where 323 324 classification can occur at a relatively young age, maturity is often dependent on reaching a threshold size (Stearns and Koella 1986). Within an early age-group, larger fish may mature 325 first and hence be distinguished earlier, which may further bias estimates of FMU-derived sex-326 specific growth compared to when all individuals are included, as proposed here. 327 Other solutions in the FMU case may include fixing the intercept at a certain value. Though this 328 may seem a simpler solution, it is not recommended given the resulting bias in the growth rate 329 for single fits (Pardo et al. 2013). In addition, for sex-specific fits, it is difficult to envisage how 330 fixing the intercept term would not bias the other sex-specific parameters where differences 331 exist. 332 Lantern shark datasets investigated demonstrated some extreme challenges to fitting with the 333 FMU approach (Figures 4). In reality, it may be possible to ascribe sex for some individuals 334 prior to maturity but these fits serve to emphasise some of the issues, as discussed below. 335

Caveats

336

337

339

340

341

342

343

344

347

349

351

353

354

357

358

359

360

361

363

364

A primary concern with the von Bertalanffy model is that the functional form of the growth curve differs between the immature and mature phases. Such valid concerns have given rise to 338 alternative mechanistic growth models such as the biphasic growth model of Roff (1983) where somatic growth is linear to the onset of maturity and thereafter a decreasing function of the gonadosomatic index. The von Bertalanffy model theoretically describes only post-maturation growth (Ricker 1975; Lester et al. 2004). We stress, however, that the form of the growth model (g in Equation 3) we use for our derivation or implementation of the algorithm is not restricted and our use of the von Bertalanffy form only reflects its common application. How-345 ever, hockey-stick implementations of sex-specific Roff biphasic growth models could be very difficult to fit to asymptotic datasets containing unclassified immature animals. The mixture 346 model developed here may contribute in this regard. Overall, it is important to recognise that the method classifies on the basis of the assumed model, so a functional form mis-match may 348 result in poor classification (e.g., E. pusillus, Figure 4). It is difficult to guard against this, other than by comparing with more local functional forms. 350 The error distribution assumption will affect the classification - we therefore recommend careful consideration of the error-distribution and the mean-variance relationship thus implied. In 352 the *lhmixr* package we provide two error distributions (normal and lognormal) but envisage other density forms could be useful in particular applications (e.g., gamma, heteroscedastic formulations). We envisage many ways of improving and highlight the E. pusillus fits (Fig-355 ure 4) as a caution of how the performance may be critiqued in given applications. 356 Of additional concern is the assumption that the sex of the individual is fixed. Fish and other taxon display a great diversity of sexual life histories (Warner 2012), including sequential hermaphroditism, where the sex of the animal can change. On first inspection, classifications to such families as Labrids or gobies, for what may be a non-constant trait, could be a less than sensible approach but again we propose a possible development via the inclusion of relevant covariates for the mixing proportions (Equation 16). 362 In terms of model choice and parameter inference, it is known that the regularity conditions of the asymptotic distribution of the deviance break down when comparing mixture models with differing numbers of components (McLachlan and Peel 2000). The same is true of AIC and BIC, though BIC has been shown to perform well in some situations for identifying the number of components (Wang et al. 1996). We recommend first using BIC to compare a fully saturated two component model (all parameters sex-specific) with a single component model, where all parameters are the same across sexes. Once the number of components is decided, and interest lies in parameter significance, nested models within that number of components can be tested using likelihood ratio tests with a chi-squared distribution. Overall, we propose that the asymptotic distribution of the deviance in this partially classified mixture model requires further work.

Management implications

Given that yields and mortality schedules are directly impacted by growth parameters, having reliable estimates of sex-specific growth parameters is paramount for management of dimorphic species. While the mixture model offers general improvements over the unclassified ommission case (Figure 3), we envisage the method assisting most where growth is more strongly asymptotic after maturation (Figure 2). Where faster growing individuals mature earlier and are therefore classifiable earlier, positive biases may exist in previously estimated sex-specific growth curves. This would imply faster growth than when all the data are used. In this setting the method could contribute to management by using all the data to estimate the relevant growth parameters and quantify uncertainty around the sex-specific growth curves. These curves can then be used, for example, in age-length based population assessment models (e.g., Methot 2000; Frøysa et al. 2002).

Future developments

While the approach developed here focusses on partially classified datasets, the mixture model could also be applied to datasets where no sexed animals are recorded, as in common applications to fully unclassified datasets (McLachlan and Peel 2000).

