

¹ An investigation of factors affecting inferences from and
² reliability of state-space age-structured assessment
³ models

⁴ Timothy J. Miller^{1,2} ~~Greg~~^{Gregory L.} Britten³

⁵ Elizabeth N. Brooks² Gavin Fay⁴ ~~Alex~~^{Alexander C.} Hansell²

⁶ Christopher M. Legault² Chengxue Li² Brandon Muffley⁵

⁷ Brian C. Stock⁶ John Wiedenmann⁷

⁸ ~~02 May 19 December~~, 2025

⁹ ¹corresponding author: timothy.j.miller@noaa.gov

¹⁰ ²Northeast Fisheries Science Center, Woods Hole Laboratory, 166 Water Street, Woods
¹¹ Hole, MA 02543 USA

¹² ³Biology Department, Woods Hole Oceanographic Institution, 266 Woods Hole Rd. Woods
¹³ Hole, MA, USA

¹⁴ ⁴Department of Fisheries Oceanography, School for Marine Science and Technology,
¹⁵ University of Massachusetts Dartmouth, 836 S Rodney French Boulevard, New Bedford,
¹⁶ MA 02740, USA

¹⁷ ⁵Mid-Atlantic Fishery Management Council, 800 North State Street, Suite 201, Dover, DE

¹⁸ 19901 USA

¹⁹ ⁶Institute of Marine Research, Nye Flødevigveien 20, 4817 His, Norway

²⁰ ⁷Department of Ecology, Evolution, and Natural Resources. Rutgers University

²¹

22 **keywords:** stock assessment, state-space, model selection, bias, convergence, retrospective
23 patterns

24 **Abstract**

25 State-space models ~~are increasingly used for stock assessment, and evaluations of their~~
26 ~~statistical reliability and best practices for selecting among process error configurations are~~
27 ~~have been promoted as the next-generation of fisheries stock assessment and evaluation of~~
28 ~~their reliability is~~ needed. We simulated 72 operating models that varied fishing pressure ~~and~~
29 ~~observation error across process errors in recruitment, survival, selectivity, catchability, and~~
30 ~~/or natural mortality. We fit estimating models with different assumptions on the process~~
31 ~~error source and whether, magnitude of observation error, and sources of process error.~~
32 ~~For each operating model, we fit a range of estimating models with correct and incorrect~~
33 ~~configurations. We measured reliability of estimating models by convergence rate, accuracy~~
34 ~~of AIC-based model selection, estimation bias, and magnitude of retrospective patterns. All~~
35 ~~reliability measures were generally better with lower observation error, contrast in fishing~~
36 ~~pressure over time, and when median natural mortality or a stock-recruit relationship were~~
37 ~~estimated. Estimating models without a stock-recruit relationship that assumed the correct~~
38 ~~rate is known. The magnitude of the log-likelihood gradients was not a reliable indicator~~
39 ~~of convergence. AIC can generally distinguish process error source and median natural~~
40 ~~mortality had high convergence rates and low bias. Bias was also low under many incorrect~~
41 ~~process error assumptions when there was contrast in fishing pressure and low observation~~
42 ~~error. Marginal AIC most accurately distinguished process errors on recruitment, survival,~~
43 ~~and selectivity, and other process error sources when variability was greater with lower~~
44 ~~observation error and higher true process error variability. Distinguishing the stock recruit~~
45 ~~relationship with AIC required large contrast in spawning biomass and low recruitment~~
46 ~~variation, but bias in stock-recruit parameter estimation was prevalent. Retrospective pat-~~

47 terns were generally small but were sizable for recruitment when observation error was high.
48 These results help establish the statistical reliability of state space assessment models and
49 pave the way for the next generation of fisheries stock assessment ~~not large for mis-specified~~
50 models. These findings improve our understanding of when results from state space models
51 will be reliable.

52 **Introduction**

53 Application of state-space models in fisheries stock assessment and management has ex-
54 panded dramatically within the International Council for the Exploration of the Sea (ICES),
55 Canada, and the Northeast US (Nielsen and Berg 2014; Cadigan 2016; Pedersen and Berg
56 2017; Stock and Miller 2021). State-space models treat latent population characteristics as
57 statistical time series with periodic observations that also may have error due to sampling or
58 other sources of measurement errormeasurement properties. Traditional assessment models
59 may use state-space approaches to account for temporal variability in population charac-
60 teristics (Legault and Restrepo 1999; Methot and Wetzel 2013), but these models treat the
61 annual parameters as penalized fixed effectseffect parameters where the variance parameters
62 controlling the penalties are assumed known (Thorson and Minto 2015). Modern state-space
63 models can estimate the annually varying parameters as random effects with variance param-
64 eters estimated using maximum marginal likelihood or corresponding Bayesian approaches.
65 These latter-random-effects approaches are considered best practice and ~~a~~are recommended
66 for the next generation of stock assessment models (Hoyle et al. 2022; Punt 2023).

67 State-space stock assessment models, with nonlinear functions of latent parameters and
68 multiple types of observations with varying distributional assumptions, are one of the most
69 complex examples of this analytical approach. Statistical aspects of state-space models and
70 their application within fisheries have been studied extensively, but previous work has focused
71 primarily on linear and Gaussian state-space models (Aeberhard et al. 2018; Auger-Méthé
72 et al. 2021). Therefore, current understanding of the reliability of state-space models does
73 not extend to usage for stock assessment.

74 As state-space models provide greater flexibility by allowing multiple processes to vary as
75 random effects (Nielsen and Berg 2014; Aeberhard et al. 2018; Stock et al. 2021), one of the
76 most immediate questions regards the implications of mis-specification among alternative
77 sources of process error. Incorrect treatment of population attributes as temporally varying

78 (Trijoulet et al. 2020; Liljestrand et al. 2024) could lead to misidentification of stock
79 status and biased population estimates, ultimately impacting fisheries management decisions
80 (Legault and Palmer 2016; Szuwalski et al. 2018; Cronin-Fine and Punt 2021). Furthermore,
81 biological, fishery, and observational processes are often confounded in catch-at-age data,
82 which may adversely affect the ability to distinguish between true process variability and
83 observational error (Li et al. In review; Punt et al. 2014; Stewart and Monnahan 2017;
84 Cronin-Fine and Punt 2021; Fisch et al. 2023; Li et al. 2025a).

85 Li et al. (2024) conducted a full-factorial simulation-estimation study to assess model reli-
86 ability when confounding random-effects processes (numbers-at-age, fishery selectivity, and
87 natural mortality) were included. Their results suggest that while state-space models can
88 generally identify sources of process error, overly complex models, even when misspecified
89 (i.e., incorporating process error that did not exist in reality), often performed similarly to
90 correctly specified models, with little to no bias in key management quantities. Similarly,
91 Liljestrand et al. (2024) found little downside in assuming process error in recruitment or
92 selectivity, even when it was absent.

93 Despite mounting efforts, several limitations increasing research on state space assessment
94 models, several uncertainties in state space assessment modeling remain. First, confound-
95 ing processes that can be treated as random effects in the model were not have not been
96 thoroughly examined or tested within a simulation-estimation framework. Second, previous
97 studies relied on operating models conditioned on specific fisheries, limiting their generaliz-
98 ability (Li-Liljestrand et al. In review; Liljestrand 2024; Li et al. 20242025a). In particular,
99 the effects of observation error and underlying fishing history have not been fully isolated in
100 simulation study designs, making it challenging to disentangle the interplay between process
101 and observation error magnitudes, as demonstrated in Fisch et al. (2023). Third, explicitly
102 modeling stock-recruit relationships (SRRs) as mechanistic drivers of population dynamics
103 is promising (Fleischman et al. 2013; Pontavice et al. 2022), but reliability of inferences
104 within integrated state-space age-structured models has not been evaluated. Evidence from

other studies suggests that when both process and observation errors are unknown, estimating density dependence parameters becomes highly uncertain (Knape 2008; Polansky et al. 2009). In particular, Knape (2008) demonstrated that stronger density dependence becomes increasingly difficult to estimate in the presence of observation error. Therefore, it is crucial to assess whether ~~density-dependence~~ density-dependent mechanisms can be estimated with sufficient precision for use in fisheries management (Auger-Méthé et al. 2016). Finally, although the importance of autocorrelation in process errors is recognized, investigations of the ability to distinguish state-space assessment models with and without autocorrelation and whether such misspecification is detrimental to estimation of important population metrics are lacking (Johnson et al. 2016; Xu et al. 2019).

In the present study, we conduct a simulation study with operating models (OMs) varying by degree of observation error, source and variability of process error, and fishing history. The simulations from these OMs are fitted with estimation models (EMs) that make alternative assumptions for sources of process error, whether ~~a-an~~ SRR was estimated, and whether natural mortality is estimated. Given the confounding nature of process errors, developing diagnostic tools to detect model misspecification is of great scientific interest and could aid the next generation of stock assessments (Auger-Méthé et al. 2021). We evaluate whether ~~convergence and~~ QM and EM attributes affect rates of convergence and the ability of Akaike Information Criterion (AIC) ~~ean to~~ correctly determine the source of process error ~~and or~~ the existence of ~~a-an~~ SRR. We also evaluate ~~when retrospective patterns oocur and the degree of bias in the outputs of the assessment model that are effects of OM and EM attributes on~~ magnitude of retrospective patterns and bias in estimation of parameters and other model outputs important for management.

128 Methods

129 We used the Woods Hole Assessment Model (WHAM) to configure OMs and EMs in our
130 simulation study ([Miller and Stock 2020](#); [Stock and Miller 2021](#); [Miller et al. 2025](#)).

131 WHAM is an R package freely available via a [github-Github](#) repository and is built on the
132 Template Model Builder package (Kristensen et al. 2016). For this study we used version
133 1.0.6.9000, commit 77bbd94. WHAM has also been used to configure OMs and EMs for
134 closed loop simulations evaluating index-based assessment methods (Legault et al. 2023)
135 and is currently used or accepted for use in management of numerous [NEUS-Northeast](#)
136 [United States \(NEUS\)](#) fish stocks (e.g., NEFSC 2022a, 2022b; NEFSC 2024).

137 We completed a simulation study with a number of OMs that can be categorized based on
138 where process error random effects were assumed: ~~recruitment (R, assumed present in all~~
139 ~~models),~~ [R OMs assume process error for recruitment only. Other OM categories assume](#)
140 [recruitment process errors along with process errors for](#) apparent survival (~~denoted R+S~~),
141 natural mortality (R+M), fleet selectivity (R+Sel), or index catchability (R+q). We refer
142 to the ~~(R+S)~~OMs as modeling apparent survival because on ~~logscale~~ [log-scale](#) the random
143 effects $(\epsilon_{a,y})$ are additive to the total mortality [\(F+M\)fishing and natural mortality](#) between
144 numbers at age, thus they modify the survival term. ~~However, as Stock and Miller (2021)~~
145 ~~note, these random effects can be due to events other than mortality, such as immigration,~~
146 ~~emigration, missreported catch, and other sources of misspecification.~~ For each OM, assump-
147 tions about the magnitude of the variance of process errors and observations are required
148 and the values we used were based on a review of the range of estimates from [Northeast](#)
149 [Unite States \(NEUS\)](#) [NEUS](#) assessments using WHAM.

150 In total, we configured 72 OMs with alternative assumptions about the source and magnitude
151 of process errors, magnitude of observation error in indices and age composition data, and
152 contrast in fishing pressure over time. ~~We fitted 20 EMs to observations generated from~~
153 ~~each of~~ For each OM, we simulated 100 simulations where process errors were also simulated

154 ~~—Each EM~~ time series of abundance at age with process errors, and for each realized time
155 series, we simulated observation data sets. For each data set, we fitted a number of EMs that
156 differed in assumptions about the source of process errors, whether natural mortality (or the
157 median for models with process error in natural mortality) was estimated, and whether a
158 Beverton-Holt SRR was estimated within the EM. Details of each of the OMs and EMs are
159 described below.

160 We did not use the log-normal bias-correction feature for process errors or observations
161 described by ~~(Stock and Miller (2021))~~ for OMs and EMs to simplify interpretation of the
162 study results (Li et al. ~~In review~~^{2025b}). All code we used to perform the simulation study
163 and summarize results can be found at <https://github.com/timjmiller/SSRTWG/tree/main/>
164 Project_0/code.

165 Operating models

166 Population

167 We intended the population demographics and observation types to represent a general
168 NEUS groundfish stock. The population consists of 10 age classes, ages 1 to 10+, with the
169 last being a plus group that accumulates ages 10 and older. ~~We assume spawning occurs~~
170 ~~annually 1/4 of the way through the year.~~ The maturity at age was a logistic curve with a_{50}
171 = 2.89 and slope = 0.88 (Figure S1, top left).

172 Weight at age (W_a) was generated with a von Bertalanffy growth function ~~defining length~~
173 ~~at age:~~

$$L_a = L_\infty \left(1 - e^{-k(a-t_0)}\right),$$

174 where $t_0 = 0$, $L_\infty = 85$, and $k = 0.3$, and a ~~L-W length-weight~~ relationship such that

$$W_a = \theta_1 L_a^{\theta_2},$$

175 where $\theta_1 = e^{-12.1}$ and $\theta_2 = 3.2$ (Figure S1, top right).

176 We assumed a Beverton-Holt SRR with constant pre-recruit mortality parameters for all
177 OMs. We assume spawning occurs annually 0.25 of each year and recruitment at age 1
178 ($N_{1,y}$). All biological inputs to calculations of spawning ~~biomass stock biomass (SSB)~~ per
179 recruit (i.e., weight, maturity, and natural mortality at age) are constant in the ~~apparent~~
180 ~~survival (R+S) selectivity (R+Sel), and survey catchability (and R+q)~~ process error
181 OMs. Therefore, steepness and unfished recruitment are also constant over the time period
182 for those OMs (Miller and Brooks 2021). We assumed a value of 0.2 for the natural mortality
183 rate in OMs without process errors on natural mortality. We specified unfished recruitment
184 equal to e^{10} and $F_{\text{MSY}} = F_{40\%} = 0.348$, which equates to a steepness of 0.69 and $a = 0.60$
185 and $b = 2.4 \times 10^{-5}$ for the Beverton-Holt parameterization

$$N_{1,y} = \frac{a\text{SSB}_{y-1}}{1 + b\text{SSB}_{y-1}}$$

186 (Figure S1, bottom right). ~~We assumed a value of 0.2 for the natural mortality rate in~~
187 ~~OMs without process errors on natural mortality and for the median rate for OMs with~~
188 ~~process errors on~~ For OMs with time-varying random effects for natural mortality, steepness
189 is not constant. However, we used the same a and b parameters as other OMs, which equates
190 to a steepness and R_0 at the median of the time series process for natural mortality. Similarly,
191 for OMs with time-varying random effects for fishery selectivity, F_{MSY} also varies temporally,
192 so equilibrium conditions for these OMs are defined for mean selectivity parameters.

193 We used two fishing scenarios for OMs. In the first scenario, the stock experiences overfishing
194 at $2.5F_{\text{MSY}}$ for the first 20 years followed by fishing at F_{MSY} for the last 20 years (denoted
195 $2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$). In the second scenario, the stock is fished at F_{MSY} for the entire time
196 period (40 years). The magnitude of the overfishing assumptions is based on average esti-
197 mates of overfishing for NEUS groundfish stocks from Wiedenmann et al. (2019) and similar
198 to the approach in Legault et al. (2023).

199 The second scenario represents the ideal situation where the stock is fished at an optimal
200 level, but provides less contrast in stock sizes over time. We specified initial population
201 abundance at age at the equilibrium distribution that corresponds to fishing at either
202 $F = 2.5 \times F_{MSY}$ $F = 2.5F_{MSY}$ or $F = F_{MSY}$. This implies that, for a deterministic model,
203 the abundance at age would not change from ~~the first year to the next.~~

204 ~~For OMs with time-varying random effects for M, steepness is not constant. However, we~~
205 ~~used the same a and b parameters as other OMs, which equates to a steepness and R0 at~~
206 ~~the median year to year at the beginning of the time series process for M. For OMs with~~
207 ~~time-varying random effects for fishery selectivity, is also not constant, but since we use the~~
208 ~~same F history as other OMs, this corresponds to at the mean selectivity parameters.~~

209 Fleets

210 We assumed a single fleet operating year round for catch observations with logistic selectivity
211 ~~for the fleet~~ ($a_{50} = 5$ and slope = 1; Figure S1, bottom left). This selectivity was used to
212 define F_{MSY} for the Beverton-Holt SRR parameters above. We assumed a logistic-normal
213 distribution with no correlation on the multivariate normal scale for the corresponding annual
214 age-composition observations~~for the fleet~~.

215 Indices

216 Two time series of fishery-independent surveys measured in numbers are generated for the
217 entire 40 year period with one occurring in the spring (0.25 of each year) and one in the
218 fall (0.75 of each year), representing current bottom trawl surveys conducted in the NEUS.
219 Catchability of both surveys are assumed to be 0.1. Like the fishing fleet, we assumed logistic
220 selectivity for both indices ($a_{50} = 5$ and slope = 1) and a logistic-normal distribution with
221 no correlation on the multivariate normal scale for the annual age-composition observations.

222 **Observation Uncertainty**

223 The standard deviation for log-aggregate catch was 0.1 for all OMs, a common assumption
224 for commercial removals in NEUS stock assessments. Two levels of observation error variance
225 (high and low) were specified for indices and all age composition observations (both indices
226 and catch). The low uncertainty specification assumed a standard deviation of 0.1 for both
227 series of log-aggregate index observations, and the standard deviation of the logistic-normal
228 for age composition observations was 0.3. In the high uncertainty specification, the standard
229 deviation for log-aggregate indices was 0.4 and that for the age composition observations
230 was 1.5. The low standard deviation for index observations is typical for fish stocks that
231 are consistently sampled across survey stations whereas the high value is typical for more
232 sporadically sampled stocks. The standard deviations for the age composition observations
233 were determined from the range of values estimated from WHAM fits to NEUS stocks that
234 assumed the logistic-normal model. For all EMs, the standard deviation for log-aggregate
235 observations was assumed known whereas that for the logistic-normal age composition ob-
236 servations was estimated.

237 **Operating models with random effects on numbers at age**

238 For operating models with random effects on recruitment ~~and(or) only and also on~~ ap-
239 parent survival (R, R+S), we assumed marginal standard deviations for recruitment of
240 $\sigma_R \in \{0.5, 1.5\}$ ~~and~~. The marginal standard deviations for apparent survival random effects
241 at older age classes ~~of were~~ $\sigma_{2+} \in \{0, 0.25, 0.5\}$. The full factorial combination of these
242 process error assumptions (~~2x3-2 x 3~~ levels) and scenarios for fishing history (2 levels) and
243 observation error (2 levels) scenarios described above results in 24 different R ($\sigma_{2+} = 0$) and
244 R+S operating models (Table S1).

245 **Operating models with random effects on natural mortality**

246 All R+M OMs treat natural mortality as constant across age, but with annually varying
247 random effects. WHAM treats natural mortality as a log-transformed parameter

$$\log M_{y,a} = \mu_M + \epsilon_{M,y}$$

248 that is a linear combination of a mean log-natural mortality parameter that is constant
249 across ages ($\mu_M = \log(0.2)$) and any annual random effects are marginally distributed as
250 $\epsilon_{M,y} \sim N(0, \sigma_M^2)$. The marginal standard deviations we assumed for log natural mortality
251 random effects were $\sigma_M \in \{0.1, 0.5\}$ and the random effects were either uncorrelated or first-
252 order autoregressive (AR1, $\rho_M \in \{0, 0.9\}$). Uncorrelated random effects were also included
253 on recruitment with $\sigma_R = 0.5$ (hence, we denote these OMs as R+M). The full factorial
254 combination of these process error assumptions and fishing history (2 levels) and observation
255 error (2 levels) scenarios described above results in 16 different R+M OMs (Table S2).

256 **Operating models with random effects on fleet selectivity**

257 WHAM treats each selectivity parameter s as a logit-transformed parameter

$$\log \left(\frac{p_{s,y} - l_s}{u_s - p_{s,y}} \right) = \mu_s + \epsilon_{s,y}$$

258 that is a linear combination of a mean μ_s and any annual random effects marginally dis-
259 tributed as $\epsilon_{s,y} \sim N(0, \sigma_s^2)$, where the lower and upper bounds of the parameter (l_s and
260 u_s) can be specified by the user. All selectivity parameters (a_{50} and slope parameters) were
261 bounded by ~~0-and-10~~ $s_L = 0$ and $s_u = 10$ for all OMs and EMs. The marginal standard
262 deviations we assumed for logit scale random effects were $\sigma_s \in \{0.1, 0.5\}$ and AR1 autocor-
263 relation parameters of $\rho_s \in \{0, 0.9\}$. Like R+M OMs, the full factorial combination of these
264 process error assumptions (2x2 levels) and scenarios described above for fishing history (2

265 levels) and observation error (2 levels) results in 16 different R+Sel OMs (Table S3).

266 Operating models with random effects on index catchability

267 Like selectivity parameters, WHAM treats catchability for an index i as a logit-transformed
268 parameter

$$\log \left(\frac{q_{i,y} - l_i}{u_i - q_{i,y}} \right) = \mu_i + \epsilon_{i,y}$$

269 that is a linear combination of a mean μ_i and any annual random effects marginally dis-
270 tributed as $\epsilon_{i,y} \sim N(0, \sigma_i^2)$ where the lower and upper bounds of the catchability (l_i and u_i)
271 can be specified by the user. We assumed bounds of 0 and 1000 for all OMs and EMs. For
272 all OMs and EMs with process errors on catchability, the temporal variation only applies
273 to the first index, which could be interpreted as capturing some unmeasured seasonal pro-
274 cess that affects availability to the survey. The marginal standard deviations we assumed
275 for logit scale random effects were $\sigma_i \in \{0.1, 0.5\}$ and AR1 autocorrelation parameters of
276 $\rho_i \in \{0, 0.9\}$. Like R+M and R+Sel OMs, the full factorial combination of these process
277 error assumptions and fishing history (2 levels) and observation error (2 levels) scenarios
278 described above results in 16 different R+q OMs (Table S4).

279 Estimation models

280 For each of the data sets simulated from an OM, 20 EMs were fit. A total of 32 different
281 EMs were fit across OMs where the subset of 20 depended on the source of process error
282 in the OM (Table S5). The EMs have different assumptions about the source of process
283 error (R+S, R+M, R+Sel, R+q) and whether or not 1) there is temporal autocorrelation,
284 2) a Beverton-Holt SRR is estimated, and 3) the natural mortality rate (μ_M , the constant
285 or mean on log scale for R+M EMs) is estimated. For simplicity we refer to the derived
286 estimate e^{μ_M} as the median natural mortality rate regardless of whether natural mortality
287 random effects are estimated in the EM.

288 Subsets of 20 EMs in Table S5 were fit to ~~simulate~~ simulated data sets from each of the
289 OM process error ~~eategories~~sources. For R and R+S OMs, fitted EMs had matching process
290 error assumptions as well as R+Sel, R+M, and R+q assumptions without autocorrelation.
291 Similarly, For other OM process error ~~eategories~~sources, we fit EMs with ~~matching~~ correct
292 process error assumptions~~as well as other process error types~~, the correct process error source
293 but incorrect correlation assumption, and the incorrect process error source without autocor-
294 relation. As such, EMs were configured correctly for the OM, or they had mis-specification
295 in assumptions of process error autocorrelation, the source of process error, and(or) the SRR
296 (Beverton-Holt or none).

297 The maturity at age, weight at age for catch and SSB, and observation error ~~variance of~~
298 standard deviations for aggregate catch and indices were all assumed known at the true
299 values. However, the variance parameters for the logistic-normal distributions for age com-
300 position observations were estimated in the EMs. As such, EMs would either be configured
301 ~~completely correctly for the OM, or there could be mis-specification in assumptions of process~~
302 ~~error autoeorellation, the type of process error, or the SRR (Beverton-Holt or none)~~.

303 Measures of reliability

304 Convergence

305 The first measure of reliability we investigated was frequency of convergence when fitting
306 each EM to the simulated data sets. There are various ways to assess convergence of the fit
307 (e.g., Carvalho et al. 2021; Kapur et al. 2025), but given the importance of estimates of un-
308 certainty when using assessment models in management, we estimated ~~probability~~probability
309 of convergence as measured by occurrence of a positive-definite ~~hessian~~Hessian matrix at
310 the optimized negative log-likelihood that could be inverted (i.e., providing Hessian-based
311 standard error estimates). We also provide results in the Supplementary Materials for ~~the~~
312 ~~maximum convergence defined by the maximum absolute gradient < 1^-6 and the maximum~~

313 of the absolute ~~values among all gradients~~ gradient values for all fits of a given EM to all
314 simulated data sets from a given OM that produced ~~hessian-based~~ Hessian-based standard
315 errors for all estimated fixed effects. This provides an indication of how poor the calculated
316 gradients can be, but still presumably converged adequately enough for parameter inferences.
317 ~~We used the Clopper-Pearson exact method for constructing 95% confidence intervals of the~~
318 ~~probabilities of convergence (Clopper and Pearson 1934; Thulin 2014).~~

319 AIC for model selection

320 We ~~estimated the probability of selection~~ investigated the reliability of AIC-based model
321 selection for two purposes. First, we analyzed selection of each process error model ~~structure~~
322 ~~source~~ (R, R+S, R+M, R+Sel, R+q) using marginal AIC. For a given ~~operating model~~ OM
323 ~~simulated data set~~, we compared AIC for EMs ~~that were all configured the same for~~ with
324 different process error assumption conditional on whether median natural mortality (~~known~~
325 ~~or estimated~~) and the SRR (~~rate and the~~ Beverton-Holt ~~or none~~).

326 ~~We also estimated the probability of correctly selecting SRR were estimated. Second, we~~
327 ~~analyzed AIC-based selection~~ between EMs with ~~Beverton-Holt SRR assumed and models~~
328 ~~without the SRR (null model)~~. We made these comparisons between models that otherwise
329 ~~assumed the same process error structure as the operating model and both of the compared~~
330 ~~models either estimate median natural mortality or assume it is known and without the~~
331 ~~Beverton-Holt SRR assumed~~. Contrast in fishing pressure and time series with recruitment
332 at low stock size ~~has have~~ been shown to improve estimation of SRR parameters (Magnus-
333 son and Hilborn 2007; Conn et al. 2010). Our preliminary inspections ~~of the proportions~~
334 ~~of simulations where the correct recruitment model was chosen indicated generally poor~~
335 ~~performance of AIC in determining the Beverton-Holt SRR model~~ for a given set of OM
336 factors (including contrast in fishing pressure) ~~indicated generally poor performance of AIC~~,
337 ~~even when the EM was configured with the correct process error source~~. Therefore, we ~~fit~~

338 logistic regression models to the indicator of Beverton-Holt models having lower AIC as
339 a function conditioned on the EMs having the correct process error assumption and also
340 considered the effect of the log-standard deviation of the true log(SSB) (log SD_{SSB}; similar
341 to the log of the coefficient of variation for SSB) on model selection since simulations with
342 realized SSB producing low and high recruitments recruitment would have larger variation
343 in realized SSB.

344 All model selection results condition on whether all of the compared estimating models
345 completed the only on completion of the optimization process without failure for all of the
346 compared EMs. We did not condition on convergence as defined by a gradient threshold
347 or invertibility of the hessian above because optimization could correctly determine an inap-
348 propriate process error assumption by estimating variance parameters at the lower bound
349 of zero. Such an optimization could indicate poor convergence but the likelihood would be
350 equivalent to that without the mis-specified random effects and the AIC would be appro-
351 priately higher because more (variance) parameters were estimated. All other measures of
352 reliability described below (bias and Mohn's ρ) use these same criteria for inclusion of EM
353 fits in the summarized results.

