

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319381159>

Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries

Article in *Canadian Journal of Fisheries and Aquatic Sciences* · August 2017

DOI: 10.1139/cjfas-2017-0143

CITATIONS

90

READS

1,554

2 authors:



Merrill B Rudd

University of Washington Seattle

24 PUBLICATIONS 583 CITATIONS

[SEE PROFILE](#)



James T Thorson

Alaska Fisheries Science Center

255 PUBLICATIONS 8,074 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



FISHGLOB: fish biodiversity under global change [View project](#)



Fisheries Assessments (NWFSC) [View project](#)

1 **Accounting for variable recruitment and fishing mortality in length-based stock**
2 **assessments for data-limited fisheries**

3
4 Merrill B. Rudd¹ and James T. Thorson²

5
6 ¹School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA
7 98195, USA, merrillrudd@gmail.com

8 ² Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center,
9 National Marine Fisheries Service, NOAA, Seattle, WA, USA, james.thorson@noaa.gov

10
11 Corresponding author: Merrill Rudd

12 Address: 1122 NE Boat Street, Box 355020, Seattle, WA 98195

13 Phone: 201-207-0958

14 Fax: 206-685-7481

15 Email: merrillrudd@gmail.com

16

17 **Abstract**

18 In fisheries with limited capacity for monitoring, it is often easier to collect length
19 measurements from fishery catch than quantify total catch. Conventional stock assessment tools
20 that rely on length measurements without total catch do not directly account for variable fishing
21 mortality and recruitment over time. However, this equilibrium assumption is likely violated in
22 almost every fishery, degrading estimation performance. We developed an extension of length-
23 only approaches to account for time-varying recruitment and fishing mortality. This Length-
24 based Integrated Mixed Effects (LIME) method at a minimum requires a single year of length
25 data and basic biological information, but can fit to multiple years of length data, catch, and an
26 abundance index if available. We use simulation testing to demonstrate that LIME can estimate
27 how much fishing has reduced spawning output in the most recent year across a variety of
28 scenarios for recruitment and fishing mortality. LIME improves data-limited fisheries stock
29 assessments by its flexibility to incorporate additional years or types of data if available, and
30 obviates the need for equilibrium assumptions.

31

32 **Introduction**

33 Many fisheries worldwide lack the quality and quantity of data used in classical stock
34 assessments, but must deal with limited information to make management decisions (Quinn et al.
35 2016). Stock assessments can provide a quantitative starting point for developing management
36 strategies and monitoring the impacts of management. For example, the Magnuson-Stevens
37 Fisheries Conservation and Management Act mandates fisheries managers in the United States to
38 set catch limits based on the ‘best available science’, involving advice from stock assessment to
39 inform those limits (Darcy and Matlock 1999; Methot et al. 2014). The Marine Stewardship
40 Council (MSC) similarly requires fisheries seeking certification to go through a stock assessment
41 process to determine their sustainability (Gulbrandsen 2009). In the case of small-scale fisheries
42 in developing nations, it is possible to manage a fishery using only harvest control rules to meet
43 management objectives, without formal stock assessment estimating status relative to reference
44 points (Mahon 1997). However, beyond their use to evaluate management strategies for fishery
45 resources (Carruthers et al. 2014), modeling tools can help with community engagement in the
46 scientific process, such as conflict resolution (Butler et al. 2006) and integration of local
47 knowledge to support cooperation between fishermen and scientists (Neis 1992; Azzurro et al.
48 2011).

49 Many stocks worldwide remain unassessed, e.g., the U.N. Food and Agriculture
50 Organization global fishery statistics database includes 19,624 unique combinations of country
51 and taxa (FAO 2016), while the RAM Legacy stock assessment database only includes 1,268
52 stock assessments (Ricard et al. 2012). The regions of the world with the fewest stock
53 assessments relative to the number of stocks perform worse across fishery management attributes
54 (Melnychuk et al. 2017), indicating the presence of fishery monitoring and assessment may have

55 some relation with successful management. In the following, we define “data-limited” as any
56 stock with uninformative data (no contrast to provide information on rates of change) or lacking
57 data types typically used in statistical estimates of stock status (e.g. fishery-independent surveys,
58 proportion-at-ages) (Costello et al. 2012; Dowling et al. 2016). Data-limited stock assessment
59 methods are in great demand in both developed and developing nations due to the ubiquity of
60 data-limited fisheries, and the need for science-based management decisions (Wetzel and Punt
61 2011; Chrysafi and Kuparinen 2016; Dowling et al. 2015).

62 Length-based assessment methods are a vital component of the data-limited stock
63 assessment toolbox because it is much easier to obtain reliable length measurements of a portion
64 of the fishery catch than to measure total catch or record effort data for many small-scale or non-
65 target species (Harley et al. 2001; Nadon et al. 2015; Prince et al. 2015b; Kokkalis et al. 2015).
66 Similarly, age information and a fishery-independent survey that are representative of total
67 abundance are prohibitively expensive or impossible to collect for most fisheries in the world.
68 Prominent length-based methods for estimating reference points in data-limited fisheries include
69 length-based spawning potential ratio (LB-SPR; Hordyk et al. 2015) and mean-length mortality
70 estimation methods (Nadon et al. 2015). LB-SPR uses length-composition data and assumptions
71 about biological parameters to make a rapid assessment of stock status relative to unfished levels
72 assuming equilibrium conditions (Hordyk et al. 2015; Prince et al. 2015b). While LB-SPR can
73 use multiple years of length data, status determination is based on one year of data at a time (i.e.
74 estimates of status over multiple years are based on that year’s length composition alone). Mean-
75 length mortality estimators (e.g., Gedamke and Hoenig 2006), first developed by Beverton and
76 Holt (1957), assume fishing mortality directly influences mean length of the catch, and have
77 been used for assessments in the U.S. South Atlantic, Pacific islands, and Caribbean (Ehrhardt

78 and Ault 1992; Ault et al. 2005, 2008; Gedamke and Hoenig 2006; Nadon et al. 2015). As
79 measures of stock status, these length-based methods derive the spawning potential ratio (SPR)
80 reference point, defined as the proportion of unfished reproductive potential at a given level of
81 fishing pressure (Goodyear 1993).

82 Length-based assessment methods estimating stock status assume that recruitment and
83 fishing mortality arise from deterministic relationships or have not changed over a period
84 significant for management and the life history of the species (termed “equilibrium
85 assumptions”). Equilibrium assumptions are often violated (Gedamke and Hoenig 2006), as
86 recruitment is quite variable for most species and fishing mortality changes with markets and
87 other socioeconomic factors in the fishing community (Thorson et al. 2013, 2014a). Stochastic
88 ocean conditions and productivity regime shifts may cause recruitment to vary erratically,
89 gradually, or periodically at any given time (Vert-pre et al. 2013; Thorson et al. 2014b;
90 Szuwalski et al. 2015). However, the violation of the equilibrium assumption may be difficult to
91 detect. For example, the equilibrium assumption may appear valid when the mean length is
92 constant over time (Gedamke and Hoenig 2006; Nadon et al. 2015). If recruitment is then
93 constant over time, increasing fishing mortality will lead to decreasing mean length as the larger
94 individuals are harvested and only smaller individuals remain in the population. However,
95 constant fishing mortality and a recruitment pulse would also lead to a decrease in mean length,
96 with more young individuals entering the population. Given both variable fishing mortality and
97 recruitment processes occurring on the same population as well as errors when measuring mean
98 length, a mean-length time series may appear constant when time-varying population processes
99 are instead cancelling each other out.

100 As an alternative to the equilibrium assumptions, a mixed-effects model can be used to
101 deal with important demographic changes by estimating random variation in recruitment, fishing
102 mortality, or other biological processes, as well as the magnitude (variance) of random variation
103 in each process (de Valpine and Hastings 2002; Buckland et al. 2004; Schnute and Haigh 2007;
104 Thorson and Minto 2015). Mixed-effects models can directly account for variation arising from
105 natural processes or measurement processes separately, and therefore improve performance in
106 nonlinear fisheries models (de Valpine and Hastings 2002; Ono et al. 2012; Thorson et al.
107 2015b). A main criticism of length-based methods with equilibrium assumptions is that with a
108 single year of length-composition data and general understanding of biological parameters, it is
109 impossible to determine whether a larger proportion of small fish in the catch is caused by strong
110 cohort in recent years, or by the removal of larger fish from the system. Accounting for random
111 variation in recruitment, fishing mortality and observation error arising from the process of
112 sampling fish lengths from the population, helps to tease apart each of these processes and better
113 identify the true state of the fish population.

114 The aim of this study is to introduce a new length-based, integrated, mixed-effects
115 (LIME) model and demonstrate its statistical performance when estimating reference points
116 assuming only length composition and basic biological information are available. This method
117 builds upon the catch-curve stock reduction analysis (CCSRA) model (Thorson and Cope 2014),
118 which includes an estimate of mortality from the age composition and at least one year of total
119 fishery catch to estimate MSY-based reference points without assuming information about final
120 biomass relative to unfished biomass (as generally used in stock reduction analysis). As an
121 extension, LIME uses samples of length in place of the more resource-intensive samples of age,
122 and can estimate the SPR reference point if catch data are unavailable. To demonstrate the LIME

123 model, we used simulation testing to (i) demonstrate that LIME is unbiased across several life-
124 history types and patterns of fishing mortality and recruitment variability and able to include
125 more years of length measurements and catch and/or abundance index data, (ii) examine the
126 sensitivity of the model to sample size of length measurements and error in input parameters, and
127 (iii) compare LIME against LB-SPR to assess performance under various violations of model
128 assumptions, including the timing of sampling within the year and modeling monthly time steps.
129

130 **Methods**

131 LIME is an age-structured population dynamics model with the ability to (1) account for
132 variable fishing mortality and recruitment when only length data are available, and (2) to treat
133 multiple years and types of data in an integrated manner to improve estimates of fishing
134 mortality changes over time. The minimum inputs for the LIME assessment method are data on
135 the length composition of the catch from a single year as well as assumed life history
136 information, including the length-at-age relationship, an assumed natural mortality rate, and
137 length at 50% maturity. LIME estimates annual fishing mortality rates, lengths at 50% and 95%
138 selectivity to the fishing gear, and the Dirichlet-multinomial parameter θ as fixed effects. The
139 effective sample size of length data is linearly related to the input sample size with intercept
140 $(1+\theta)^{-1}$ and slope $\theta(1+\theta)^{-1}$ (Thorson et al. 2017). LIME can be differentiated from other age-
141 structured models (e.g. Stock Synthesis) in that annual recruitments are treated as random
142 effects, where mean and standard deviation of a distribution for recruitment are additionally
143 estimated as fixed effects. Another key difference is that LIME does not require catch data: if no
144 information on the scale of population size is available, recruitment will be estimated relative to
145 average levels for an unfished population. As measures of stock status, we derived the SPR

146 reference point to compare results with LB-SPR. We also derived the $F_{30\%}$ and $F_{40\%}$ reference
147 points (the fishing mortality rates that would result in SPR of 30% and 40%, respectively; Clark
148 2002). We derived MSY by finding the fishing mortality rate that results in the highest yield per
149 recruit. When total catch data were available (thereby providing information on scale of the
150 population size), LIME would estimate equilibrium recruitment, which would scale the MSY
151 based on a per-recruit equation to a scale appropriate for the population size.

152 We developed an operating model to simulate true populations and generate data under a
153 variety of fishing, recruitment, and life history scenarios. We then used this operating model to
154 explore the estimation performance of LIME for different scenarios regarding recruitment,
155 fishing mortality, data-availability, and life history (Figure 1). We conducted all simulation
156 modeling using the open-source statistical software R (R Core Team 2016), and all estimation in
157 the R package *Template Model Builder* (TMB; Kristensen et al. 2015) as implemented in our R
158 package *LIME* (<https://github.com/merrillrudd/LIME>, doi: 10.5281/zenodo.834404, version
159 1.0.0).

160

161 *Operating model*

162 We tested LIME for three different life-history types chosen to reflect the types of taxa
163 for which length-based assessments are commonly demanded (Figure 2). These life-history types
164 were: (a) a short-lived fish, mimicking rabbitfish (*Siganus sutor*, L_{∞} =36.2 cm, k =0.87, M =1.49,
165 L_m^{50} =20.2 cm, maximum age =4, Hicks and McClanahan 2012), (b) a medium-lived fish,
166 mimicking spotted rose snapper (*Lutjanus guttatus*, L_{∞} =64.6 cm, k =0.21, M =0.43, L_m^{50} =34.0 cm,
167 maximum age=15, Bystrom 2015), and (c) a longer-lived fish, mimicking red grouper
168 (*Epinephelus morio*, L_{∞} =90.0 cm, k =0.13, M =0.18, L_m^{50} =50.0 cm, maximum age=26, Heemstra

169 and Randall 1993) (Table 1). Each simulated population began with biomass at a fraction of
170 unfished biomass, drawn from a uniform distribution between 0.05 and 0.95.

171 For each life history scenario, we tested LIME performance under three scenarios of
172 fishing mortality and recruitment variability (Figure 3). The first is the “equilibrium scenario”,
173 which matches the non-variable fishing mortality and recruitment assumptions of LB-SPR. The
174 equilibrium scenario involved fishing mortality and recruitment constant over a 20-year period,
175 with a standard deviation for fishing mortality and recruitment set to a negligible 0.01. The
176 second scenario, the “two-way base scenario”, involved a linear change from the fishing
177 mortality that would result in the randomly chosen initial depletion to the rate associated with
178 20% SPR ($F_{20\%}$) over the first seven years of the 20-year time series. This change could be
179 positive or negative depending on the F associated with randomly selected value for initial
180 depletion and $F_{20\%}$. Over the next seven years of the time series, F was constant at $F_{20\%}$, then the
181 fishing rate decreased linearly down to half of $F_{20\%}$ for the last six years of the time series. $F_{20\%}$
182 was calculated deterministically based on the biological information and selectivity associated
183 with each life-history type. Thus, this value would not vary between different scenarios of
184 variability within a life-history type. The fishing mortality time series in the “two-way base
185 scenario” varied between simulation iterations in a) the randomly chosen initial depletion and b)
186 lognormally distributed deviations around this two-way trip (following the same equation for
187 lognormal recruitment deviates in Table 3, eq. 1 except using the standard deviation for fishing
188 mortality, σ_F , equal to 0.2). While we calculated $F_{20\%}$ for the medium-lived and longer-lived life-
189 history types, we fixed $F_{20\%}$ to 3.0 for the short-lived life-history type. Based on the assumption
190 of instantaneous annual fishing mortality, $F_{20\%}$ for the short-lived fish was calculated to be much
191 greater than could be supported even by very high fishing capacity. Recruitment was variable

192 and autocorrelated over a 20-year period (Table 3, eq. 2). This scenario included a standard
193 deviation of recruitment residuals σ_R equal to 0.737 and a first-order autoregressive coefficient
194 equal to 0.426, the mean of the predictive distribution from a meta-analysis of recruitment
195 variability in global fish orders (Thorson et al. 2014b). A third scenario, the “one-way base
196 scenario”, involved the same recruitment variability and autocorrelation as the “two-way”
197 scenario, but with fishing mortality changing linearly from the rate that would result in the
198 randomly chosen initial depletion to $F_{20\%}$ over the 20-year period. Thus, this change could be
199 positive or negative depending on the F associated with the randomly chosen initial depletion
200 relative to $F_{20\%}$ for the life-history type. This combination of scenarios tests the ability to track
201 how the population processes are changing over time with variability in the system.