391 Extending the method longitudinally would allow for the life histories of unclassified individ-

uals to be understood in the context of sex-specific differences. Longitudinal features could 392 be built in by modelling the correlation structure of the errors (e.g., using generalized least 393 squares, estimating equations or random effects (Diggle et al. 2003)). 394 Other sex-specific life history relationships may also be amenable to the methods developed 395 here. We are currently developing a similar approach for the treatment of maturity. Perceivably 396 these could be coupled into a multivariate partially classified mixture model, where it is explic-397 itly recognised that animals develop along multiple coupled processes. 398 399 The estimation time is short (seconds) and *lhmixr* von Bertalanffy complete data log-likelihoods are also accompanied by their analytical gradients to provide more stable fitting. The EM al-400 gorithm can, however, be comparatively slow in terms of the number of iterations taken to 401 convergance. This could be improved by implementing the loops in a compiled language (e.g., 402 C++). It would also be of particular interest to start the model at the M-step (e.g., assuming an 403 equal probability of 0.5 for unclassified individuals) and test performance compared to start-404 ing at the E-step with a grid of starting values (as done here) or starting at the FMU solution. 405 Bayesian estimation of finite mixture models is an active area of development (Frühwirth-406 Schnatter 2006), which would assist straightforward estimation of parameter uncertainty using 407 MCMC, for example, but with associated chain convergence and runtime overheads. 408 Despite the long history of application of mixture models to ecological problems (Pearson 409 1895; Bhattacharya 1967; Macdonald and Pitcher 1979) and recent applications (Thorson et al. 410 2011; Cosgrove et al. 2014), we note in closing that the EM algorithm is at present a con-411 412 siderably under-recognised algorithm in ecological modelling. The EM algorithm contributes to areas as diverse as state space modelling (Shumway and Stoffer 1982), structural equa-413 tions (Ullman and Bentler 2003), hidden Markov systems (McLachlan and Krishnan 2008), 414 all of which are important topics in modern ecological modelling, yet other routines (MCMC, 415 Laplace, quadrature approximations to the marginal likelihood) are more commonly used when 416 estimating complex ecological models (e.g., Bolker et al. 2013; Kristensen et al. 2015). While 417 recognising the central importance of these methods, we contend that for many users there is a 418 lack of understanding as to how the methods work. In contrast, the steps of the EM algorithm 419 420 are transparent when estimating unobserved constructs. We believe that the EM algorithm de-

- veloped here demonstrates a small amount of the potential of this family of algorithms, which
- 422 deserve more recognition in modern ecological modelling.

423 Acknowledgements

- 424 We are grateful to Dr Guillaume Bal, Marine Institute, Ireland for valuable feedback. Two
- anonymous reviewers are thanked for greatly improving the manuscript. We thank John Corco-
- 426 ran for initial discussions on sex-specific growth models for perch that instigated development
- of the method. CM thanks Dr Conor Nolan for encouraging insights on animal growth.

28 References

- Bayes, T. 1764. An essay toward solving a problem in the doctrine of chances. Philosophical
- Transactions of the Royal Society of London **53**: 370–418.
- Beverton, R.J.H. and Holt, S.J. 1957. On the Dynamics of Exploited Fish Populations. Fisheries
- Investigations Series 2, Vol. 19. Ministry of Agriculture Fisheries and Food, London, UK.
- Bhattacharya, C.G. 1967. A simple method of resolution of a distribution into Gaussian com-
- 434 ponents. Biometrics **23**: 115–135.
- 435 Bolker, B.M., Gardner, B., Maunder, M., Berg, C.W., Brooks, M., Comita, L., Crone, E.,
- Cubaynes, S., Davies, T., Valpine, P. et al. 2013. Strategies for fitting nonlinear ecological
- models in r, ad model builder, and bugs. Methods in Ecology and Evolution 4(6): 501–512.
- 438 Brody, S. 1926. Time relations of growth: I. Genetic growth constants of animals. The Journal
- 439 of General Physiology **8**(3): 233–251.
- 440 Brunel, T. and Dickey-Collas, M. 2010. Effects of temperature and population density on
- von bertalanffy growth parameters in atlantic herring: a macro-ecological analysis. Marine
- Ecology Progress Series **405**: 15–28.
- 443 Chen, Y., Jackson, D.A., and Harvey, H.H. 1992. A comparison of von bertalanffy and poly-
- 444 nomial functions in modelling fish growth data. Canadian Journal of Fisheries and Aquatic
- 445 Sciences **49**(6): 1228–1235.