354 Bias

355 We also investigated bias in estimation of various model attributes as a measure of reliability.
356 For a given model attribute we calculated the relative error

$$RE(\theta_i) = \frac{\hat{\theta}_i - \theta_i}{\theta_i}$$

357

$$RE(\theta_j) = \frac{\hat{\theta}_j - \theta_j}{\theta_j} \quad (1)$$

358 from fitting a given estimating model EM to simulated data set $i \rightarrow j$ configured for a given
359 OM where $\hat{\theta}_i$ and θ_i , $\hat{\theta}_j$ and θ_j are the estimated and true values for simulation i . We

360 estimated bias as the median of the relative errors across all simulations for a given OM
361 and EM combination. We constructed 95% confidence intervals for the median relative bias
362 using the binomial distribution approach as in Miller and Hyun (2018) and Stock and Miller
363 (2021). We present results for bias in terminal year estimates of j . We analyzed simulation
364 results for estimates of terminal year SSB and recruitment, Beverton-Holt stock recruit SRR
365 parameters (a and b), and median natural mortality rate. Results for terminal year fishing
366 mortality were strongly negatively correlated with those for SSB and are provided in the
367 Supplementary Materials.

368 Mohn's ρ

369 Finally, we investigated presence of retrospective patterns in fitted models as a measure of
370 reliability. We calculated Mohn's ρ for SSB, fully selected fishing mortality fishing mortality
371 (averaged over all age classes), and recruitment for each EM fit to each OM simulated data
372 set (Mohn 1999). We fit 7 peels for each EM and calculated median 95% confidence intervals
373 for $P = 7$ peels to each simulated data set and calculated Mohn's ρ using the same methods
374 as that for relative bias. for a given attribute θ as

$$\rho(\theta) = \frac{1}{P} \sum_{p=1}^P \frac{\hat{\theta}_{Y-p,Y-p} - \hat{\theta}_{Y-p,Y}}{\hat{\theta}_{Y-p,Y}} \quad (2)$$

375 where Y is last year of the full set of observations and $\hat{\theta}_{y,y'}$ is the estimate for attribute θ in
376 year y from a model fit using data up to year $y' \geq y$. Thus, θ terms where $y' = Y$ refer to
377 estimates from the fit to all years of data.

378 Results

379 Summarizing results across OM and EM attributes

380 For many R and R+S OMs, convergence rate declined when either the median natural

381 mortality rate or the Beverton-Holt SRR was estimated even when the process error
 382 assumptions of the EMs and OM_s matched (Figure S17, A). When there was high
 383 observation error and constant fishing pressure ($F = F_{MSY}$ for all 40 years), convergence
 384 was poor for all. Because the OM and EM attributes that we investigated are numerous,
 385 we used two methods to summarize the most important factors for differences in results
 386 within a given OM process error source. The first method was fitting regression models
 387 with the response being each of the measures of reliability described above and predictor
 388 variables were defined based on OM and EM characteristics (e.g., MacKinnon et al.
 389 1995; Wang et al. 2017; Harwell et al. 2018). For the binary indicators of convergence
 390 and AIC-based selection of an SRR, we performed logistic regressions. For indicators of
 391 AIC-based selection of EM process error configurations other than R EMs when fitted to R
 392 OM_s ($\sigma_{2+} = 0$) regardless of whether median natural mortality and SRRs were estimated.
 393 Convergence of R EMs was high for all R and R+S OM_s except when there was high
 394 observation error and constant fishing pressure, source (multiple categories) we performed
 395 multinomial regressions. For other measures of reliability we fit linear regression models to
 396 transformed responses. Because relative errors (Eq. 1) and when median natural mortality
 397 and SRRs were estimated. R+S EMs fit to R OM_s exhibited poor convergence regardless of
 398 whether natural mortality or a SRR was estimated. R+S EMs fit to R+S OM_s had highest
 399 convergence rates when there was contrast in fishing pressure and low observation error
 400 Mohn's ρ for the various parameters are bounded below at -1, we used a transformation of
 401 these values

$$y_j = \log [f(\hat{\theta}_j, \theta_j) + 1] \quad (3)$$

402 where f is either the relative error (Eq. 1) or Mohn's ρ (Eq. 2) for simulation j , so that
 403 values are unbounded. For relative errors, y_j is the log-scale error. We omitted simulations
 404 where estimated attributes equal to zero (RE = -1). For all regressions we fit separate models
 405 with just individual OM and EM factors included, with all factors included, with all second
 406 order interactions, and with all third order interactions. For the multinomial regression,

407 we used the `vglm` function from the VGAM package (Yee 2008; Yee 2015). We tabulated
408 percent reduction in residual deviance for each of the regression fits. We did not perform
409 formal statistical analyses of effects of OM and EM attributes on results (e.g., ANOVA)
410 because of the lack of independence of the “observations” that results from fitting multiple
411 EMs to each simulated data set.

412 The second method involved fitting classification and regression trees (Breiman et al. 1984)
413 to show how the OM and EM attributes, and their interactions, partition the values for each
414 measure of reliability (e.g., Gonzalez et al. 2018; Collier et al. 2022). We used classification
415 trees for categorical measures (convergence and AIC) and regression trees for the other
416 measures with continuous scales (relative error and Mohn’s ρ). The response variables were
417 the same as the regressions for the deviance reduction analyses. We used the `rpart` function
418 in the `rpart` package (Therneau and Atkinson 2025) to fit trees. Full trees were determined
419 using default settings except that we increased the number of cross-validations to 100. For
420 clarity, we manually pruned the full trees to show just the primary branches. Convergence
421 rates were high for all EMs when fit to data from R+S OMs with lower observation error
422 except those where median natural mortality and /or SRRs were estimated.

423 Convergence of all EMs fitted to R+M OMs was highest when the OMs had higher natural
424 mortality process error variability, low observation error, and contrast in fishing pressure
425 (Figure S17, B). We also provide detailed results for all measures of reliability at each
426 combination of OM and EM attributes in the Supplementary Materials. For confidence
427 intervals of probability of convergence, we used the Clopper-Pearson exact method (Clopper
428 and Pearson 1934; Thulin 2014). For AIC selection of process error source we provide
429 estimates of the proportions of simulations where each EM type was selected. For AIC
430 selection of the SRR (a binary indicator for each simulated data set), we fit logistic regressions
431 and present resulting predicted probabilities of correctly selecting the SRR as a function
432 of SSB variability ($\log SD_{SSB}$ described above). We estimated bias as the median of the
433 relative errors across all simulations for a given OM and EM combination. We constructed

434 95% confidence intervals for the median relative bias, and Mohn's ρ using the binomial
435 distribution approach (Thompson 1936) as in Miller and Hyun (2018) and Stock and Miller
436 (2021).

437

Results

438

Convergence performance

439 For probability of convergence, the EM process error assumption was the single attribute that
440 resulted in the largest percent reduction in deviance (14-28%) for all OM process error sources
441 other than R+M EMs that estimated autocorrelation of process errors had poor convergence
442 for R+M OMs when there was low natural mortality process error variability regardless of
443 autocorrelation of the simulated process errors. R+S EMs fitted to data generated from R+M
444 OMs always converged poorly whether or not OMs where the EM median natural mortal-
445 ity and the Beverton-Holt SRR were estimated. rate assumption (estimated or known)
446 explained the most residual deviance (>11%; Table 1). However, including interactions of
447 OM and EM factors also provided large reductions in residual deviance (35-47%), suggesting
448 successful convergence depended on a combination OM and EM attributes.

449 The RClassification trees for each OM process error source all had the primary branch
450 defined using the same attribute that provided the largest reduction in deviance (Figure 1).
451 EMs that assumed R+S EMs, in particular, had poor convergence when fit to data generated
452 from RS process errors converged poorly for all OMs that were simulated with the alternative
453 process error assumptions (R, R+Sel OMs with lower selectivity process error variability or
454 higher observation error (Figure S17, C). M, R+Sel EMs generally converged better than
455 other EMs for Sel, and R+Sel OMs with higher process error variability, lower observation
456 error, and contrast q OMs). For all trees, branches based on the OM fishing mortality history
457 showed better convergence when the OM included a change in fishing pressure regardless of

458 whether. Branches based on whether the Beverton-Holt SRR was assumed or not, showed
459 better convergence when it was not estimated and branches based on the median natural
460 mortality or a SRR was estimated.

461 For rate assumption showed better convergence when it was treated as known. For some
462 R+q OMs, convergence of RM and R+q EMs was generally better than that of other EMs
463 Sel OMs, better convergence was also observed when there was contrast in fishing pressure
464 (Figure (S17, D). Convergence of R+S EMs was generally worse than that of all other EMs
465 across all OMs whether or not median natural mortality or a SRR was estimated. Again,
466 convergence probability generally declined for all EMs when median natural mortality or a
467 SRR was estimated. lower observation uncertainty.

468 When convergence is defined by a gradient threshold, the primary factor explaining deviance
469 reduction is the same that for Hessian-based convergence for all OM process error sources,
470 but there are some differences in deviance reduction for secondary factors (Table S6), and
471 probability of convergence, overall, was lower (Figure S2). We found a wide range of maxi-
472 mum absolute values of gradients for models that converged had invertible Hessians (Figure
473 S3). The largest value observed for a given EM and OM combination was typically $< 10^{-3}$,
474 but many converged models had values greater than 1. For many OMs, EMs that assumed
475 the correct process error type source and did not estimate median natural mortality or the
476 Beverton-Holt SRR produced the lowest gradient values.

477 AIC performance

478 Marginal AIC accurately determined

479 Process error source

480 For AIC selection of the correct process error assumptions in EMs when data were generated
481 from R and configuration, the magnitude of observation and process error variation were the

482 attributes that resulted in the largest percent reductions in deviance across OM process error
483 sources other than R OMs (Table 2). Both sources of variation explained large reductions
484 in deviance for R+S OMs, regardless of whether median natural mortality or (17-22%) and
485 R+Sel (8-26%) OMs, whereas variance of process errors provided the major reductions for
486 R+M (>9%) and R+q (>13%) OMs. Comparatively, none of the OM or EM attributes
487 explained particularly large reductions in deviance for R OMs, but fishing history, whether
488 a SRR was estimated (Figure S4, A). Attempting to estimate , and whether median natu-
489 ral mortality or a SRR separately had a negligible effect on the accuracy of determining
490 the correct process error assumption. When both were estimated, there was a noticeable
491 reduction in accuracy when OMs had a constant fishing pressure, low observation error, and
492 larger variability in recruitment process errors. was known or estimated provided similar and
493 the largest reductions (approximately 5-6%). Inclusion of second and third order interactions,
494 did not provide large reductions in deviance for any of the OM process error sources.

495 For all OM process error sources other than R OMs, the attributes defining the primary
496 branches of classification trees matched those that provided the largest reductions in deviance
497 (Figure 2). Across all OMs, AIC was more accurate for the process error source when process
498 error variability was greater and when observation error was lower. For R+M OMs, marginal
499 AIC only accurately determined the correct process error model and correlation structure
500 when observation error was low and variability in natural mortality process errors was high
501 (Figures S4, B). Of these OMs, estimating the median natural mortality rate only reduced
502 the accuracy of AIC when natural mortality process errors were independent and fishing
503 pressure was constant. For OMs with poor model selection accuracy, AIC most frequently
504 selected EMs with process errors in catchability (R+q) or selectivity (R+Sel). Selection of
505 R+S EMs was generally unlikely.

506 Marginal AIC most accurately determined the correct source of process error and correlation
507 structure for R+Sel OMs with low observation error (Figures S4, C). When there was low
508 variability in selectivity process errors and high observation error , R+q or R+S EMs were

509 more likely to have the best AIC. Whether median natural mortality or SRRs were estimated
510 appeared to have little effect on the performance of AIC.

511 Marginal AIC most accurately determined the correct source of process error and S OMs,
512 there was a tendency to select R OMs when observation error was higher and apparent
513 survival variation was lower ($\sigma_{2+} = 0.25$), but accuracy for the process error source was
514 otherwise highly accurate. Larger variability of process error relative to observation error
515 was also required for accurate identification of the correct correlation structure for R+q OMs
516 with high variability in catchability process errors (Figures S4,D). The R M, R+q, and R+q
517 OMs with low variability in catchability process errors and high observation error had the least
518 model selection accuracy. However, for these OMs, the marginal AIC accurately determined
519 the correct source of process error (but not correlation structure) except when Sel OMs
520 (Figure S4). No branches were estimated for classification trees fit to the R OMs assumed
521 a likely because accuracy was high across all simulations (0.94), although inspection of
522 the fine-scale results shows there is some degradation in AIC selection when an SRR and
523 median natural mortality rate are estimated for R OMs with constant fishing pressure and
524 EMs estimated both median natural mortality and the SRR at high observation error (Figure
525 S4, top left).

526 AIC performance for the stock-recruit

527 Stock-recruit relationship

528 Our comparisons of model performance conditioned on assuming the true process error
529 configuration is known (EM and OM process error types match) and we focus on results where
530 the EMs assume median natural mortality is known because there was little difference in
531 results when the EMs estimated this parameter. Broadly, we found generally poor accuracy
532 of AIC in selecting models assuming a Logistic regressions for AIC selection of the Beverton-
533 Holt SRR over the null model without an SRR for all OMs SRR, showed OM fishing history

534 and log SD_{SSB} provided substantial reductions in deviance for R+M (>13%), R+Sel (>26%),
535 and R+q (>24%) OM_s (Table 3). For R OM_s, fishing history provided the largest reduction
536 in deviance (>9%), whereas none of the attributes individually provided large reductions
537 in deviance for R+S OM_s (all <5%). However, we also found increased accuracy of AIC
538 in determining the Beverton-Holt SRR when the simulated population exhibited greater
539 variation in spawning biomass for nearly every OM (Figure S19).

540 With inclusion of all attributes provided larger reductions in deviance than the sum of
541 individual contributions for both R (>30%) and R+S (~19%) OM_s. Further fits for R and
542 R+S process error assumptions, probability of lowest AIC for the B-H SRR as a function
543 of SSB variability were greatest for OM_s with contrast in fishing pressure and lower process
544 variability in recruitment (Figure S19, A). The largest variation in SSB occurred in OM_s with
545 larger recruitment variability ($\sigma_R = 1.5$; Figure S19, A, right column group), but the same
546 high AIC accuracy was achieved for OM_s with lower recruitment variability at lower levels
547 of SSB variation. The level of observation error had little effect on AIC accuracy OM_s that
548 including different combinations of two factors additively showed fits that included log SD_{SSB}
549 and recruitment variation only provided essentially the same reduction in deviance as the
550 models with all factors. For all OM process error sources, inclusion of interaction terms
551 provided relatively little reduction in residual deviance.

552 For R+M OM_s, probability of lowest AIC for the Beverton-Holt SRR increased steeply
553 with variation in SSB whether it was induced by Attributes defining the primary branches
554 of classification trees for AIC selection of the SRR assumption were the same as those
555 explaining the largest reductions in deviance for the logistic regression models (Figure 3).
556 All branches based on log SD_{SSB} showed better accuracy with larger variability in SSB and
557 all branches based on fishing history showed better accuracy when there was contrast in
558 fishing or variation in natural mortality process error. (Figure S19, B). There was little
559 difference in AIC accuracy whether the natural mortality process errors were correlated and
560 similar to pressure. Branches based on OM observation error or recruitment variability (R

561 and R+S OMs, there was also little effect due to level of observation error.

562 For R+Sel OMs, showed better accuracy when they were lower. For R OMs, a combination
563 of lower recruitment variability, contrast in fishing pressure over time was the primary source
564 of variation in SSB and these are the OMs where AIC accuracy for the Beverton-Holt SRR
565 was greatest (Figure S19, C). There was little effect of variability or correlation of selectivity
566 process errors or the level of observation error on AIC accuracy.

567 Like the, and higher SSB variability produced AIC accuracy over 0.8. For R+S OMs,
568 lower recruitment variability and observation error and higher SSB variability produced
569 AIC accuracy of 0.79. For R+Sel OMs, the greatest accuracy for AIC in selecting the
570 Beverton-Holt SRR occurred for M, R+Sel, and R+q OMs where there was contrast in fishing
571 pressure over time which is also where there was the greatest variation in SSB (Figure S19,
572 D). There was also little effect of variability or correlation of catchability process errors or
573 the level of observation error on AIC accuracy, accuracy of 0.87 to 0.94 was observed with
574 just increased SSB variability.

575 Bias

576 Terminal year spawning stock biomass, fishing mortality, and recruitment

577 For R OMs ($\sigma_{2+} = 0$), there was no indication of bias (95% confidence intervals included
578 0) in terminal year SSB. Regression models for log-scale errors in SSB that included the
579 various OM and EM factors showed little reduction in deviance (<5%) for any of the
580 estimating models regardless of process error assumptions, except when no SR assumption
581 was made, recruitment variability was low, and there was contrast in fishing mortality and
582 high observation error (Figure S10, A). However, errors in terminal SSB estimates were
583 highly variable when factors across all OM process error sources (Table 4). The attributes
584 producing the largest reductions were the EM assumption for median natural mortality was

585 estimated and there was constant fishing pressure and high observation error (Figure S10,
586 A, second row).

587 For (known or estimated) for R, R+S OMs, the EMs with matching process error
588 assumptions generally produced unbiased estimation of terminal SSB except when median
589 natural mortality was estimated and there was high observation error. In M, R+S OMs
590 with low observation error, EMs with incorrect process error assumptions typically provided
591 biased estimation of terminal year SSB. Estimating the Beverton-Holt SRR had little
592 discernible effect on bias of terminal year SSB estimation whereas estimating median M
593 tended to produce more variability in errors in terminal SSB estimation similar to ROMs.

594 For RSel, and R+M OMs with low variability in natural mortality process errors, low
595 observation error and contrast in fishing mortality over time all EMs produced low variability
596 in SSB estimation error that indicated unbiasedness (Figure S10, B, third row). However,
597 larger variability in natural mortality process errors increased bias of EMs without the
598 correct process error type. Estimating median natural mortality increased variability of
599 SSB estimation error particularly for OMs with high observation error and constant fishing
600 pressure over time. It also increased bias in SSB estimation for many Rq OMs (1-3%), EM
601 process error assumption for R+M OMs. Like Rand RS OMs (4%) and fishing history for all
602 OM process error sources (1-5%). Including second order interactions provided the largest
603 reductions in residual deviance (10- 26%). Including third order interactions also provided
604 further reductions for R, R+SOMs, estimating a SRR had little discernible effect on SSB
605 bias.

606 For, and R+Sel OMs, there was no evidence of bias for any EMs when variability in
607 selectivity process error and observation error was low, and q OMs between 5 and 11%.

608 In all regression trees, branches based on fishing history and level of observation error
609 generally showed less bias in SSB with contrast in fishing mortality (Figure S10, C). The
610 largest bias occurred for any EMs that estimated median natural mortality when the OMs

611 had high observation error, constant fishing pressure, and greater variability in selectivity
612 process errors ($\sigma_{Sel} = 0.5$) or low selectivity process errors ($\sigma_{Sel} = 0.1$) and low observation
613 error. However, there was no evidence of SSB bias for correctly specified R+Sel EMs when
614 observation error was low and variation in selectivity process errors was larger, whether
615 median natural mortality was estimated or not (Figure S10, C, third row). We only observed
616 an effect of estimating the Beverton-Holt SRR for R+Sel OM_s that had high observation
617 error and contrast in fishing pressure where estimating the SRR produced less biased SSB
618 estimation for many EMs (Figure S10, C, top row).

619 All EMs fit to data from R+q OM_s with low observation error and contrast in fishing
620 pressure exhibited little evidence of bias in terminal SSB estimation except for R+M EMs
621 when there was no AR1 correlation in catchability process errors (Figure S10, D). Many EMs
622 also performed well in and lower observation error (Figure 4). For scenarios where there was
623 bias, it was generally positive (over-estimation). For branches based on treatment of median
624 natural mortality rate, bias was generally less when it was known rather than estimated. For
625 some R+q OM_s with low observation error, but no contrast in fishing pressure. For Sel and
626 R+q OM_s with high observation error and contrast in fishing pressure, EMs that estimated
627 the Beverton-Holt SRR exhibited less SSB bias than those that did not. Estimating median
628 natural mortality in the EMs only resulted in much more variable SSB estimation errors
629 when there was no contrast in fishing pressure (Figure S10, D, first and third rows), less bias
630 in SSB was shown when the EM process error assumption was correct.

631 For all OM process error types, relative errors in terminal year recruitment were generally
632 more variable than SSB, but effects of R and R+S Results for bias in fishing mortality
633 and recruitment generally matched those for SSB, except that directions of bias for fishing
634 mortality were opposite to those for SSB and recruitment. Effects of individual OM and
635 EM attributes on bias (i.e., negative or positive or none) were similar (Figure S12, A).
636 Furthermore, for EM configurations where bias in terminal SSB was evident, median relative
637 errors in recruitment often indicated stronger bias in recruitment of the same sign. factors on

638 regression models were similarly small as measured by reduction in deviance (Tables S7 and
639 S8). Factors defining the primary branches of regression trees were in most cases identical
640 to those for SSB (Figures S5 and S6).

641 Stock-recruit parameters

642 Across all OMs, there was generally less bias and (or) lower variability in estimation of the
643 Beverton-Holt a parameter than the Regression models for log-scale errors of estimates of
644 both the Beverton-Holt a and b parameter. In Rand R parameters showed none of the factors
645 explained large percent reductions in deviance (Table 5). The OM fishing history provided
646 the largest deviance reduction for most OM process error sources for both parameters, but
647 reductions were generally less than 6%. Exceptions were the R+S OMs, EMs with the correct
648 assumptions about process errors provided the least biased estimation of Beverton-Holt SRR
649 parameters when there was a change in fishing pressure over time and lower variability of
650 recruitment process errors, but there was little effect of estimating median natural mortality
651 and a small increase in bias for those OMs that had high observation error (Figure S7,
652 A). For other R and Sel OMs where a and b were reduced by approximately 11% and 8%,
653 respectively, and the R+S OMs, estimating natural mortality often resulted in less biased
654 estimation of SRR parameters. There was generally large variability in relative errors of
655 the SRR q OMs where b was reduced by 10%. The EM process error assumption provided
656 similar reductions in deviance for both parameters for R OMs. Including interactions also
657 did not produce important reductions in deviance.

658 For regression trees of log-scale errors in Beverton-Holt a and b parameter estimates, but
659 the lowest variability occurred with low variability in recruitment and little or no variability
660 in survival process errors ($\sigma_{2+} \in \{0, 0.25\}$), and contrast in fishing pressure.

661 In R+M OMs, the most accurate estimation of SRR parameters for all EM process error
662 assumptions occurred when there was a change in fishing pressure, greater variability in

663 natural mortality process errors, and lower observation error (Figure S7, B). Relative to the
664 R, less bias was indicated with contrast in OM fishing pressure for all branches in trees
665 for each OM process error source (Figures 5 and 6). For all branches based on recruitment
666 variability in trees for R and R+S OMs, there was even less effect of estimating median
667 natural mortality on estimation bias for the SRR parameters.

668 Bias for SRR parameters was large and variability in relative errors was greatest for most
669 EMs fit to R+Sel OMs with constant fishing pressure (Figure S7, C). Less bias in parameter
670 estimation occurred for OMs with a change less bias in both a and b was observed with less
671 recruitment variability. For R OMs with contrast in fishing pressure and the best accuracy
672 occurred for those OMs that had low observation error and more variable and uncorrelated
673 selectivity process errors, and when the EMs had with the correct process error assumption.
674 There was little effect of estimating natural mortality on relative errors for SRR parameters.

675

676 Like greater recruitment variability EMs that assumed the incorrect R+Sel OMs, relative
677 errors in SRR parameters for R+q OMs were more accurate for most EM process error
678 types when OMs had contrast in fishing pressure and lower observation error (Figure S7, D).
679 However, the best accuracy occurred for those OMs that had M process errors produced less
680 bias in both a and b than other process error assumptions. Across all combinations of OM
681 and EM attributes, some bias was observed for both parameters, but there was generally
682 less bias and (or) lower variability in catchability process errors. The worst accuracy of SRR
683 parameter estimation regardless of EM type occurred when R+q OMs had low observation
684 error and constant fishing pressure estimation of the a parameter than the b parameter
685 (Figure S7, D, fourth row).

686 Median natural mortality rate

687 Across all OMs and EMs there was little effect of estimating SRRs on the bias in estimation

688 of Fitted regression models for log-scale errors in median natural mortality (Figure S13).
689 Median natural mortality rate was estimated accurately by all rate showed largest percent
690 reductions in residual deviance for R+S and R+M models (Table 6). The largest reductions
691 for a single attribute was the EM process error types for all R OM^s except those with
692 high observation error and constant fishing pressure, in which case relative errors were high
693 (Figure S13, A, $\sigma_{2+} = 0$). For assumption ($>20\%$) and fishing history ($>15\%$) for R+S
694 OM^s estimation of median natural mortality rate was most accurate when observation error
695 was low and there was contrast in fishing pressure and the EM process error type was correct.

696

697 For . Fishing history also provided $>10\%$ reduction for R+M OM^s, median natural mortality
698 was estimated most accurately, regardless of EM process error type, when OM^s had a change
699 in fishing pressure and low observation error (Figure S13, B). However, those but reductions
700 for all factors in R, R+M OM^s that also had greatest variability in AR1 correlated natural
701 mortality process errors only had unbiased estimation when the EM process error type was
702 correct Sel, and R+q OM^s were relatively low ($<6\%$). Interactions of OM and EM factors
703 also provided substantial further reductions for R+S and R+M OM^s (between 8 and 15%
704 for second order interactions).

705 All EM process error types accurately estimated Regression trees with branches based on
706 fishing history showed less bias in median natural mortality rate for R+Sel OM^s that had
707 with contrast in fishing pressure , low observation error , and low selectivity process error
708 variability (Figure S13, C). When selectivity process error variability increased, the incorrect
709 EM process errors produce more biased estimation of median natural mortality rate. The
710 least accurate estimation occurred for all EM process error types when observation error was
711 high and fishing pressure was constant.

712 Like and branches based on level of observation error showed less bias with more precise
713 observations (Figure 7). For R OM^s, branches based on EM process error assumption showed

714 less bias with EMs assuming the correct R and the incorrect R+S assumption. For R+Sel
715 OMs, all EM process error types produced accurate estimation of median natural mortality
716 rate when fit to S and R+q OMswith contrast in fishing pressure, low observation error and
717 low catchability process error variability (Figure S13, D). Most M OMs, branches based on
718 EM process error showed only the correct EM process error types produced biased estimation
719 of median natural mortality when R+q OMs had high observaiton error and constant fishing
720 pressureassumption with less bias.

721 Mohn's ρ

722 Regression models for Mohn's ρ for SSB was small in absolute value for all R and R+S OMs,
723 regardless of EM process error types, and whether median natural mortality rate or SRRs
724 were estimated (Figure S14, Aof SSB showed little reduction in deviance for any of the OM
725 an EM attributes (<2%; Table 7). The strongest retrospective patterns (highest absolute
726 lack of explanatory power is also reflected in the regression trees where median Mohn's ρ
727 values)occurred in OMs with the largest apparent survival process error variabilityare near
728 zero unless a large combinations of OM and EM conditions occur (Figure 8). For example,
729 in R+S OMs, with constant fishing pressure, high observation error, and contrast in fishing
730 pressure, but only for EMs with the incorrect process errortype and where median natural
731 mortality rate was assumed known (median ρ was approximately -0.15). For higher apparent
732 survival process error, EMs that assume R+M ,R+Sel, and R+q OMs, process errors have
733 a median Mohn's $\rho = -0.068$.

734 Similarly, poor explanatory power of the OM and EM attributes occurred when we fit
735 regression models for Mohn's ρ was also small in absolute value, but median values were
736 all closer to 0 than the largest values in the R and R+S OMs (Figure S14,B-D). For these
737 OMs, there was no noticeable effect of estimation of median natural mortality rate or SRRs
738 on of fishing mortality and recruitment (Tables S9 and S10). Regression trees for Mohn's ρ

739 for any EM process error types.

740 of fishing mortality were similar to those for SSB in that median values of Mohn's ρ for
741 recruitment was small in absolute value for all R OM^s with low variability in recruitment
742 process errors, regardless of EM process error type, and whether median natural mortality
743 rate or SRRs were estimated (Figure S16, A) were close to zero for most combinations of OM
744 and EM attributes (Figure S8). However, R and R+S OM^s with greater recruitment process
745 variability and higher observation error had we observed median Mohn's ρ for recruitment
746 greater than zero for most EMs even when the EM process error type was correct. In R+S
747 OM^s with lower 0.1 at branches much closer to the base of the trees with fewer interactions of
748 the OM and EM attributes (Figure S9). These branches with consistently large retrospective
749 patterns were typically defined by larger OM observation error, EMs with the correct process
750 error type exhibited better median Mohn's ρ close to 0 than EMs with the incorrect process
751 error type. For R+M, R+Sel, and R+q OM^s, results for OM constant fishing pressure,
752 or incorrect EM process error configuration. Comparing regression model and regression
753 tree fits, attributes defining the primary branches for all regression trees of all Mohn's ρ for
754 recruitment are similar to those for SSB, but the range in median values and variation in
755 Mohn's ρ values for a given OM are generally larger for recruitment (Figure S16, B-D) values
756 (SSB, fishing mortality, and recruitment) generally matched those that explained the largest
757 reductions in deviance.