202 In the operating model, we assumed the natural mortality rate was constant, known, and
203 independent of size or age. We modeled individual growth using a von Bertalanffy growth
204 function (Table 2, eq. 1; Figure 2). We assumed maturity at length m_j was based on a one-
205 parameter logistic function (Table 2, eq. 2; Figure 2). We converted maturity at length to
206 maturity at age using a normal distribution with standard deviation a function of the coefficient
207 of variation of the age-length curve (Table 2, eq. 3). Selectivity-at-length follows a two-
208 parameter logistic model (Table 2, eq. 4; Figure 2), with estimated parameter length at 50%
209 selectivity and a second parameter δ representing the difference between length at 95% and 50%
210 selectivity. We modeled weight at age as an allometric function of individual length at age
211 (Table 2, eq. 5). We calculated annual total biomass as a function of the abundance and weight-
212 at-age (Table 2, eq. 6). Spawning biomass was a function of the total annual biomass and the
213 proportion mature at age (Table 2, eq. 7). These processes contributed to an underlying age-
214 structured model (Table 2, eq. 8). Parameter definitions and input values are listed in Table 1.

215

216 *Data generation*

217 We generated length data by simulating underlying age-structured dynamics and then
218 sampling length composition from the vulnerable population instantaneously at the beginning of
219 each year. First we calculated the probability of being in a length bin for individuals of each age
220 (Table 3, eq. 3). We then calculated the probability of harvest in each length bin each year as the
221 proportion of the abundance-at-age vulnerable to the fishing gear each year multiplied by the
222 probability of being in a length bin given age (Table 3, eq. 4). We used a multinomial probability
223 distribution to generate the length frequency of samples of the catch in each length bin over time
224 (Table 3, eq. 5).

225 We assumed a sample size n of 200 individuals were measured annually, and that this
226 was the “true” sample size of the length data. We compared the base case of 200 length
227 measurements annually to model performance under sample sizes of 1,000, 500, 100, 50, and 20.
228 We used the large sample size of 1,000 to confirm the model is unbiased and precise across data
229 availability scenarios under ideal circumstances, and then tested the alternate, lower sample sizes
230 to assess performance under more realistic sample sizes.

231 We tested seven different scenarios of data availability. Two scenarios included only
232 length data, assuming one or ten years of length data were available (the “one length
233 composition”, “ten length compositions” scenarios). Data from these two scenarios were used to
234 demonstrate the value of additional years of length data for each life-history type. We also
235 explored five scenarios of additional data availability to demonstrate the integrated nature of
236 LIME. The data availability scenarios included (i) a “data-rich” scenario with 20 years of total
237 catch, 20 years of an abundance index, and 20 years of length data, used as proof-of-concept that

238 LIME works when a high amount of informative data exists; (ii) “index plus 10” with 20 years of
239 an abundance index and 10 years of length data; (iii) “index plus 1” with 20 years of an
240 abundance index and 1 year of length data; (iv) “catch plus 10” with 20 years of total catch data
241 and 10 years of length data; and (v) “catch plus 1” with 20 years of total catch data and 1 year of
242 length data. We assume that catch and abundance indices are lognormally distributed with a log-
243 standard deviation of 0.2 (Table 1). The “catch plus 1” scenario is essentially a stock-reduction
244 analysis while replacing the assumed information regarding final biomass (as used in stock-
245 reduction analysis) with a length-based catch-curve to estimate fishing mortality in the final year,
246 which is analogous to Thorson and Cope (2014), except using length-composition samples. We
247 calculated the expected catch-at-age, using the Baranov catch equation (Table 3, eq. 6), summing
248 to obtain total annual catch (Table 3, eq. 7), based on the true fishing mortality time series and
249 selectivity specified in the operating model. The generated standardized abundance index I_t was
250 proportional to spawning biomass (Table 3, eq. 8).

251 When not otherwise stated, we generated data on an annual time step, which assumes
252 length-composition data are collected instantaneously at the beginning of the year. In fisheries
253 where only length data and biological information are available (i.e. no catch or abundance
254 index) it is possible length data would be available on less than an annual time step. We tested
255 scenarios where length data were collected on a monthly time step, then either pooled into an
256 annual time step or kept on a monthly time step. The monthly data collection scenario is more
257 representative of fisheries occurring year-round, with no specific season. Furthermore, length
258 data on a monthly time step would be more representative of short-lived fish growth, to account
259 for fish observed between the midpoints of each age class.

260

261 *Estimation model*

262 The structure of LIME follows the structure of the operating model using equations 1-8 in
263 Table 2 and equations 3-4, and 6-8 in Table 3 to derive the predicted catch, index, and length
264 composition. LIME requires at least one year of length data, but is flexible to include annual
265 fishery catch and/or an abundance index. During our simulation testing, we assumed that LIME
266 had the correct values for the von Bertalanffy length-at-age relationship, including asymptotic
267 length (L_{∞}), growth coefficient (k), and age at length=0 (t_0), length-weight parameters, natural
268 mortality (M), and the one-parameter logistic maturity-at-length schedule. Future studies can
269 obtain these values from local studies, FishBase (Froese 1990), or global meta-analyses of fish
270 life-history parameters (Thorson et al. In press). We tested LIME performance under violations
271 of these assumptions in sensitivity analyses. We also fixed the values for catch and abundance
272 index observation error and coefficient of variation in the process error for the age-length curve.
273 For all data availability scenarios, the model estimates as fixed effects the annual fishing
274 mortality, lengths at 50% and 95% selectivity, the recruitment standard deviation, and the
275 Dirichlet-multinomial parameter θ related to the effective sample size of length measurements in
276 each year. In scenarios when catch data are unavailable, there is no information on the scale of
277 the population. In these scenarios, we fixed mean recruitment to a relative value of 1.0 so that the
278 model does not estimate the scale of the population but only the annual deviations in recruitment,
279 and estimates of relative reference points (e.g. SPR) can be derived but reference points based on
280 the scale of the population (e.g. MSY) are not meaningful. By contrast, when total catch data are
281 available for at least one year, we can estimate mean recruitment and derive spawning biomass.
282 When an index of abundance was available, we estimated the catchability coefficient for that

283 index as an additional parameter q . A list of parameters estimated and fixed for each scenario is
284 presented in Table 1.

285 We treated annual recruitment as a random effect in LIME, where recruitment each year
286 is a function of an expected recruitment based on a Beverton-Holt stock recruitment relationship
287 (Table 4, eq. 1), and the estimated recruitment standard deviation (Table 4, eq. 2). For the
288 simulation experiments in this paper, we fixed the steepness parameter h of the Beverton-Holt
289 stock-recruitment function at 1.0, meaning that expected recruitment is constant among years and
290 independent of the spawning stock biomass the previous year. It is possible to fix the steepness
291 parameter at a value less than 1.0, where the mean of the lognormal distribution is the predicted
292 number of recruits from the Beverton-Holt stock-recruitment relationship, as opposed to 1.0 or
293 the equilibrium unfished recruitment. We chose to test LIME with 1.0 for the mean of the
294 lognormal distribution to determine how well the model can estimate annual recruitment with all
295 variation from recruitment deviates. Unfished spawning biomass is calculated using the same
296 equation as fished spawning biomass (Table 2, eq. 7) but without the fishing mortality and
297 selectivity terms (Table 2, eq. 9). Unlike the operating model, we did not account for
298 autocorrelation in recruitment in the LIME estimation model.

299 We applied a random-walk penalty on annual estimates of fishing mortality, which
300 shrinks the estimate of fishing mortality in year $t + 1$ towards its estimate in year t (Table 4, eq.
301 3). Drastic changes in fishing mortality between years are unlikely in the real world given costs
302 of entering or leaving a fishery, but the random-walk process accommodates gradual changes in
303 fishing mortality (Nielsen and Berg 2014). However, a fixed value of 0.2 for the standard
304 deviation of the fishing mortality penalty does allow the model to estimate variability in fishing
305 mortality if supported by the data (e.g. exit and entry from the fishery).

306 We also placed a lognormal penalty on the standard deviation of recruitment deviations,
307 σ_R . Based on the meta-analysis conducted by Thorson et al. (2014), the log of the mean was set
308 to the log of 0.737, with a log standard deviation of 0.353 (Table 4, eq. 4). The prior aids in the
309 convergence of the σ_R parameter estimation, preventing the estimate from going to an unlikely
310 value. There is also an upper bound on σ_R at 2.0.

311 The joint log-likelihood of the observed data is the sum of the log-likelihoods of the
312 observed length data, log-probability of fishing mortality and recruitment variation, and the log-
313 likelihood of the catch and abundance index, if available. We include bias-correction for
314 recruitment deviations using the TMB bias-correction feature (Thorson and Kristensen 2016).
315 For scenarios that include abundance index and catch data, a lognormal probability distribution
316 was assumed to describe error in both data types (Table 4, eq. 7, eq. 8). The respective
317 observation errors σ_I and σ_C are fixed *a priori* (not estimated as parameters). We assumed the
318 length data arose from a Dirichlet-multinomial probability distribution with estimated parameters
319 θ_c related to the effective sample size of length measurements each year. Many stock assessment
320 methods use a multinomial distribution to fit age or length data, but the effective sample size
321 must be calculated externally (Francis 2014). By contrast, the Dirichlet distribution can represent
322 variability in the proportions in each length bin, but the parameters do not correspond to the
323 easily interpretable effective sample size of length data for which model results are highly
324 sensitive. The Dirichlet-multinomial is an alternative to these two distributions, estimating an
325 additional parameter θ within the integrated model (Thorson et al. 2016). The effective sample
326 size is a nonlinear function of input sample size (Table 4, eq. 6; Thorson et al. 2016). As θ
327 approaches infinity, the effective sample size is equal to the observed sample size, and the
328 multinomial distribution is a special case of the Dirichlet-multinomial distribution.

329 In model runs, we assessed that the model had converged if the final gradient for all
330 parameters was less than 0.001. If the initial model run did not converge (resulting in NAs or a
331 high final gradient), the model would be run up to ten additional times with starting values equal
332 to the estimates from the non-converged model plus a random number drawn from a normal
333 distribution with mean zero and standard deviation 0.2. For each combination of life-history
334 type, data availability scenario, fishing mortality pattern, and recruitment dynamics, we obtained
335 100 iterations of generated data and ran the estimation model for each set.

336

337 *Comparison to LB-SPR*

338 We ran LB-SPR from the R package *LBSPR* (Hordyk et al. 2015) with one year (LBSPR
339 1) and ten years (LBSPR 10) of length data using the operating model described above. LB-SPR
340 requires as input the length data in each year and the ratio of natural mortality to the von
341 Bertalanffy growth coefficient (M/k), as well as inputs similar to those required for LIME: the
342 von Bertalanffy asymptotic length parameter, coefficient of variation of the asymptotic length,
343 length at 50% and 95% maturity, length-weight parameters, and starting values for the length at
344 50% and 95% selectivity (Table 1). We assumed the coefficient of variation of the asymptotic
345 length was equivalent to the coefficient of variation of the entire age-length curve. For the base
346 runs, we used the true value for the length at 95% maturity and selectivity from the two-
347 parameter logistic curves used in the operating model. LB-SPR uses these inputs to calculate the
348 abundance at relative age at equilibrium. LB-SPR estimates the ratio of fishing mortality to
349 natural mortality and the lengths at 50% and 95% selectivity to best fit the predicted and
350 observed length composition proportions, and derives SPR, outputting estimates for these four
351 values for each year with length data (Hordyk et al. 2015).

352 We note that this simulation experiment is comparing performance of LIME and LB-SPR
353 using an operating model that is more similar to LIME. To further compare the strengths and
354 limitations of each method, we used the *LBSPR* package to simulate length data under the LB-
355 SPR equilibrium, length-based structure, so that we could also test LIME and LB-SPR using an
356 operating model that is more similar to LB-SPR.

357 A major issue in the comparison of models was the time step used. To test model
358 performance, we tested both the LIME and LB-SPR methods using the monthly data generation
359 feature in the operating model, where sampling is spread out over all twelve months instead of
360 collected instantaneously at one point. We then compared LIME and LB-SPR performance
361 pooling the monthly length data into an annual distribution and then running each model on an
362 annual time step. These additional simulation tests were designed to compare these two methods
363 under different data-generating models, with the goal of providing guidance for which methods
364 may be appropriate under a variety of real-world conditions.

365

366 *Reference points*

367 We calculated SPR (Table 2, eq. 10-12) as a biological reference point, used as a proxy
368 for MSY when information on the scale of population size is not available, and for comparison
369 with LB-SPR. A harvest strategy that targets a fishing mortality rate that is expected to result in
370 40% of unfished spawning output (termed “40% SPR”), is considered risk averse for many
371 species (Clark 2002). Therefore, we calculated these values as examples of possible fishing
372 mortality reference points that could be used to compare to other length-based assessment
373 methods.

374

375 *Sensitivity tests*

376 Preliminary exploration suggested that LIME performed poorly for short-lived species
377 when using an annual time step, therefore we explored LIME performance on a monthly time
378 step. We generated monthly length data for the short-lived life-history type and ran LIME on a
379 monthly time step by specifying the number of years as number of months, ages as fractions of a
380 year, and dividing the input natural mortality rate M by twelve so that mortality and growth
381 occurs in each month (Figure A1).