- 446 Coelho, R. and Erzini, K. 2007. Population parameters of the smooth lantern shark, *Etmopterus*
- pusillus, in southern portugal (NE Atlantic). Fisheries Research **86**(1): 42–57.
- 448 Coelho, R. and Erzini, K. 2008. Life history of a wide-ranging deepwater lantern shark in the
- north-east Atlantic, Etmopterus spinax (chondrichthyes: Etmopteridae), with implications
- for conservation. Journal of Fish Biology **73**: 1419–1443.
- 451 Coelho, R., Erzini, K., Bentes, L., Correia, C., Lino, P.G., Monteiro, P., Ribeiro, J., and
- Gonçalves, J.M. 2005. Semi-pelagic longline and trammel net elasmobranch catches in
- southern portugal: catch composition, catch rates and discards. Journal of Northwest At-
- 454 lantic Fishery Science **35**: 531–537.
- 455 Cope, J. and Punt, A. 2007. Admitting ageing error when fitting growth curves: an example
- using the von Bertalanffy growth function with random effects. Canadian Journal of Fisheries
- and Aquatic Sciences **64**(2): 205–218.
- 458 Cosgrove, R., Sheridan, M., Minto, C., and Officer, R. 2014. Application of finite mixture
- models to catch rate standardization better represents data distribution and fleet behavior.
- 460 Fisheries Research **153**: 83–88.
- Dean, N., Murphy, T.B., and Downey, G. 2006. Using unlabelled data to update classication
- rules with applications in food authenticity studies. Journal of the Royal Statistical Society.
- 463 Series C **55**(1): 1–14.
- Dempster, A., Laird, N., and Rubin, D. 1977. Maximum likelihood from incomplete data via
- the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**:
- 466 1–38.
- 467 Diggle, P., Heagerty, P., Liang, K., and Zeger, S. 2003. Analysis of Longitudinal Data. Oxford
- 468 University Press, Oxford, U.K.
- 469 Fairbairn, D.J., Blanckenhorn, W.U., and Székely, T. 2007. Sex, size, and gender roles: evolu-
- 470 tionary studies of sexual size dimorphism. Oxford University Press Oxford.

- 471 Ford, E. 1933. An account of the herring investigations conducted at Plymouth during the years
- from 1924 to 1933. Journal of the Marine Biological Association of the United Kingdom
- 473 **19**(01): 305–384.
- 474 Francis, R.I.C.C. 1988. Are growth parameters estimated from tagging and age-length data
- comparable? Canadian Journal of Fisheries and Aquatic Sciences **45**(6): 936–942.
- 476 Frøysa, K.G., Bogstad, B., and Skagen, D.W. 2002. Fleksibesth age-length structured fish
- stock assessment model. Fisheries Research **55**(1): 87–101.
- 478 Frühwirth-Schnatter, S. 2006. Finite mixture and Markov switching models. Springer Science
- 479 & Business Media.
- 480 Gallucci, V. and Quinn, II, T.J. 1979. Reparameterizing, fitting, and testing a simple growth
- model. Transactions of the American Fisheries Society **108**(1): 14–25.
- 482 Helser, T.E. and Lai, H.L. 2004. A bayesian hierarchical meta-analysis of fish growth: with an
- example for north american largemouth bass, *Micropterus salmoides*. Ecological Modelling
- 484 **178**(3): 399–416.
- Jones, T.T., Hastings, M.D., Bostrom, B.L., Pauly, D., and Jones, D.R. 2011. Growth of captive
- leatherback turtles, *Dermochelys coriacea*, with inferences on growth in the wild: Implica-
- 487 tions for population decline and recovery. Journal of Experimental Marine Biology and
- 488 Ecology **399**(1): 84 92.
- 489 Kendall, N.W. and Quinn, T.P. 2012. Size-selective fishing affects sex ratios and the opportu-
- 490 nity for sexual selection in alaskan sockeye salmon *Oncorhynchus nerka*. Oikos pp. DOI:
- 491 10.1111/j.1600-0706.2012.20319.x.
- 492 Kimura, D. 1980. Likelihood methods for the von Bertalanffy growth curve. Fish. Bull 77(4):
- 493 765–776.
- 494 Krebs, C. and Cowan, I.M. 1962. Growth studies of reindeer fawns. Canadian Journal of
- 495 Zoology **40**(5): 863–869.