758 Discussion

759 Assessing convergence

760 Analyses of model convergence across simulations can be useful for understanding the
761 utility of alternative convergence criteria used in applications to real data for directing the
762 practitioner to more appropriate random effects configurations. It is common during the

763 ~~assessment model fitting process to check that the maximum absolute gradient component~~
764 ~~is less than some threshold prior to inspecting the Hessian of the optimized likelihood for~~
765 ~~invertibility (Carvalho et al. 2021). However, there is no accepted standard for the gradient~~
766 ~~threshold (e.g., Lee et al. 2011; Hurtado-Ferro et al. 2014; Rudd and Thorson 2018) and~~
767 ~~some thresholds would exclude models that in fact have an invertible Hessian. We found the~~
768 ~~Hessian at the optimized log-likelihood can often be invertible when the maximum absolute~~
769 ~~gradient was much larger than what would perceived to be a sensible threshold.~~

770 Poor convergence was common in our results when the incorrect process error source was
771 assumed. Li et al. (2024) found that convergence ~~rate~~ could be a useful diagnostic espe-
772 ~~cially for separating the correct model simpler process error assumption from overly complex~~
773 ~~models. However, the criteria for convergence used in their study may also lead to limited~~
774 ~~ability to distinguish the correct model from overly simplistic models, a pattern that was~~
775 ~~also noted by Liljestrand et al. (2024) in which one process error may absorb all sources of~~
776 ~~process error when the magnitude of other process errors are low.~~

777 ~~Often poor convergence result when Poor convergence often occurs when~~ parameter estimates
778 ~~are at their bounds (Carvalho et al. 2021), and this also applies to variance parameters for~~
779 ~~random effects with state-space assessment models. Even. However, even~~ when the Hes-
780 ~~sian is invertible for a converged model, parameters that are poorly informed will have~~
781 ~~extremely large variance estimates. This further inspection can lead to a more appropri-~~
782 ~~ate and often more parsimonious model configuration where the problematic parameters~~
783 ~~are not estimated. For example, process error variance parameters in state-space models~~
784 ~~that are estimated close to 0 indicates that the random effects are estimated to have lit-~~
785 ~~tle or no variability and removing these process errors is warranted. Generally, our results~~
786 ~~suggest we can expect lower probability of convergence of state-space assessment models~~
787 ~~when estimating natural mortality or SRRs because of the difficulty distinguishing these~~
788 ~~parameters from others being estimated in assessment model with data that are typically~~
789 ~~available. Our experiments did not aim to emulate the practitioner decision process in~~

790 ~~developing model configurations (e.g. removing a source of process error and refitting the~~
791 ~~model when process error variance parameters were estimated close to 0).~~ Evaluating the
792 ~~determining an appropriate model configurations, but evaluating the~~ efficacy of such a deci-
793 sion process when applying EMs might be important in closed loop simulations (e.g. MSE)
794 aimed at quantifying management performance (e.g., a management strategy evaluation).

795 It is common during the assessment model fitting process to check that the maximum
796 absolute gradient component is less than some threshold prior to inspecting the Hessian
797 of the optimized likelihood for invertibility (Carvalho et al. 2021), but we found reliance on
798 magnitude of the gradient values for fitted models as a convergence criterion questionable.
799 There is no accepted standard for the gradient threshold (e.g., Lee et al. 2011; Hurtado-Ferro
800 et al. 2014; Rudd and Thorson 2018), but the Hessian at the optimized log-likelihood was
801 often invertible when the maximum absolute gradient was much larger than what might be
802 perceived to be a sensible threshold in some of our simulations. Therefore, the gradient
803 criterion could exclude models that in fact have an invertible Hessian.

804 A factor affecting the convergence criteria, particularly for maximum likelihood estimation
805 of models with random effects, is numerical accuracy. All optimizations performed in these
806 simulations are of the Laplace approximation of the marginal likelihood and, therefore, gra-
807 dients and Hessians are also with respect to this approximation (see TMB::sdreport in the
808 Template Model Builder package). Functionality within the Template Model Builder pack-
809 age exists (i.e., TMB::checkConsistency) to check the validity of the Laplace approximation
810 and the utility of this as a diagnostic for state-space assessment models should be explored
811 further. Furthermore, numerical methods are used to calculate and invert the Hessian for
812 variance estimation for models with random effects. ~~Along with our results, Our results,~~
813 along with the potential lack of accuracy imposed by these approximations, ~~suggests suggest~~
814 at least investigating whether the Hessian is positive definite when the calculated absolute
815 gradients are not terribly large (e.g, < 1).

816 **Configuring process error**

817 Of the OM process error configurations we considered, we found AIC to be accurate for
818 selecting models with process errors on recruitment and apparent survival (Rand We found
819 accuracy of marginal AIC-based selection for the correct process error source required only
820 low observation error for R, R+S). Fitting models to other OMs rarely preferred R+S
821 EMsSel, and Rand +q OMs. R+S EMs were nearly always selected for the matching OMs;
822 a similar result was reported by Liljestrand M OMs further required higher process error
823 variability, but this also improved accuracy for the other OM process errors sources when
824 there was higher observation error. These results seem consistent with Li et al. (2024). For
825 other sources of process error, Their simulation studies investigated models with multiple
826 process error sources and found good accuracy of AIC was improved when there was larger
827 variability in the process errors and/or lower observation error in detecting correct process
828 error assumptions for simulations based on two stocks (Gulf of Maine cod and Southern
829 New England-Mid-Atlantic yellowtail founder) that are well sampled by NEUS bottom trawl
830 surveys used as indices in the respective assessments and poor accuracy for Atlantic mackerel,
831 a semi-pelagic species that is observed relatively poorly.

832 Across all OM process error configurations, AIC performed poorly in identifying that the
833 presence of

834 **Stock recruitment relationships**

835 Variation in SSB was the most important factor for using marginal AIC to correctly
836 distinguish the Beverton-Holt SRR in the OM unless there was contrast in fishing pressure
837 possibly in combination with other factors such as lower variability in recruitment process
838 errors (in R from the null model without an SRR. For R+M, R+Sel, and R+S models) or
839 greater variation in natural mortality process errors (for R+M OMs , Fig. S19) . As such,
840 properly accounting for process error in natural mortality could be important (Li et al.

841 2024) when evaluating SRRs in state-space models. Curiously, we did not find a marked
842 effect of the level of observation error on ability q OMs, the SRR was accurately detected
843 when the CV of SSB over the time series was at least 40 to detect the SRR, but it is
844 possible that AIC would perform better if observations have even lower uncertainty than
845 we considered 50% ($\log SD_{SSB} = -0.9$ to -0.7) regardless of any other OM or EM attributes.
846 Detection of the SRR for R and R+S OMs required lower recruitment variability, but
847 this lower level ($\sigma_R = 0.5$) was assumed for all of the other OM process error source and
848 represents the lower range of estimates from recent NEUS stock assessments. Our results
849 assumed that the EM process error configuration was correct, but this may not be a strong
850 limitation given the ability of AIC to distinguish the process error source in many scenarios.

851 Although we did not compare models with alternative SRRs (e.g., Ricker and vs. Beverton-
852 Holt), we do not expect AIC to perform any better distinguishing between relationships and
853 may be more difficult than distinguishing from the null model even with larger variability in
854 SSB. Our finding that AIC tended to choose simpler recruitment models in most many cases
855 contrasts with the noted bias in AIC for more complex models (Shibata 1976; Katz 1981;
856 Kass and Raftery 1995), but, whereas those. However, these earlier findings apply to the
857 much more common comparison of models that are fit to raw and independent observations,
858 here we are comparing whereas our comparisons of state-space models which account for
859 observation error and separately estimate process errors in latent variables.

860 Our results comport with those of de Valpine and Hastings (2002) who found AIC could not
861 distinguish among state-space SRRs that were fit just to SSB and recruitment observations
862 (i.e., not within an assessment model). Similarly, Britten et al. (In review) found AIC
863 could not reliably distinguish the Beverton-Holt SRR from no SRR, nor identify alternative
864 environmental effects on SRR parameters. However, Miller et al. (2016) did find AIC to
865 prefer a-an SRR with environmental effects when applied to data for the SNEMA Southern
866 New England-Mid-Atlantic yellowtail flounder stock and AIC also selected an environmental
867 covariate on a-an SRR for the most recent stock assessment of Georges Bank yellowtail

868 flounder (NEFSC 2025). Both of these yellowtail flounder stocks have large changes in
869 stock size and the values of environmental covariates over time. Additionally, this species is
870 well-observed by the bottom trawl survey that is used for an index in assessment models.

871 ~~As expected, bias in all parameters and assessment output was generally improved with~~
872 ~~lower observation error.~~ Estimation of SRR parameters was only moderately reliable in
873 ideal scenarios of low observation error and contrast in fishing for ~~some~~ R+Sel and R+M
874 OMs ~~, but generally with large temporal variability in process errors.~~ Otherwise, SRR
875 parameter estimation was biased and(or) highly variable. We found substantial bias in esti-
876 mated SRR parameters in R and R+S OMs particularly with high variability in recruitment
877 and apparent survival process errors, suggesting that practitioners should be cautious with
878 SRR inferences when fitted assessment models have these properties. We only evaluated
879 effects of SSB variability on accuracy of AIC in identifying the SRR, but those results
880 suggests we might find less bias for the SRR parameters in such cases as well. Another
881 condition that could improve perception of bias in our simulation studies is restricting
882 results to fits that converged with Hessian-based standard errors for all parameters, but
883 Britten et al. (In review) did not find less SRR parameter bias when restricting estimates
884 using a gradient-based criterion. A simulation study by Stock and Miller (2021) examining
885 configurations of environmental covariate effects on a Beverton-Holt SRR for the previously
886 mentioned, well-observed, Southern New England-Mid-Atlantic yellowtail flounder stock
887 found little or no bias for the density-independent mortality parameter a , but still biased
888 estimation of the density-dependent parameter b .

889 ~~On the other hand, estimation of~~

890 Estimating assessment model quantities

891 As expected, bias in parameters, SSB, and other assessment output was generally improved
892 with lower observation error. Estimation of median natural mortality was reliable in many

893 OM scenarios with contrast in fishing pressure, consistent with Hoenig et al. (2025). In
894 some OMs, when EMs estimated the SRR parameters and median natural mortality, bias
895 for those parameters was improved. Conversely, for some R+Sel and R+q OMs where there
896 was bias in natural mortality due to high observation error, estimating the SRR reduced the
897 bias in median natural mortality rate. However, estimating median natural mortality did
898 cause However, we found poor accuracy in SSB estimation terminal SSB estimation when
899 estimating median natural mortality in many OMs without when there was no contrast in
900 fishing pressure over time and with higher observation error. Thus Therefore, estimating me-
901 dian natural mortality should be approached with caution in state-space assessment models,
902 particularly given its significant impact on determination of reference point and stock status
903 (Li et al. 2024).

904 Negligible retrospective patterns

905 Incorrect EM process error assumptions did not produce strong retrospective patterns for
906 SSB for any OMs regardless of whether median natural mortality or a-an SRR was estimated,
907 but some weak retrospective patterns occur although some weak patterns occurred when
908 observation error was high and there was contrast in fishing pressure. However, retrospective
909 patterns tended to be more variable for recruitment and were sometimes large even when
910 the EM was correct. Therefore, we recommend emphasis-de-emphasis on inspection of
911 retrospective patterns primarily for SSB and F patterns for recruitment, but further research
912 on retrospective patterns in other assessment model parameters, management quantities such
913 as biological reference points, and projections may be beneficial (Brooks and Legault 2016).

914 The general lack of retrospective patterns with mis-specified process errors is perhaps to be
915 expected. Retrospective patterns are often induced in simulation studies by rapid changes
916 in a quantity such as index catchability, natural mortality, or perceived catch during years
917 toward the end of the time series (Legault 2009; Miller and Legault 2017; Huynh et al. 2022;

918 Breivik et al. 2023). In our simulations, the process errors changing over time may have
919 trends in ~~particular~~certain simulations, particularly when strong autocorrelation is imposed,
920 but the random effects have no trend on average across simulations. Szuwalski et al. (2018)
921 and ~~Li et al.~~Li et al. (2024) also found relatively small retrospective patterns when the source
922 of mis-specification was temporal variation in ~~demography~~demographic attributes. Indeed,
923 it is common for the flexibility provided by temporal random effects to reduce retrospective
924 patterns (Miller et al. 2018; Stock et al. 2021; Stock and Miller 2021), though it does not
925 necessarily indicate a more accurate assessment model (Perretti et al. 2020; Li et al. 2024;
926 Liljestrand et al. 2024). Our results together with the existing literature seem to suggest
927 that when a strong retrospective pattern is observed in an assessment it is more likely to
928 be due to a mis-specification of a rapid shift in some model attribute rather than whether a
929 particular process is assumed to be randomly varying temporally.

930 Summarization approach

931 ~~Our simulation study examined the importance of several factors for reliable inferences from~~
932 ~~state-space age-structured assessment models. Contrast in fishing pressure was consistently~~
933 ~~an important factor across all~~ We found the use of regression models and classification
934 and regression trees extremely useful in understanding the most important OM and EM
935 attributes explaining variation in the measures of reliability we examined. ~~AIC accurately~~
936 distinguished models with process errors on recruitment only (R) or on recruitment and
937 apparent survival (R+S). Accuracy for other process error types required a strong signal
938 (high process variability) with low noise (low observation uncertainty). Therefore, we expect
939 practitioners will find R+S configurations to provide satisfactory diagnostics across a range of
940 life history and data quality scenarios. AIC generally performed poorly for selecting the SRR,
941 but performance was improved with ~~across all simulations. The classification and regression~~
942 trees are generally a good tool for determining the OM and EM attributes that produce better

943 or worse measures of reliability. However, determining the combination of attributes that
944 produce the best or worst measures of reliability can be challenging using the trees alone. For
945 example, in the regression tree for median natural mortality rate estimates in R OM_s (Figure
946 7), both of the first branches imply bias is low regardless of OM fishing history, but when
947 OM fishing pressure is constant, results are much better when OM observation error is low
948 (median RE about -6%) than when OM observation error is high (median RE about 40%).
949 The default pruning of the trees can exclude these lower branches. However, inspection of
950 deviance explained by various regression models shows the ~9% reduction in residual deviance
951 by including second order interaction of all OM and EM factors (Table 6), indicating that the
952 interaction of factors may be important, thereby complimenting the regression tree analysis.
953 Higher order interactions of some factors could also provide reductions in deviance and,
954 therefore, inspection of results for each combinations of OM and EM factors, as provided in
955 the Supplementary Materials, can also be important.

956 Recommendations and conclusions

957 Our findings regarding model convergence suggests practitioners using state-space models
958 and maximum marginal likelihood for estimation should not heavily weight the magnitude
959 of the gradient values in determining convergence as long as the maximum absolute value
960 is around 1 or lower. Instead, positive-definiteness of the Hessian of the minimized negative
961 log-likelihood should be evaluated.

962 Unfortunately, whether the practitioner includes a Beverton-Holt SRR will often depend
963 on biological plausibility of this particular SRR because using AIC to determine its validity
964 required a combination of low recruitment variability and contrast in fishing pressure. Some
965 large variation in SSB over time, and lower observation error, which applies to a limited
966 number of managed stocks. Furthermore, some bias in estimation in at least one of the
967 SRR parameters existed in nearly all OM-EM combinations (and MSY-based reference points

968 should be expected. Because bias in terminal SSB and retrospective patterns were indifferent
969 to whether or not the SRR was estimated, ~~and convergence was slightly better~~ the prevalence
970 of bias in SRR parameter estimation, and often better convergence without the SRR, we
971 recommend a sensible default ~~would be to fit models without an assumed SRR is to exclude~~
972 an SRR when fitting assessment models, as also suggested by Brooks (2024).

973 We found marginal AIC can, in many cases, accurately distinguished models with process
974 errors. We saw the best accuracy for models with process errors on recruitment only
975 (R), recruitment and apparent survival (R+S), and recruitment and selectivity (R+Sel),
976 especially with lower observation error. However, AIC could also distinguish R+M and R+q
977 process errors when variability of those processes was greater. The R+S assumption for
978 process errors is common in applications of WHAM in the NEUS and the SAM assessment
979 framework (Nielsen and Berg 2014) in ICES, and we can have some confidence that
980 practitioners are correctly arriving at this assumption over other sources of process error
981 using marginal AIC.

982 Acknowledgements

983 This work was funded by NOAA Fisheries Northeast Fisheries Science Center. We thank
984 Jon Deroba, two anonymous reviewers, and the associate editor for helpful comments on ~~an~~
985 ~~earlier version~~ earlier versions of this manuscript ~~that markedly improved its clarity~~.

986 **References**

- 987 Aeberhard, W.H., Flemming, J.M., and Nielsen, A. 2018. Review of State-Space Models
988 for Fisheries Science. *Annual Review of Statistics and Its Application* **5**(1): 215–235.
989 doi:10.1146/annurev-statistics-031017-100427.
- 990 Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D.,
991 and Mills Flemming, J. 2016. State-space models' dirty little secrets: Even simple
992 linear Gaussian models can have estimation problems. *Scientific reports* **6**(1): 26677.
993 doi:10.1038/srep26677.
- 994 Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A.A., Leos-
995 Barajas, V., Mills Flemming, J., Nielsen, A., Petris, G., and others. 2021. A guide to
996 state-space modeling of ecological time series. *Ecological Monographs* **91**(4): e01470.
997 doi:10.1002/ecm.1470.
- 998 Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and regres-
999 sion trees. Chapman; Hall/CRC, New York, NY USA. doi:10.1201/9781315139470.
- 1000 Breivik, O.N., Aldrin, M., Fuglebakke, E., and Nielsen, A. 2023. Detecting significant retro-
1001 spective patterns in state space fish stock assessment. *Canadian Journal of Fisheries and*
1002 *Aquatic Sciences* **80**(9): 1509–1518. doi:10.1139/cjfas-2022-0250.
- 1003 Britten, G., Brooks, E.N., and Miller, T.J. In review. Identification and performance of
1004 environmentally-driven stock-recruitment relationships in state space assessment models.
1005 *Canadian Journal of Fisheries and Aquatic Sciences*.
- 1006 Brooks, E.N. 2024. Pragmatic approaches to modeling recruitment in fisheries stock assess-
1007 ment: A perspective. *Fisheries Research* **270**: 106896. doi:10.1016/j.fishres.2023.106896.
- 1008 Brooks, E.N., and Legault, C.M. 2016. Retrospective forecasting – evaluating performance
1009 of stock projections for New England groundfish stocks. *Canadian Journal of Fisheries*
1010 *and Aquatic Sciences* **73**(6): 935–950. doi:10.1139/cjfas-2015-0163.
- 1011 Cadigan, N.G. 2016. A state-space stock assessment model for northern cod, including under-

- 1012 reported catches and variable natural mortality rates. Canadian Journal of Fisheries and
1013 Aquatic Sciences **73**(2): 296–308. doi:10.1139/cjfas-2015-0047.
- 1014 Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M.,
1015 Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel, C.R., Doering,
1016 K., Johnson, K.F., and Methot, R.D. 2021. A cookbook for using model diagnostics in
1017 integrated stock assessments. Fisheries Research **240**: 105959. doi:<https://doi.org/10.1016/j.fishres.2021.105959>.
- 1018 Clopper, C.J., and Pearson, E.S. 1934. The use of confidence or fiducial limits illustrated in
1019 the case of the binomial. Biometrika **26**(4): 404–413. doi:10.1093/biomet/26.4.404.
- 1020 Collier, Z.K., Zhang, H., and Soyoye, O. 2022. Alternative methods for interpreting Monte
1021 Carlo experiments. Communications in Statistics - Simulation and Computation: 1–16.
1022 doi:10.1080/03610918.2022.2082474.
- 1023 Conn, P.B., Williams, E.H., and Shertzer, K.W. 2010. When can we reliably estimate the
1024 productivity of fish stocks? Canadian Journal of Fisheries and Aquatic Sciences **67**(3):
1025 511–523. doi:10.1139/F09-194.
- 1026 Cronin-Fine, L., and Punt, A.E. 2021. Modeling time-varying selectivity in size-structured
1027 assessment models. Fisheries Research **239**: 105927. Elsevier.
- 1028 de Valpine, P., and Hastings, A. 2002. Fitting population models incorporating process noise
1029 and observation error. Ecological Monographs **72**(1): 57–76.
- 1030 Fisch, N., Shertzer, K., Camp, E., Maunder, M., and Ahrens, R. 2023. Process and sampling
1031 variance within fisheries stock assessment models: Estimability, likelihood choice, and the
1032 consequences of incorrect specification. ICES Journal of Marine Science **80**(8): 2125–
1033 2149. doi:10.1093/icesjms/fsad138.
- 1034 Fleischman, S.J., Catalano, M.J., Clark, R.A., and Bernard, D.R. 2013. An age-structured
1035 state-space stock–recruit model for Pacific salmon (*Oncorhynchus spp.*). Canadian Jour-
1036 nal of Fisheries and Aquatic Sciences **70**(3): 401–414. doi:10.1139/cjfas-2012-0112.
- 1037 Gonzalez, O., O’Rourke, H.P., Wurpts, I.C., and Grimm, K.J. 2018. Analyzing Monte Carlo

- 1039 simulation studies with classification and regression trees. Structural Equation Modeling:
1040 A Multidisciplinary Journal **25**(3): 403–413. doi:10.1080/10705511.2017.1369353.
- 1041 Harwell, M., Kohli, N., and Peralta-Torres, Y. 2018. A survey of reporting practices of
1042 computer simulation studies in statistical research. The American Statistician **72**(4):
1043 321–327. doi:10.1080/00031305.2017.1342692.
- 1044 Hoenig, J.M., Hearn, W.S., Leigh, G.M., and Latour, R.J. 2025. Principles for estimating
1045 natural mortality rate. Fisheries Research **281**: 107195. doi:10.1016/j.fishres.2024.107195.
- 1046 Hoyle, S.D., Maunder, M.N., Punt, A.E., Mace, P.M., Devine, J.A., and A'mar, Z.T. 2022.
1047 Preface: Developing the next generation of stock assessment software. Fisheries Research
1048 **246**: 106176. doi:10.1016/j.fishres.2021.106176.
- 1049 Hurtado-Ferro, F., Szwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., John-
1050 son, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K.,
1051 Vert-Pre, K.A., Whitten, A.R., and Punt, A.E. 2014. Looking in the rear-view mirror:
1052 Bias and retrospective patterns in integrated, age-structured stock assessment models.
1053 ICES Journal of Marine Science **72**(1): 99–110. doi:10.1093/icesjms/fsu198.
- 1054 Huynh, Q.C., Legault, C.M., Hordyk, A.R., and Carruthers, T.R. 2022. A closed-loop
1055 simulation framework and indicator approach for evaluating impacts of retrospective
1056 patterns in stock assessments. ICES Journal of Marine Science **79**(7): 2003–2016.
1057 doi:10.1093/icesjms/fsac066.
- 1058 Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., and Punt, A.E.
1059 2016. Can autocorrelated recruitment be estimated using integrated assessment mod-
1060 els and how does it affect population forecasts? Fisheries Research **183**: 222–232.
1061 doi:10.1016/j.fishres.2016.06.004.
- 1062 Kapur, M.S., Ducharme-Barth, N., Oshima, M., and Carvalho, F. 2025. Good practices,
1063 trade-offs, and precautions for model diagnostics in integrated stock assessments. Fish-
1064 eries Research **281**: 107206. doi:10.1016/j.fishres.2024.107206.
- 1065 Kass, R.E., and Raftery, A.E. 1995. Bayes factors. Journal of the American Statistical

- 1066 Association **90**(430): 773–795. doi:10.1080/01621459.1995.10476572.
- 1067 Katz, R.W. 1981. On some criteria for estimating the order of a Markov chain. *Technometrics* **23**(3): 243–249. doi:10.1080/00401706.1981.10486293.
- 1068
- 1069 Knape, J. 2008. Estimability of density dependence in models of time series data. *Ecology* **89**(11): 2994–3000. doi:10.1890/08-0071.1.
- 1070
- 1071 Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software* **70**(5): 1–21. doi:10.18637/jss.v070.i05.
- 1072
- 1073
- 1074 Lee, H.-H., Maunder, M.N., Piner, K.R., and Methot, R.D. 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. *Fisheries Research* **109**(1): 89–94. doi:10.1016/j.fishres.2011.01.021.
- 1075
- 1076
- 1077
- 1078 Legault, C.M. 2009. Report of the retrospective working group, 14-16 january 2008. US Department of Commerce Northeast Fisheries Science Center Reference Document 09-01. US Department of Commerce Northeast Fisheries Science Center. Woods Hole, MA.
- 1079
- 1080
- 1081
- 1082 Legault, C.M., and Palmer, M.C. 2016. In what direction should the fishing mortality target change when natural mortality increases within an assessment? *Canadian Journal of Fisheries and Aquatic Sciences* **73**(3): 349–357. doi:10.1139/cjfas-2015-0232.
- 1083
- 1084
- 1085 Legault, C.M., and Restrepo, V.R. 1999. A flexible forward age-structured assessment program. *Col. Vol. Sci. Pap. ICCAT* **49**(2): 246–253.
- 1086
- 1087 Legault, C.M., Wiedenmann, J., Deroba, J.J., Fay, G., Miller, T.J., Brooks, E.N., Bell, R.J., Langan, J.A., Cournane, J.M., Jones, A.W., and Muffley, B. 2023. Data-rich but model-resistant: An evaluation of data-limited methods to manage fisheries with failed age-based stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* **80**(1): 27–42. doi:10.1139/cjfas-2022-0045.
- 1088
- 1089
- 1090
- 1091
- 1092 Li, C., Deroba, J.J., Berger, A.M., Goethel, D.R., Langseth, B.J., Schueler, A.M., and

- 1093 Miller, T.J. 2025a. Random effects on numbers-at-age transitions implicitly account for
1094 movement dynamics and improve performance within a state-space stock assessment.
1095 Canadian Journal of Fisheries and Aquatic Sciences **82**. doi:10.1139/cjfas-2025-0092.
- 1096 Li, C., Deroba, J.J., Miller, T.J., Legault, C.M., and Perretti, C. 2025b. Guidance on
1097 bias-correction of log-normal random effects and observations in state-space assessment
1098 models. Canadian Journal of Fisheries and Aquatic Sciences. doi:10.1139/cjfas-2025-
1099 0093.
- 1100 Li, C., Deroba, J.J., Miller, T.J., Legault, C.M., and Perretti, C.T. 2024. An evalua-
1101 tion of common stock assessment diagnostic tools for choosing among state-space
1102 models with multiple random effects processes. Fisheries Research **273**: 106968.
1103 doi:10.1016/j.fishres.2024.106968.
- 1104 Liljestrand, E.M., Bence, J.R., and Deroba, J.J. 2024. The effect of process variability and
1105 data quality on performance of a state-space stock assessment model. Fisheries Research
1106 **275**: 107023. doi:10.1016/j.fishres.2024.107023.
- 1107 MacKinnon, D.P., Warsi, G., and Dwyer, J.H. 1995. A simulation study of mediated effect
1108 measures. Multivariate Behavioral Research **30**(1): 41–62. doi:10.1207/s15327906mbr3001_3.
- 1109 Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish and
1110 Fisheries **8**(4): 337–358. doi:10.1111/j.1467-2979.2007.00258.x.
- 1111 Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical frame-
1112 work for fish stock assessment and fishery management. Fisheries Research **142**: 86–99.
1113 doi:10.1016/j.fishres.2012.10.012.
- 1114 Miller, T.J., and Brooks, E.N. 2021. Steepness is a slippery slope. Fish and Fisheries **22**(3):
1115 634–645. doi:10.1111/faf.12534.
- 1116 Miller, T.J., Curti, K.L., and Hansell, A.C. 2025. Space for WHAM: A multi-region, multi-
1117 stock generalization of the woods hole assessment model with an application to black sea
1118 bass. Canadian Journal of Fisheries and Aquatic Sciences **82**: 1–26. doi:10.1139/cjfas-
1119 2025-0097.