382 We also included sensitivity tests to LIME base models to understand biases associated
383 with imperfect knowledge about species biology, fishery characteristics, and low sample sizes.
384 We assessed the performance of LIME with (a) parameter misspecification of +/- 25% for each
385 of the life history inputs (M , CV_L , L_∞ , k , and L_m^{50}) and (b) sample sizes of length data of 20, 50,
386 100, 500, and 1,000 independent samples annually.

387

388 *Model performance*

389 To assess the ability of the model to accurately and precisely estimate quantities of
390 management interest, we consider bias and precision (Table 4, eq. 9-10) between estimated and
391 true SPR in the last year of data across the 100 iterations of simulated data. We used median
392 relative error (MRE; Table 4, eq. 9) to quantify bias, and median absolute relative error (MARE;
393 Table 4, eq. 10) to quantify precision. To understand the ability of the model to accurately
394 capture uncertainty, we computed the “interval coverage”, the proportion of iterations out of 100
395 where the true value of a population parameter in the terminal year is within the 50% confidence
396 intervals. A well-performing model would have close to nominal coverage (i.e., a 50% coverage

397 interval will contain the true value in 50% of simulation replicates). Coverage is presented to
398 illustrate whether confidence intervals accurately capture model uncertainty.

399

400 **Results**

401 *LIME performance across life history, variability, and data scenarios*

402 Based on the Monte Carlo simulations, LIME can estimate unbiased SPR when length
403 data are available and biological characteristics are correctly specified across various scenarios
404 of fishing mortality and recruitment patterns (Figure 4). When tested using the same model
405 structure and assumptions in the operating model, LIME performed best for short-lived fish, with
406 increasing bias and decreasing precision for medium-lived fish, and a further deterioration for
407 longer-lived fish (Table 5). The bias in SPR in the terminal year was 0.002 for short-lived, -0.003
408 for medium-lived, and 0.016 for longer-lived fish across all data availability and population
409 variability scenarios with 200 length measurements annually. Additional scenarios illustrate
410 further LIME strengths and weaknesses (Table 5). With only one year of length data, LIME
411 estimated SPR with bias of -0.038 for the short-lived life history, -0.186 for medium-lived, and -
412 0.152 for longer-lived across variability scenarios. Integrating ten years of length data improved
413 accuracy in the estimation of SPR in the terminal year on average for each life history type
414 (short-lived from -0.038 to -0.016; medium-lived from -0.186 to -0.048; longer-lived from -
415 0.152 to 0.102). The only exception within variability scenarios was decreased accuracy with ten
416 years of length data for the longer-lived life-history type under the one-way variability scenario
417 (from -0.181 to 0.200). While ten years of length data increased precision of SPR estimates for
418 the short-lived life-history type over one year of length data (from 0.072 to 0.041), the additional

419 length data decreased precision on average for the medium- and longer-lived life-history types
420 (medium-lived from 0.334 to 0.391; longer-lived from 0.529 to 0.626; Table 5).

421 LIME provides the added advantage over other length-based assessment methods of
422 including a catch time series and abundance index, if available, which further increases accuracy
423 and precision under most cases. Across all scenarios, including an abundance index decreased
424 bias (from -0.047 to 0.020) compared to when only length data were available. An abundance
425 index particularly improved LIME performance for the short- and medium-lived life-history
426 types. Breaking down by life-history types, bias decreases with an abundance index included for
427 the short-lived life-history type (from -0.026 to 0.005) and medium-lived life-history type (from
428 -0.126 to 0.025), but increases for the longer-lived life-history type (from -0.049 to 0.072) on
429 average. Across all life-history types, including an abundance index increases precision (short-
430 lived from 0.054 to 0.052; medium-lived from 0.346 to 0.190; longer-lived from 0.594 to 0.257).

431 Adding a catch time series to length data decreased bias minimally on average across all
432 scenarios (from -0.047 to 0.031). The improvement in accuracy is strongest under equilibrium
433 conditions (from -0.036 to -0.007). Given variable fishing mortality and recruitment, including
434 catch data with length composition increased bias (from -0.055 to 0.118). This was mainly due to
435 high bias in the “catch plus one” scenario for short-lived fish (Table 5, Figure 4). Excluding this
436 anomalous scenario, including a catch time series decreased bias for a population with variability
437 (from -0.055 to 0.047). The “catch plus ten” scenario removed the bias from the “catch plus one”
438 scenario for the short-lived life-history type under variability scenarios. However, the “catch plus
439 ten” scenario did not necessarily improve bias nor precision over the “catch plus one” scenario
440 for the medium- and longer-lived life-history types under variability scenarios (Table 5). This
441 indicates that it may be advisable to include a catch time series only when more than one year of

442 length data are available for a short-lived life-history type, and consider the possibility of over-
443 estimating SPR when fishing mortality is changing over time for medium- and longer-lived life-
444 history types when catch data are included.

445 LIME converged for 95% of iterations of generated data across life-history types,
446 variability scenarios, data availability scenarios, and sample sizes of length data. LIME
447 converged for 100% of iterations of generated data for the data-rich scenario (Figure 5). With
448 only length data, LIME converged for 95% of iterations across scenarios. The LIME
449 convergence rate was 92% including catch data and 94% including an abundance index. Non-
450 convergence was always due to a high final gradient in parameter estimation, as opposed to any
451 parameter being estimated at the upper or lower bound.

452 The interval coverage for a 50% confidence interval of LIME was 61% across all
453 scenarios of life history, variability scenarios, data availability, and sample sizes of length data.
454 This indicates confidence interval estimates from LIME are generally informative about model
455 uncertainty, but may be wider (more conservative) on average. LIME estimated confidence
456 intervals that were too narrow or wide for individual life history, variation, and data availability
457 scenarios (Figure 5). The “catch plus one” scenario had the lowest coverage across life-history
458 and variability scenarios (42%), likely because this data availability scenario had the most bias,
459 and thus the true SPR would likely not fall within the confidence intervals at the expected rate.
460 For the short-lived life-history type, LIME interval coverage was only 43%. While the method is
461 expected to be unbiased for this life-history type on average, the confidence intervals are often
462 too tight to include the true SPR within the 50% intervals at a rate of 50%. LIME estimated
463 wider confidence intervals for the medium-lived life-history type, with interval coverage 72%,
464 and longer-lived life-history types, with an interval coverage of 67%.

465

466 *LIME sensitivity to sample size, input parameters, and selectivity estimation*

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

With one year of length data, LIME estimated SPR with greater precision with more length measurements, but not necessarily with greater accuracy (Table 6). Across all life history types, accuracy and precision improved greatly in the equilibrium scenario between 20 and 1,000 length measurements collected annually. The short-lived life-history type experienced the strongest improvements in accuracy and precision (Table 6). This pattern was strongest with a population in equilibrium, but was generally maintained for scenarios of variability as well. Accuracy generally reached a plateau with greater than 100 samples of length measurements for the short-lived life-history type. Alternatively, accuracy and precision improved for the medium- and longer-lived life-history types with at least 500 samples of length measurements for a population at equilibrium. With variability, there were no clear patterns in accuracy improvement for the medium- and longer-lived life-history types with increasing sample size of length measurements. This indicates accuracy in SPR estimates is more likely to be improved with more data types than increasing sample size, particularly for medium- and longer-lived life histories (Table 5, Table 6).

Like all age or length-based methods, LIME performance is sensitive to the correct specification of life history information (Figure 6). When all biological parameters were correctly specified, LIME estimated unbiased SPR on average with ten years of length data (bias = -0.017). When L_{∞} was misspecified as 25% greater than the truth, bias increased (to -0.554), meaning that on average LIME estimated SPR to be lower than the truth. In this case, we would expect to see larger fish in the observed data. If those fish are not present in the length data, length-based models attribute the difference to a higher fishing mortality, and thus a lower SPR

488 than the truth. On the other hand, when L_{∞} was misspecified as 25% lower than the truth, LIME
489 estimated SPR higher than the truth on average (bias = 0.761). Assuming k was 25% lower than
490 the truth, LIME estimated SPR higher than the truth on average (bias = 0.478). When k was
491 assumed to be 25% higher than the truth, LIME estimated SPR lower than the truth on average,
492 but to a lower degree than the other biases due to life history misspecification (bias = -0.146).
493 However, this lack of severe bias is simply due some model runs resulting in estimates much
494 greater than or lower than the truth, while none are unbiased. When M was fixed 25% higher
495 than the truth, LIME estimated SPR to be higher than the truth (bias = 0.285). A fish that is faster
496 to die can generally sustain a higher fishing pressure, and the interpretation of the higher SPR is
497 that the population has more of its potential spawning biomass than it truly does. When M was
498 assumed to be 25% lower than the truth, LIME estimated SPR lower than the truth (bias = -
499 0.352).

500 Fixing length at 50% maturity at a value 25% higher than the truth resulted in negative
501 bias in SPR (bias = -0.383). In this case the LIME model attributes a lower proportion of the
502 population as being mature, leading the estimated SPR to be lower than the truth. The opposite is
503 true when length at maturity was 25% lower than the truth; estimates of SPR were biased higher
504 than the truth (bias = 0.152). LIME was relatively insensitive to misspecification of CV_L for the
505 age-length curve, but a higher CV_L resulted in a relatively lower estimate of SPR on average
506 (bias = -0.044), and a higher CV_L resulted in a relatively higher estimate of SPR on average (bias
507 = 0.047).

508

509 *Comparing LIME and LB-SPR*

510 Simulation testing demonstrated that the LIME and LB-SPR methods perform well under
511 different conditions. LB-SPR is biased when the equilibrium conditions are violated (Table 5).
512 However, testing of LB-SPR against the age-structured LIME operating model with
513 instantaneous sampling at the beginning of each year resulted in poor performance for the short-
514 lived life-history type under equilibrium conditions (Table 5, Table 7, Figure 7). We verified that
515 LB-SPR performs well across life-history types when tested using an operating model that
516 matches its model assumptions. LIME, on the other hand, over-estimated SPR for the short-lived
517 life history type with one year of length data and, similar to the performance under its own
518 operating model, had low precision for the medium- and longer-lived life history types with ten
519 years of length data (Table 7). When length data were collected monthly and pooled into an
520 annual length composition, LB-SPR performance improved over the annual model for the short-
521 lived life history type, but LIME over-estimated SPR (Table 7, Figure 7). Running LIME on a
522 monthly time step with monthly length data decreased bias compared to running on an annual
523 time step (from 0.468 to -0.040 with ten years of monthly length data and 0.484 to 0.034 with
524 one year of monthly length data; Table 7).

525

526 Discussion

527 This study demonstrates that the length-based integrated mixed-effects (LIME) method
528 can be a valuable tool for fisheries stock assessment when at least one year of length data and
529 basic biological information are available for the species (Table 8). LIME estimates reference
530 points more accurately than LB-SPR under many common scenarios. The LIME assessment
531 method has several characteristics that make it particularly useful in situations where an
532 abundance index and/or catch data cannot be collected reliably, as data collection programs

533 improve to include more data types, and when recruitment and fishing mortality are believed to
534 be changing more rapidly than would match the equilibrium assumptions required for existing
535 methods with the same data requirements.

536 The *length-based* aspect of the model setup allows the analyst to fit to length, rather than
537 age, composition data from the catch. The current toolbox of data-limited stock assessment
538 methods includes length-only methods with equilibrium assumptions (Hordyk et al. 2015; Nadon
539 et al. 2015; Kokkalis et al. 2015), *ad hoc* assumptions about changes in mortality (Gedamke and
540 Hoenig 2006), or age-structured models that cannot fit to length-composition data (Martell and
541 Froese 2013; Thorson and Cope 2014). MULTIFAN (Fournier et al. 1990) and Stock Synthesis
542 (Methot and Wetzel 2013) can use length instead of age data to gain information on cohort
543 strength and total mortality, but require a catch time series to estimate stock status. The
544 flexibility of LIME to fit to length composition, as opposed to age-composition, is more realistic
545 for capacity-limited fisheries. LIME can be used in conjunction with other length-based methods
546 as a diagnostic tool to see if variations in recruitment or fishing mortality are being predicted by
547 the model based on the length data, and how the results of the multiple assessment types would
548 vary in light of those possibly violated assumptions.

549 The *integrated* nature of LIME is useful in situations where monitoring programs are
550 continuing to be developed over time, incorporating more years of length data or other data types
551 in the same assessment framework. LB-SPR is a quick way of conducting an assessment with
552 only one year of length data, but as monitoring programs improve, the method must be applied
553 independently over multiple years (Prince et al. 2015a, 2015b). There are also many catch-based
554 methods that assume a catch time series is available, but they are not thoroughly tested in fitting
555 to length, rather than age, composition data (Sabater and Kleiber 2013; Martell and Froese 2013;

556 Thorson and Cope 2014). LIME allows for these data to be included into the same assessment
557 framework, rather than requiring a switch between assessment models as monitoring programs
558 develop and new data types become available (Maunder and Punt 2013). Also, few existing
559 models can incorporate an abundance index and length data without catch data. LIME can be
560 used in this scenario to estimate a relative SPR reference point with no measure of scale. This
561 scenario is common for small-scale fisheries, bycatch species, or other fisheries where trends in
562 abundance are available but total harvest is not.

563 Finally, the *mixed effects* aspect of LIME extends length-based methods by estimating
564 changes in recruitment and fishing mortality over time. Estimating random effects has been
565 integrated into existing stock assessment models, such as Stock Synthesis (Thorson et al. 2015a)
566 and the state-space assessment model (SAM) framework (Nielsen and Berg 2014). Many
567 assessments for European stocks now use SAM to separate process and observation errors as an
568 objective method of weighting data in age-structured models (Berg et al. 2014). Computing the
569 marginal likelihood for mixed-effects models was previously too computationally challenging,
570 hindering wide application of mixed-effects models. Now, programs such as TMB can do these
571 computations much faster than previous options (Nielsen and Berg 2014). To illustrate the
572 benefits of LIME, a vignette is available on the repository site that walks the user through a
573 simple example, with tips for model interpretation and convergence. An app using the R package
574 *shiny* (Chang et al. 2017) is also available on the repository site
575 (https://github.com/merrillrudd/LIME_shiny) that allows analysts to run the model in a graphical
576 user interface.