- 496 Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B. 2015. TMB: automatic differ-
- entiation and laplace approximation. arXiv preprint arXiv:1509.00660.
- 498 Lacointe, A. 2000. Carbon allocation among tree organs: a review of basic processes and
- representation in functional-structural tree models. Annals of Forest Science **57**(5): 521–
- 500 533.
- Lester, N.P., Shuter, B.J., and Abrams, P.A. 2004. Interpreting the von bertalanffy model of
- somatic growth in fishes: the cost of reproduction. Proceedings of the Royal Society of
- 503 London. Series B: Biological Sciences **271**(1548): 1625–1631.
- 504 Macdonald, P.D.M. and Pitcher, T.J. 1979. Age-groups from size-frequency data: a versatile
- and efficient method of analyzing distribution mixtures. Journal of the Fisheries Board of
- 506 Canada **36**(8): 987–1001.
- 507 McLachlan, G. and Peel, D. 2000. Finite Mixture Models. John Wiley & Sons, Inc., New York,
- 508 USA.
- McLachlan, G. and Krishnan, T. 2008. The EM algorithm and extensions. John Wiley & Sons,
- Inc., Hoboken, NJ, USA, 2nd edition.
- Methot, R.D. 2000. Technical description of the stock synthesis assessment program. Technical
- 512 report, NOAA.
- 513 Myers, R.A. and Hoenig, J.M. 1997. Direct estimates of gear selectivity from multiple tagging
- experiments. Canadian Journal of Fisheries and Aquatic Sciences **54**(1): 1–9.
- 515 Nentwig, W. 2012. Ecophysiology of spiders. Springer Science & Business Media.
- Pardo, S.A., Cooper, A.B., and Dulvy, N.K. 2013. Avoiding fishy growth curves. Methods in
- 517 Ecology and Evolution **4**(4): 353–360. doi:10.1111/2041-210x.12020.
- Parker, G.A. 2006. The evolution of sexual size dimorphism in fish. Journal of Fish Biology
- 519 **41**(sB): 1–20.

- 520 Pauly, D. 1980. On the interrelationships between natural mortality, growth parameters, and
- mean environmental temperature in 175 fish stocks. Journal du Conseil **39**(2): 175–192.
- 522 Pearson, K. 1895. Contributions to the mathematical theory of evolution. ii. skew variation
- in homogeneous material. Philosophical Transactions of the Royal Society of London **186**:
- 524 343–414.
- 525 Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations.
- Bulletin of the Fisheries Research Board of Canada **191**: 382.
- Robertson, T.B. 1923. The chemical basis of growth and senescence. JB Lippincott company,
- 528 Philadelphia, USA.
- 829 Roff, D.A. 1982. Reproductive strategies in flatfish: a first synthesis. Can. J. Fish. Aquat. Sci.
- 530 **39**: 1686–1698.
- Roff, D. 1983. An allocation model of growth and reproduction in fish. Can. J. Fish. Aquat.
- 532 Sci. **40**(9): 1395–1404.
- 533 Schnute, J. 1981. A versatile growth model with statistically stable parameters. Canadian
- Journal of Fisheries and Aquatic Sciences **38**(9): 1128–1140.
- 535 Shumway, R.H. and Stoffer, D.S. 1982. An approach to time series smoothing and forecasting
- using the em algorithm. Journal of time series analysis 3(4): 253–264.
- 537 Starck, J.M. and Ricklefs, R.E. 1998. Avian growth and development: evolution within the
- altricial-precocial spectrum. 8. Oxford University Press.
- 539 Stearns, S.C. and Koella, J.C. 1986. The evolution of phenotypic plasticity in life-history traits:
- predictions of reaction norms for age and size at maturity. Evolution pp. 893–913.
- 541 Thorson, J.T., Stewart, I.J., Punt, A.E., and Jech, J.M. 2011. Accounting for fish shoals in
- single-and multi-species survey data using mixture distribution models. Canadian Journal of
- Fisheries and Aquatic Sciences **68**(9): 1681–1693.
- 544 Ullman, J.B. and Bentler, P.M. 2003. Structural equation modeling. Wiley Online Library.