- 1120 Miller, T.J., Hare, J.A., and Alade, L. 2016. A state-space approach to incorporating envi-
1121 ronmental effects on recruitment in an age-structured assessment model with an appli-
1122 cation to Southern New England yellowtail flounder. Canadian Journal of Fisheries and
1123 Aquatic Sciences **73**(8): 1261–1270. doi:10.1139/cjfas-2015-0339.
- 1124 Miller, T.J., and Hyun, S.-Y. 2018. Evaluating evidence for alternative natural mortality
1125 and process error assumptions using a state-space, age-structured assessment model.
1126 Canadian Journal of Fisheries and Aquatic Sciences **75**(5): 691–703. doi:10.1139/cjfas-
1127 2017-0035.
- 1128 Miller, T.J., and Legault, C.M. 2017. Statistical behavior of retrospective patterns and
1129 their effects on estimation of stock and harvest status. Fisheries Research **186**: 109–120.
1130 doi:10.1016/j.fishres.2016.08.002.
- 1131 Miller, T.J., O'Brien, L., and Fratantoni, P.S. 2018. Temporal and environmental variation
1132 in growth and maturity and effects on management reference points of Georges Bank
1133 Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences **75**(12): 2159–2171.
1134 doi:10.1139/cjfas-2017-0124.
- 1135 Mohn, R. 1999. The retrospective problem in sequential population analysis: An investi-
1136 gation using cod fishery and simulated data. ICES Journal of Marine Science **56**(4):
1137 473–488. doi:10.1006/jmsc.1999.0481.
- 1138 NEFSC. 2022a. Final report of the haddock research track assessment working group. Avail-
1139 able at https://s3.us-east-1.amazonaws.com/nefmc.org/14b_EGB_Research_Track_Haddock_WG_I
- 1140 NEFSC. 2022b. Report of the American plaice research track working group. Available at
1141 https://s3.us-east-1.amazonaws.com/nefmc.org/2_American-Plaice-WG-Report.pdf.
- 1142 NEFSC. 2024. Butterfish research track assessment report. US Dept Commer Northeast
1143 Fish Sci Cent Ref Doc. 24-03; 191 p.
- 1144 NEFSC. 2025. Yellowtail flounder research track working group report. Available at
1145 <https://d23h0vhsm26o6d.cloudfront.net/10c.-Yellowtail-Flounder-RT-WG-Report.pdf>.
- 1146 Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments

- 1147 using state-space models. *Fisheries Research* **158**: 96–101. doi:10.1016/j.fishres.2014.01.014.
- 1148 Pedersen, M.W., and Berg, C.W. 2017. A stochastic surplus production model in continuous
1149 time. *Fish and Fisheries* **18**(2): 226–243. doi:10.1111/faf.12174.
- 1150 Perretti, C.T., Deroba, J.J., and Legault, C.M. 2020. Simulation testing methods for es-
1151 timating misreported catch in a state-space stock assessment model. *ICES Journal of*
1152 *Marine Science* **77**(3): 911–920. doi:10.1093/icesjms/fsaa034.
- 1153 Polansky, L., De Valpine, P., Lloyd-Smith, J.O., and Getz, W.M. 2009. Likelihood ridges
1154 and multimodality in population growth rate models. *Ecology* **90**(8): 2313–2320.
1155 doi:10.1890/08-1461.1.
- 1156 Pontavice, H. du, Miller, T.J., Stock, B.C., Chen, Z., and Saba, V.S. 2022. Ocean model-
1157 based covariates improve a marine fish stock assessment when observations are limited.
1158 *ICES Journal of Marine Science* **79**(4): 1259–1273. doi:10.1093/icesjms/fsac050.
- 1159 Punt, A.E. 2023. Those who fail to learn from history are condemned to repeat it: A per-
1160 spective on current stock assessment good practices and the consequences of not following
1161 them. *Fisheries Research* **261**: 106642. doi:10.1016/j.fishres.2023.106642.
- 1162 Punt, A.E., Hurtado-Ferro, F., and Whitten, A.R. 2014. Model selection for selectivity in
1163 fisheries stock assessments. *Fisheries Research* **158**: 124–134. doi:10.1016/j.fishres.2013.06.003.
- 1164 Rudd, M.B., and Thorson, J.T. 2018. Accounting for variable recruitment and fishing mor-
1165 tality in length-based stock assessments for data-limited fisheries. *Canadian Journal of*
1166 *Fisheries and Aquatic Sciences* **75**(7): 1019–1035. doi:10.1139/cjfas-2017-0143.
- 1167 Shibata, R. 1976. Selection of the order of an autoregressive model by Akaike's information
1168 criterion. *Biometrika* **63**(1): 117–126. doi:10.1093/biomet/63.1.117.
- 1169 Stewart, I.J., and Monnahan, C.C. 2017. Implications of process error in selectivity for
1170 approaches to weighting compositional data in fisheries stock assessments. *Fisheries*
1171 *Research* **192**: 126–134. doi:10.1016/j.fishres.2016.06.018.
- 1172 Stock, B.C., and Miller, T.J. 2021. The Woods Hole Assessment Model (WHAM): A general
1173 state-space assessment framework that incorporates time- and age-varying processes via

- random effects and links to environmental covariates. *Fisheries Research* **240**: 105967. doi:10.1016/j.fishres.2021.105967.
- Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., and Nye, J.A. 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. *Fisheries Research* **237**: 105873. doi:10.1016/j.fishres.2021.105873.
- Szuwalski, C.S., Ianelli, J.N., and Punt, A.E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. *ICES Journal of Marine Science* **75**(2): 596–609. doi:10.1093/icesjms/fsx159.
- Therneau, T., and Atkinson, B. 2025. Rpart: Recursive Partitioning and Regression Trees. Available from <https://github.com/bethatkinson/rpart>.
- Thompson, W.R. 1936. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Annals of Mathematical Statistics* **7**(3): 122–128. doi:10.1214/aoms/1177732502.
- Thorson, J.T., and Minto, C. 2015. Mixed effects: A unifying framework for statistical modelling in fisheries biology. *ICES Journal of Marine Science* **72**(5): 1245–1256. doi:10.1093/icesjms/fsu213.
- Thulin, M. 2014. The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics* **8**(1): 817–840. doi:10.1214/14-EJS909.
- Trijoulet, V., Fay, G., and Miller, T.J. 2020. Performance of a state-space multispecies model: What are the consequences of ignoring predation and process errors in stock assessments? *Journal of Applied Ecology* **57**(1): 121–135. doi:10.1111/1365-2664.13515.
- Wang, S., Cadigan, N.G., and Benoît, H.P. 2017. Inference about regression parameters using highly stratified survey count data with over-dispersion and repeated measurements. *Journal of Applied Statistics* **44**(6): 1013–1030. doi:10.1080/02664763.2016.1191622.
- Wiedenmann, J., Free, C.M., and Jensen, O.P. 2019. Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks:

- 1201 A case study using northeast U.S. Fish stocks. *Fisheries Research* **209**(1): 129–142.
- 1202 doi:10.1016/j.fishres.2018.09.018.
- 1203 Xu, H., Thorson, J.T., Methot, R.D., and Taylor, I.G. 2019. A new semi-parametric method
1204 for autocorrelated age- and time-varying selectivity in age-structured assessment models.
1205 Canadian Journal of Fisheries and Aquatic Sciences **76**(2): 268–285. doi:10.1139/cjfas-
1206 2017-0446.
- 1207 Yee, T.W. 2008. The VGAM package. *R News* **8**(2): 28–39. Available from <https://journal.r-project.org/articles/RN-2008-014/>.
- 1208
- 1209 Yee, T.W. 2015. Vector generalized linear and additive models: With an implementation in
1210 R. Springer, New York, NY USA. doi:10.1007/978-1-4939-2818-7.

Fig. 1. ~~Estimated probability of fits~~ Classification trees indicating primary factors determining convergence as defined by providing ~~hessian-based~~ Hessian-based standard errors for EMs assuming alternative process error (colored points and lines) R, and median natural mortality (estimated or known) and Beverton-Holt stock-recruit relationships (estimated or not; along x-axis) when fitted to operating models that have Rand R+S(A), R+Sel (B)M, R+M (C), or Sel and R+q OMs. Nodes denote percent convergence (D_{top}) process error structures. Circled values indicate results where the EM process error structure matches that and number of fits (bottom) for the operating model and vertical lines represent 95% confidence intervals corresponding subset.

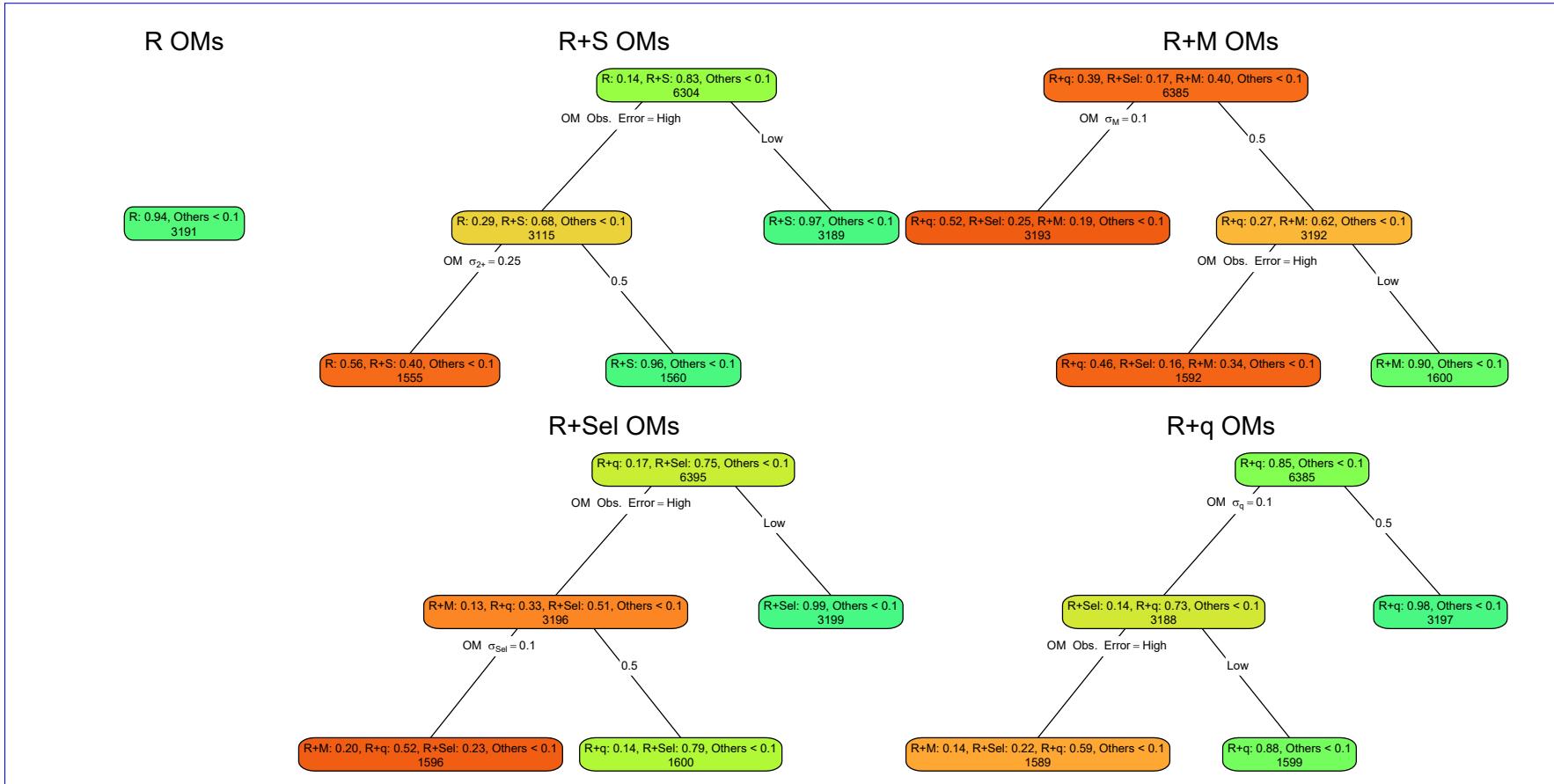


Fig. 2. Classification trees indicating primary factors determining which EM process error assumption provides the lowest AIC for R+S, R+M, R+Sel and R+q OM. Each node shows the proportion of EM process error models with lowest AIC (top) and number of observations (bottom) for the corresponding subset. Lower or higher accuracy of the process error assumption are indicated by more red or green polygons, respectively.

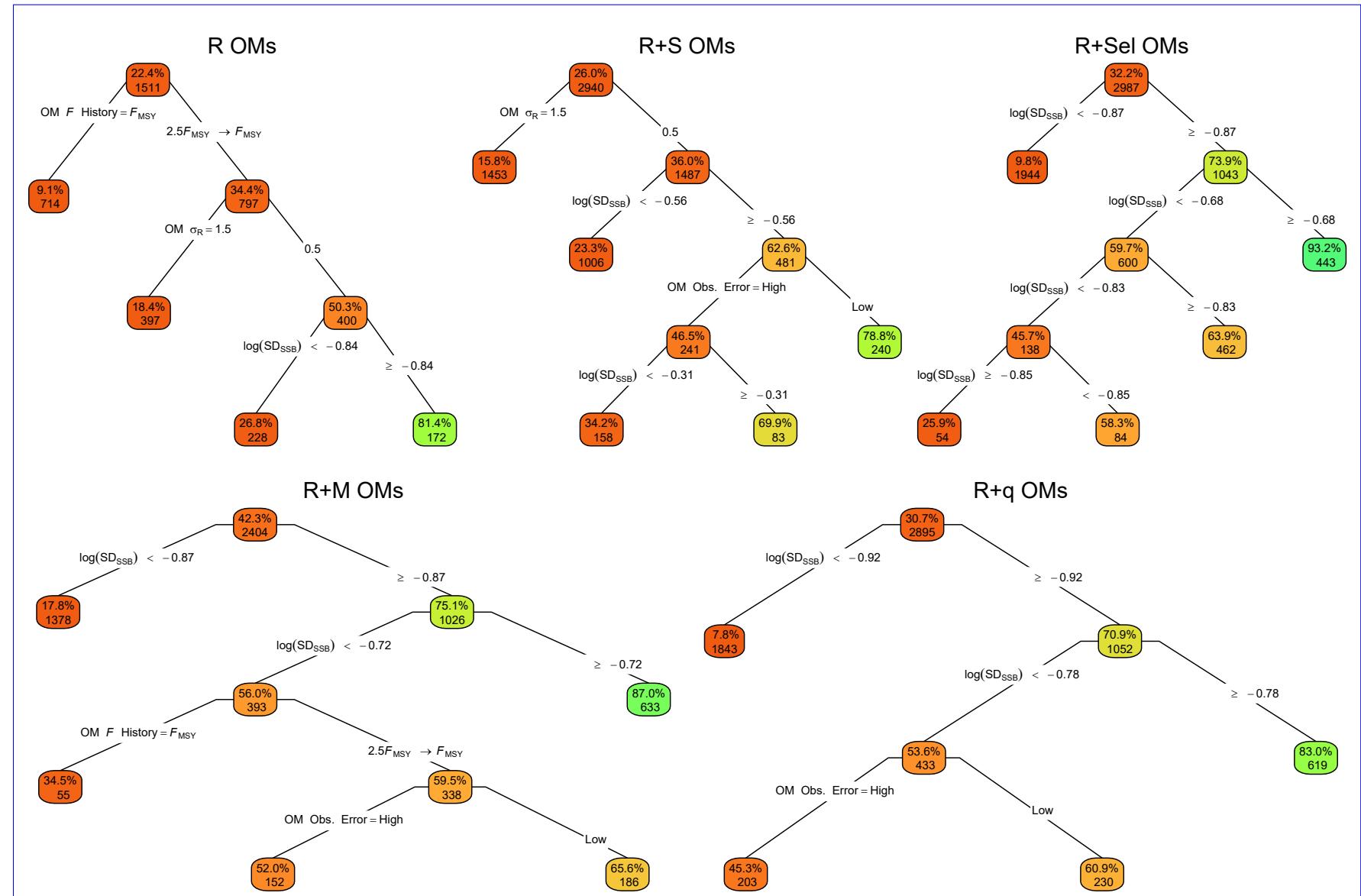


Fig. 3. Estimated probability of lowest AIC for EMs assuming alternative process error structures. Classification trees indicating primary factors determining which EM SRR assumption (colored bars) conditional on alternative assumptions for median natural mortality (estimated none or known) and Beverton-Holt stock-recruit relationships (estimated or not; along x-axis) when fitted to operating models that provide the lowest AIC for Rand (A), R+S (B), R+Sel (C), R+M (D), or Sel and R+q (E) process error structures OM. Striped bars indicate results where All EMs assume the EM correct process error structure matches source. Nodes denote the percentage of EMs that assume the SRR with lowest AIC (top) and number of observations

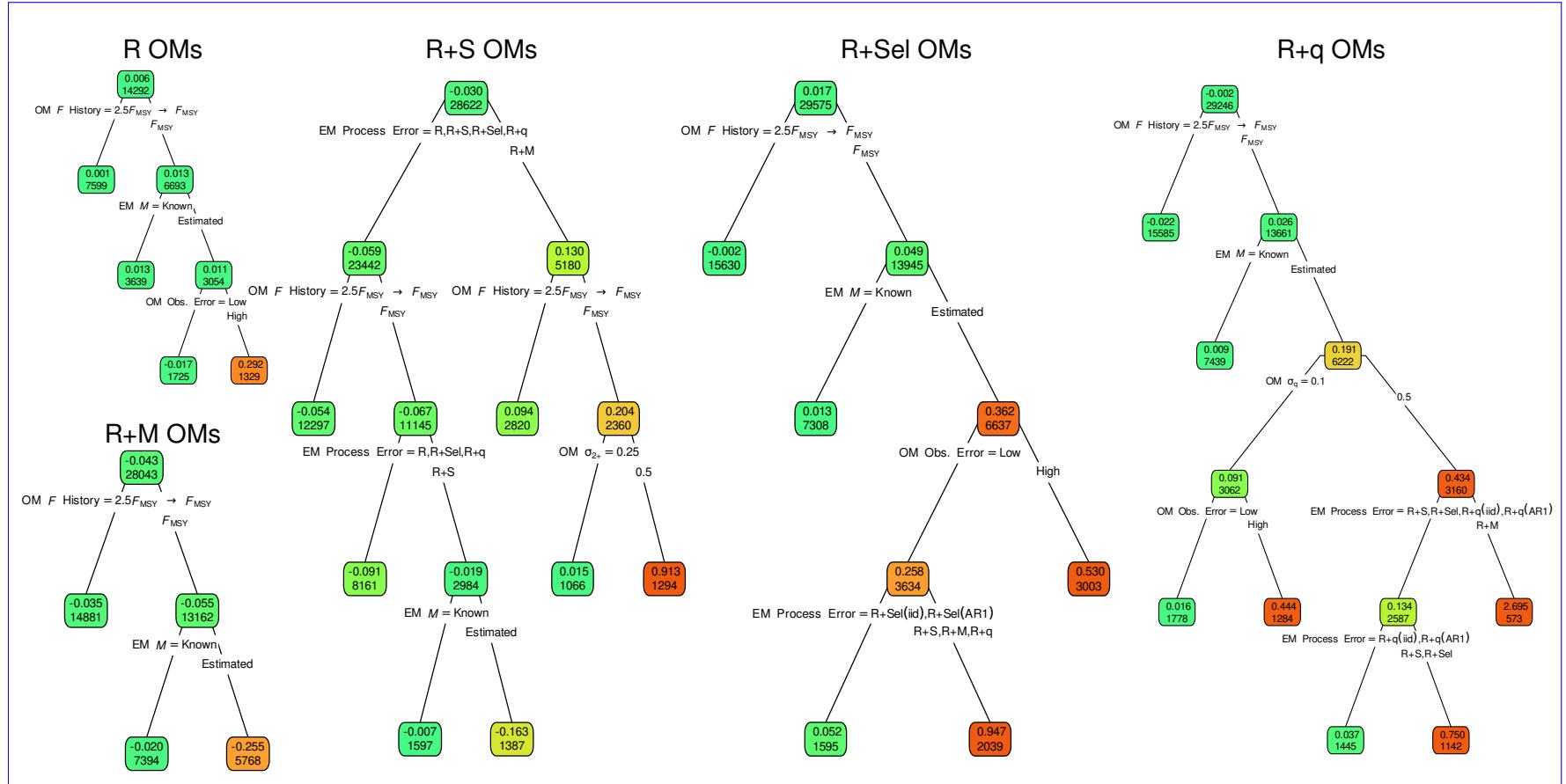


Fig. 4. Regression trees indicating primary factors determining reductions in sums of squares of errors in estimation measured by Eq. 3 for terminal year SSB for R+S, R+M, R+Sel and R+q OM. Each node shows the median error (top) and number of observations (bottom) for the corresponding subset. Median errors closer to or further from zero are indicated by more green or red polygons, respectively.

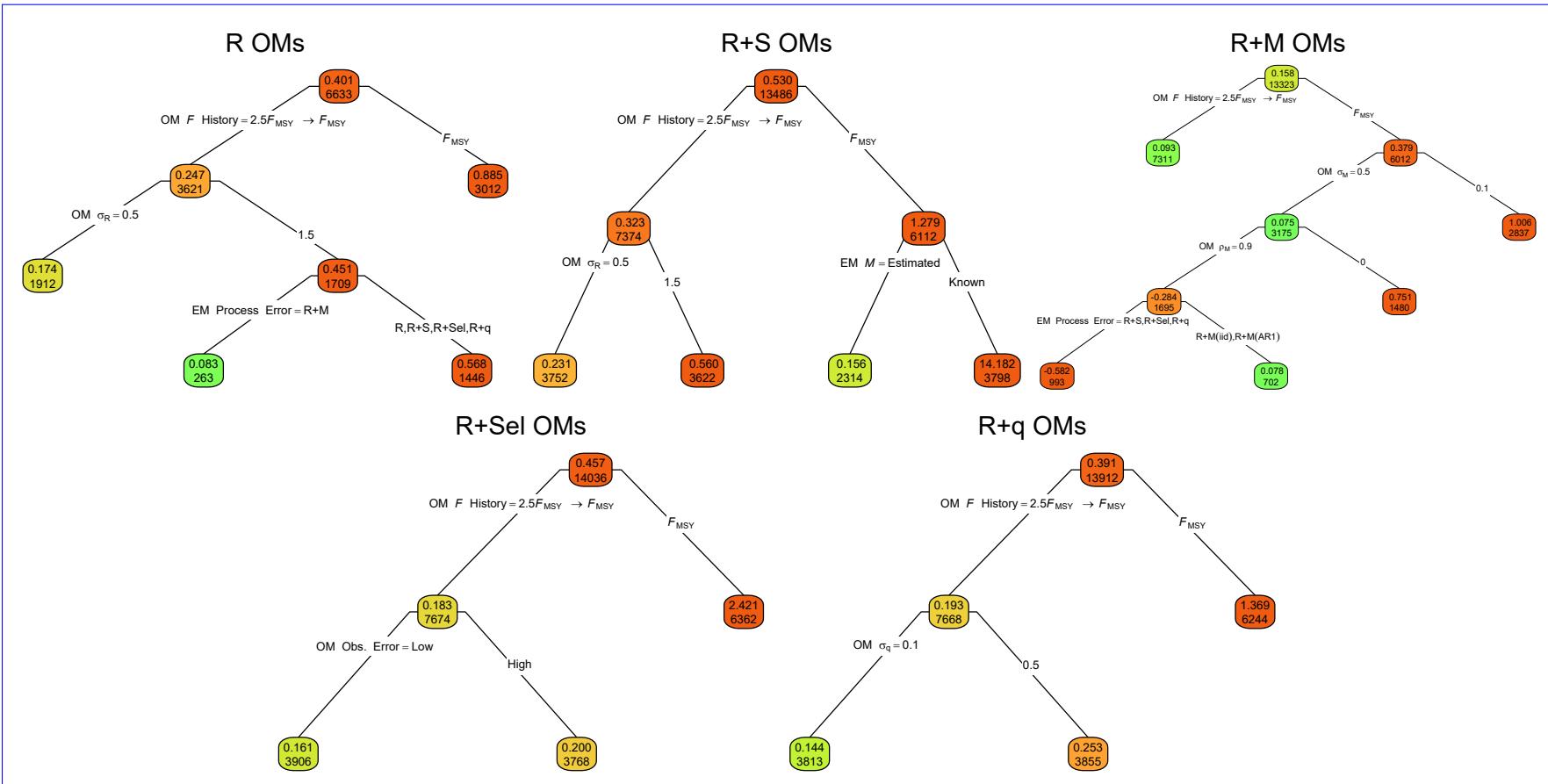


Fig. 5. Estimated probability Regression trees indicating primary factors determining reductions in sums of lowest AIC from logistic regression on the log-standard deviation squares of the true log(SSB) errors in each simulation estimation measured by Eq. 3 for estimating model with Beverton-Holt stock-recruit relationships, rather than the otherwise equivalent EM without the stock-recruit relationship. Results are conditional on median M is known in the EM and alternative assumptions EMs having the correct process error structure: Beverton-Holt SRR parameter a for R and R+S(A), R+Sel(B), R+M(C), or Sel and R+q OMs. Each node shows the median error (D_{top}), and median M is assumed known in the EM. Solid and dashed lines are for OMs with and without temporal contrast in fishing pressure, respectively, and polygons represent 95% confidence intervals. Range number of results indicates the range of log-standard deviation of logobservations (SSBbottom) for simulations the corresponding subset. Lower or higher median absolute errors of the particular OM process error assumption are indicated by more green or red polygons, respectively.

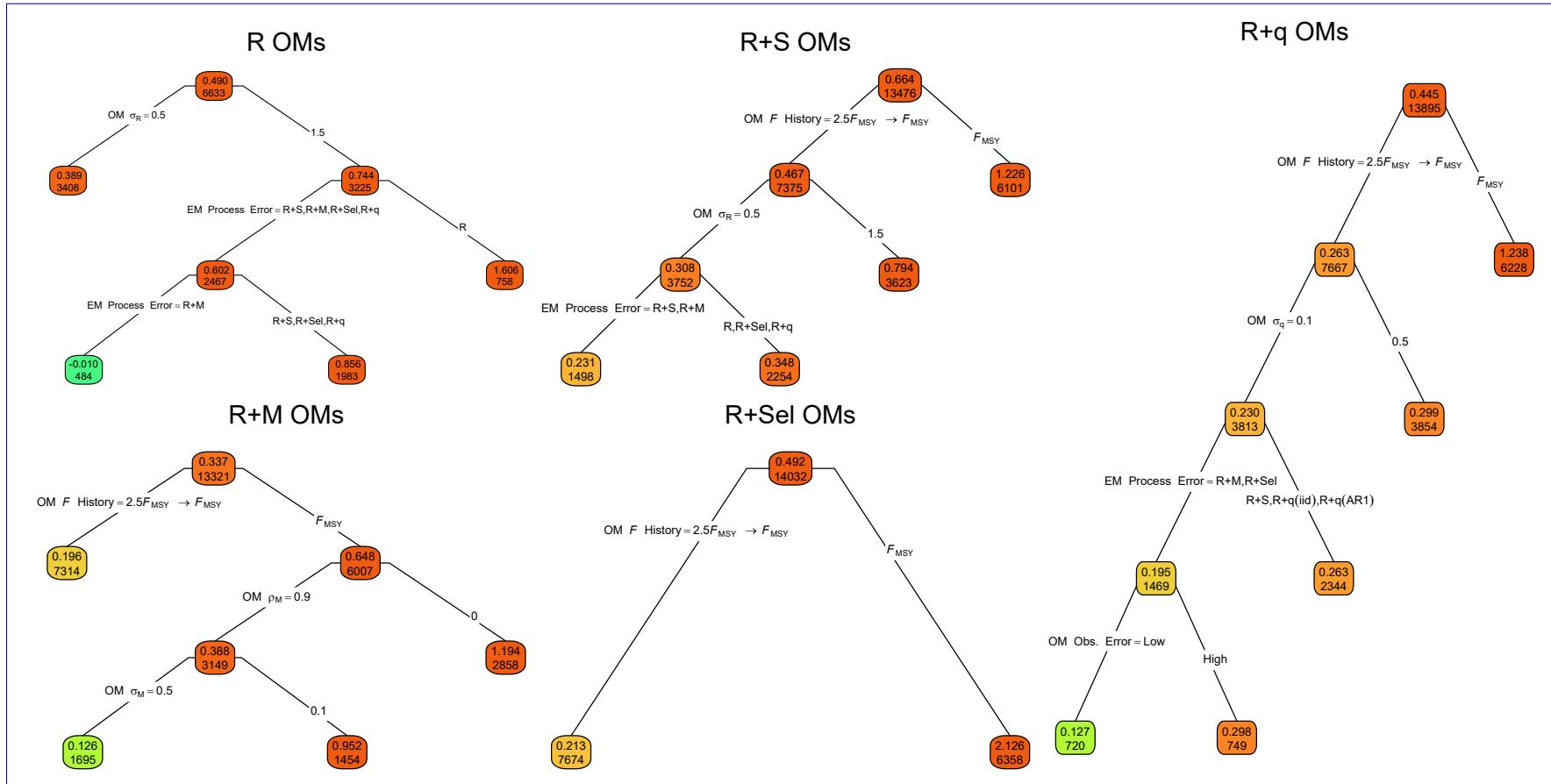


Fig. 6. **Median relative error** Regression trees indicating primary factors determining reductions in sums of terminal year SSB squares of errors in estimation measured by Eq. 3 for estimating models fitted to data sets simulated with alternative process error structures: the Beverton-Holt SRR parameter b for R and R+S(A), R+Sel(B), M, R+M(C), or Sel and R+q OMs. Each node shows the median error (D_{top}) and number of observations (bottom) for the corresponding subset. Circled values indicate results where Lower or higher median absolute errors of the EM process error structure matches that of the operating model and vertical lines represent 95% confidence intervals assumption are indicated by more green or red polygons, respectively.