577 Simulation testing in this study demonstrated the best LIME performance with only
578 length data for the short-lived life-history type, with comparatively lower performance for

579 medium- and longer-lived fish. With only length data, LIME performs well for the shorter-lived
580 fish likely because the model is tracking cohorts through the length data to estimate recruitment
581 deviations. This is likely difficult for the longer-lived life histories due to variation in the age-
582 length relationship (especially for older fish), and because each individual cohort represents a
583 small proportion of total abundance (hence requiring more samples to track each individual
584 cohort). The increasing uncertainty in a longer-lived fish's age with increasing length blurs the
585 cohorts as they age, making it difficult to track recruitment events without a much longer time
586 series of length data. With ten years of length data, accuracy in estimating SPR increases but
587 precision decreases for the medium- and longer-lived life-history types, providing support for the
588 increased ability to track recruitment events but propagating uncertainty in the age-length curve
589 for older individuals. With short-lived fish that only live to four years old, each length more
590 clearly matches up with an age group. Regardless of fish growth, one year of length data holds
591 information on an entire generation for a short-lived fish. Even if recruitment is occurring
592 monthly or seasonally and not annually, a ten-year time series of length data can inform variation
593 in cohort strength across several generations of the short-lived fish.

594 By testing LIME and LB-SPR using multiple operating models, we identified scenarios
595 in which each method excels. LB-SPR performs better than LIME for estimating SPR if the
596 population is at equilibrium and only length data are available. If the population is not in
597 equilibrium, then LIME performs better. If the data are collected continuously throughout the
598 year (e.g. monthly), LIME is expected to estimate higher SPR than the truth for short-lived fish.
599 In this case, analysts should run LIME using monthly (or shorter) time steps to account for fish
600 growth during the year. This bias does not occur for medium- or longer-lived fish, where the
601 growth during the year is less rapid. On the other hand, LB-SPR likely under-estimates SPR for

602 short-lived fish if data are collected instantaneously, but is unbiased if data are collected
603 continuously.

604 An important question for management is often how many years or how much data are
605 enough, particularly for fisheries where funding for monitoring is limited. The simulation testing
606 in this study demonstrated that we can get accurate but highly uncertain estimates of SPR,
607 fishing mortality, and recruitment from only one year of length data on average (Figure 8). As
608 expected, collecting more data (ten years of length data instead of one) resulted in greater
609 accuracy and precision in LIME estimates of SPR (Figure 4, Table 5). When using only length
610 data, LIME performance is best with at least 500 independent length measurements per year for
611 medium- and longer-lived life histories, but is expected to perform well with 100 or more length
612 measurements per year for short-lived life-history types (Table 6). However, including an
613 abundance index or catch time series improves bias and precision to a greater extent than
614 collecting more independent length measurements (Table 5, Table 6). Managers must determine
615 how much uncertainty they are comfortable with to set monitoring goals. Collecting more years
616 of length data, taking more independent length measurements during each year of length data
617 collection, conducting surveys, and monitoring catch data hold varying levels of information
618 used to estimate variable fishing mortality and recruitment. Their costs and benefits should be
619 weighed to help managers prioritize data collection.

620 Including an abundance index or catch time series improves LIME performance in most
621 scenarios. Catch data provides useful information on the scale of the population, which is not
622 possible with length data alone. The abundance index provides useful information to help inform
623 the trajectory of the population from which the length data arose. The abundance index could be
624 more informative on the state of the population than length data alone, as long as the abundance

625 index is proportional to abundance and any significant changes in fishing mortality and
626 recruitment occurred during the surveyed time series. Our simulation study shows that the biases
627 associated with including a time series with one year of length data are often overcome with
628 more years of length data. The exception is for the longer-lived life history type, for which LIME
629 performs worse when including an abundance index with ten years of length composition data
630 than if the abundance index was excluded.

631 However, the simulation study also identified some potential issues with data conflict
632 when an abundance index or catch time series are included with length data. Data conflict can be
633 diagnostic of poor data quality, such as catch misreporting, indices from spatial areas with
634 ontogenetic differences, or length data not representative of the fishery. Data quality is an
635 equally important issue as limited data types and should be considered in LIME applications. In
636 the case of this study, however, poor data quality was not the culprit behind any data conflict
637 because we used the true catch data, abundance index from a single area, and representative
638 length data to test LIME. To avoid data conflict, Maunder and Piner (2015) recommend
639 modeling process error explicitly, most commonly via time-varying recruitment, as a better
640 alternative to down-weighting or eliminating data conflicts. LIME takes this approach, but data
641 conflict may still occur, particularly because recruitment deviations are treated as a random
642 effect even if this is not the correct model process for which the data hold conflicting information
643 (Maunder and Piner 2015). Other options to avoid data conflict are to estimate the variance
644 parameter for observed data outside of the stock assessment model (Lee et al. 2014; Maunder
645 and Piner 2015). Alternatively, an analyst could identify whether data conflict is occurring in the
646 LIME model by likelihood profiling individual data components (Ichinokawa et al. 2014) or

647 retrospective analysis, which quantifies the impact of additional years of data on the stock
648 assessment output (Hurtado-Ferro et al. 2014).

649 Like all length-based assessment methods, fixing the biological parameters at their true
650 values is an important first step for estimating unbiased reference points. LB-SPR circumvents
651 this strong assumption by using Beverton-Holt life history invariants to estimate the expected
652 age or length structure, as opposed to assuming known values of natural mortality and growth
653 (Prince et al. 2015a). Sensitivity tests in this study demonstrating the impact of misspecifying
654 biological parameters provide support for local studies of species growth rates for stocks
655 assessed using LIME. Analysts should be aware of how SPR is expected to be biased given their
656 assumptions on fixed values for input parameters or model structure. A next step for LIME is to
657 use Bayesian priors on biological parameters to more thoroughly represent the uncertainty in
658 population parameter estimates relevant to management (e.g., from FishLife; Thorson et al. In
659 press). Sensitivity tests and likelihood profiles should be conducted on different levels of dome-
660 shaped selectivity to understand how SPR is expected to be biased if the model structure is
661 misspecified.

662 Length-based stock assessments are good starting-points for making management
663 decisions with limited data and monitoring capacity. The shrinkage of poorly estimated
664 parameters towards an estimated distribution (as implemented within mixed effects models) has
665 been shown to increase accuracy and precision in stock assessments (Thorson et al. 2013;
666 Nielsen and Berg 2014), but high uncertainty in estimates of stock status will always result if
667 there are only one or two years of length data. Any stock assessment deals with data limitation
668 and uncertainty, and it is vital to appropriately represent and communicate this uncertainty to
669 managers.

670 Ideally, conclusions on stock status should be drawn using an ensemble of assessment
671 models with varying structure and assumptions (Stewart and Martell 2015; Anderson et al.
672 2017). In the data-limited context, existing models represent a relatively small range of
673 alternative model assumptions. We have shown that LIME presents a way to represent both
674 process and observation uncertainty. LIME can complement other length-based (Hordyk et al.
675 2015; Nadon et al. 2015; Gedamke and Hoenig 2006) and catch-only methods (Rosenberg et al.
676 2014; Carruthers et al. 2014), with the strength of estimating recruitment variability. More
677 accurate and precise estimates of recruitment variability can help decipher whether decreased
678 mean length is due to fishing pressure or recruitment variability, and understanding the range of
679 possible levels of recruitment into the future. We therefore conclude that LIME is a step forward
680 in dealing with uncertainty in decision-making for fisheries where length data are collected.

681

682 *Acknowledgements*

683 M.B.R. was funded by the National Science Foundation IGERT Program on Ocean Change and
684 a University of Washington School of Aquatic and Fishery Sciences fellowship. We thank J.
685 Cope, T. Branch, R. Hilborn, J. Hastie, M. McClure, and T.R. McClanahan for their early review
686 of this manuscript and A. Hordyk for discussions and feedback throughout the method
687 development process.

688

689 **References**

690 Ault, J.S., Smith, S.G., and Bohnsack, J.A. 2005. Evaluation of average length as an estimator of
691 exploitation status for the Florida coral-reef fish community. *ICES J. Mar. Sci.* **62**(3): 417–
692 423. doi: 10.1016/j.icesjms.2004.12.001.

- 693 Ault, J.S., Smith, S.G., Luo, J., Monaco, M.E., and Appeldoorn, R.S. 2008. Length-based
694 assessment of sustainability benchmarks for coral reef fishes in Puerto Rico. *Environ.*
695 *Conserv.* **35**(3): 221–231. doi: 10.1017/S0376892908005043.
- 696 Azzurro, E., Moschella, P., and Maynou, F. 2011. Tracking signals of change in mediterranean
697 fish diversity based on local ecological knowledge. *PLoS One* **6**(9): e24885. doi:
698 10.1371/journal.pone.0024885.
- 699 Berg, C.W., Nielsen, A., and Kristensen, K. 2014. Evaluation of alternative age-based methods
700 for estimating relative abundance from survey data in relation to assessment models. *Fish.*
701 *Res.* **151**: 91–99. doi: 10.1016/j.fishres.2013.10.005.
- 702 Beverton, R.J.H., and Holt, S.J. 1957. On the dynamics of exploited fish populations. Ministry of
703 Agriculture, Lowestoft, UK.
- 704 Buckland, S.T., Newman, K.B., Thomas, L., and Koesters, N.B. 2004. State-space models for the
705 dynamics of wild animal populations. *Ecol. Modell.* **171**: 157–175. doi:
706 10.1016/j.ecolmodel.2003.08.002.
- 707 Butler, J.R.A., Middlemas, S.J., Graham, I.M., Thompson, P.M., and Armstrong, J.D. 2006.
708 Modelling the impacts of removing seal predation from Atlantic salmon, *Salmo salar*, rivers
709 in Scotland: A tool for targeting conflict resolution. *Fish. Manag. Ecol.* **13**(5): 285–291. doi:
710 10.1111/j.1365-2400.2006.00504.x.
- 711 Bystrom, A.B. 2015. Análisis de características biológico-pesqueras del pargo manchado
712 (*Lutjanus guttatus* (Steindachner, 1869) y tendencias socio-ecológicas de la pesca artesanal
713 con líneas de fondo en el distrito de Bejuco, Pacífico de Costa Rica. M.Sc. thesis, Manejo
714 de Recursos Naturales, Universidad Estatal a Distancia, San Jose, Costa Rica.
- 715 Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., and Cope,

- 716 J. 2014. Evaluating methods for setting catch limits in data-limited fisheries. *Fish. Res.* **153**:
717 48–68. doi: 10.1016/j.fishres.2013.12.014.
- 718 Chang, W., Cheng, J., Allaire, J.J., Xie, Y., and McPherson, J. 2017. Shiny: Web Application
719 Framework for R. R package version 1.0.0. <https://CRAN.R-project.org/package=shiny>
- 720 Chrysafi, A., and Kuparinen, A. 2016. Assessing abundance of populations with limited data:
721 Lessons learned from data-poor fisheries stock assessment. *Environ. Rev* **24**: 25-38. doi:
722 10.1017/CBO9781107415324.004.
- 723 Clark, W.G. 2002. F35% revisited ten years later. *North Am. J. Fish. Manag.* **22**(1): 251–257.
- 724 Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status
725 and solutions for the world's unassessed fisheries. *Science* **338**: 517–520. doi:
726 10.1126/science.1223389.
- 727 Darcy, G.H., and Matlock, G.C. 1999. Application of the precautionary approach in the national
728 standard guidelines for conservation and management of fisheries in the United States.
729 *ICES J. Mar. Sci.* **56**(6): 853–859. doi: 10.1006/jmsc.1999.0533.
- 730 Dowling, N.A., Dichmont, C.M., Haddon, M., Smith, D.C., Smith, A.D.M., and Sainsbury, K.
731 2015. Empirical harvest strategies for data-poor fisheries: A review of the literature. *Fish.*
732 *Res.* **171**: 141–153. doi: 10.1016/j.fishres.2014.11.005.
- 733 Dowling, N.A., Wilson, J.R., Rudd, M.B., Babcock, E.A., Dougherty, D., and Gleason, M. 2016.
734 FishPath: A Decision Support System for Assessing and Managing Data- and Capacity-
735 Limited Fisheries. *In* Assessing and managing data-limited fish stocks. *Edited by* T.J.I.
736 Quinn, J.L. Armstrong, M.R. Baker, J. Heifetz, and D. Witherell. Alaska Sea Grant,
737 University of Alaska Fairbanks. pp. 59–96. Available from
738 <http://doi.org/10.4027/amdlfs.2016.03>.

- 739 Ehrhardt, N.M., and Ault, J.S. 1992. Analysis of Two Length-Based Mortality Models Applied
740 to Bounded Catch Length Frequencies. *Trans. Am. Fish. Soc.* **121**: 115–122.
- 741 FAO. 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security
742 and nutrition for all. Rome. 200 pp.
- 743 Fournier, D.A., Sibert, J.R., Majkowski, J., and Hampton, J. 1990. MULTIFAN a Likelihood-
744 Based Method for Estimating Growth Parameters and Age Composition from Multiple
745 Length Frequency Data Sets Illustrated using Data for Southern Bluefin Tuna (*Thunnus*
746 *maccoyii*). *Can. J. Fish. Aquat. Sci.* **47**: 301–317.
- 747 Francis, R.I.C.C. 2014. Replacing the multinomial in stock assessment models: A first step. *Fish.*
748 *Res.* **151**: 70–84. doi: 10.1016/j.fishres.2013.12.015.
- 749 Froese, R. 1990. FishBase: An information system to support fisheries and aquaculture research.
750 *ICLARM Fishbyte* **8**(3): 21–24.
- 751 Gedamke, T., and Hoenig, J.M. 2006. Estimating mortality from mean length data in
752 nonequilibrium situations, with application to the assessment of goosfish. *Trans. Am. Fish.*
753 *Soc.* **135**(2): 476–487. doi: 10.1577/T05-153.1.
- 754 Goodyear, C.P. 1993. Spawning stock biomass per recruit in fisheries management: foundation
755 and current use Canadian Special Publicatio. *In Risk Evaluation and Biological Reference*
756 *Points for Fisheries Management. Edited by S.J. Smith, J.J. Hunt, and D. Rivard. Canadian*
757 *Special Publication on Fisheries and Aquatic Science.*
- 758 Gulbrandsen, L.H. 2009. The emergence and effectiveness of the Marine Stewardship Council.
759 *Mar. Policy* **33**(4): 654–660. doi: 10.1016/j.marpol.2009.01.002.
- 760 Harley, S.J., Myers, R. a., and Dunn, A. 2001. Is catch-per-unit-effort proportional to
761 abundance? *Can. J. Fish. Aquat. Sci.* **58**(9): 1760–1772. doi: 10.1139/cjfas-58-9-1760.