- Vitale, F., Svedäng, H., and Cardinale, M. 2006. Histological analysis invalidates macroscop-
- ically determined maturity ogives of the kattegat cod (Gadus morhua) and suggests new
- proxies for estimating maturity status of individual fish. ICES Journal of Marine Science
- **63**(3): 485–492.
- von Bertalanffy, L. 1938. A quantitative theory of organic growth. Human Biology 10: 181–
- 550 213.
- Walford, L.A. 1946. A new graphic method of describing the growth of animals. Biological
- 552 Bulletin **90**(2): 141–147.
- Wang, P., Puterman, M., Cockburn, I., and Le, N. 1996. Mixed poisson regression models with
- covariate dependent rates. Biometrics **52**(2): 381–400.
- Warner, R.R. 2012. Review of reproduction and sexuality in marine fishes: Patterns and pro-
- cesses. Quarterly Review of Biology 87: 157.
- 557 Weisberg, S., Spangler, G., and Richmond, L.S. 2010. Mixed effects models for fish growth.
- Canadian Journal of Fisheries and Aquatic Sciences **67**(2): 269–277. doi:10.1139/F09-181.
- Werner, E.E. 1986. Amphibian metamorphosis: Growth rate, predation risk, and the optimal
- size at transformation. The American Naturalist **128**(3): 319–341.

Table 1: Root-median-square error by: parameter, method and whether maturation occurred early or late in the lifespan of the simulated species. *Asymptotic* and *linear* refer to the form of the simulated curves. Full, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively.

| | L_{∞} | | K | | t | 0 | (| 7 |
|------------|--------------|-------|-------|-------|-------|-------|-------|-------|
| | Early | Late | Early | Late | Early | Late | Early | Late |
| Asymptotic | | | | | | | | |
| FULL | 0.899 | 0.901 | 0.019 | 0.020 | 0.060 | 0.061 | 0.007 | 0.007 |
| FMU | 1.107 | 1.547 | 0.045 | 0.094 | 0.377 | 0.999 | 0.010 | 0.014 |
| MIX | 0.912 | 0.987 | 0.022 | 0.028 | 0.076 | 0.092 | 0.009 | 0.011 |
| Linear | | | | | | | | |
| FULL | 2.013 | 2.067 | 0.008 | 0.008 | 0.045 | 0.045 | 0.007 | 0.007 |
| FMU | 2.689 | 3.759 | 0.014 | 0.026 | 0.235 | 0.997 | 0.009 | 0.013 |
| MIX | 2.090 | 2.348 | 0.008 | 0.010 | 0.055 | 0.063 | 0.008 | 0.010 |

Table 2: Proportion of time correct model chosen. True model refers to the true simulated model (e.g., true $L_{\infty,s}$ means the sexes differ in their asymptotic size), "Chosen" refers to the model chosen via BIC in fitting. Full, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively. Early and late denote maturation age. All combinations of bound and unbound parameters by sex were fitted (such that each setting sub-column sums to unity) but only the four true model results are presented