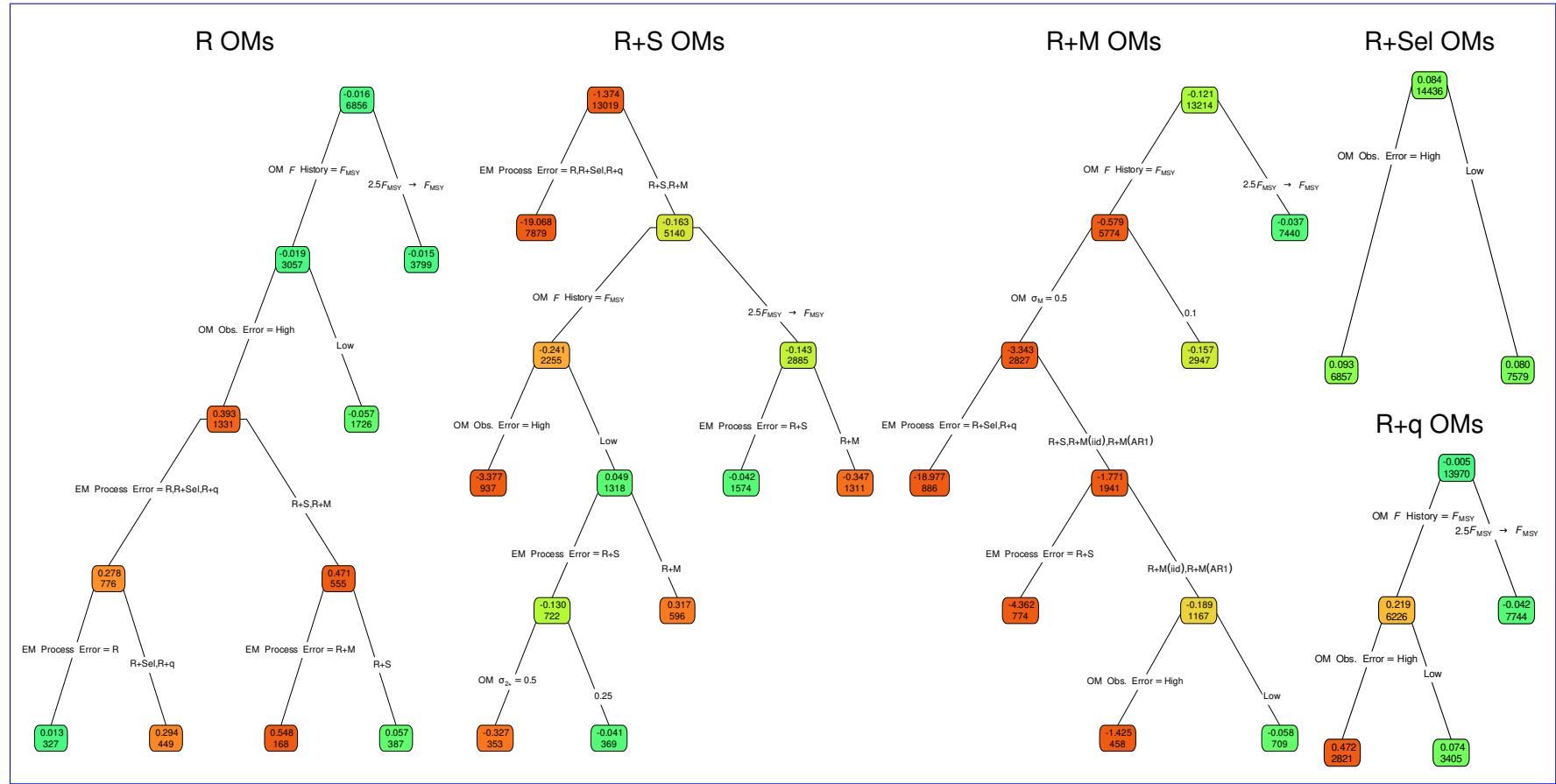


Fig. 7. ~~Median Mohn's rho~~ Regression trees indicating primary factors determining reductions in sums of squares of errors in estimation measured by Eq. 3 for SSB the median natural mortality rate for estimating models fitted to data sets simulated with alternative process error structures: Rand R+S(A), R+Sel (B)M, R+M (C), or Sel and R+q OMs. Each node shows the median error (D_{top}) and number of observations (bottom) for the corresponding subset. ~~Circled values indicate results where Lower or higher median absolute errors of the EM process error structure matches that of the operating model and vertical lines represent 95% confidence intervals assumption are indicated by more green or red polygons, respectively.~~

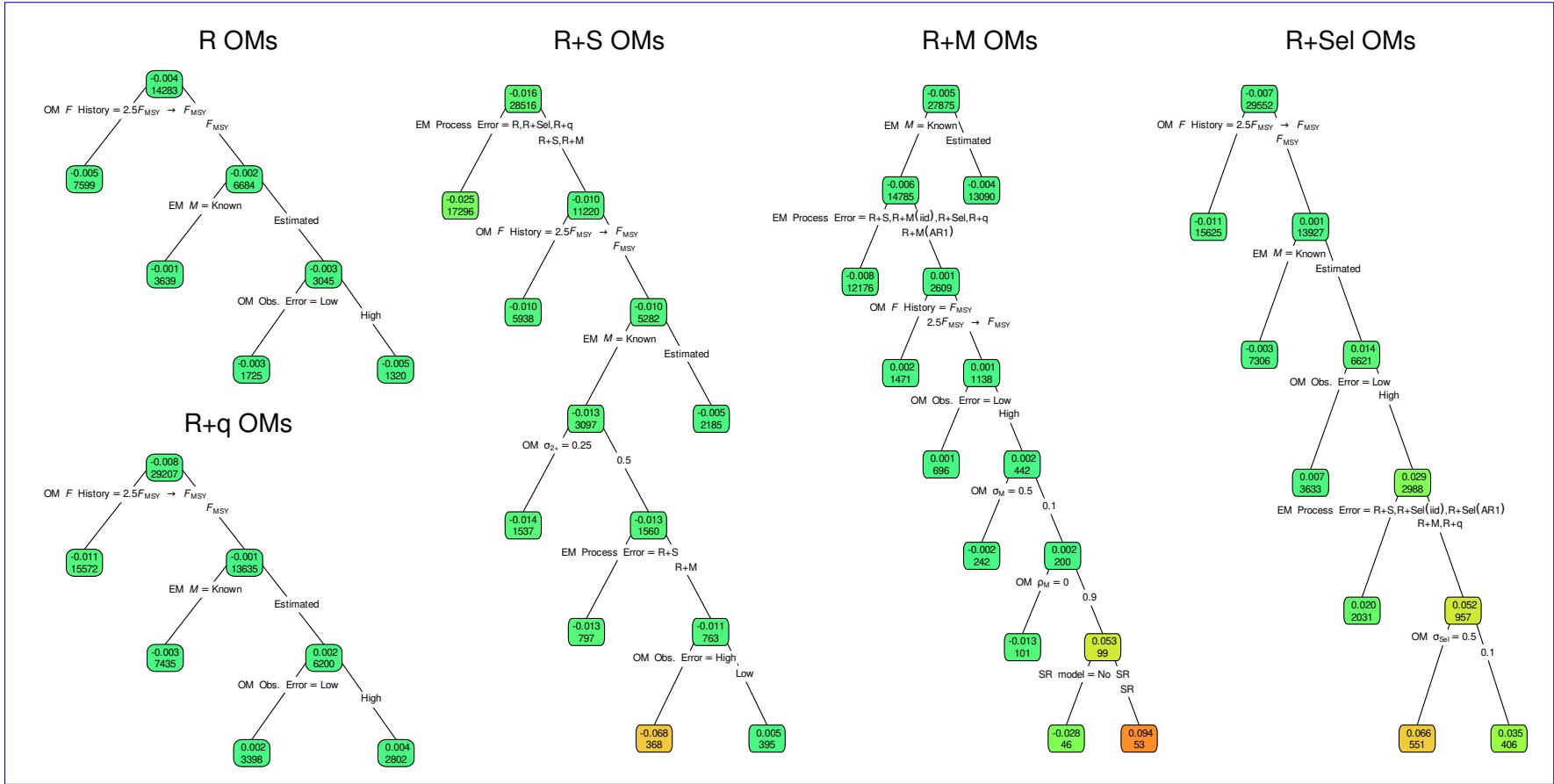


Fig. 8. Regression trees indicating primary factors determining reductions in sums of squares of errors in transformed Mohn's ρ (Eq. 3) for SSB for R+S, R+M, R+Sel and R+q OMs. Each node shows the median Mohn's ρ (top) and number of observations (bottom) for the corresponding subset. Median Mohn's ρ closer to or further from zero are indicated by more green or red polygons, respectively.

Table 1. For each OM process error source (columns), percent reduction in deviance for logistic regression models fit to indicators of convergence (providing Hessian-based standard errors) with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM Process Error	27.95	4.58	14.68	17.24	24.66
EM M Assumption	1.07	11.43	2.45	0.56	1.46
EM SR Assumption	2.88	3.30	1.24	2.47	1.59
OM Obs. Error	0.75	4.64	2.06	4.54	1.60
OM F History	2.32	3.37	1.63	3.30	2.59
OM σ_R	0.10	0.02	—	—	—
OM σ_{2+}	—	0.40	—	—	—
OM σ_M	—	—	0.22	—	—
OM ρ_M	—	—	0.17	—	—
OM σ_{Sel}	—	—	—	1.81	—
OM ρ_{Sel}	—	—	—	0.02	—
OM σ_q	—	—	—	—	0.34
OM ρ_q	—	—	—	—	<0.01
All factors	39.54	31.46	24.85	34.83	36.31
± All Two Way	45.03	39.89	35.20	42.81	43.70
± All Three Way	47.02	44.57	37.88	45.51	46.87

Table 2. For each OM process error source (columns), percent reduction in deviance for multinomial logistic regression models fit to indicators of EM process error assumption with lowest AIC with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	5.52	1.05	0.52	0.61	1.32
EM SR Assumption	5.60	0.75	1.13	0.93	1.95
OM Obs. Error	2.96	22.46	3.42	25.67	5.03
OM F History	5.77	0.62	0.94	0.91	2.05
OM σ_R	0.10	0.66	—	—	—
OM σ_{2+}	—	16.86	—	—	—
OM σ_M	—	—	9.06	—	—
OM ρ_M	—	—	0.38	—	—
OM σ_{Sel}	—	—	—	7.59	—
OM ρ_{Sel}	—	—	—	0.60	—
OM σ_q	—	—	—	—	13.50
OM ρ_q	—	—	—	—	0.75
All factors	20.98	46.12	16.58	40.83	25.99
± All Two Way	22.02	48.94	21.63	44.08	30.17
± All Three Way	22.05	49.98	22.36	44.54	31.38

Table 3. For each OM process error source (columns), percent reduction in deviance for logistic regression models fit to indicators of EM SRR assumption (none or Beverton-Holt) with lowest AIC with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	0.04	0.21	0.18	0.02	0.01
OM Obs. Error	<0.01	0.65	0.14	0.04	0.02
OM F History	9.17	3.79	13.08	26.56	24.60
OM σ_R	3.54	4.74	—	—	—
OM σ_{2+}	—	0.14	—	—	—
OM σ_M	—	—	1.14	—	—
OM ρ_M	—	—	0.05	—	—
OM σ_{Sel}	—	—	—	0.02	—
OM ρ_{Sel}	—	—	—	0.17	—
OM σ_q	—	—	—	—	0.36
OM ρ_q	—	—	—	—	0.02
log(SD _{SSB})	4.11	1.59	33.39	41.36	39.23
All factors	31.52	18.99	34.23	43.77	42.31
± All Two Way	34.79	22.24	35.99	45.84	44.04
± All Three Way	35.41	23.09	37.57	46.39	44.63

Table 4. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. 3 for the terminal year SSB with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	2.28	1.15	1.04	2.92	3.26
EM SR assumption	0.10	0.06	0.08	0.06	0.08
EM Process Error	0.43	4.28	0.40	0.11	1.05
OM Obs. Error	1.63	0.07	0.78	0.32	<0.01
OM F History	2.62	3.15	1.28	3.22	4.72
OM σ_R	0.03	0.01	—	—	—
OM σ_{2+}	—	0.93	—	—	—
OM σ_M	—	—	0.18	—	—
OM ρ_M	—	—	0.01	—	—
OM σ_{Sel}	—	—	—	0.16	—
OM ρ_{Sel}	—	—	—	0.04	—
OM σ_q	—	—	—	—	1.02
OM ρ_q	—	—	—	—	0.06
All factors	7.59	9.86	3.93	7.04	10.64
± All Two Way	17.99	25.56	10.06	13.44	22.43
± All Three Way	23.39	36.74	13.76	16.55	31.11

Table 5. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. 3 for the Beverton-Holt SRR parameters with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	Beverton-Holt <i>a</i>					Beverton-Holt <i>b</i>				
	R	R+S	R+M	R+Sel	R+q	R	R+S	R+M	R+Sel	R+q
EM <i>M</i> Assumption	0.02	1.05	0.02	0.11	0.02	0.05	1.06	0.03	0.01	0.40
EM Process Error	2.74	0.18	0.20	1.25	1.90	2.29	1.21	0.12	1.40	3.06
OM Obs. Error	0.16	<0.01	0.01	0.04	<0.01	<0.01	0.01	0.05	0.01	0.01
OM <i>F</i> History	3.15	3.34	5.60	11.37	10.00	1.16	1.17	2.01	7.97	3.87
OM σ_R	2.31	0.74	—	—	—	1.67	0.52	—	—	—
OM σ_{2+}	—	0.29	—	—	—	—	0.01	—	—	—
OM σ_M	—	—	0.30	—	—	—	—	0.13	—	—
OM ρ_M	—	—	0.51	—	—	—	—	0.22	—	—
OM σ_{Sel}	—	—	—	0.13	—	—	—	—	0.05	—
OM ρ_{Sel}	—	—	—	0.07	—	—	—	—	0.04	—
OM σ_q	—	—	—	—	0.04	—	—	—	—	0.10
OM ρ_q	—	—	—	—	<0.01	—	—	—	—	<0.01
All factors	8.07	5.15	6.73	12.64	11.79	4.91	3.75	2.55	9.12	7.22
± All Two Way	9.96	7.37	9.76	13.59	13.65	7.55	7.15	4.32	10.08	12.16
± All Three Way	11.22	8.15	11.13	14.48	14.87	9.78	9.02	5.26	11.08	14.73

Table 6. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. 3 for the median natural mortality rate parameter with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM SR assumption	0.21	0.38	0.11	0.26	0.43
EM Process Error	1.98	20.36	3.16	0.94	1.31
OM Obs. Error	4.74	0.79	0.40	2.23	1.88
OM F History	5.07	15.11	10.65	0.24	2.38
OM σ_R	≤ 0.01	0.01	—	—	—
OM σ_{2+}	—	5.04	—	—	—
OM σ_M	—	—	5.32	—	—
OM ρ_M	—	—	0.85	—	—
OM σ_{Sel}	—	—	—	1.30	—
OM ρ_{Sel}	—	—	—	0.37	—
OM σ_q	—	—	—	—	0.46
OM ρ_q	—	—	—	—	0.06
All factors	12.64	40.10	21.29	5.54	6.52
± All Two Way	21.17	48.12	36.19	9.87	11.71
± All Three Way	23.03	50.38	42.82	11.58	14.64

Table 7. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to transformed Mohn's ρ values for each simulation (Eq. 3) for SSB with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	0.79	0.18	0.15	0.95	1.24
EM SR assumption	<0.01	0.01	<0.01	<0.01	<0.01
EM Process Error	<0.01	0.22	0.14	0.08	0.04
OM Obs. Error	0.12	0.03	0.05	0.18	0.21
OM F History	0.84	0.14	0.07	1.08	1.56
OM σ_R	0.01	0.01	—	—	—
OM σ_{2+}	—	0.02	—	—	—
OM σ_M	—	—	0.01	—	—
OM ρ_M	—	—	<0.01	—	—
OM σ_{Sel}	—	—	—	0.01	—
OM ρ_{Sel}	—	—	—	0.02	—
OM σ_q	—	—	—	—	0.01
OM ρ_q	—	—	—	—	0.01
All factors	1.89	0.63	0.43	2.43	3.29
± All Two Way	3.63	1.10	0.91	4.75	6.22
± All Three Way	4.27	1.65	1.50	5.73	7.53

1211 **Supplementary Materials**

1212 Referenced Figures

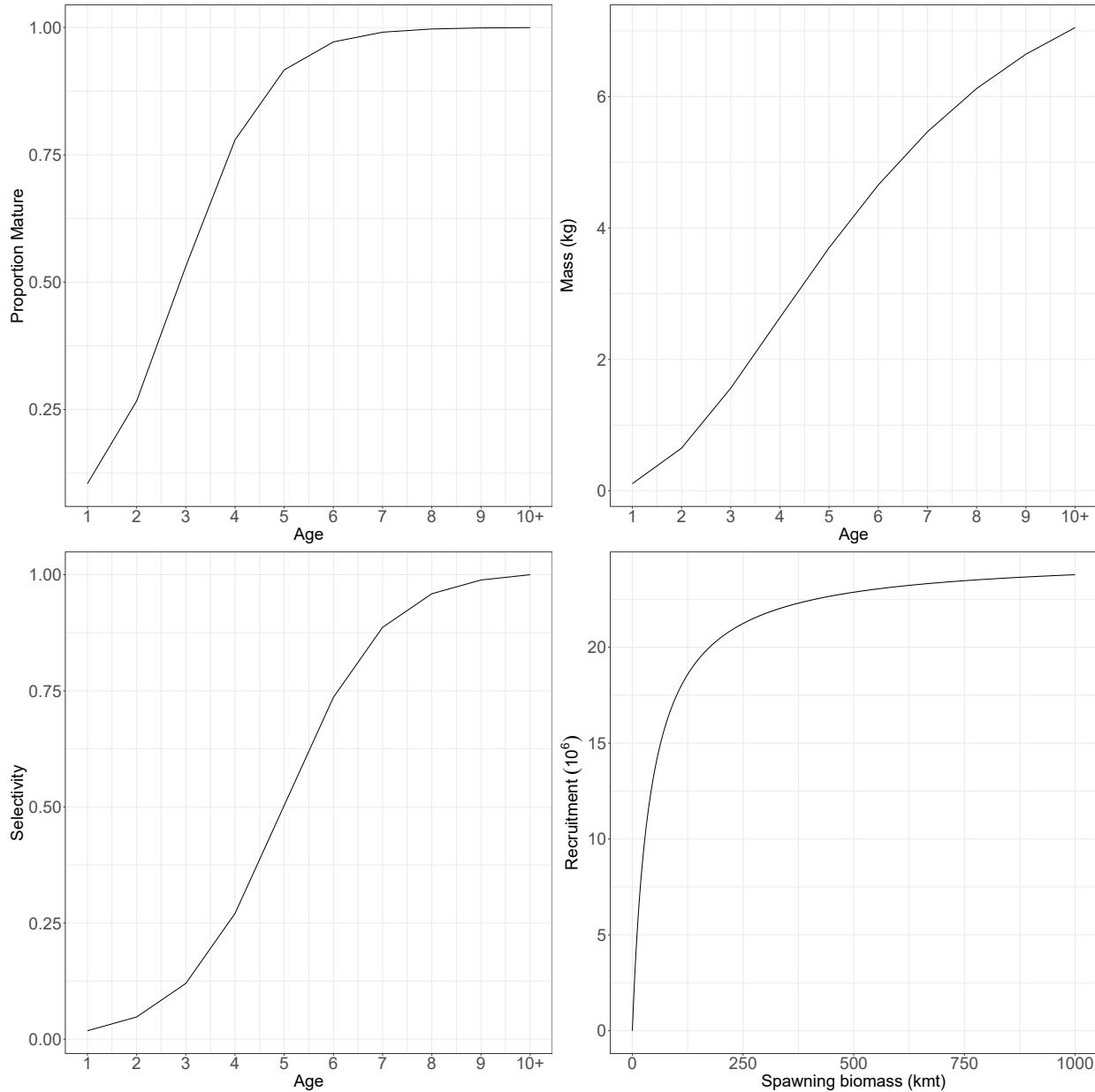


Fig. S1. The proportion mature at age, weight at age, fleet and index selectivity at age, and Beverton-Holt stock-recruit relationship-SRR assumed for the population in all operating modelsOMs. For operating models-OMs with random effects on fleet selectivity, this represents the selectivity at the mean of the random effects.

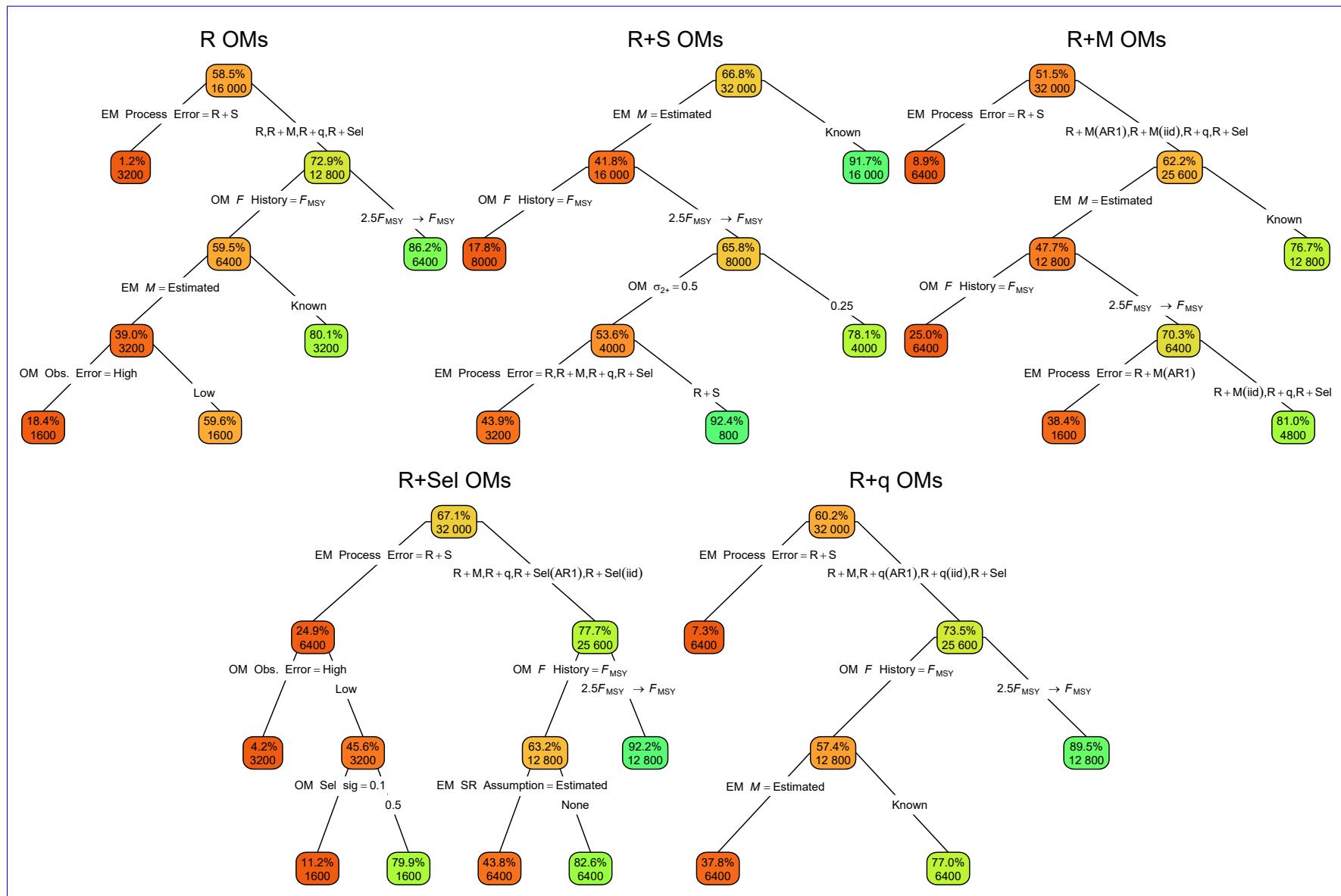


Fig. S2. Classification trees indicating primary factors determining convergence as defined by a maximum absolute gradient $< 10^{-6}$ for R, R+S, R+M, R+Sel and R+q OMs. Nodes denote percent convergence (top) and number of fits (bottom) for the corresponding subset. Lower or higher convergence rates are indicated by more red or green polygons, respectively

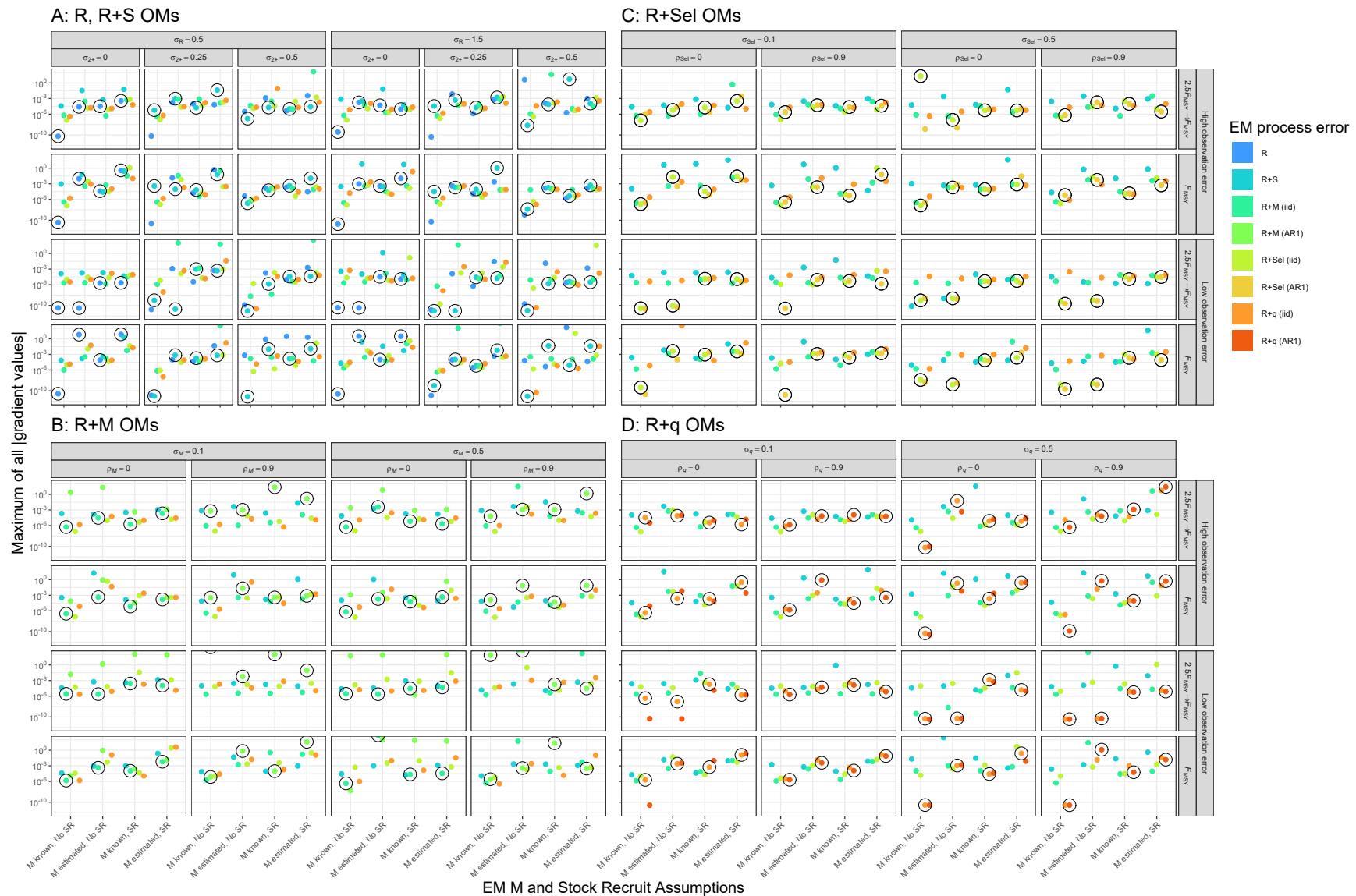


Fig. S3. The maximum of the absolute values of all gradient values for all fits that provided Hessian-based standard errors across all simulated data sets of a given OM configuration (A: R and R+S, B: R+M, C: R+Sel, or D: R+q). Results are conditional on EM fits with alternative process error assumptions (colored points and lines), median natural mortality (estimated or known) and recruitment assumptions (Beverton-Holt SRR or not). Circled values indicate results where the EM process error structure matches that of the OM and vertical lines represent 95% confidence intervals.