- 762 Heemstra, P.C., and Randall, J.E. 1993. FAO Species Catalogue. Vol. 16. Groupers of the world
763 (family *Serranidae*, subfamily *Epinephelinae*). An annotated and illustrated catalogue of the
764 grouper, rockcod, hind, coral grouper, and lyretail species known to date. Rome: FAO. FAO
765 Fish. Synop. **125**(16): 382 pp.
- 766 Hicks, C.C., and McClanahan, T.R. 2012. Assessing gear modifications needed to optimize
767 yields in a heavily exploited, multi-species, seagrass and coral reef fishery. PLoS ONE **7**(5):
768 e36022.
- 769 Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015. A novel length-based
770 empirical estimation method of spawning potential ratio (SPR) and tests of its performance,
771 for small-scale, data-poor fisheries. ICES J. Mar. Sci. **72**(1): 217-231.
- 772 Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson,
773 K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre,
774 K.A., Whitten, A.R., and Punt, A.E. 2014. Looking in the rear-view mirror: Bias and
775 retrospective patterns in integrated, age-structured stock assessment models. ICES J. Mar.
776 Sci. **72**(1): 99–110. doi: 10.1093/icesjms/fsu198.
- 777 Ichinokawa, M., Okamura, H., and Takeuchi, Y. 2014. Data conflict caused by model mis-
778 specification of selectivity in an integrated stock assessment model and its potential effects
779 on stock status estimation. Fish. Res. **158**: 147–157. doi: 10.1016/j.fishres.2014.02.003.
- 780 Kokkalis, A., Thygesen, U.H., Nielsen, A., and Andersen, K.H. 2015. Limits to the reliability of
781 size-based fishing status estimation for data-poor stocks. Fish. Res. **171**: 4–11.
782 doi:10.1016/j.fishres.2014.10.007.
- 783 Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B. 2016. TMB: Automatic

- 784 Differentiation and Laplace Approximation. *J. Stat. Softw.* **70**: 1–21. doi:
785 10.18637/jss.v070.i05.
- 786 Lee, H.H., Piner, K.R., Methot, R.D., and Maunder, M.N. 2014. Use of likelihood profiling over
787 a global scaling parameter to structure the population dynamics model: AN example using
788 blue marlin in the Pacific Ocean. *Fish. Res.* **158**: 138–146. doi:
789 10.1016/j.fishres.2013.12.017.
- 790 Mahon, R. 1997. Does fisheries science serve the needs of managers of small stocks in
791 developing countries. *Can. J. Fish. Aquat. Sci.* **54**(9): 2207–2213. doi: 10.1139/cjfas-54-9-
792 2207.
- 793 Martell, S., and Froese, R. 2013. A simple method for estimating MSY from catch and resilience.
794 *Fish Fish.* **14**(4): 504–514.
- 795 Maunder, M.N., and Piner, K.R. 2017. Dealing with data conflicts in statistical inference of
796 population assessment models that integrate information from multiple diverse data sets.
797 *Fish. Res.* **192**: 16–27. doi: 10.1016/j.fishres.2016.04.022.
- 798 Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock
799 assessment. *Fish. Res.* **142**: 61–74. doi: 10.1016/j.fishres.2012.07.025.
- 800 Melnychuk, M.C., Peterson, E., Elliott, M., and Hilborn, R. 2017. Fisheries management impacts
801 on target species status. *Proc. Natl. Acad. Sci.* **14**: 178-183.
- 802 Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for
803 fish stock assessment and fishery management. *Fish. Res.* **142**: 86-99.
- 804 Methot, R.D., Tromble, G.R., Lambert, D.M., and Greene, K.E. 2014. Implementing a science-
805 based system for preventing overfishing and guiding sustainable fisheries in the United
806 States. *ICES J. Mar. Sci.* **71**(2): 183–194. doi: 10.1093/icesjms/fst034.

- 807 Nadon, M.O., Ault, J.S., Williams, I.D., Smith, S.G., and DiNardo, G.T. 2015. Length-Based
808 Assessment of Coral Reef Fish Populations in the Main and Northwestern Hawaiian
809 Islands. *PLoS One* **10**(8): e0133960. doi: 10.1371/journal.pone.0133960.
- 810 Neis, B. 1992. Fishers' Ecological Knowledge and Stock Assessment in Newfoundland.
- 811 Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments
812 using state-space models. *Fish. Res.* **158**: 96–101. doi: 10.1016/j.fishres.2014.01.014.
- 813 Prince, J., Hordyk, A., Valencia, S.R., Loneragan, N., and Sainsbury, K. 2015a. Revisiting the
814 concept of Beverton-Holt life-history invariants with the aim of informing data-poor
815 fisheries assessment. *ICES J. Mar. Sci.* **72**(1): 194-203.
- 816 Prince, J., Victor, S., Kloulchad, V., and Hordyk, A. 2015b. Length based SPR assessment of
817 eleven Indo-Pacific coral reef fish populations in Palau. *Fish. Res.* **171**: 42–58. doi:
818 10.1016/j.fishres.2015.06.008.
- 819 Quinn, T.J.I., Armstrong, J.L., Baker, M., Heifetz, J., and Witherell, D. 2016. Assessing and
820 Managing Data-Limited Fish Stocks. Alaska Sea Grant, University of Alaska Fairbanks.
- 821 R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for
822 Statistical Computing, Vienna, Austria. Available from <https://www.r-project.org/>.
- 823 Ricard, D., Minto, C., Jensen, O.P., Baum, J.K. 2012. Examining the knowledge base and status
824 of commercially exploited marine species with the RAM legacy stock assessment database.
825 *Fish. Fish.* **13**(4): 380-398.
- 826 Rosenberg, A.A., Fogarty, M.J., Cooper, A.B., Dickey-Collas, M., Fulton, E.A., Gutierrez, N.L.,
827 Hyde, K.J.W., Kleisner, K.M., Kristiansen, T., Longo, C., Minto-Vera, C., Minto, C.,
828 Mosqueira, I., Chato Osio, G., Ovando, D., Selig, E.R., Thorson, J.T., and Ye, Y. 2014.
829 Developing new approaches to global stock status assessment and fishery production

- 830 potential of the seas. *In* FAO Fisheries and Aquaculture Circular No. 1086. Rome, FAO.
- 831 Sabater, M., and Kleiber, P. 2014. Augmented catch-MSY approach to fishery management in
832 coral-associated fisheries. *In* Interrelationships Between Corals and Fisheries. *Edited by*
833 S.A. Bortone. CRC Press. pp. 199–218. doi: 19.1201/b17159-12.
- 834 Schnute, J.T., and Haigh, R. 2007. Compositional analysis of catch curve data, with an
835 application to *Sebastes maliger*. *ICES J. Mar. Sci.* **64**(2): 218–233. doi:
836 10.1093/icesjms/fsl024.
- 837 Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., and Hilborn, R. 2015. Examining
838 common assumptions about recruitment: A meta-analysis of recruitment dynamics for
839 worldwide marine fisheries. *Fish Fish.* **16**(4): 633–648. doi: 10.1111/faf.12083.
- 840 Thorson, J.T., J.M. Cope, and W.S. Patrick. 2014a. Assessing the quality of life history
841 information in publicly available databases. *Ecological Applications* 24:217–226.
- 842 Thorson, J.T., A. C. Hicks, and R. D. Methot. 2015b. Random effect estimation of time-varying
843 factors in Stock Synthesis. *ICES Journal of Marine Science* **72**: 178–185.
- 844 Thorson, J.T., and Kristensen, K. 2016. Implementing a generic method for bias correction in
845 statistical models using random effects, with spatial and population dynamics examples.
846 *Fish. Res.* **175**: 66–74. doi:10.1016/j.fishres.2015.11.016.
- 847 Thorson, J.T., Skaug, H., Kristensen, K., Shelton, A., Ward, E., Harms, J., and Benante, J.
848 2015a. The importance of spatial models for estimating the strength of density dependence.
849 *Ecology* **96**(5): 1202–1212.
- 850 Thorson, J.T., and Cope, J.M. 2014. Catch curve stock-reduction analysis: An alternative
851 solution to the catch equations. *Fish. Res.* **171**: 33–41. doi: 10.1016/j.fishres.2014.03.024.
- 852 Thorson, J.T., Jensen, O.P., and Zipkin, E.F. 2014b. How variable is recruitment for exploited

- 853 marine fishes? A hierarchical model for testing life history theory. *Can. J. Fish. Aquat. Sci.*
854 **71**(7): 973–983.
- 855 Thorson, J.T., Johnson, K.F., Methot, R.D., and Taylor, I.G. 2017. Model-based estimates of
856 effective sample size in stock assessment models using the Dirichlet-multinomial
857 distribution. *Fish. Res.* **192**: 84-93.
- 858 Thorson, J.T., and Minto, C. 2015. Mixed effects: a unifying framework for statistical modelling
859 in fisheries biology. *ICES J. Mar. Sci.* **72**(5): 1245–1256. doi: 10.1093/icesjms/fst048.
- 860 Thorson, J.T., Minto, C., Minto-vera, C. V, Kleisner, K.M., and Longo, C. 2013. A new role for
861 effort dynamics in the theory of harvested populations and data-poor stock assessment. *Can.*
862 *J. Fish. Aquat. Sci.* **70**(12): 1829–1844.
- 863 Thorson, J.T., Munch, S.B., Cope, J.M., and Gao, J. In press. Predicting life history parameters
864 for all fishes worldwide. *Ecol. Appl.*
- 865 de Valpine, P., and Hastings, A. 2002. Fitting population models incorporating process noise and
866 observation error. *Ecol. Monogr.* **72**(1): 57–76.
- 867 Vert-pre, K.A., Amoroso, R.O., Jensen, O.P., and Hilborn, R. 2013. Frequency and intensity of
868 productivity regime shifts in marine fish stocks. *Proc. Natl. Acad. Sci.* **110**(5): 1779–84.
869 doi: 10.1073/pnas.1214879110.
- 870 Wetzel, C.R., and Punt, A.E. 2011. Model performance for the determination of appropriate
871 harvest levels in the case of data-poor stocks. *Fish. Res.* **110**(2): 342–355. doi:
872 10.1016/j.fishres.2011.04.024.

873

874

875

876

877 Table 1. Parameter definitions, including parameter input values for the base scenario for each of
 878 the three life history types, and whether the parameter is fixed (“Fixed”) or estimated (“Est.”)
 879 depending on the data availability scenario, or if the parameter is used in data generation only
 880 (“Sim.”). Values separated by commas indicate alternate values for sensitivity analysis. Note that
 881 scenarios including “LC” are each divided into two scenarios with either one or ten years of
 882 length data. Parameter input values are based on short-lived *Siganus sutor* (Hicks and
 883 McClanahan 2012), medium-lived *Lutjanus guttatus* (Bystrom 2015), and longer-lived
 884 *Epinephelus morio* (Heemstra and Randall 1993).

Symbol	Description	Input value			Data Scenario			
		Life history type			Index	Catch	LC	
Biological		Short	Medium	Longer	Rich	+ LC	+ LC	only
L_{∞}	Asymptotic length	36.2cm	64.6cm	90.0cm	Fixed	Fixed	Fixed	Fixed
k	Brody growth coefficient	0.87	0.21	0.13	Fixed	Fixed	Fixed	Fixed
t_0	Age at length=0	-0.01	-0.01	-0.01	Fixed	Fixed	Fixed	Fixed
L_m^{50}	Length at 50% maturity	20.2cm	34.0cm	50.0cm	Fixed	Fixed	Fixed	Fixed
α	Length-weight scalar	0.0597	0.0245	0.0264	Fixed	Fixed	Fixed	Fixed
β	Length-weight allometric	2.75	2.79	2.96	Fixed	Fixed	Fixed	Fixed
A	Maximum age	4	18	26	Fixed	Fixed	Fixed	Fixed
M	Natural mortality	1.49	0.43	0.18	Fixed	Fixed	Fixed	Fixed
h	Steepness parameter	1	1	1	Fixed	Fixed	Fixed	Fixed
CV_L	Coefficient of variation for the length-age curve	0.1	0.1	0.1	Fixed	Fixed	Fixed	Fixed
J	Maximum length bin	54cm	97cm	135cm	Fixed	Fixed	Fixed	Fixed
R_0	Equilibrium recruitment	1	1	1	Est.	Fixed	Est.	Fixed
σ_R	Recruitment standard deviation	0.737	0.737	0.737	Est.	Est.	Est.	Est.
ρ	Recruitment autocorrelation	0.426	0.426	0.426	Sim.	Sim.	Sim.	Sim.
Non-biological								
L_s^{50}	Length at 50% selectivity	11.3cm	20.0cm	25.0cm	Est.	Est.	Est.	Est.
δ	Difference $L_s^{95} - L_s^{50}$ (expressed here as ratio L_s^{95} / L_s^{50} to compare across life histories)	1.3	1.3	1.3	Est.	Est.	Est.	Est.
q	Catchability coefficient	1e-5	1e-5	1e-5	Est.	Est.	Fixed	Fixed
σ_F	Fishing mortality penalty standard deviation	0.2	0.2	0.2	Fixed	Fixed	Fixed	Fixed
σ_c	Standard deviation for observed catch	0.2	0.2	0.2	Fixed	Fixed	Fixed	Fixed
σ_I	Standard deviation for observed abundance index	0.2	0.2	0.2	Fixed	Fixed	Fixed	Fixed
θ	Dirichlet-multinomial parameter related to effective sample size	10	10	10	Est.	Est.	Est.	Est.
n	Sample size of length measurements	200	200	200	Fixed	Fixed	Fixed	Fixed
ϕ	Cumulative normal probability distribution							

885

886

887 Table 2. Population dynamic equations used in the operating model and LIME estimation model.

Equation		Description
1	$L_a = L_\infty(1 - \exp(-k(a - t_0)))$	Von Bertalanffy length-at-age a
2	$m_j = 1/(1 + \exp(L_m^{50} - j))$	Maturity-at-length
3	$m_a = \sum_{j=1}^J m_j \left(\frac{1}{L_a CV_L \sqrt{2\pi}} \right) \exp\left(-\frac{(j - L_a)^2}{2L_a CV_L^2}\right)$	Maturity-at-age
4	$S_j = 1/(1 + \exp\left(-\frac{\ln(19)(j - L_s^{50})}{L_s^{95} - L_s^{50}}\right))$	Logistic selectivity-at-length
5	$w_a = \alpha L_a^\beta$	Weight-at-age
6	$B_t = \sum_{a=1}^A N_{a,t} w_a$	Annual total population biomass
7	$SB_t = \sum_{a=0}^A N_{a,t} w_a m_a$	Annual spawning biomass
8	$N_{a,t} = \begin{cases} R_t, & a = 0 \\ N_{a-1,t} \exp(-M - F_t S_{a-1}), & 0 < a < A \text{ and } t=1 \\ \frac{N_{a-1,t} \exp(-M - F_t S_{a-1})}{1 - \exp(-M - F_t S_{a-1})}, & a = A \text{ and } t=1 \\ N_{a-1,t-1} \exp(-M - F_{t-1} S_{a-1}), & 0 < a < A \text{ and } t > 1 \\ (N_{a-1,t-1} + N_{a,t-1}) \exp(-M - F_{t-1} S_{a-1}), & a = A \text{ and } t > 1 \end{cases}$	Abundance at age over time
9	$SB_0 = \sum_{a=0}^A R_0 \exp(-aM) w_a m_a$	Unfished spawning biomass
10	$E_0 = \sum_{a=0}^A \exp(-aM) w_a m_a$	Expected lifetime egg production (unfished)
11	$E_f = \sum_{a=0}^A \exp(-a(M + FS_a)) w_a m_a$	Expected lifetime egg production (fished)
12	$SPR = \frac{E_f}{E_0}$	Spawning potential ratio

888

889

890 Table 3. Functions for generating data in the operating model. All except Equations 2 and 5 are
 891 also used in the LIME estimation model to predict values for the observed data.