| | | | | | | | True model | nodel | | | | | |
|------------|----------------|------|----------------|-------|-----------|------|----------------|-------|-----------|------|----------------|-------|-----------|
| | | | FUI | T | | | FMC | Ω | | | MIX | × | |
| Setting | Chosen | None | $L_{\infty,s}$ | K_s | $t_{0,s}$ | None | $L_{\infty,s}$ | K_s | $t_{0,s}$ | None | $L_{\infty,s}$ | K_s | $t_{0,s}$ |
| Asymptotic | | | | | | | | | | | | | |
| Early | None | | 0 | 0 | 0.03 | 0.88 | 0 | 0.07 | 0.79 | 0.87 | 0 | 0.02 | 0.45 |
| | $L_{\infty,s}$ | 0.02 | 6.0 | 90.0 | 0.02 | 0.02 | 0.89 | 0.16 | 0.04 | 0.02 | 0.88 | 0.00 | 0.02 |
| | K_s | | 0 | 0.82 | 0.13 | 0.02 | 0.01 | 0.48 | 0.04 | 0.03 | 0 | 0.75 | 0.1 |
| | $t_{0,s}$ | | 0 | 90.0 | 0.75 | 0.02 | 0 | 0.23 | 0.07 | 0.02 | 0 | 0.05 | 0.36 |
| | | | | | | | | | | | | | |
| Late | None | 6.0 | 0 | 0 | 0.02 | 0.87 | 0 | 0.61 | 0.87 | 0.87 | 0 | 0.3 | 0.5 |
| | $L_{\infty,s}$ | 0.02 | 0.88 | 90.0 | 0.02 | 0.04 | 0.79 | 0.00 | 0.04 | 0.03 | 0.87 | 0.08 | 0.02 |
| | K_s | 0.02 | 0 | 8.0 | 0.13 | 0.02 | 0.1 | 0.14 | 0.02 | 0.03 | 0 | 0.46 | 0.16 |
| | $t_{0,s}$ | 0.02 | 0 | 90.0 | 0.75 | 0.02 | 0.04 | 0.1 | 0.02 | 0.01 | 0 | 0.08 | 0.24 |
| Linear | | | | | | | | | | | | | |
| Early | None | 0.88 | 0 | 0 | 0.01 | 0.88 | 0 | 0 | 0.7 | 0.88 | 0 | 0 | 0.32 |
| | $L_{\infty,s}$ | 0.02 | 0.82 | 0.16 | 0.01 | 0.02 | 0.75 | 0.25 | 0.05 | 0.02 | 0.79 | 0.18 | 0.02 |
| | K_s | 0.01 | 0.1 | 0.75 | 90.0 | 0.01 | 0.15 | 0.61 | 90.0 | 0.02 | 0.1 | 0.72 | 90.0 |
| | $t_{0,s}$ | 0.03 | 0 | 0 | 0.85 | 0.02 | 0 | 0.09 | 0.14 | 0.02 | 0 | 0 | 0.54 |
| | | | | | | | | | | | | | |
| Late | None | | 0 | 0 | 0 | 0.87 | 0 | 0.03 | 98.0 | 0.89 | 0 | 0.02 | 0.44 |
| | $L_{\infty,s}$ | _ | 0.81 | 0.19 | 0.01 | 0.02 | 0.67 | 0.33 | 0.03 | 0.02 | 0.78 | 0.2 | 0.01 |
| | K_s | 0.01 | 0.1 | 0.72 | 0.05 | 0.02 | 0.22 | 0.33 | 0.01 | 0.03 | 0.1 | 69.0 | 0.04 |
| | $t_{0,s}$ | _ | 0 | 0 | 0.85 | 0.02 | 0.03 | 0.26 | 0.03 | 0.02 | 0 | 0.01 | 0.42 |

Table 3: Median maximum a posteriori classification rate across iterations. *Asymptotic* and *linear* refer to the form of the simuated curves. FULL, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively.

| | $L_{\infty,s}$ | | K | s | t_0 | ,s | No | ne |
|------------|----------------|------|-------|------|-------|------|-------|------|
| | Early | Late | Early | Late | Early | Late | Early | Late |
| Asymptotic | | | | 2 | | | | |
| FULL | 0.88 | 0.87 | 0.82 | 0.78 | 0.73 | 0.66 | 0.53 | 0.53 |
| FMU | 0.77 | 0.64 | 0.69 | 0.52 | 0.48 | 0.48 | 0.47 | 0.48 |
| MIX | 0.86 | 0.85 | 0.79 | 0.72 | 0.60 | 0.52 | 0.47 | 0.48 |
| Linear | | | | | | | | |
| FULL | 0.87 | 0.87 | 0.85 | 0.84 | 0.76 | 0.68 | 0.54 | 0.53 |
| FMU | 0.79 | 0.66 | 0.77 | 0.64 | 0.51 | 0.48 | 0.48 | 0.48 |
| MIX | 0.85 | 0.86 | 0.84 | 0.82 | 0.69 | 0.58 | 0.47 | 0.48 |

Table 4: Best fitting (lowest AIC) von Bertalanffy parameter estimates by method for two deepwater lantern shark species *Etmopterus spinax* and *Etmopterus pusillus*. Full, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively. Standard errors are provided in parentheses.

| Species | Sex | Method | $\ln L_{\infty}$ | $\ln K$ | $\ln -t_0$ | $\ln \sigma$ | Class rate |
|-------------|--------|------------|------------------|--------------|-------------|--------------|------------|
| E. spinax | Female | FULL | 4.03 (0.04) | -2.16 (0.07) | 0.7 (0.05) | 0.65 (0.03) | 0.65 |
| | | FMU | 3.94 (0.17) | -2.41 (0.55) | 1.89 (0.44) | 0.2 (0.04) | 0.08 |
| | | MIX | 3.85 (0.03) | -1.74 (0.07) | 0.42 (0.08) | 0.41 (0.03) | 0.35 |
| | Male | FULL | 4.03 (0.04) | -2.34 (0.06) | 0.92 (0.04) | 0.44 (0.04) | 0.46 |
| | | FMU | 3.82 (0.17) | -2.41 (0.55) | 1.89 (0.44) | 0.2 (0.04) | 1.00 |
| | | MIX | 4.43 (0.2) | -2.97 (0.27) | 1.14 (0.08) | 0.41 (0.03) | 0.80 |
| E. pusillus | Female | FULL | 3.96 (0.01) | -1.92 (0.05) | 0.92 (0.06) | 0.42 (0.03) | 0.33 |
| | | FMU | 4.07 (0.08) | -2.65 (0.29) | 2.35 (0.23) | 0 (0.06) | 0.12 |
| | | MIX | 3.94 (0.01) | -1.77 (0.04) | 0.88 (0.05) | -0.19 (0.11) | 0.21 |
| | Male | FULL | 3.94 (0.02) | -1.92 (0.05) | 0.92 (0.06) | 0.42 (0.03) | 0.84 |
| | | FMU | 4.07 (0.08) | -2.77 (0.27) | 2.35 (0.23) | 0 (0.06) | 0.98 |
| | | MIX | 3.94 (0.01) | -1.91 (0.03) | 0.88 (0.05) | 0.34 (0.04) | 0.91 |
| | | | | 75 | | | |