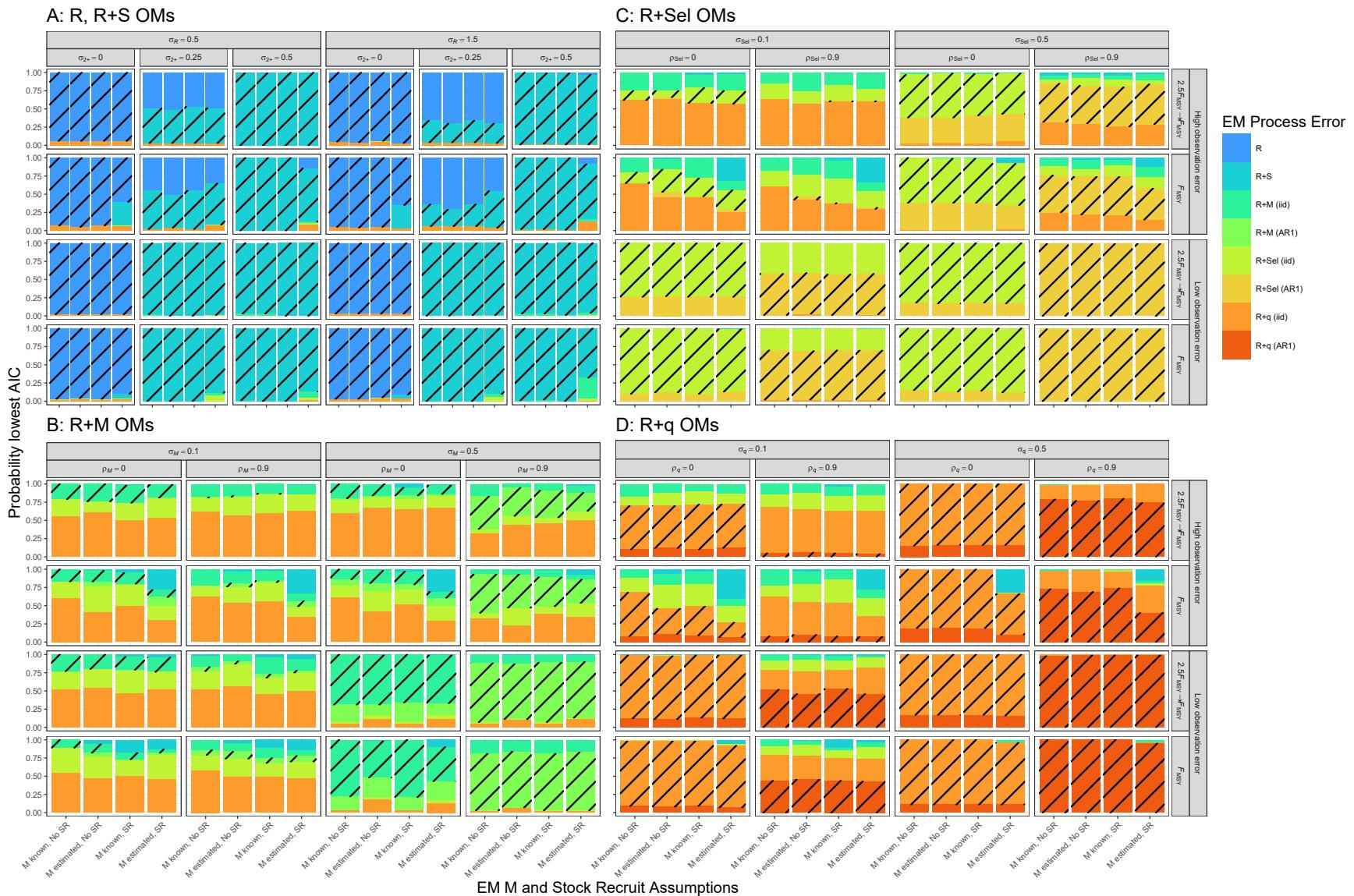


Fig. S4. Estimated probability of lowest AIC for EMs assuming alternative process error assumptions (colored bars) conditional on alternative assumptions for median natural mortality (estimated or known) and Beverton-Holt SRR (estimated or not; along x-axis) when fitted to OMs that have R and R+S (A), R+Sel (B), R+M (C), or R+q (D) process error sources. Striped bars indicate results where the EM process error structure matches that of the OM.

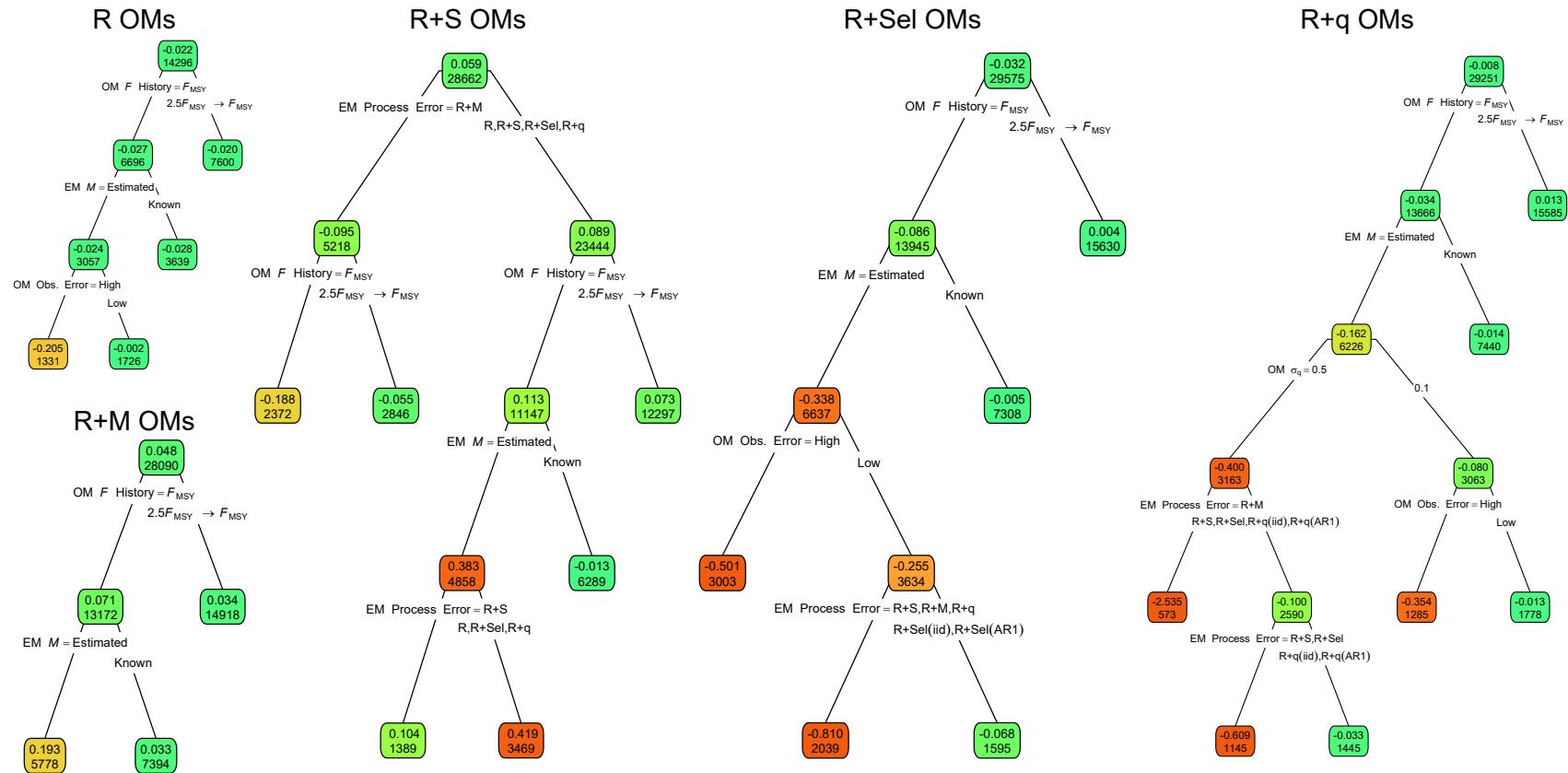


Fig. S5. Regression trees indicating primary factors determining reductions in sums of squares of errors in estimation measured by Eq. 3 for terminal year fully-selected fishing mortality for R+S, R+M, R+Sel and R+q OMs. Each node shows the median error (top) and number of observations (bottom) for the corresponding subset. Median errors closer to or further from zero are indicated by more green or red polygons, respectively.

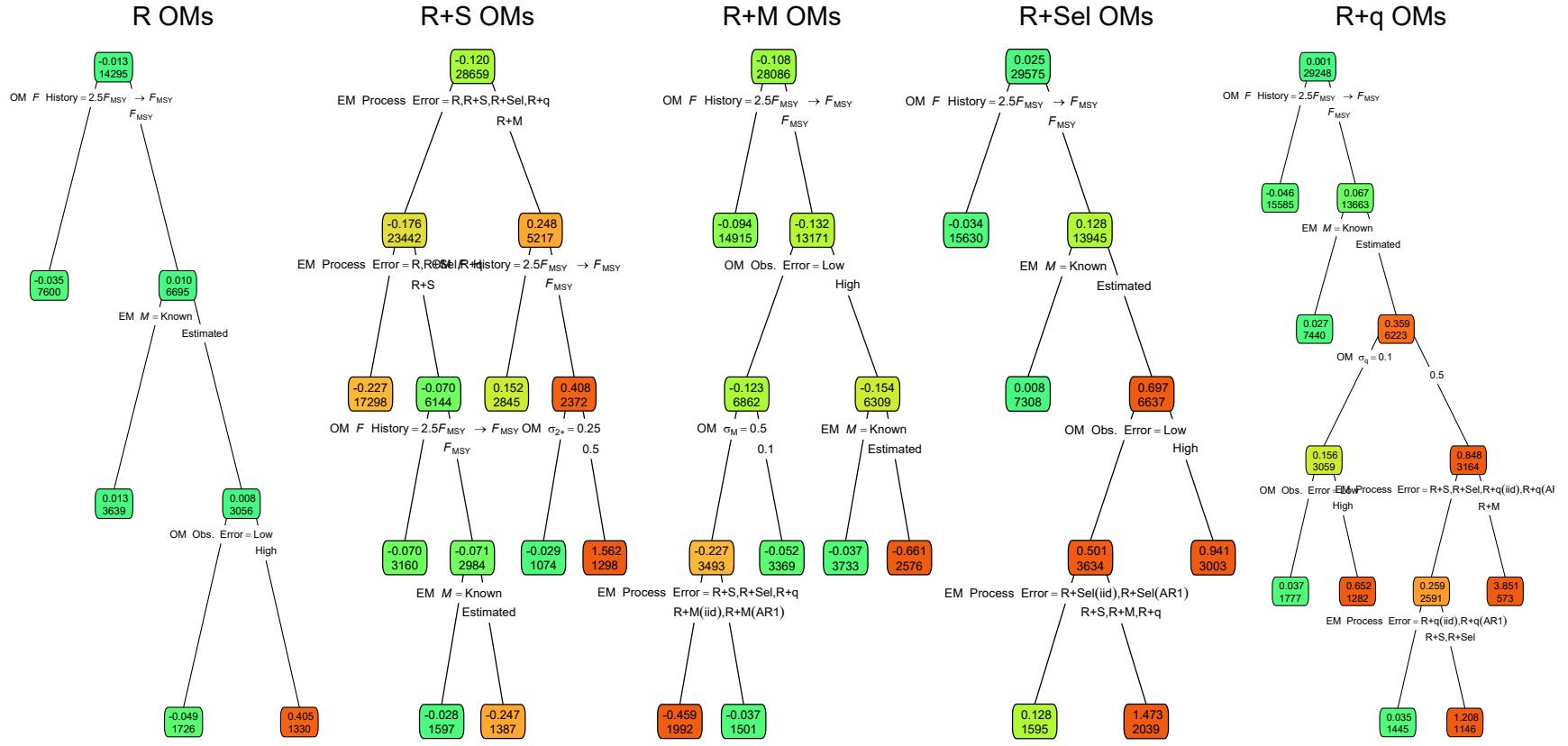


Fig. S6. Regression trees indicating primary factors determining reductions in sums of squares of errors in estimation measured by Eq. 3 for terminal year recruitment for R+S, R+M, R+Sel and R+q OM. Each node shows the median error (top) and number of observations (bottom) for the corresponding subset. Median errors closer to or further from zero are indicated by more green or red polygons, respectively.

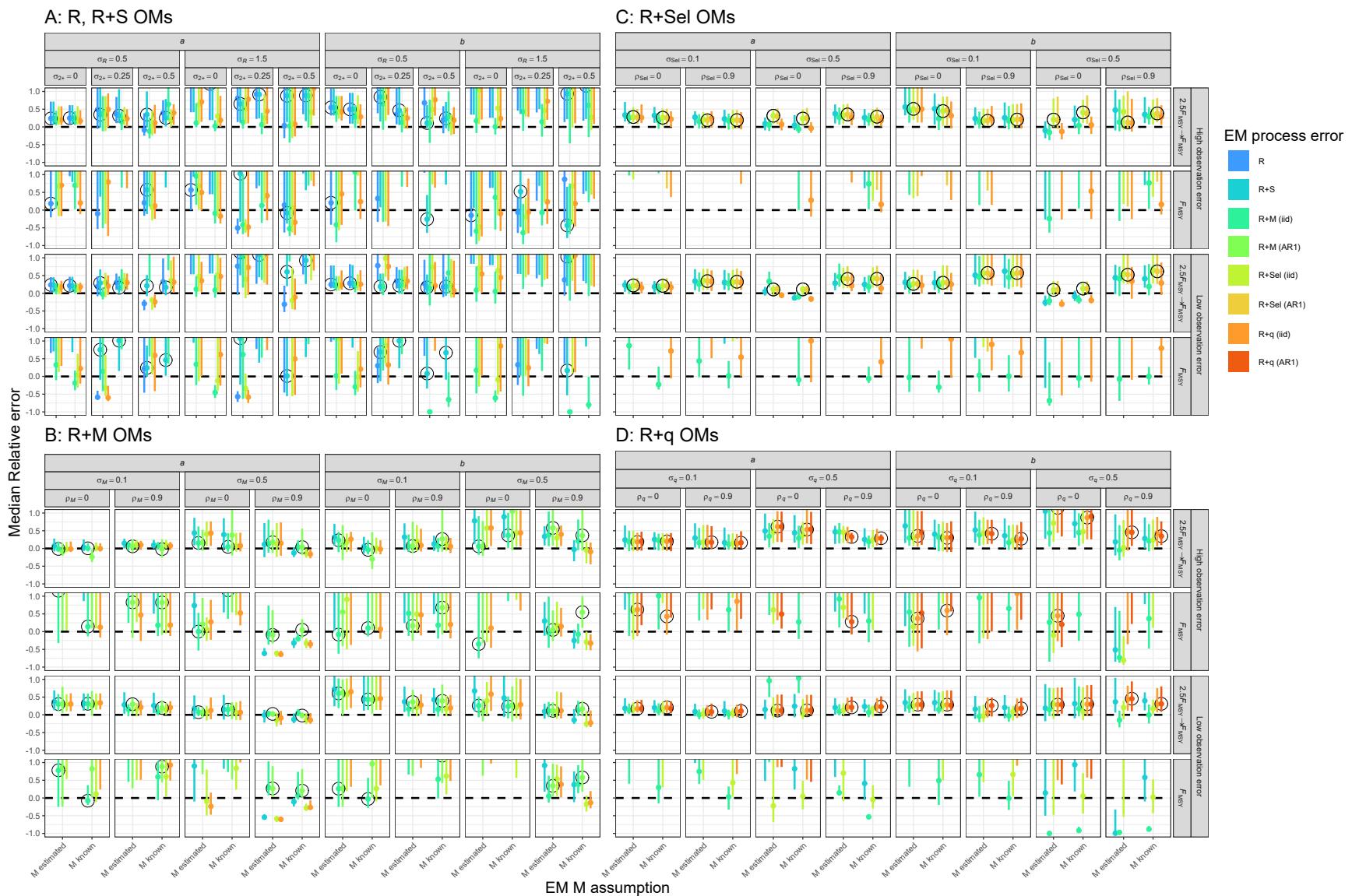


Fig. S7. Median relative error of Beverton-Holt SRR parameters (*a* and *b*) for EMs fitted to data sets simulated with alternative process error structures: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where the EM process error structure matches that of the OM and vertical lines represent 95% confidence intervals.

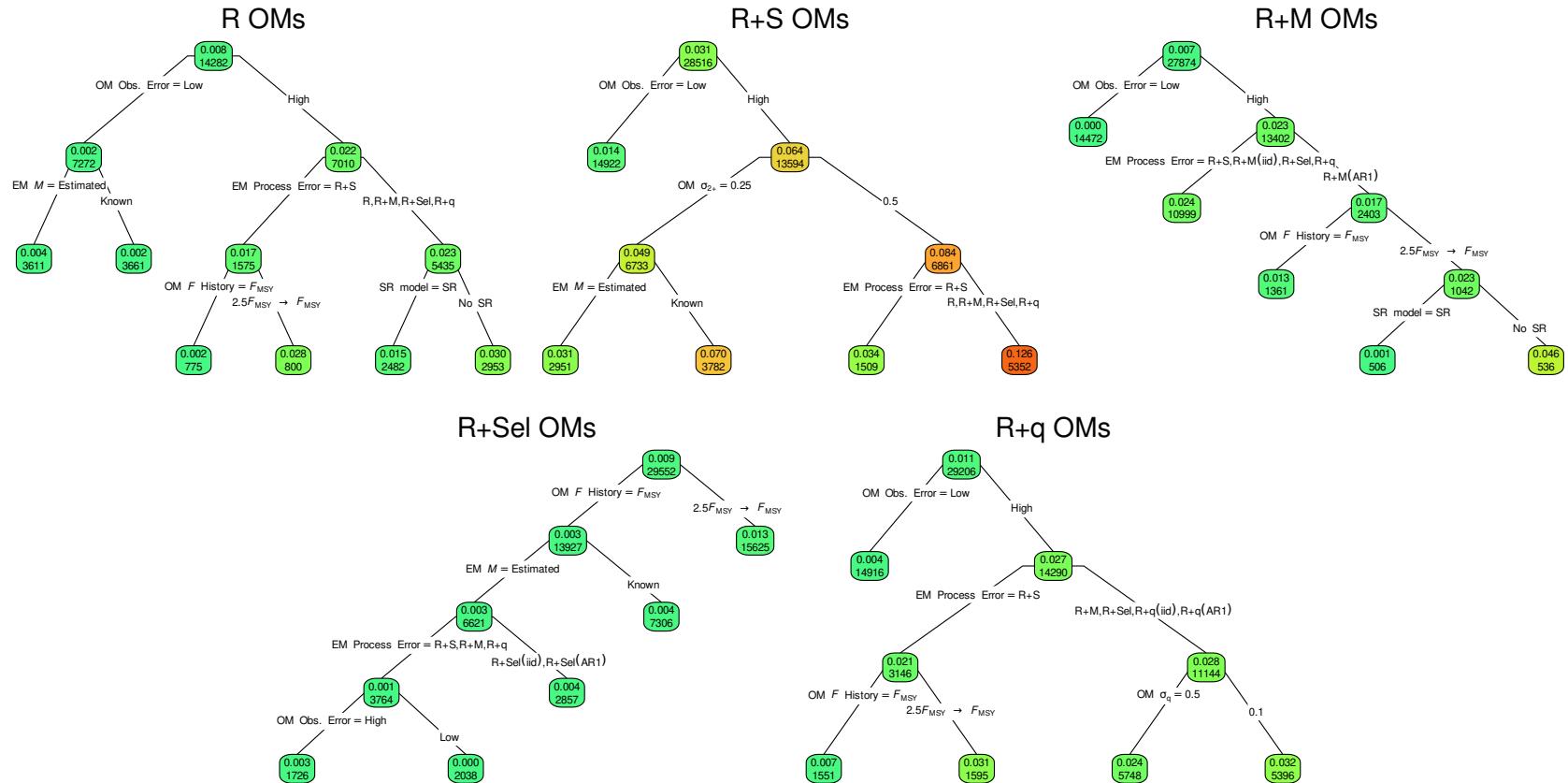


Fig. S8. Regression trees indicating primary factors determining reductions in sums of squares of errors in transformed Mohn's ρ (Eq. 3) for fishing mortality averaged over all age classes for R+S, R+M, R+Sel and R+q OM models. Each node shows the median Mohn's ρ (top) and number of observations (bottom) for the corresponding subset. Median Mohn's ρ closer to or further from zero are indicated by more green or red polygons, respectively.

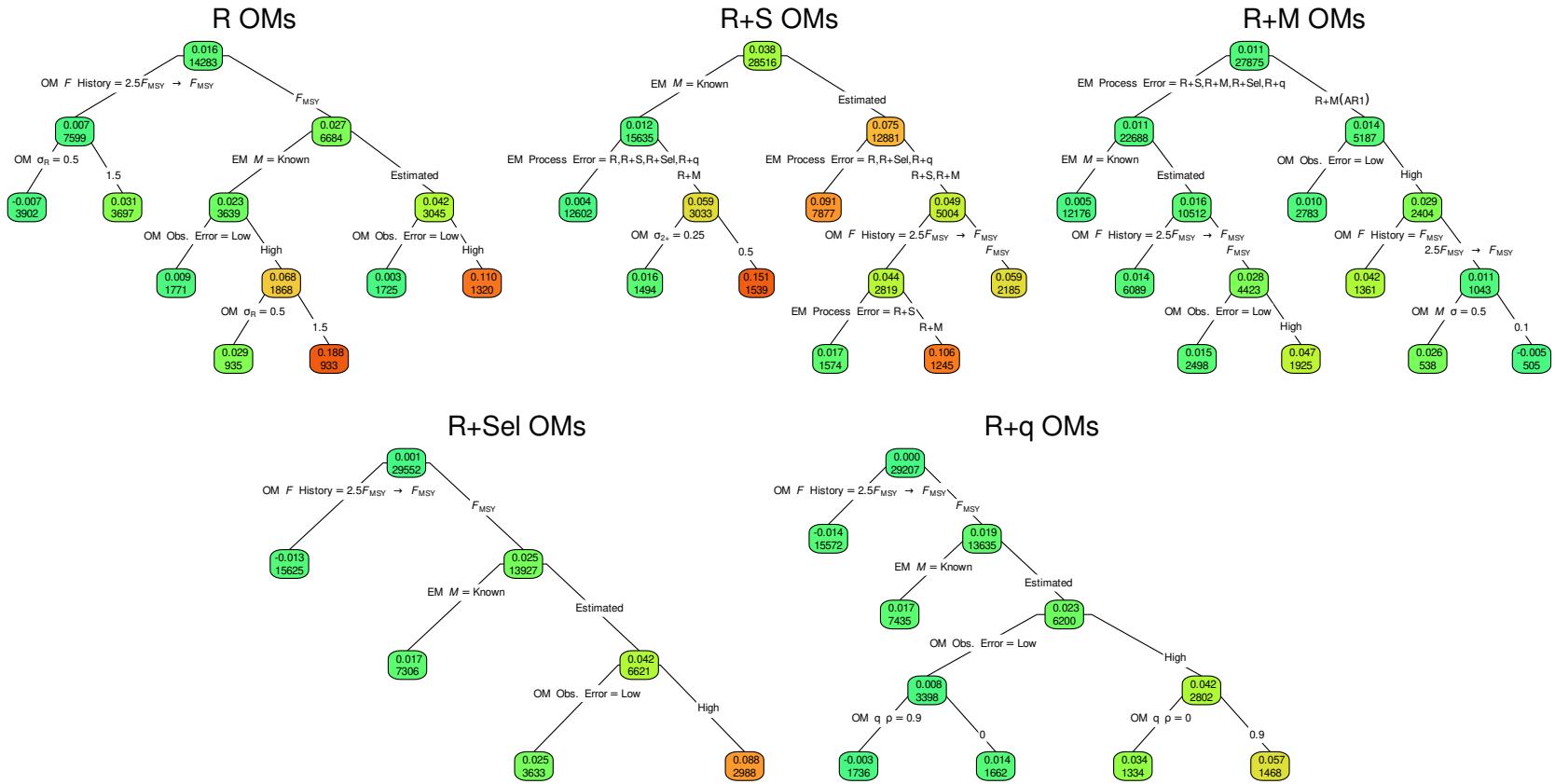


Fig. S9. Regression trees indicating primary factors determining reductions in sums of squares of errors in transformed Mohn's ρ (Eq. 3) for recruitment for R+S, R+M, R+Sel and R+q OM. Each node shows the median Mohn's ρ (top) and number of observations (bottom) for the corresponding subset. Median Mohn's ρ closer to or further from zero are indicated by more green or red polygons, respectively.

1213 Referenced Tables

Table S1. Distinguishing characteristics of the ~~operating models~~ QMs with random effects on recruitment and apparent survival (R, R+S). Standard When observation uncertainty is low, standard deviations (SD) are for log-normal distributed indices and logistic normal distributed age composition observations (fleet are 0.1 and indices) 0.3, respectively, and when it is high, standard deviations are 0.4 and 1.5, respectively. Fishing mortality either changes from $2.5F_{MSY}$ to F_{MSY} after year 20 (of 40) for fishing histories where fishing mortality or is not constant at F_{MSY} over all years.