Equation		Description
1	$\varepsilon_t \sim \text{Lognormal}(0, \sigma_R)$	Non-autocorrelated recruitment deviations
2	$\tau_t = \varepsilon_t \quad t = 1$ $\tau_t = \varepsilon_t \tau_{t-1} \rho \sqrt{1 - \rho^2} \quad t > 1$	Autocorrelated recruitment deviations
3	$p_{j,a} = \begin{cases} \phi\left(\frac{j - L_a}{L_a CV_L}\right), & j = 1 \\ \phi\left(\frac{j - L_a}{L_a CV_L}\right) - \phi\left(\frac{j - 1 - L_a}{L_a CV_L}\right), & 1 < j < J \\ 1 - \phi\left(\frac{j - 1 - L_a}{L_a CV_L}\right), & j = J \end{cases}$	Probability of being in a length bin given age
4	$\pi_j = p_{j,a} \frac{\sum_{a=0}^A N_{a,t} S_a}{N_t}$	Predicted probability of harvest by length bin
5	$\tilde{\pi}_j \sim \text{Multinomial}(n, \pi_j)$	Generated probability of harvest by length bin
6	$C_{a,t} = \frac{F_t S_a}{M + F_t S_a} N_{a,t} (1 - \exp(-M - F_t S_a))$	Annual catch at age
7	$C_t = \sum_{a=0}^A C_{a,t}$	Annual catch
8	$I_t = q B_t$	Abundance index

892

893

894 Table 4. Components of the joint likelihood function in the LIME model, as well as performance
 895 metrics across iterations of generated data.

Equation	Description
1	$\mu_{Rt} = \frac{4hR_0SB_{t-1}}{SB_0(1-h) + SB_{t-1}(5h-h)}$ Expected annual recruitment based on Beverton-Holt stock-recruit relationship
2	$R_t \sim \text{Lognormal}(\mu_{Rt}, \sigma_R^2)$ Annual recruitment R_t arising from a lognormal distribution
3	$F_t \sim \text{Normal}(F_{t-1}, \sigma_F^2)$ Penalty on annual fishing mortality F_t
4	$\sigma_R \sim \text{Lognormal}(0.7, 0.2^2)$ Recruitment standard deviation penalty
5	$\log L(\pi_j, \theta \tilde{\pi}_j, n) = \log \Gamma(n+1) - \sum (\log \Gamma(n\tilde{\pi}_j + 1)) + \log \Gamma(\theta n) - \log \Gamma(n + \theta n) + \sum (\log \Gamma(n\tilde{\pi}_j + \theta n\pi_j) - \log \Gamma(\theta n\pi_j))$ Dirichlet-multinomial log-likelihood
6	$n_{eff} = (1 + \theta n) / (1 + \theta)$ Effective sample size
7	$C_t^{obs} \sim \text{Lognormal}(\sum_{a=0}^A C_t, \sigma_C^2)$ Lognormal likelihood (catch)
8	$I_t^{obs} \sim \text{Lognormal}(qB_t, \sigma_I^2)$ Lognormal likelihood (abundance index)
9	$MRE = \text{median} \left(\frac{x_{estimated} - x_{true}}{x_{true}} \right)$ Median relative error to quantify bias
10	$MARE = \text{median} \left(\left \frac{x_{estimated} - x_{true}}{x_{true}} \right \right)$ Median absolute relative error to quantify precision

896

Table 5. LIME performance in estimation of SPR in the terminal year across life-history types, scenarios of data availability, and patterns in fishing mortality and recruitment variability, compared with LB-SPR. Bias (top) is measured as median relative error and precision (bottom) is measured as median absolute relative error from 100 iterations of generated data using the LIME age-structured operating model assuming instantaneous sampling with 200 length measurements annually. Dark green indicates bias/precision less than 5%, light green indicates bias/precision less than 10%, yellow indicates bias/precision less than 20%, light blue indicates bias/precision less than 30%, and dark blue indicates bias/precision greater than 30%.

Data availability scenario	Short			Medium			Longer		
	Equilibrium	Variability, two-way F	Variability, one-way F	Equilibrium	Variability, two-way F	Variability, one-way F	Equilibrium	Variability, two-way F	Variability, one-way F
<i>Bias (MRE)</i>									
Rich	-0.019	-0.017	0.023	-0.004	0.005	0.034	0.001	0.012	0.023
Index+LC10	-0.014	-0.039	0.023	-0.049	-0.001	0.094	0.125	0.173	0.076
Index+LC1	0.012	-0.006	0.033	-0.003	0.060	0.065	0.040	-0.061	0.029
Catch+LC10	-0.018	-0.045	0.018	-0.021	0.150	0.165	-0.002	0.124	0.280
Catch+LC1	0.029	0.796	0.972	-0.056	0.119	0.193	0.010	-0.146	0.236
LC10	-0.022	-0.056	0.016	-0.086	-0.137	0.072	0.104	-0.006	0.200
LC1	-0.045	-0.082	-0.025	-0.140	-0.355	-0.154	0.166	-0.468	-0.181
LBSPR10	-0.570	-0.676	-0.708	0.006	-0.405	0.196	0.062	-0.359	0.452
LBSPR1	-0.603	-0.698	-0.738	-0.030	-0.372	-0.181	0.059	-0.403	0.296
<i>Precision (MARE)</i>									
Rich	0.027	0.069	0.028	0.038	0.044	0.101	0.038	0.042	0.089
Index+LC10	0.039	0.051	0.038	0.185	0.176	0.295	0.244	0.449	0.352
Index+LC1	0.046	0.079	0.062	0.105	0.207	0.235	0.180	0.223	0.223
Catch+LC10	0.025	0.055	0.035	0.080	0.382	0.476	0.077	0.423	0.630
Catch+LC1	0.060	0.796	0.972	0.129	0.429	0.496	0.161	0.313	0.714
LC10	0.028	0.063	0.031	0.246	0.499	0.470	0.498	0.743	0.702
LC1	0.072	0.126	0.039	0.216	0.437	0.362	0.301	0.705	0.549
LBSPR10	0.570	0.676	0.708	0.066	0.412	0.404	0.078	0.367	0.486
LBSPR1	0.603	0.698	0.738	0.113	0.478	0.432	0.120	0.501	0.536

Table 6. LIME bias and precision in estimating SPR in the terminal year for a variety of sample size of independent length measurements. In all cases one year of length data are provided to the model. Bias (top) is measured as median relative error and precision (bottom) is measured as median absolute relative error from 100 iterations of generated data using the LIME age-structured operating model assuming instantaneous sampling. Dark green indicates bias/precision less than 5%, light green indicates bias/precision less than 10%, yellow indicates bias/precision less than 20%, light blue indicates bias/precision less than 30%, and dark blue indicates bias/precision greater than 30%.

Sample size	Short			Medium			Longer		
	Equilibrium	Variability, two-way F	Variability, one-way F	Equilibrium	Variability, two-way F	Variability, one-way F	Equilibrium	Variability, two-way F	Variability, one-way F
<i>Bias (MRE)</i>									
1000	-0.033	-0.070	-0.029	0.034	-0.369	-0.201	-0.061	-0.449	-0.201
500	-0.028	-0.117	-0.041	-0.001	-0.505	-0.157	-0.031	-0.621	-0.145
200	-0.045	-0.082	-0.025	-0.140	-0.355	-0.154	0.166	-0.468	-0.181
100	-0.027	-0.070	-0.028	-0.055	-0.370	-0.201	-0.127	-0.480	-0.206
50	-0.082	-0.124	-0.041	-0.182	-0.499	-0.163	-0.141	-0.621	-0.107
20	-0.171	-0.153	-0.043	-0.295	-0.431	-0.214	-0.465	-0.751	-0.451
<i>Precision (MARE)</i>									
1000	0.070	0.148	0.048	0.110	0.443	0.422	0.240	0.648	0.474
500	0.053	0.172	0.047	0.125	0.514	0.423	0.245	0.663	0.519
200	0.072	0.126	0.039	0.216	0.437	0.362	0.301	0.705	0.549
100	0.087	0.147	0.046	0.217	0.453	0.429	0.397	0.665	0.503
50	0.108	0.178	0.046	0.343	0.510	0.455	0.421	0.665	0.506
20	0.228	0.168	0.045	0.374	0.469	0.610	0.547	0.754	0.615

Table 7. Comparison of LIME and LB-SPR performance in estimation of SPR in the terminal year between equilibrium scenarios with length data collected annually and using an annual model, length data collected monthly but pooled annually for an annual model, and using the LB-SPR operating model. We also compared the annual model runs with monthly length data run on a monthly time step for the short-lived life-history type. Bias (top) is measured as median relative error and precision (bottom) is measured as median absolute relative error from 100 iterations of generated data. Dark green indicates bias/precision less than 5%, light green indicates bias/precision less than 10%, yellow indicates bias/precision less than 20%, light blue indicates bias/precision less than 30%, and dark blue indicates bias/precision greater than 30%.

Data availability scenario	Short				Medium			Longer		
	Annual data and model	Monthly data, annual model	Monthly data, monthly model	LB-SPR	Annual data and model	Monthly data, annual model	LB-SPR	Annual data and model	Monthly data, annual model	LB-SPR
<i>Bias (MRE)</i>										
LC10	-0.022	0.468	-0.040	0.068	-0.086	0.072	-0.133	0.104	0.184	-0.112
LC1	-0.045	0.484	0.034	0.195	-0.140	-0.104	-0.097	0.166	0.168	-0.113
LBSPR10	-0.570	0.136	NA	-0.005	0.006	0.090	0.002	0.062	0.117	0.010
LBSPR1	-0.603	0.111	NA	0.021	-0.030	0.056	-0.003	0.059	0.114	-0.030
<i>Precision (MARE)</i>										
LC10	0.028	0.468	0.102	0.140	0.246	0.272	0.260	0.498	0.442	0.429
LC1	0.072	0.484	0.130	0.233	0.216	0.209	0.133	0.301	0.311	0.167
LBSPR10	0.570	0.138	NA	0.050	0.066	0.094	0.048	0.078	0.117	0.050
LBSPR1	0.603	0.154	NA	0.152	0.113	0.139	0.138	0.120	0.157	0.149

Table 8. Summary table of key questions and conclusions.

Objective	Conclusion
1. <i>Is LIME unbiased:</i>	
a. Across life-history types?	Yes, if the assumption of instantaneous length sampling is met, and growth parameters are known with no error. Performance is best for life histories with longevity less than 20 years under this assumption.
b. With population variability?	Yes. It is recommended to use more than one year of length data for short- and medium-lived life-history types, and include catch or an abundance index if longer-lived.
c. With inclusion of catch and/or an abundance index?	Yes, with more than one year of length data for short-lived species. May over-estimate SPR for medium and longer-lived species if the population is not in equilibrium.
2. <i>Is LIME sensitive to:</i>	
a. Sample size of length measurements?	Yes, accuracy and precision generally improve with higher annual sample size. There is a performance plateau for 100 samples for short-lived fish and 500 samples for medium- or longer-lived fish.
b. Error in input parameters?	Relatively insensitive to error in the CV of the age-length curve. If asymptotic length, von Bertalanffy growth coefficient, or length at 50% maturity are input lower than the truth, or natural mortality input higher than the truth, SPR will be estimated higher than the truth (and vice versa).
3. <i>Comparison of LIME and LB-SPR under equilibrium conditions</i>	
a. Instantaneous annual sampling from age-structured model	LIME unbiased across life-history types; LB-SPR estimates lower SPR for short-lived life history.
b. Continuous sampling (monthly) from age-structured model	LIME should use monthly time steps for short-lived species, but can use annual time steps for medium and longer-lived species.
c. Length-structured operating model	LIME over-estimates SPR for short-lived and under-estimates SPR for medium- and longer-lived. LB-SPR unbiased across life-history types.