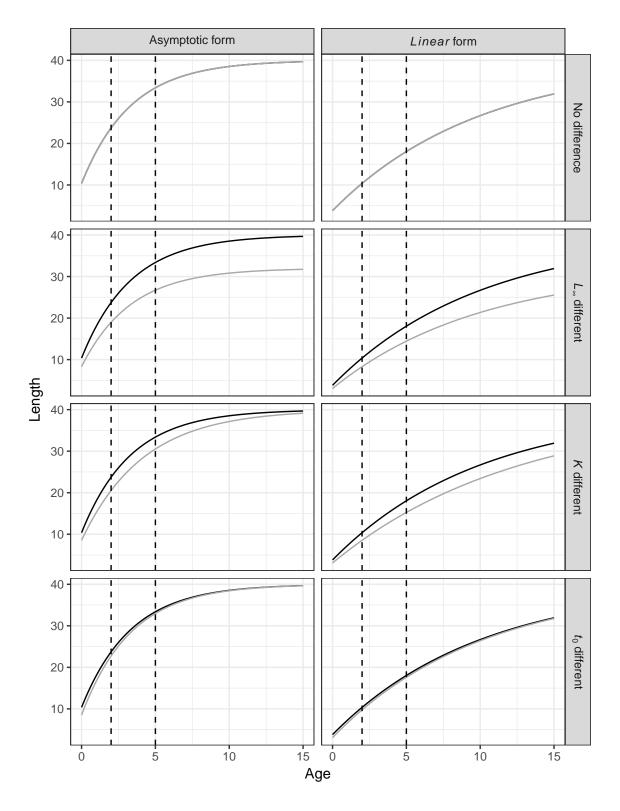


Figure 1: Simulation framework illustration. Asymptotic and linear (weakly asymptotic) refer to the form of the simulated curves. Rows denote parameter differences between females and males. Female and male mean curves are shown in black and grey, respectively. Dashed vertical lines denote early and late maturation ages.

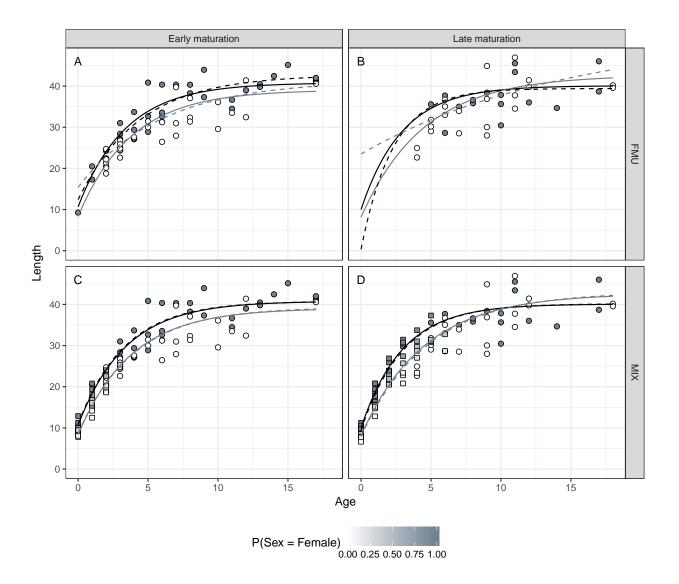


Figure 2: Example simulations for an early maturing (left column) and late maturing (right column) population. FMU and MIX refer to the mature animals only and mixture models, respectively. Full data curves are displayed as solid lines (females black, males grey). Estimated FMU and MIX curves are shown as short-dashed and long-dashed curves, respectively. Known-sex individuals are shown as circles and unknown-sex individuals as squares. The fill colour of the points denotes the probability of the individual being female (from the MIX model for unknowns).