Model	σ_R	σ_{2+}	Fishing History	Observation Uncertainty
1	0.5		$2.5F_{MSY} \rightarrow F_{MSY}$	Low
2	1.5		$2.5F_{MSY} \rightarrow F_{MSY}$	Low
3	0.5	0.25	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
4	1.5	0.25	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
5	0.5	0.50	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
6	1.5	0.50	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
7	0.5		F_{MSY}	Low
8	1.5		F_{MSY}	Low
9	0.5	0.25	F_{MSY}	Low
10	1.5	0.25	F_{MSY}	Low
11	0.5	0.50	F_{MSY}	Low
12	1.5	0.50	F_{MSY}	Low
13	0.5		$2.5F_{MSY} \rightarrow F_{MSY}$	High
14	1.5		$2.5F_{MSY} \rightarrow F_{MSY}$	High
15	0.5	0.25	$2.5F_{MSY} \rightarrow F_{MSY}$	High
16	1.5	0.25	$2.5F_{MSY} \rightarrow F_{MSY}$	High
17	0.5	0.50	$2.5F_{MSY} \rightarrow F_{MSY}$	High
18	1.5	0.50	$2.5F_{MSY} \rightarrow F_{MSY}$	High
19	0.5		F_{MSY}	High
20	1.5		F_{MSY}	High
21	0.5	0.25	F_{MSY}	High
22	1.5	0.25	F_{MSY}	High
23	0.5	0.50	F_{MSY}	High
24	1.5	0.50	F_{MSY}	High

Table S2. Distinguishing characteristics of the ~~operating models~~ OMs with random effects on recruitment and natural mortality (R+M). ~~Standard~~ When observation uncertainty is low, standard deviations (~~SD~~) are for log-normal distributed indices and logistic normal distributed age composition observations (~~fleet~~ are 0.1 and ~~indiees~~ 0.3, respectively, and when it is high, standard deviations are 0.4 and 1.5, respectively). Fishing mortality either changes from $2.5F_{MSY}$ to F_{MSY} after year 20 (of 40) for fishing histories where fishing mortality or is not constant at F_{MSY} over all years. For AR1 process errors, σ_M is defined for the marginal distribution of the processes.

Model	σ_R	σ_M	ρ_M	Fishing History	Observation Uncertainty
1	0.5	0.1	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
2	0.5	0.5	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
3	0.5	0.1	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
4	0.5	0.5	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
5	0.5	0.1	0.0	F_{MSY}	Low
6	0.5	0.5	0.0	F_{MSY}	Low
7	0.5	0.1	0.9	F_{MSY}	Low
8	0.5	0.5	0.9	F_{MSY}	Low
9	0.5	0.1	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	High
10	0.5	0.5	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	High
11	0.5	0.1	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	High
12	0.5	0.5	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	High
13	0.5	0.1	0.0	F_{MSY}	High
14	0.5	0.5	0.0	F_{MSY}	High
15	0.5	0.1	0.9	F_{MSY}	High
16	0.5	0.5	0.9	F_{MSY}	High

Table S3. Distinguishing characteristics of the ~~operating models~~ OMs with random effects on recruitment and selectivity (R+Sel). ~~Standard~~ When observation uncertainty is low, standard deviations (~~SD~~) are for log-normal distributed indices and logistic normal distributed age composition observations (~~fleet~~ are 0.1 and ~~indiees~~ 0.3, respectively, and when it is high, standard deviations are 0.4 and 1.5, respectively). Fishing mortality either changes from $2.5F_{MSY}$ to F_{MSY} after year 20 (of 40) ~~for fishing histories where fishing mortality or is not constant at F_{MSY} over all years~~. For AR1 process errors, σ_{Sel} is defined for the marginal distribution of the processes.

Model	σ_R	σ_{Sel}	ρ_{Sel}	Fishing History	Observation Uncertainty
1	0.5	0.1	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
2	0.5	0.5	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
3	0.5	0.1	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
4	0.5	0.5	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	Low
5	0.5	0.1	0.0	F_{MSY}	Low
6	0.5	0.5	0.0	F_{MSY}	Low
7	0.5	0.1	0.9	F_{MSY}	Low
8	0.5	0.5	0.9	F_{MSY}	Low
9	0.5	0.1	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	High
10	0.5	0.5	0.0	$2.5F_{MSY} \rightarrow F_{MSY}$	High
11	0.5	0.1	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	High
12	0.5	0.5	0.9	$2.5F_{MSY} \rightarrow F_{MSY}$	High
13	0.5	0.1	0.0	F_{MSY}	High
14	0.5	0.5	0.0	F_{MSY}	High
15	0.5	0.1	0.9	F_{MSY}	High
16	0.5	0.5	0.9	F_{MSY}	High

Table S4. Distinguishing characteristics of the ~~operating models~~ OMs with random effects on recruitment and catchability ($R+q$). ~~Standard~~ When observation uncertainty is low, standard deviations (SD) are for log-normal distributed indices and logistic normal distributed age composition observations (fleet are 0.1 and ~~indiees~~ 0.3, respectively, and when it is high, standard deviations are 0.4 and 1.5, respectively). Fishing mortality either changes from $2.5F_{\text{MSY}}$ to F_{MSY} after year 20 (of 40) for fishing histories where fishing mortality or is not constant at F_{MSY} over all years. For AR1 process errors, σ_{α} is defined for the marginal distribution of the processes.

Model	σ_R	σ_q	ρ_q	Fishing History	Observation Uncertainty
1	0.5	0.1	0.0	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	Low
2	0.5	0.5	0.0	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	Low
3	0.5	0.1	0.9	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	Low
4	0.5	0.5	0.9	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	Low
5	0.5	0.1	0.0	F_{MSY}	Low
6	0.5	0.5	0.0	F_{MSY}	Low
7	0.5	0.1	0.9	F_{MSY}	Low
8	0.5	0.5	0.9	F_{MSY}	Low
9	0.5	0.1	0.0	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	High
10	0.5	0.5	0.0	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	High
11	0.5	0.1	0.9	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	High
12	0.5	0.5	0.9	$2.5F_{\text{MSY}} \rightarrow F_{\text{MSY}}$	High
13	0.5	0.1	0.0	F_{MSY}	High
14	0.5	0.5	0.0	F_{MSY}	High
15	0.5	0.1	0.9	F_{MSY}	High
16	0.5	0.5	0.9	F_{MSY}	High

Table S5. Distinguishing characteristics of the ~~estimating models~~ EMs and ~~operating model indication (+) of which OM process error categories sources~~ (R, R+S, R+M, R+Sel, R+q) ~~where used each EM configuration was fit.~~

Model	Recruitment model	Median M	Process error	R,R+S OM _s	R+M OM _s	R+Sel OM _s	R+q OM _s
1	Mean recruitment	0.2	R ($\sigma_{2+} = 0$)	+			
2	Beverton-Holt	0.2	R ($\sigma_{2+} = 0$)	+			
3	Mean recruitment	Estimated	R ($\sigma_{2+} = 0$)	+			
4	Beverton-Holt	Estimated	R ($\sigma_{2+} = 0$)	+			
5	Mean recruitment	0.2	R+S (σ_{2+} estimated)	+	+	+	+
6	Beverton-Holt	0.2	R+S (σ_{2+} estimated)	+	+	+	+
7	Mean recruitment	Estimated	R+S (σ_{2+} estimated)	+	+	+	+
8	Beverton-Holt	Estimated	R+S (σ_{2+} estimated)	+	+	+	+
9	Mean recruitment	0.2	R+M ($\rho_M = 0$)	+	+	+	+
10	Beverton-Holt	0.2	R+M ($\rho_M = 0$)	+	+	+	+
11	Mean recruitment	Estimated	R+M ($\rho_M = 0$)	+	+	+	+
12	Beverton-Holt	Estimated	R+M ($\rho_M = 0$)	+	+	+	+
13	Mean recruitment	0.2	R+Sel ($\rho_{Sel} = 0$)	+	+	+	+
14	Beverton-Holt	0.2	R+Sel ($\rho_{Sel} = 0$)	+	+	+	+
15	Mean recruitment	Estimated	R+Sel ($\rho_{Sel} = 0$)	+	+	+	+
16	Beverton-Holt	Estimated	R+Sel ($\rho_{Sel} = 0$)	+	+	+	+
17	Mean recruitment	0.2	R+q ($\rho_q = 0$)	+	+	+	+
18	Beverton-Holt	0.2	R+q ($\rho_q = 0$)	+	+	+	+
19	Mean recruitment	Estimated	R+q ($\rho_q = 0$)	+	+	+	+
20	Beverton-Holt	Estimated	R+q ($\rho_q = 0$)	+	+	+	+
21	Mean recruitment	0.2	R+M (ρ_M estimated)		+		
22	Beverton-Holt	0.2	R+M (ρ_M estimated)		+		
23	Mean recruitment	Estimated	R+M (ρ_M estimated)		+		
24	Beverton-Holt	Estimated	R+M (ρ_M estimated)		+		
25	Mean recruitment	0.2	R+Sel (ρ_{Sel} estimated)			+	
26	Beverton-Holt	0.2	R+Sel (ρ_{Sel} estimated)			+	
27	Mean recruitment	Estimated	R+Sel (ρ_{Sel} estimated)			+	
28	Beverton-Holt	Estimated	R+Sel (ρ_{Sel} estimated)			+	
29	Mean recruitment	0.2	R+q (ρ_q estimated)				+
30	Beverton-Holt	0.2	R+q (ρ_q estimated)				+
31	Mean recruitment	Estimated	R+q (ρ_q estimated)				+
32	Beverton-Holt	Estimated	R+q (ρ_q estimated)				+

Table S6. For each OM process error source (columns), percent reduction in deviance for logistic regression models fit to indicators of convergence (maximum absolute gradient $< 10^{-6}$) with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM Process Error	30.40	0.45	17.57	16.04	24.03
EM M Assumption	2.38	24.11	4.42	1.02	2.66
EM SR Assumption	1.80	0.32	0.96	3.38	2.13
OM Obs. Error	0.12	0.77	0.33	1.76	0.28
OM F History	3.51	6.33	2.36	5.86	5.30
OM σ_R	≤ 0.01	≤ 0.01	—	—	—
OM σ_{2+}	—	≤ 0.01	—	—	—
OM σ_M	—	—	0.39	—	—
OM ρ_M	—	—	0.09	—	—
OM σ_{Sel}	—	—	—	1.08	—
OM ρ_{Sel}	—	—	—	0.01	—
OM σ_q	—	—	—	—	0.06
OM ρ_q	—	—	—	—	≤ 0.01
All factors	43.69	35.72	29.33	34.57	40.69
± All Two Way	50.53	42.99	43.91	45.93	48.62
± All Three Way	52.30	48.41	46.81	47.71	50.40

Table S7. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. 3 for the terminal year fully-selected fishing mortality with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	2.26	1.33	1.26	2.93	3.26
EM SR assumption	0.11	0.07	0.08	0.07	0.09
EM Process Error	0.46	4.18	0.38	0.13	1.02
OM Obs. Error	1.61	0.06	0.86	0.41	<0.01
OM F History	2.49	3.23	1.42	3.22	4.55
OM σ_R	0.02	0.02	—	—	—
OM σ_{2+}	—	0.87	—	—	—
OM σ_M	—	—	0.16	—	—
OM ρ_M	—	—	0.01	—	—
OM σ_{Sel}	—	—	—	0.24	—
OM ρ_{Sel}	—	—	—	0.05	—
OM σ_q	—	—	—	—	1.03
OM ρ_q	—	—	—	—	0.05
All factors	7.42	9.96	4.37	7.26	10.43
± All Two Way	17.63	25.76	10.94	13.88	22.07
± All Three Way	22.97	37.03	14.74	17.32	30.74

Table S8. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to errors in estimation measured by Eq. 3 for the terminal year recruitment with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	1.96	0.40	0.69	3.52	3.03
EM SR assumption	0.06	0.02	0.05	0.02	0.05
EM Process Error	0.39	4.74	0.41	0.12	1.16
OM Obs. Error	1.47	0.08	0.64	0.18	<0.01
OM F History	2.54	2.66	1.11	4.18	5.06
OM σ_R	0.03	0.01	—	—	—
OM σ_{2+}	—	1.05	—	—	—
OM σ_M	—	—	0.36	—	—
OM ρ_M	—	—	0.02	—	—
OM σ_{Sel}	—	—	—	0.23	—
OM ρ_{Sel}	—	—	—	0.06	—
OM σ_q	—	—	—	—	1.09
OM ρ_q	—	—	—	—	0.06
All factors	6.90	9.01	3.43	8.58	10.90
± All Two Way	16.48	24.64	9.73	15.76	22.75
± All Three Way	21.46	35.60	13.56	19.07	31.15

1214 [Further Results](#)

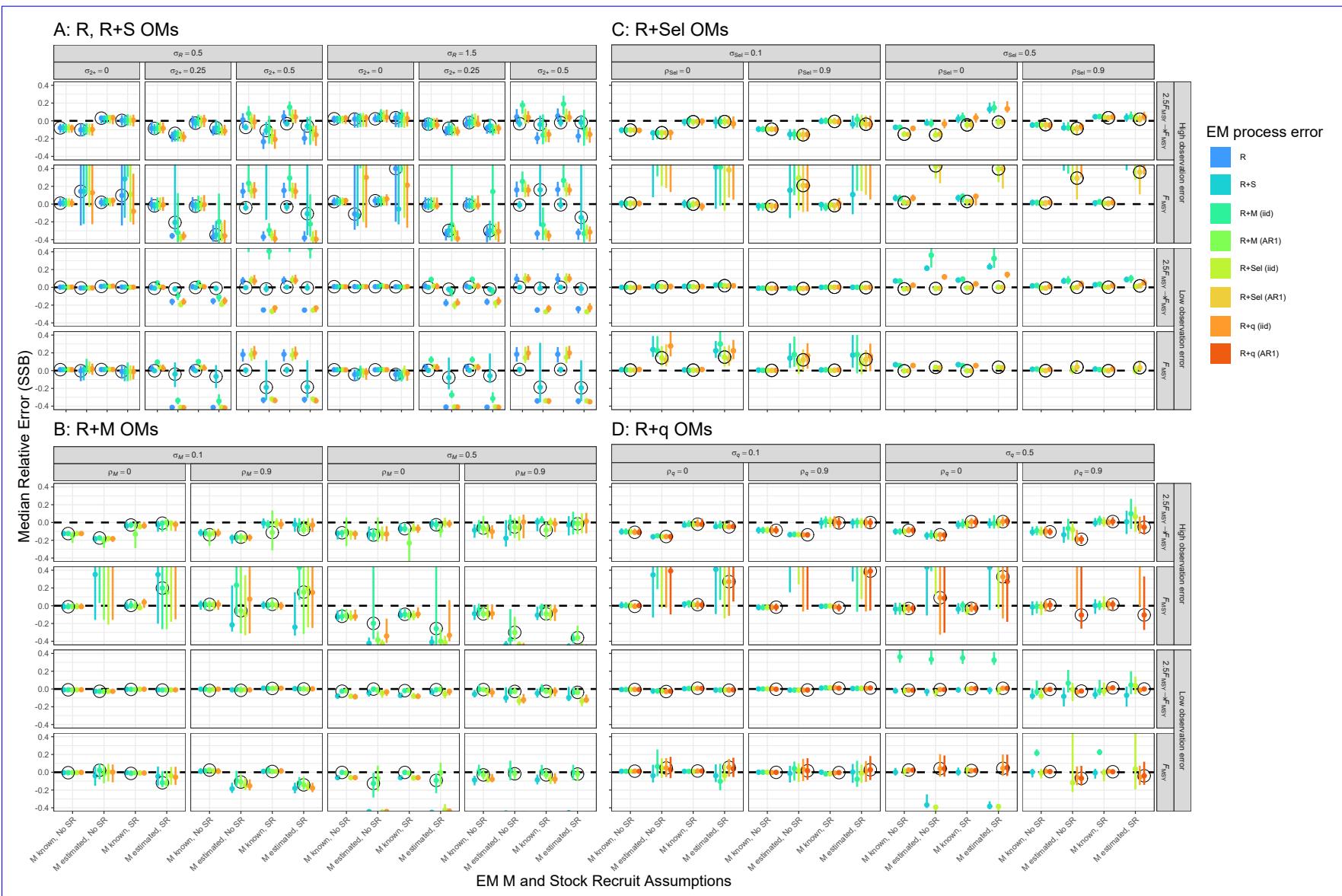


Fig. S10. The maximum Median relative error of the absolute values of all gradient values terminal year SSB for all fits that provided hessian-based standard errors across all simulated EMs fitted to data sets of a given OM configuration (A simulated with alternative process error sources: R and R+S (A), B: R+MSel (B), C: R+SelM (C), or D: R+q). Results are conditional on EM fits with alternative process error type (colored points and linesD), median natural mortality (estimated or known) and recruitment assumptions (Beverton-Holt stock-recruit relationship or not). Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.

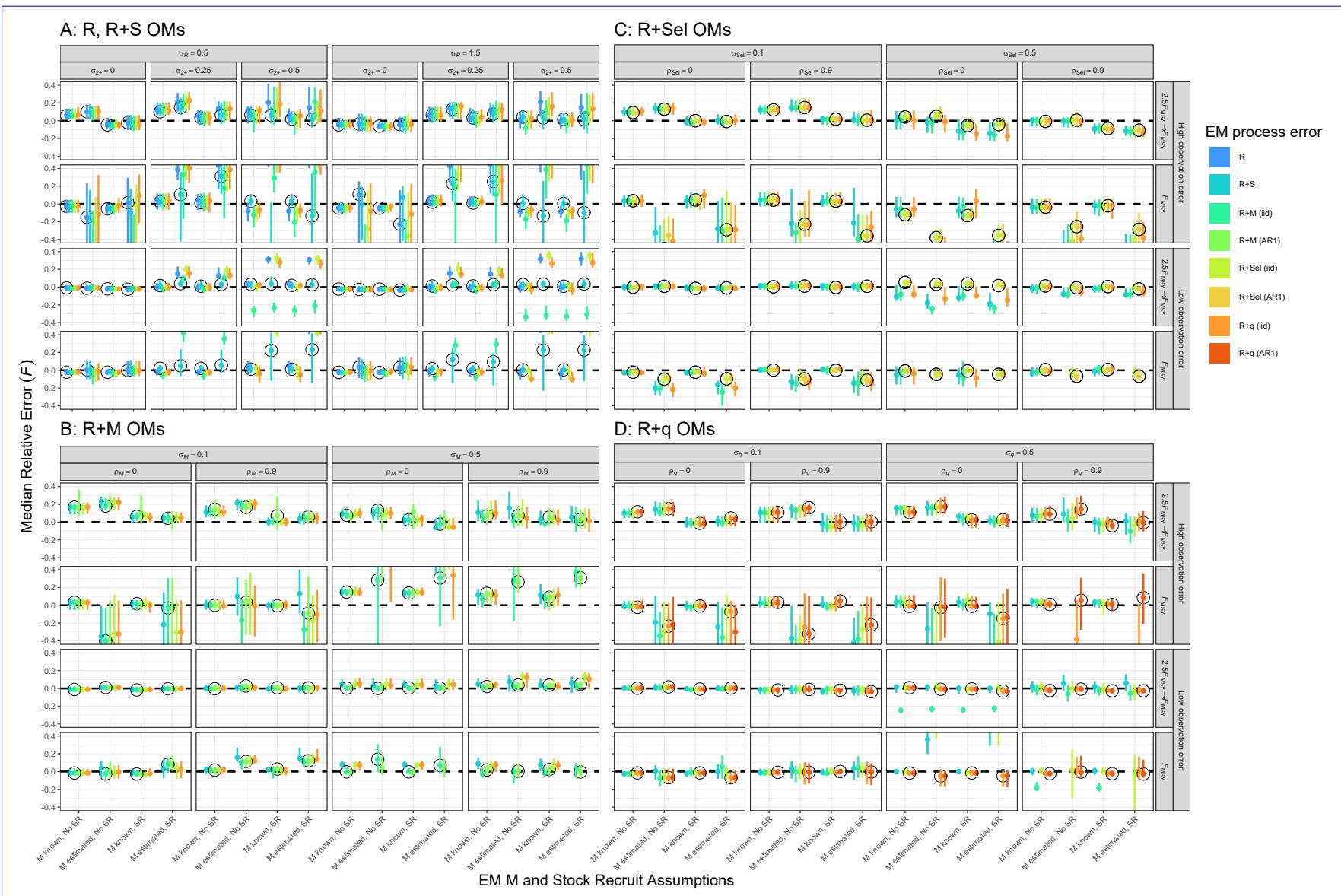


Fig. S11. Probability of estimating models providing maximum absolute values of gradients less than 10^{-6} assuming alternative process Median relative error (colored points and lines), and median natural of terminal year fully-selected fishing mortality (estimated or known) and Beverton-Holt stock-recruit relationships (estimated or not; along x-axis) when for EMs fitted to operating models that have data sets simulated with alternative process error sources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D) process error structures. Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.

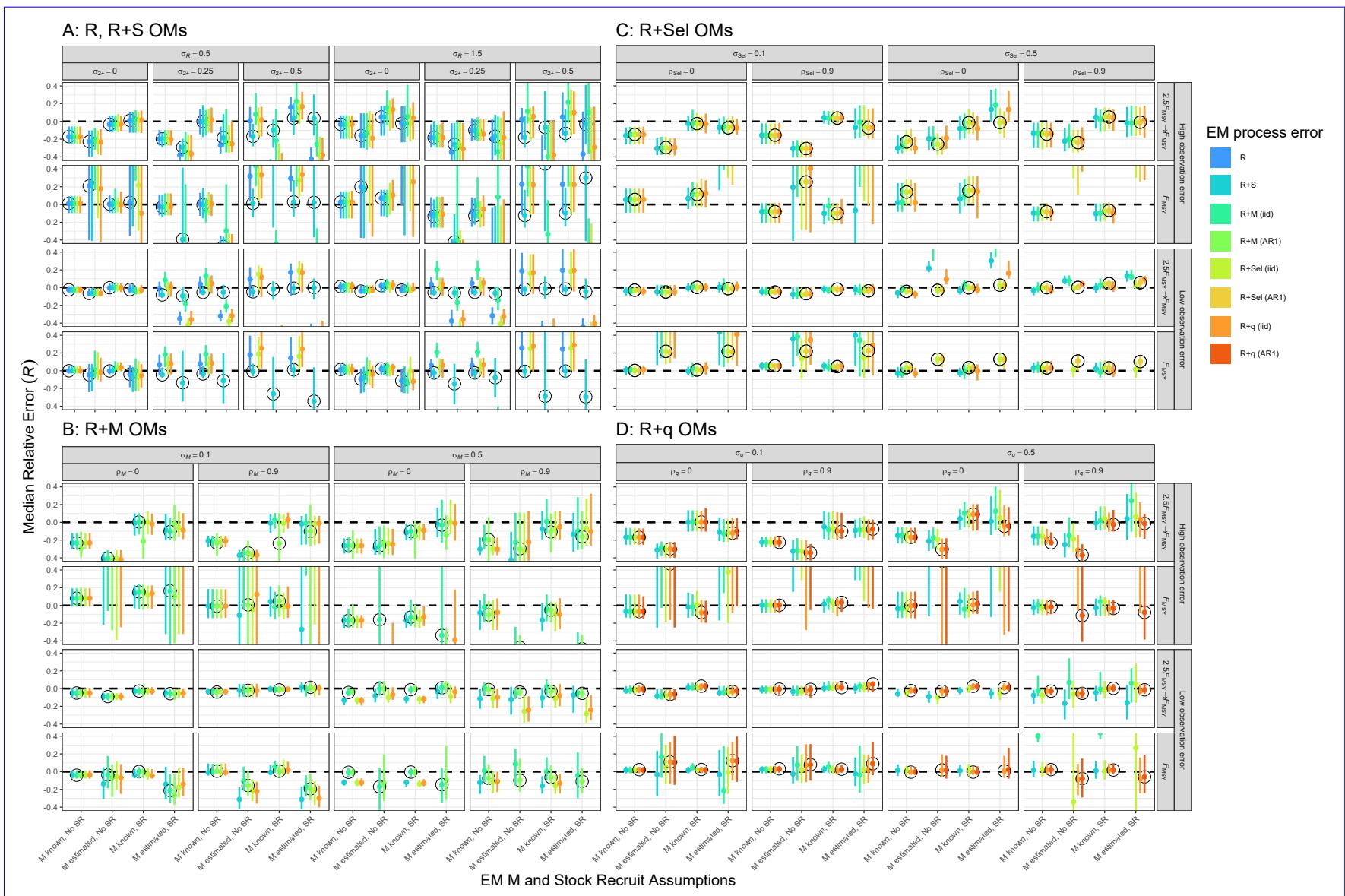


Fig. S12. Estimated probability Median relative error of lowest AIC from logistic regression on the log-standard deviation of the true log(SSB) in each simulation terminal year recruitment for estimating model EMs fitted to data sets simulated with Beverton-Holt stock-recruit relationships, rather than the otherwise equivalent EM without the stock-recruit relationship. Results are conditional on alternative assumptions for median natural mortality (estimated or known) and on EMs having the correct process error structure sources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Rug along x-axis denotes $SD(\log(SSB))$. Circled values for each simulation indicate results where the EM process error structure matches that of the

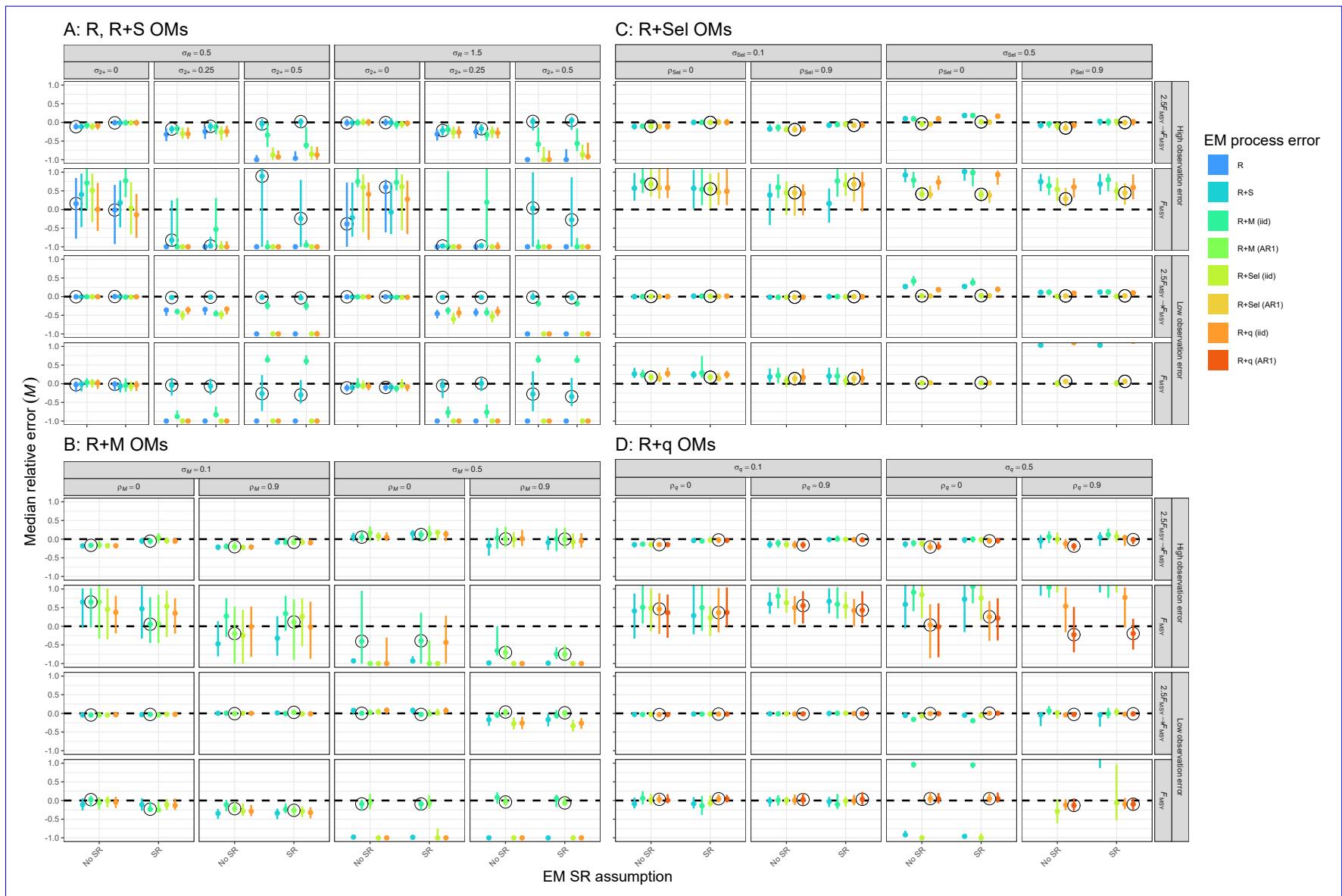


Fig. S13. Median relative error of median natural mortality for EMs fitted to data sets simulated with alternative process error sources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where the EM process error structure matches that of the OM and vertical lines represent 95% confidence intervals.