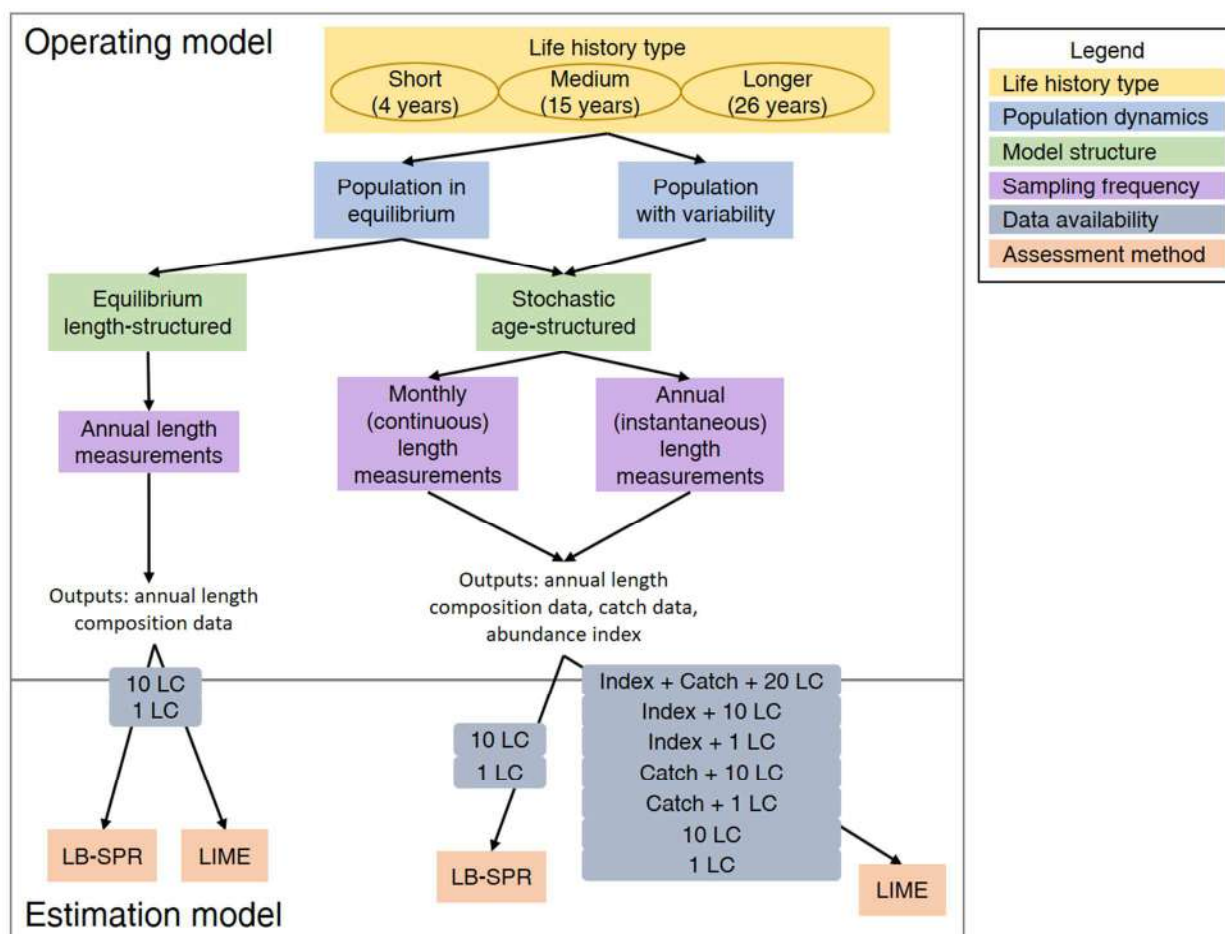


Figure 1. Diagram of the simulation study. “LC” stands for length data, numbers indicate number of years in the 20-year time-period modeled. In the case of the instantaneous, annual length measurement scenario, we compared multiple sample sizes of length measurements annually (1,000, 500, 200, 100, 50, and 20). For the other scenarios, we assumed the base case of 200 length measurements annually.

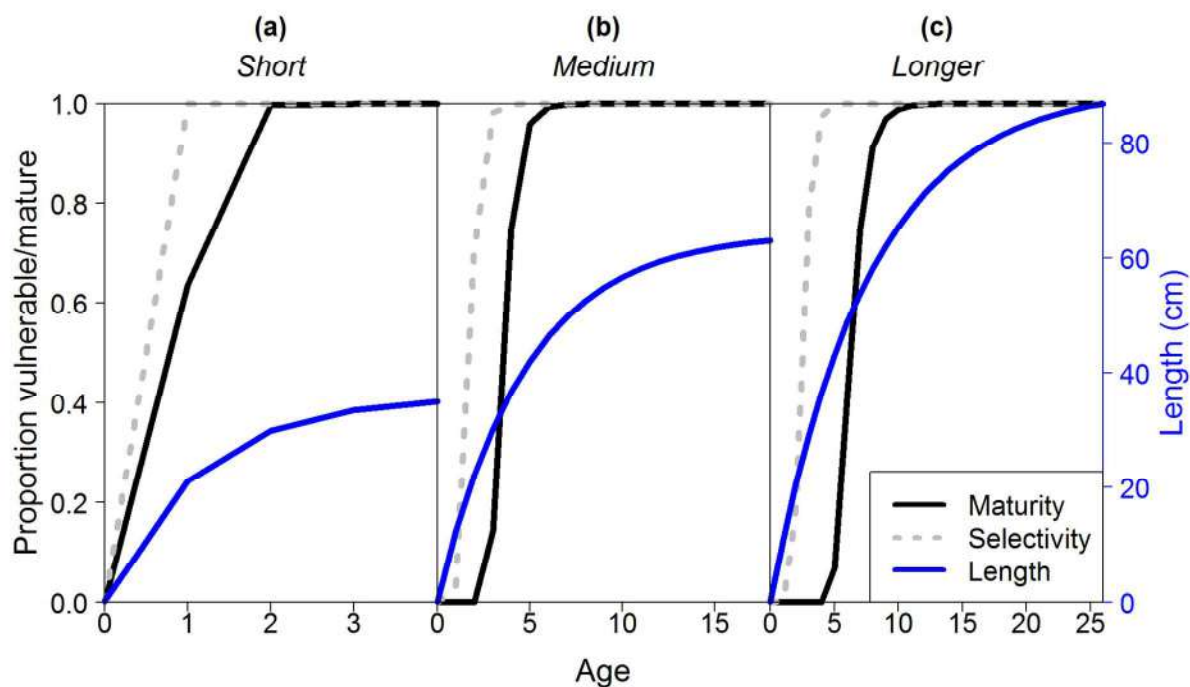


Figure 2. Selectivity, maturity, and length at age curves for the three life-history types tested in the simulation study, mimicking: (a) rabbitfish (*Siganus sutor*), a short-lived fish ($L_{\infty}=36.2$ cm, $k=0.87$, $M=1.49$, $a_m^{50}=1$ year, maximum age $A=4$ years, Hicks and McClanahan 2012), (b) spotted rose snapper (*Lutjanus guttatus*), a medium-lived fish (asymptotic length $L_{\infty}=64.6$ cm, von Bertalanffy $k=0.21$, natural mortality $M=0.43$, length at 50% maturity $a_m^{50}=4$ years, maximum age $A=15$ years; Bystrom 2015), and (c) grouper (*Epinephelus morio*), a longer-lived fish ($L_{\infty}=90$ cm, $k=0.13$, $M=0.18$, $a_m^{50}=7$ years, maximum age $A=26$ years; Heemstra and Randall 1993).

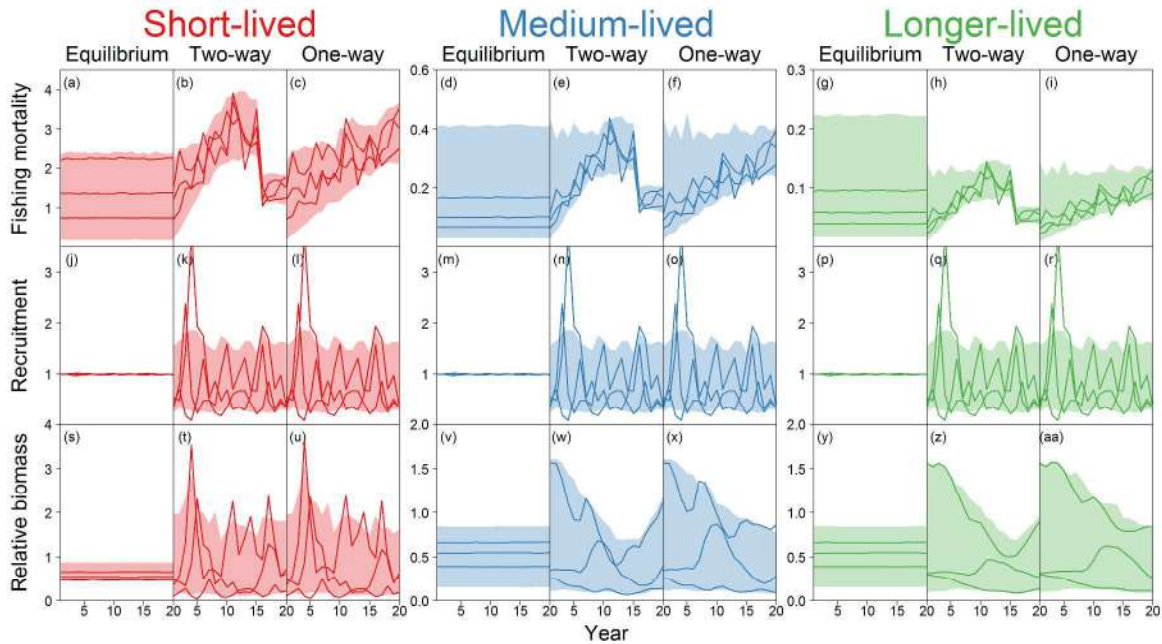


Figure 3. Scenarios of fishing mortality and recruitment under equilibrium and variable conditions for the three life-history types, with corresponding trajectories of relative spawning biomass. Shaded regions represent the area between the 5th and 95th percentile of generated data, and the lines show three randomly chosen iterations out of 100, as examples. Scenarios labeled “equilibrium” demonstrate that the initial depletion may start between 0.05 and 0.95, but fishing mortality remains constant to produce that level of depletion, and recruitment is constant at 1.0 over time. Scenarios labeled “two-way” demonstrate fishing mortality that increases to a fishing mortality rate $F_{20\%}$ that results in 20% SPR over the first seven years, stays at $F_{20\%}$ for seven years, and then drops to half of $F_{20\%}$ for the last six years. Scenarios labeled “one-way” demonstrate a change from equilibrium fishing mortality resulting in an initial depletion between 0.05 to 0.95 to the fishing mortality $F_{20\%}$ over the 20-year period.

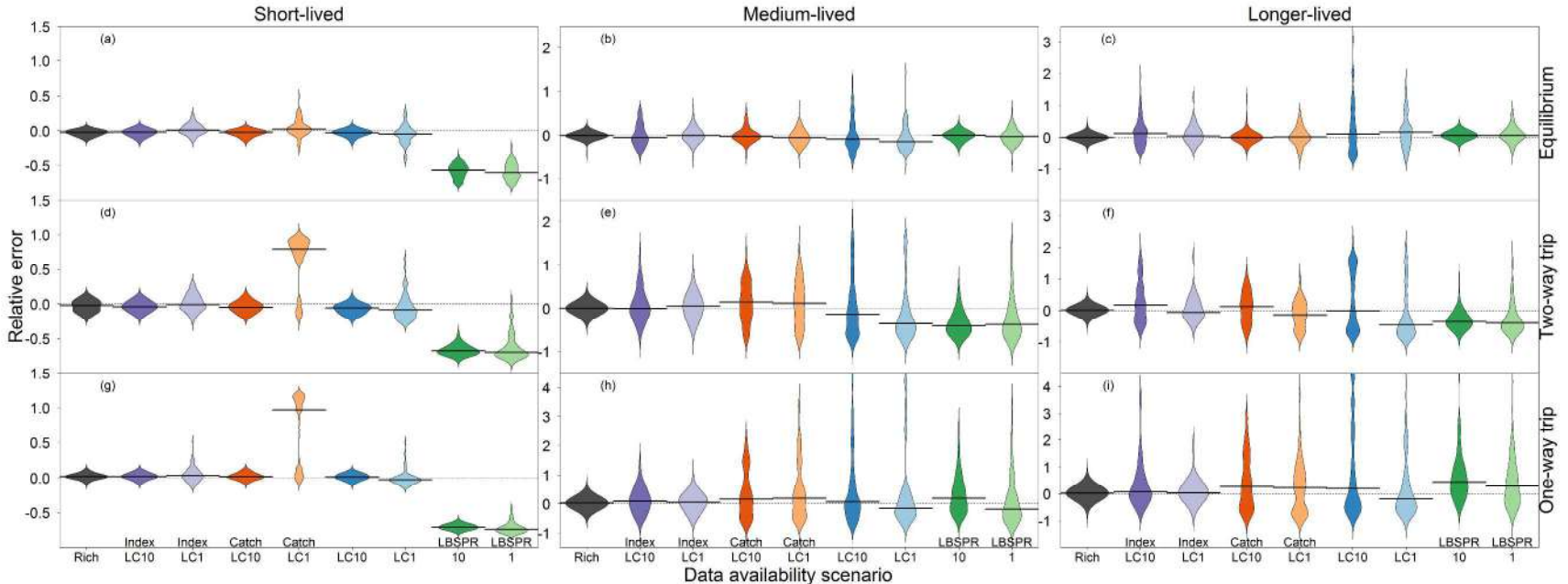


Figure 4. Distribution of relative error ($(estimated - true) / true$) for spawning potential ratio (SPR) in the current year for 100 iterations of simulated populations across the LIME and LB-SPR data availability scenarios for the three life-history types and scenarios of equilibrium and variable fishing mortality and recruitment with 200 samples of length measurements annually. The gray bean represents the “data-rich” scenario, which verifies LIME is unbiased and most precise when an unrealistically high amount of data are available. Darker colors represent data availability scenarios with 10 years of length data, and lighter colors represent the scenario with one year of length data available. Each life-history type has a different y-range.