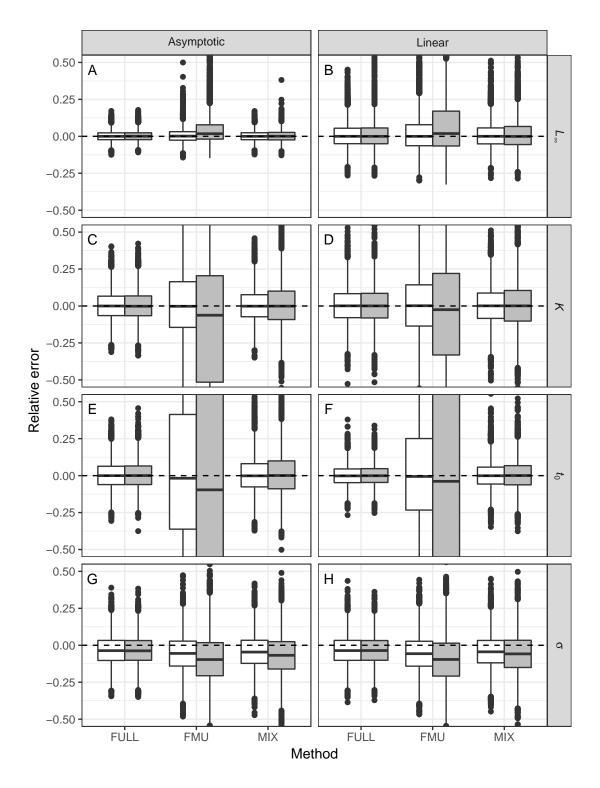


Figure 3: Boxplots displaying the distribution of the relative error by: parameter, method, and whether maturation occurred early (white) or late (grey) in the lifespan of the simulated species. Boxes correspond to the 25th and 75th percentiles, whiskers extend to the closest points to one and a half times the interquartile range. Full, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively. Asymptotic and linear refer to the form of the simuated curves (Figure 1). Y-axis range is restricted to facilitate comparison of most distributions.

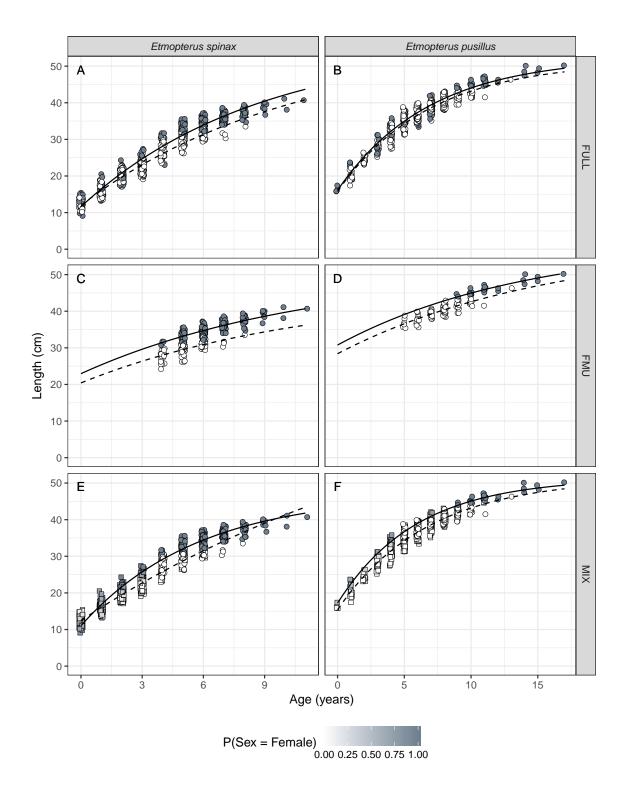


Figure 4: Best fitting (lowest BIC) von Bertalanffy fits to age-length data from two deepwater lantern sharks species *Etmopterus spinax* and *Etmopterus pusillus*. Full, FMU and MIX refer to the complete knowledge, mature animals only and mixture model, respectively. Estimated curves for females and males are shown as solid and dashed lines, respectively. Known-sex individuals are shown as circles and unknown-sex individuals as squares. The fill colour of the triangles in the MIX model case denote the probability of the individual being female. Age values are jittered for visualization.