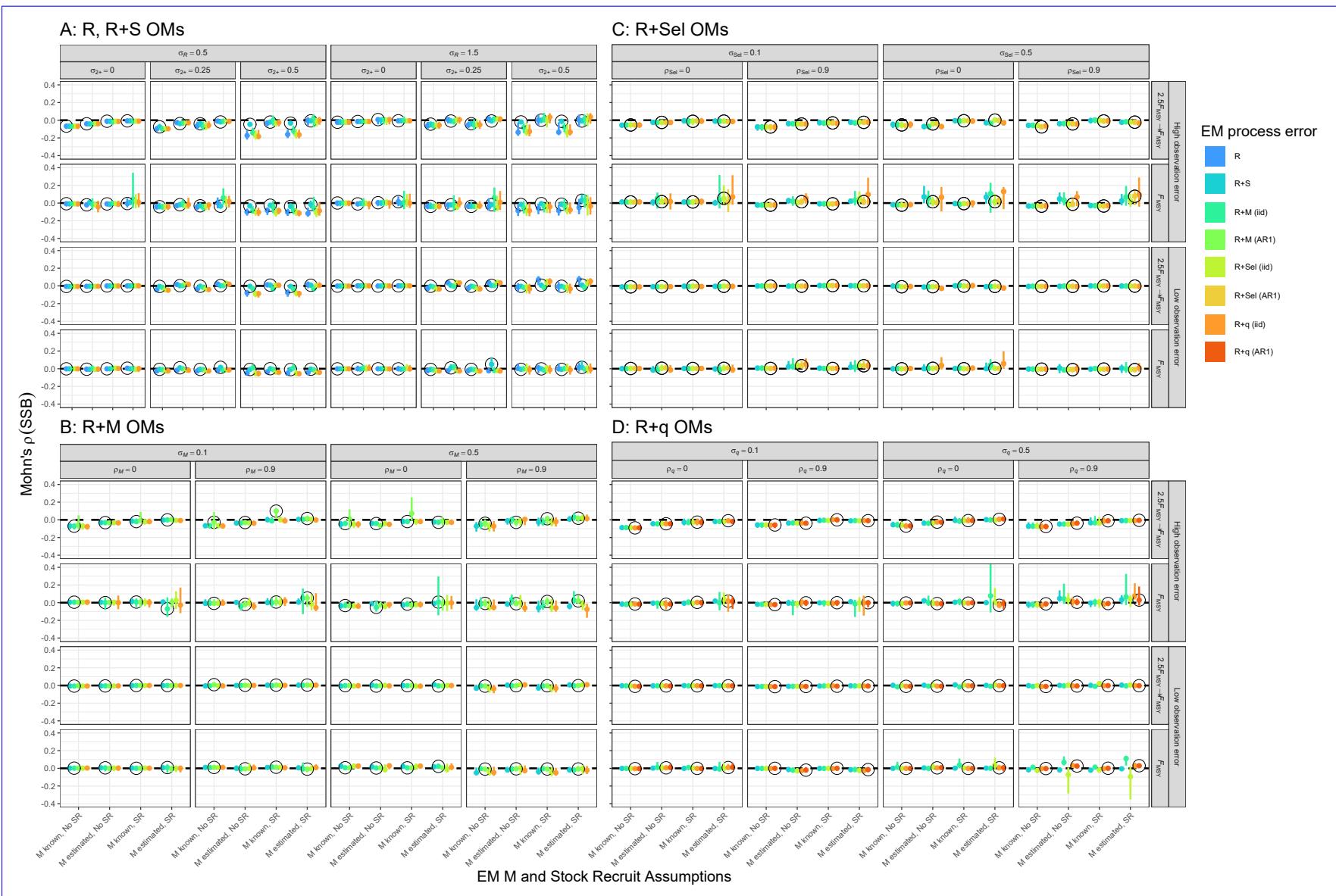


Fig. S14. Median relative error of terminal year fully selected fishing mortality Mohn's ρ for estimating models SSB for EMs fitted to data sets simulated with alternative process error structures sources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.

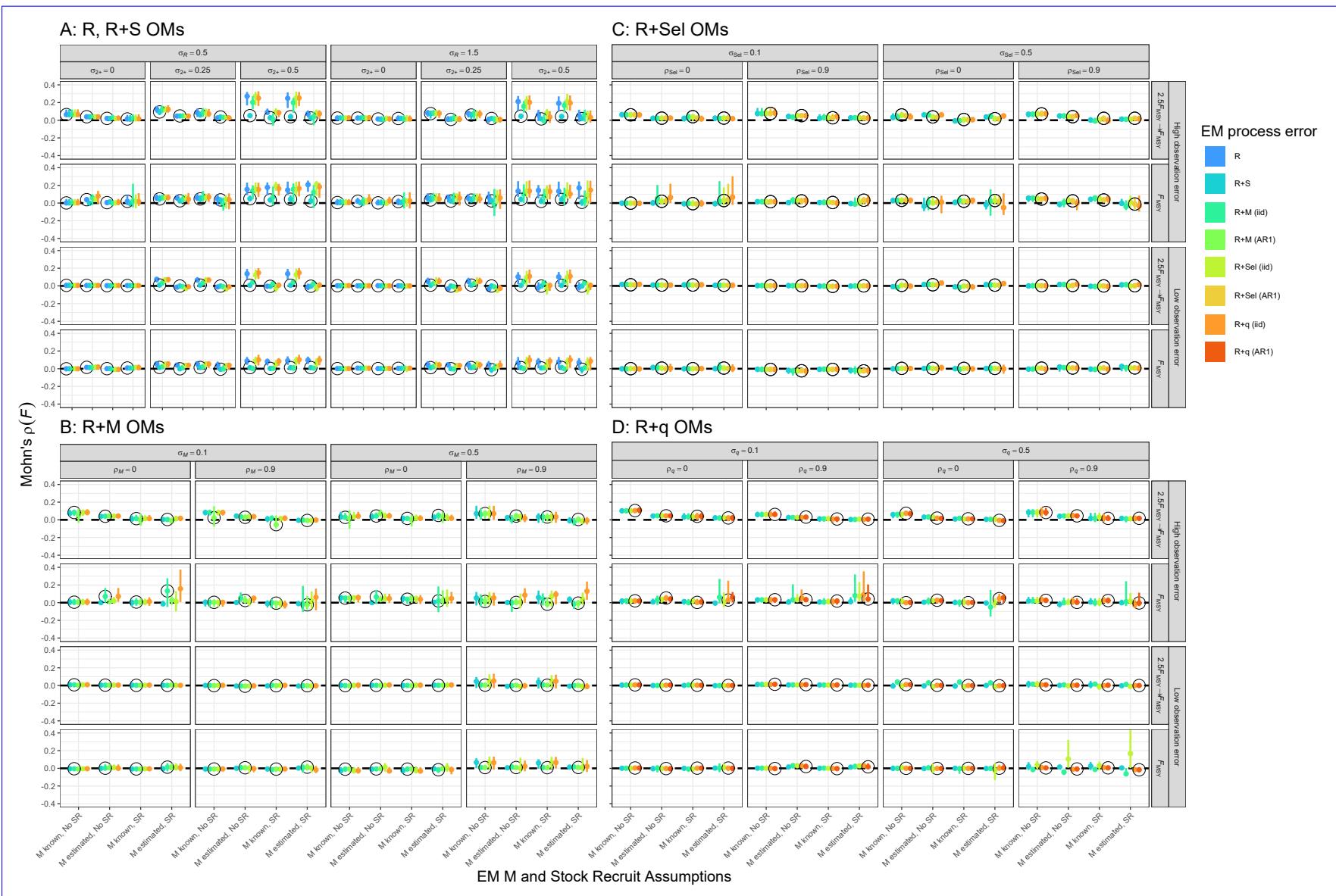


Fig. S15. Median relative error—Mohn's ρ of terminal year recruitment-fishing mortality averaged over all age classes for estimating models EMs fitted to data sets simulated with alternative process error structures sources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.

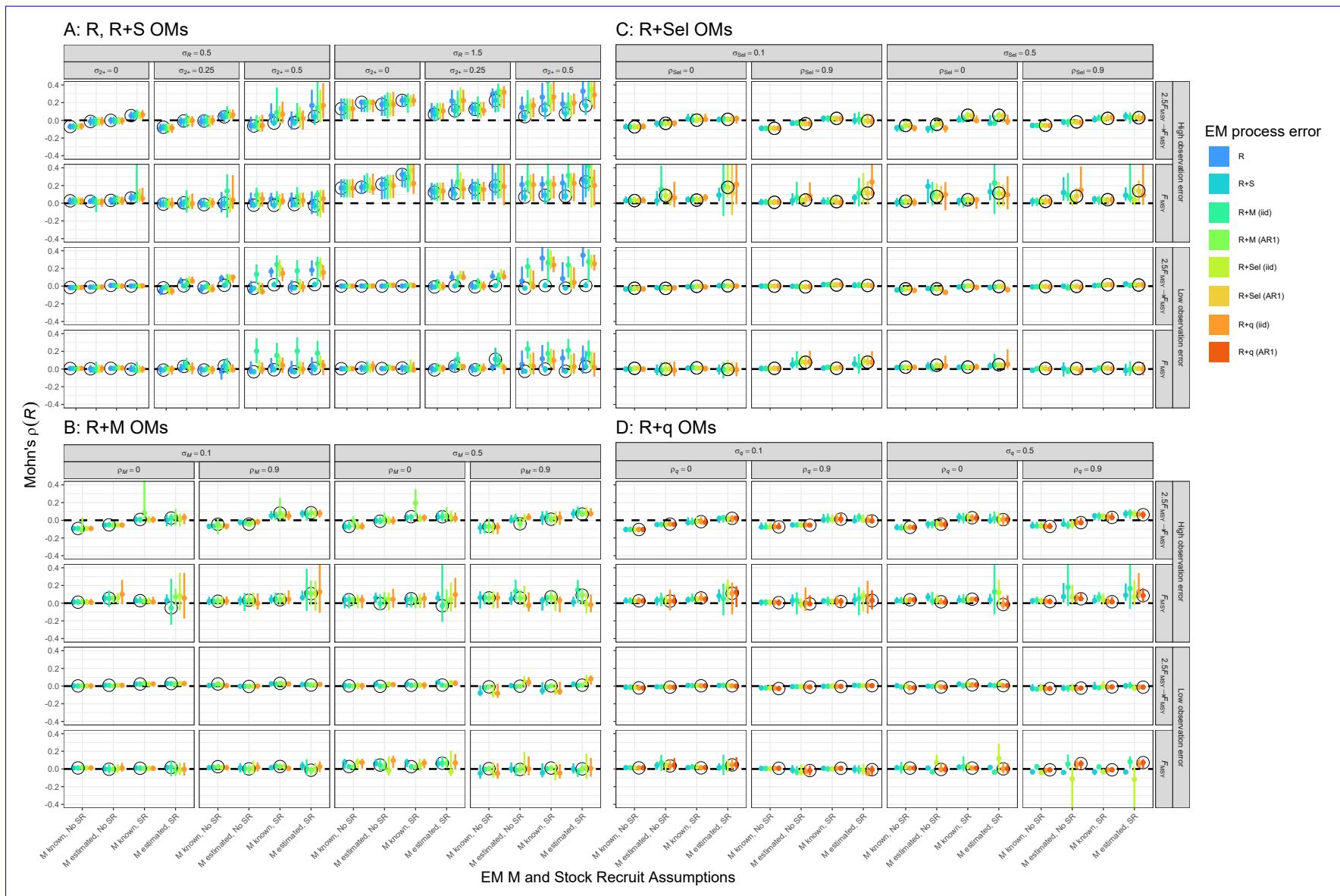


Fig. S16. Median relative error Mohn's ρ of Beverton-Holt stock-recruit parameters (*a* and *b*) recruitment for estimating models EMs fitted to data sets simulated with alternative process error structuresources: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.



Fig. S17. Median relative error Probability of EMs providing Hessian-based standard errors with alternative process error (colored points and lines), and median natural mortality for estimating models (estimated or known) and Beverton-Holt SRR (estimated or not; along x-axis) assumptions when fitted to data sets simulated with alternative process error structures: OMs that have R and R+S (A), R+Sel (B), R+M (C), or R+q (D) process error sources. Circled values indicate results where the EM process error structure matches that of the operating model OM, and vertical lines represent 95% confidence intervals.

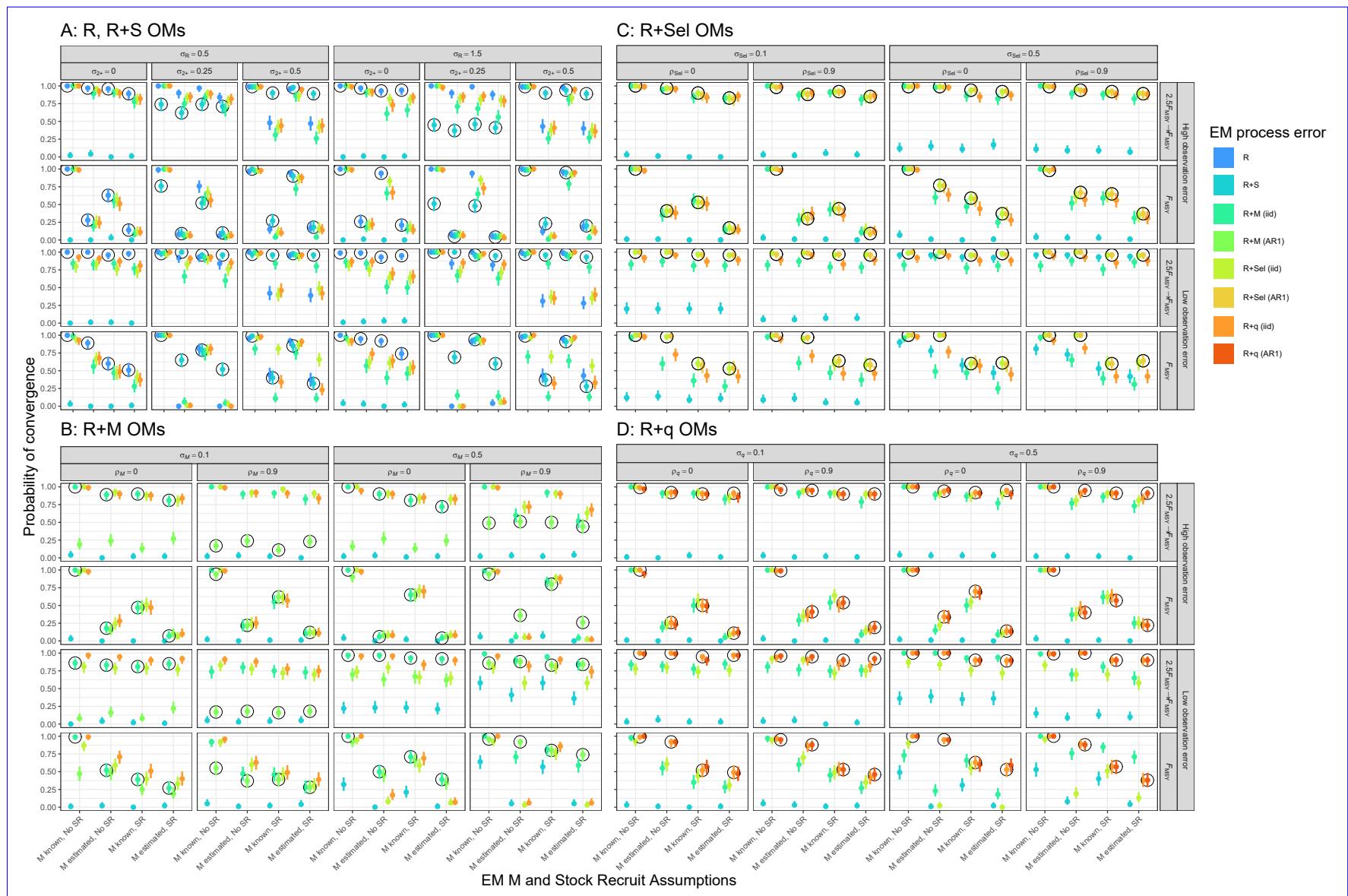


Fig. S18. Median Mohn's ρ -Probability of fishing mortality averaged over all age classes for estimating models fitted to data sets simulated EMs providing maximum absolute values of gradients less than 10^{-6} with alternative process error structures (colored points and lines), and median natural mortality (estimated or known) and Beverton-Holt SRR (estimated or not; along x-axis) assumptions when fitted to OMs that have R and R+S (A), R+Sel (B), R+M (C), or R+q (D) process error sources. Circled values indicate results where the EM process error structure matches that of the operating model OM and vertical lines represent 95% confidence intervals.

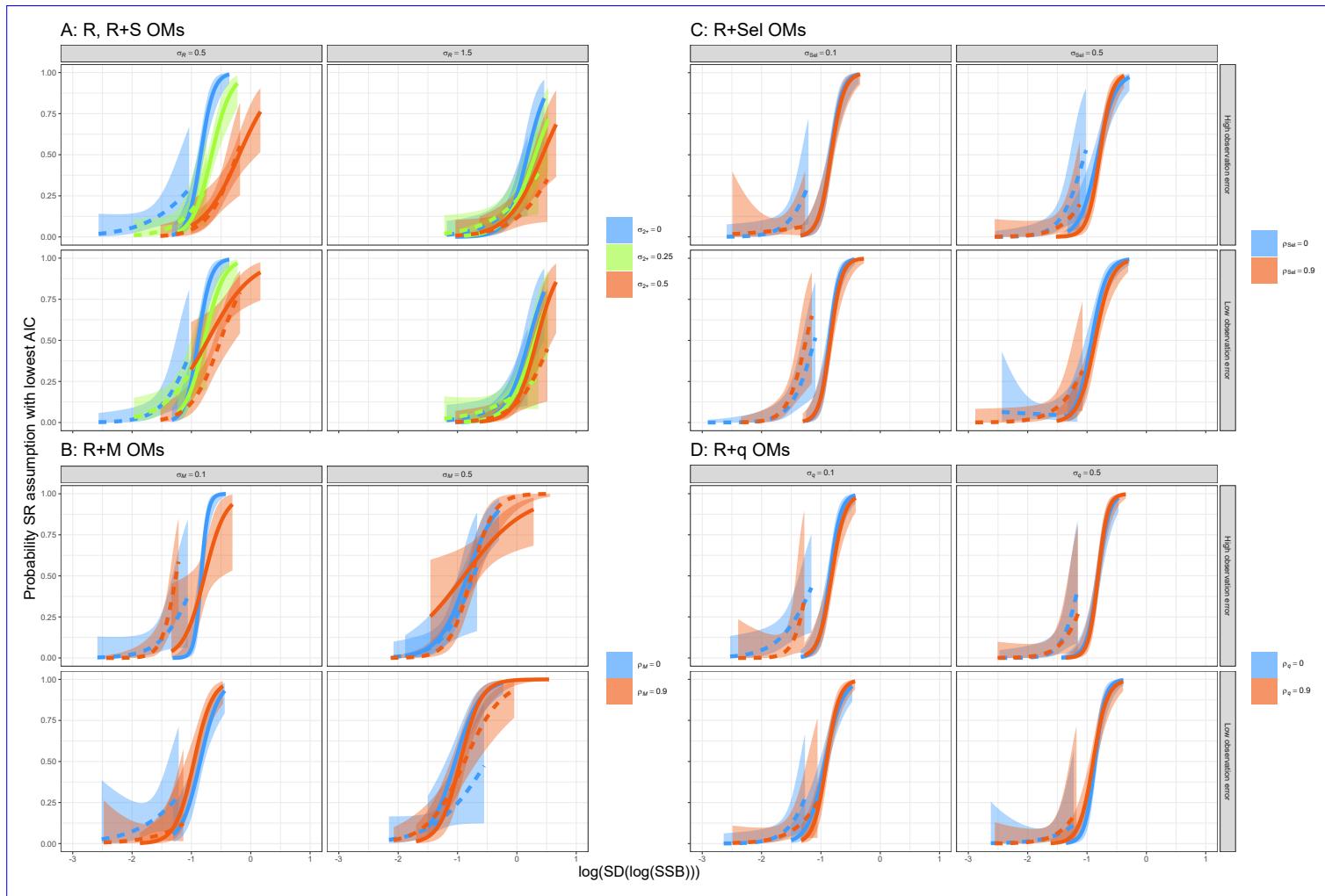


Fig. S19. Median Mohn's ρ -Probability of recruitment lowest AIC from logistic regression on the log-standard deviation of the true $\log(\text{SSB})$ in each simulation for estimating models fitted to data sets simulated EM with Beverton-Holt SRRs, rather than the otherwise equivalent EM without the SRR. Results are conditional on median M is known in the EM and alternative assumptions EMs having the correct process error structure: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Circled values indicate results where M is assumed known in the EM process error structure matches that of the operating model. Solid and vertical dashed lines are for OMs with and without temporal contrast in fishing pressure, respectively, and polygons represent 95% confidence intervals. Range of results indicates the range of log-standard deviation of $\log(\text{SSB})$ for simulations of the particular OM.

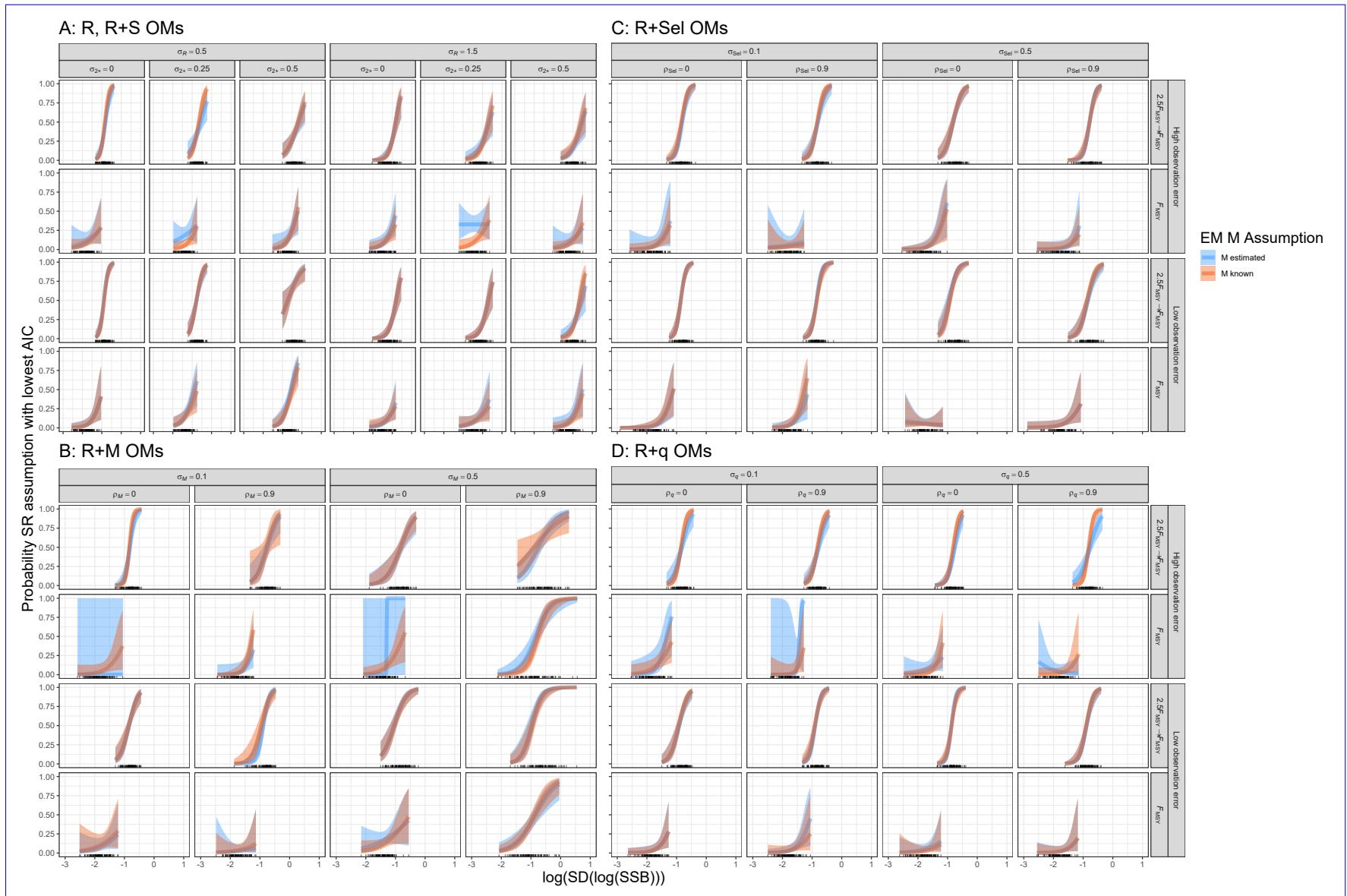


Fig. S20. Estimated probability of lowest AIC from logistic regression on the log-standard deviation of the true $\log(\text{SSB})$ in each simulation for EM with Beverton-Holt SRRs, rather than the otherwise equivalent EM without the SRR. Results are conditional on alternative assumptions for median natural mortality (estimated or known) and on EMs having the correct process error structure: R and R+S (A), R+Sel (B), R+M (C), or R+q (D). Rug along x-axis denotes $\text{SD}(\log(\text{SSB}))$ values for each simulation and polygons represent 95% confidence intervals.

Table S9. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to transformed Mohn's ρ values for each simulation (Eq. 3) for fishing mortality averaged over all age classes with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	0.06	0.09	0.01	0.12	0.01
EM SR assumption	0.01	<0.01	0.01	0.02	0.01
EM Process Error	0.03	0.07	0.02	0.06	0.03
OM Obs. Error	0.16	0.10	0.05	0.02	0.07
OM F History	0.07	0.02	0.03	0.24	0.03
OM σ_R	<0.01	0.01	~	~	~
OM σ_{2+}	~	0.09	~	~	~
OM σ_M	~	~	<0.01	~	~
OM ρ_M	~	~	<0.01	~	~
OM σ_{Sel}	~	~	~	0.01	~
OM ρ_{Sel}	~	~	~	<0.01	~
OM σ_q	~	~	~	~	<0.01
OM ρ_q	~	~	~	~	0.01
All factors	0.32	0.38	0.12	0.48	0.15
+ All Two Way	0.65	0.67	0.30	0.95	0.43
+ All Three Way	1.18	1.11	0.63	1.34	0.90

Table S10. For each OM process error source (columns), percent reduction in deviance for linear regression models fit to transformed Mohn's ρ values for each simulation (Eq. 3) for recruitment with each OM and EM factor (rows) included individually, combined, and with second and third order interactions.

Factor	R	R+S	R+M	R+Sel	R+q
EM M Assumption	0.86	0.56	0.16	1.00	1.27
EM SR assumption	<0.01	0.02	0.01	0.01	0.01
EM Process Error	0.01	0.59	0.18	0.07	0.04
OM Obs. Error	0.34	0.01	0.08	0.24	0.27
OM F History	0.91	0.22	0.06	1.20	1.67
OM σ_R	<0.01	0.14	—	—	—
OM σ_{2+}	—	0.11	—	—	—
OM σ_M	—	—	0.01	—	—
OM ρ_M	—	—	<0.01	—	—
OM σ_{Sel}	—	—	—	0.01	—
OM ρ_{Sel}	—	—	—	0.01	—
OM σ_q	—	—	—	—	0.01
OM ρ_q	—	—	—	—	0.01
All factors	2.28	1.74	0.51	2.66	3.51
+ All Two Way	4.20	2.74	1.08	5.08	6.51
+ All Three Way	4.83	3.79	1.79	6.03	7.82