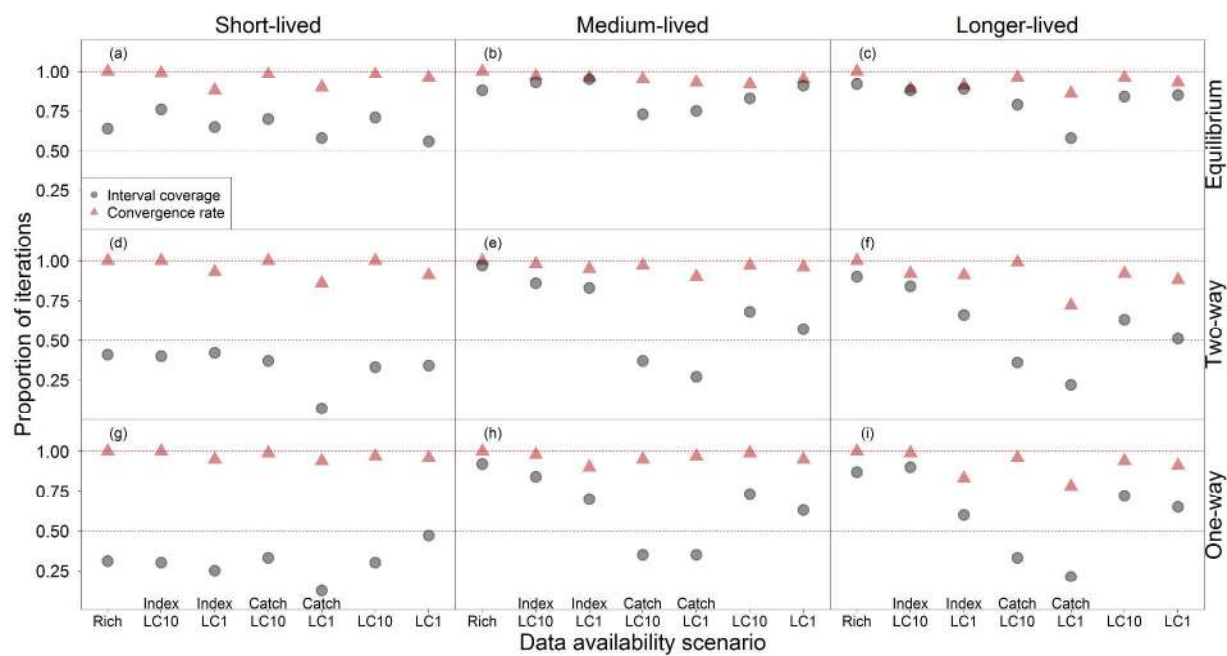


Figure 5. Proportion of iterations out of 100 where the true value of the spawning potential ratio (SPR) lies within the 50% confidence intervals for each life-history type across various patterns of fishing mortality and recruitment (gray circles) compared to the convergence rates for the same scenario (pink triangles).

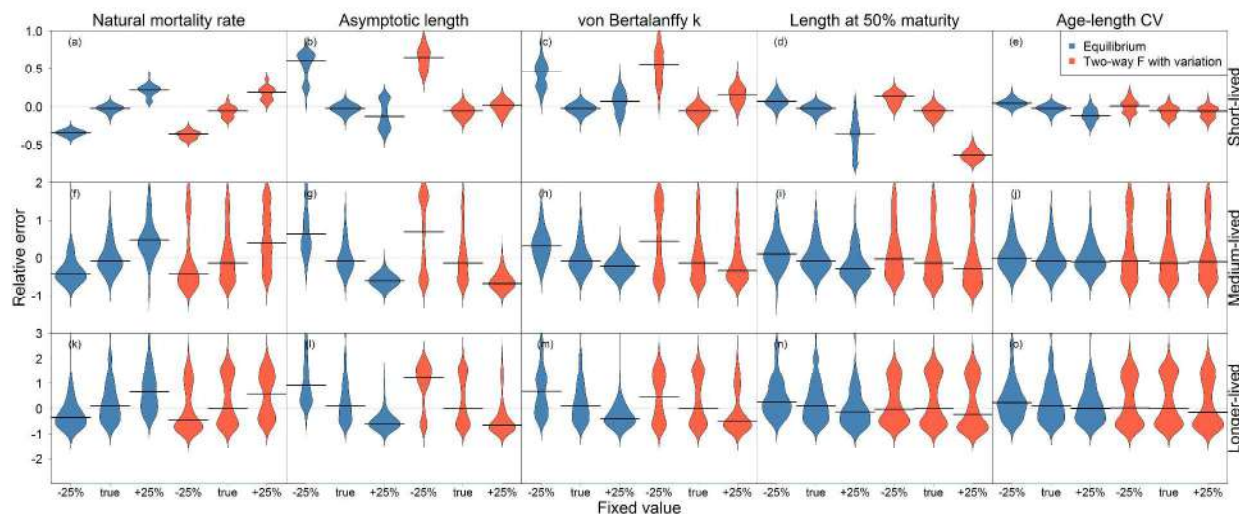


Figure 6. Comparison of the distribution of estimation error for spawning potential ratio (SPR) in the terminal year with the life history parameters fixed at $\pm 25\%$ of their true value from 100 iterations of generated data across the various life history scenarios for the equilibrium (blue) and the two-way fishing mortality scenario with variable and autocorrelated recruitment (red), with other biological parameters fixed at their true values, including 10 years of length data with 200 length measurements annually. Life history parameters include natural mortality, asymptotic length, von Bertalanffy growth coefficient, length at 50% maturity, and the coefficient of variation for the age-length curve.

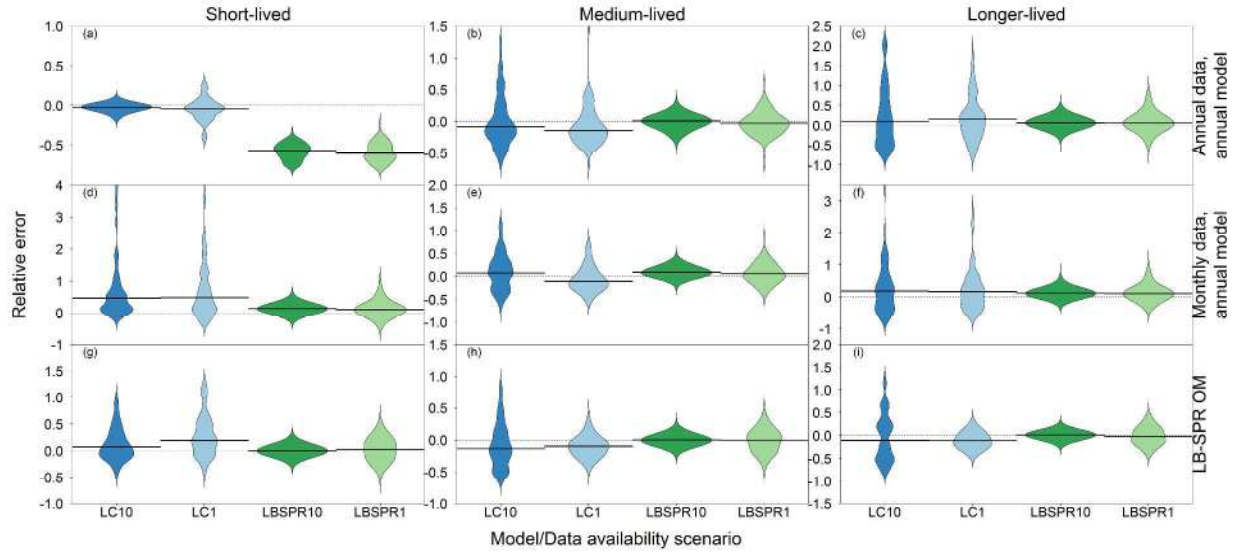


Figure 7. Comparison of the LIME and LB-SPR methods in the equilibrium state via the distribution of relative error ($estimated - true / true$) for spawning potential ratio (SPR) in the current year for 100 iterations of simulated populations from the LIME age-structured operating model using annual length composition collected instantaneously in the year (a-c), monthly length data collection pooled into annual length compositions (d-f), and the LB-SPR operating model based on relative ages. The methods are compared across life-history types. Darker colors represent data availability scenarios with 10 years of length data, and lighter colors represent the scenario with one year of length data available. Each scenario has a different y-range.

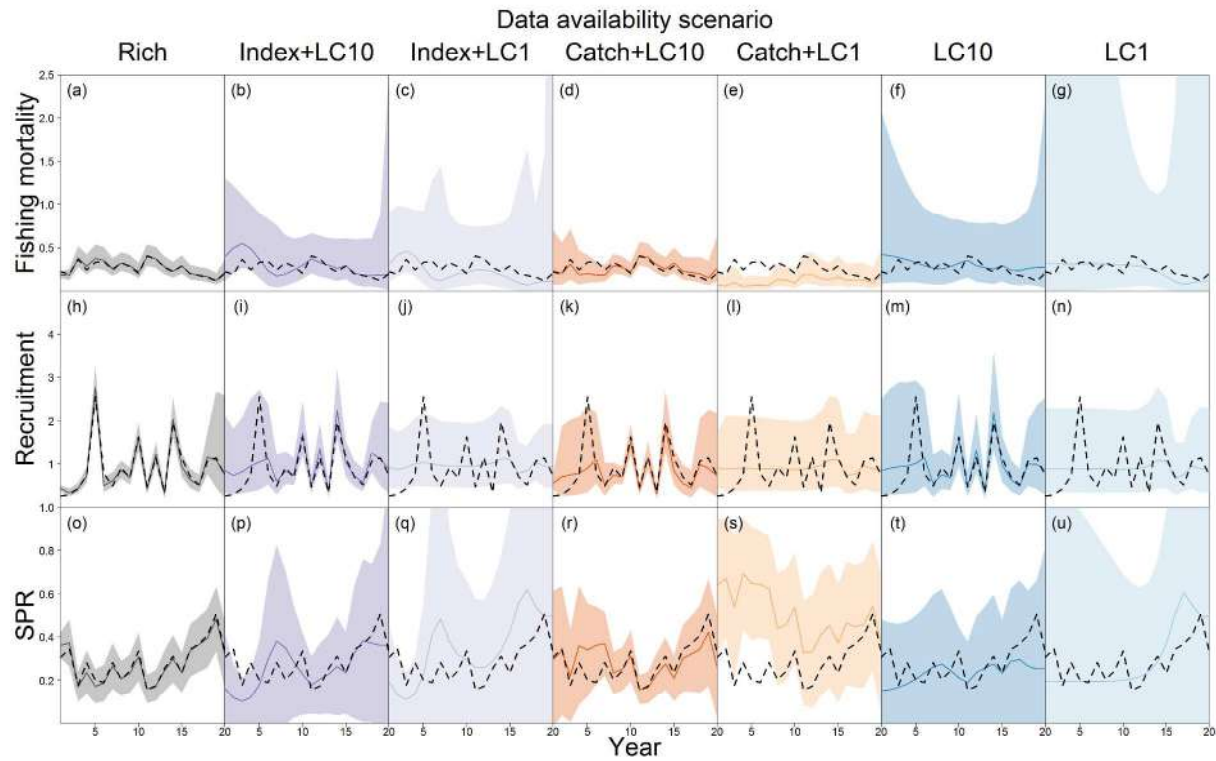


Figure 8. Example of model estimates (color lines) with 95% confidence intervals (shaded areas) and the true simulated population dynamics (black dotted line) for each of the data availability scenarios for one iteration of a simulated population with medium-lived life history, 200 length measurements annually in the length data, for the base variation scenario.

Appendix A.

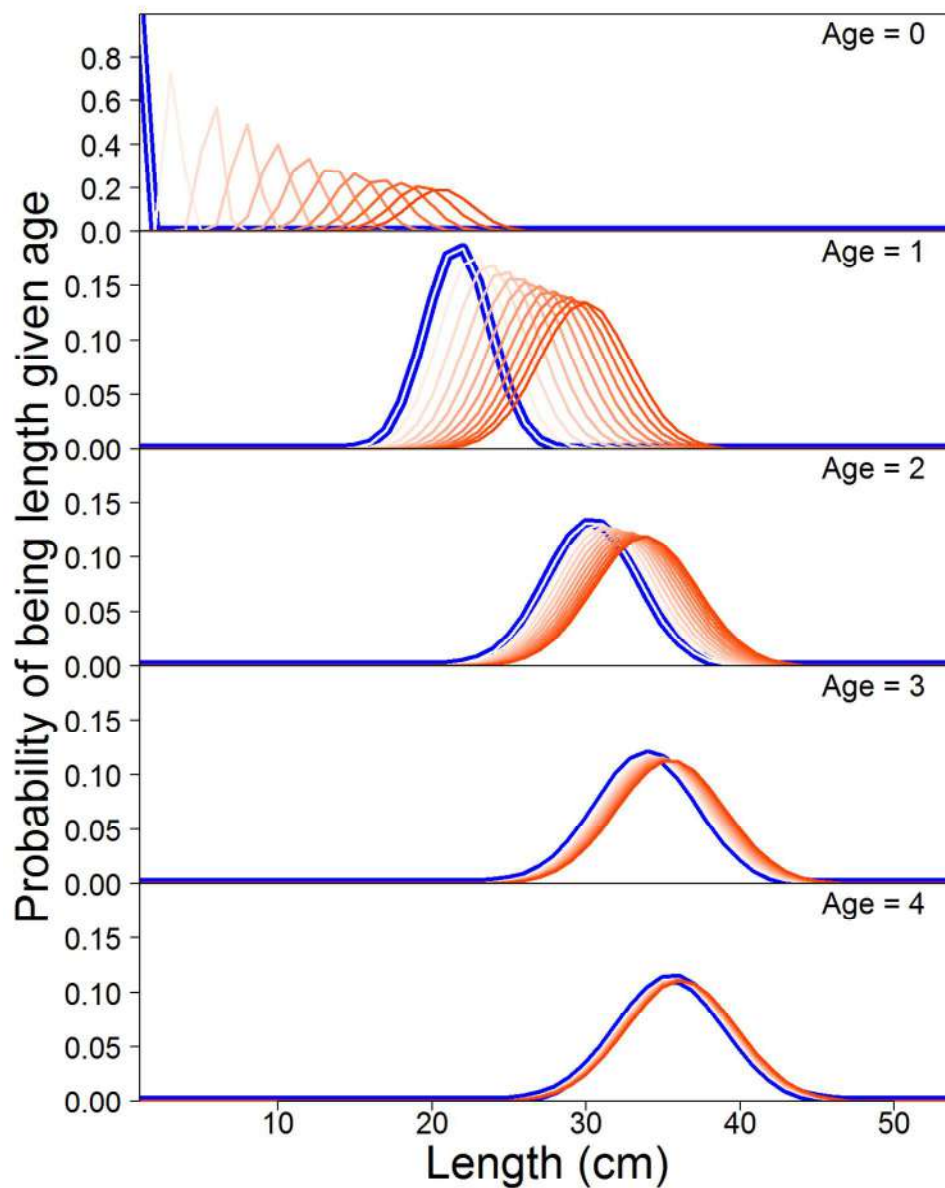


Figure A1. Size at age for the yearly (blue) vs. monthly (white-orange) time steps for the short-lived life-history type. White lines (monthly) overlapping blue lines (yearly) demonstrate that size-at-age is identical at the start to each year, and orange lines increasing in density indicate the size-at-age distributions during the following months within each year.