The Woods Hole Assessment Model (WHAM): a general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates

Brian C. Stock[1], Timothy J. Miller [1]

[1]brian.stock@noaa.gov, timothy.j.miller@noaa.gov, Northeast Fisheries Science Center, National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543, USA

Abstract

The rapid changes observed in many marine ecosystems that support fisheries pose a challenge to stock assessment and management predicated on time-invariant productivity and considering species in isolation. In single-species assessments, two main approaches have been used to account for productivity changes: allowing biological parameters to vary stochastically over time (empirical), or explicitly linking population processes such as recruitment (R) or natural mortality (M) to environmental covariates (mechanistic). Here, we describe the Woods Hole Assessment Model (WHAM) framework and software package, which combines these two approaches. WHAM can estimate time- and age-varying random effects on annual transitions in numbers at age (NAA), M, and selectivity, as well as fit environmental time-series with process and observation errors, missing data, and nonlinear links to R and M. WHAM can also be configured as a traditional statistical catch-at-age (SCAA) model in order to easily replicate status quo models and test them against models with state-space and environmental effects in a single framework.

We fit models with and without (independent or autocorrelated) random effects on NAA, M, and selectivity to data from five stocks with a broad range of life history, status, number of ages, and time-series length. Models that included random effects performed well across stocks and processes, especially random effects models with a two dimensional (2D) first-order autoregressive (AR1) covariance structure over age and year. We conducted simulation tests and found negligible or no bias in estimation of important assessment outputs (SSB, F, stock status, and catch) when the operating and estimation models matched. However, bias in SSB and F was often non-trivial when the estimation model was less complex than the operating model, especially when models without random effects were fit to data simulated from models with random effects. Bias of the variance and correlation parameters controlling random effects was also negligible or slightly negative as expected. Our results suggest that WHAM can be a useful tool for stock assessment when environmental effects on R or M, or stochastic variation in NAA transitions, M, or selectivity are of interest. In the U.S. Northeast, where the productivity of several groundfish stocks has declined, conducting assessments in WHAM with time-varying processes via random effects or environment-productivity links may account for these trends and potentially reduce retrospective bias.

Keywords

state-space; stock assessment; random effects; time-varying; environmental effects; recruitment; natural mortality; Template Model Builder (TMB)

1 Introduction

The last two decades have increasingly seen a push for more holistic, ecosystem-based fisheries management (Larkin, 1996; Link, 2002). In part, this is a recognition that considering single species in isolation produces riskier and less robust outcomes long-term (Patrick and Link, 2015). In several high-profile cases, fisheries management has failed to prevent collapses because they did not reduce fishing pressure in responses to changes in natural mortality (M), recruitment, or migration patterns caused by dynamics external to the stock in question (Northern cod: Shelton et al., 2006; Rose and Rowe, 2015; Gulf of Maine cod: Pershing et al., 2015; Pacific sardine: Zwolinski and Demer, 2012). This is particularly concerning in the context of climate change and the wide range of biological processes—often assumed to be constant—in stock assessments that are likely to be affected (Stock et al., 2011).

One approach to account for changing productivity is to explicitly link population processes to environmental covariates in single-species stock assessments, i.e. the mechanistic approach *sensu* Punt et al. (2014). Traditional single-species assessments are based on internal population dynamics and the effect of fishing mortality (F), even though fisheries scientists have long known that these are important drivers of time-varying population processes, e.g. recruitment, mortality, growth, and movement (Garstang, 1900; Hjort, 1914). Effects of the environment or interactions with other species can be considered contextually, rather than explicitly as estimated parameters (although temporal variation in empirical weight and maturity at age can affect reference points). Despite how counterintuitive this may seem to ecologists and oceanographers who study such relationships, the evidence for direct linkages to specific environmental covariates is often weak and can break down over time (McClatchie et al., 2010; Myers, 1998). Additionally, the primary goal of most assessments is to provide management advice on near-term sustainable harvest levels—not to explain ecological relationships. Even if an environmental covariate directly affects fish productivity, including the effect in an assessment may not improve management advice if the effect is weak (De Oliveira and Butterworth, 2005). Worse, including environmental effects in an assessment or management system has been shown to actually provide worse management in some cases (De Oliveira and Butterworth, 2005; Punt et al., 2014; Walters and Collie, 1988). This can be true even in cases of relatively well-understood mechanistic links between oceanic conditions and fish populations, as in the case of sea surface temperature and Pacific sardine (Hill et al., 2018; Zwolinski and Demer, 2012). Still, incorporating mechanistic environment-productivity links in assessments does have the potential to reduce residual variance, particularly in periods when few demographic data exist (Miller et al., 2016; Shotwell et al., 2014).

An alternative approach is to allow biological parameters to vary stochastically over time, i.e. the empirical

approach *sensu* Punt et al. (2014). In this case, the variation is caused by a range of sources that are not explicitly modeled. Statistical catch-at-age (SCAA) models typically only estimate year-specific recruitment (R_t) and F_t, often as deviations from a mean that may be a function of spawning biomass, e.g. $\log R_t = \log R_0 + \epsilon_t$. The main reason that other parameters are assumed constant is simply that there are not enough degrees of freedom to estimate many time-varying parameters. One common solution is to penalize the deviations, e.g. $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$, although the penalty terms, σ_ϵ^2, must be fixed or iteratively tuned and are therefore somewhat subjective (Aeberhard et al., 2018; Methot and Taylor, 2011; Methot and Wetzel, 2013; Xu et al., 2020). State-space models that treat parameters as unobserved states can, in principle, avoid such subjectivity by estimating the penalty terms as variance parameters constraining random effects and maximizing the marginal likelihood (Thorson, 2019). In this way, state-space models can allow processes to vary in time while simultaneously estimating fewer parameters.

Although state-space stock assessments have existed for some time (Gudmundsson, 1994; Mendelssohn, 1988; Sullivan, 1992), the recent development of Template Model Builder (TMB, Kristensen et al., 2016) software to perform efficient Laplace approximation has greatly expanded their use (Cadigan, 2016; Miller et al., 2016; Nielsen and Berg, 2014). In addition to the key advantage of objectively estimating variance, or "data weighting", parameters, state-space models naturally predict unobserved states, and therefore handle missing data and short-term projections in a straightforward way (ICES, 2020). In comparisons with SCAA models, state-space models have been shown to have larger, more realistic, uncertainty and lower retrospective bias (Miller and Hyun, 2018; Stock et al., n.d.).

Retrospective bias can occur when changing environmental conditions lead to changes in productivity that are unaccounted for in stock assessments, and this is a concern common to several groundfish stocks on the Northeast U.S. Shelf (Brooks and Legault, 2016; Tableau et al., 2018). The Northeast U.S. Shelf ecosystem is rapidly changing, and this has motivated managers to make the "continue[d] development of stock assessment models that include environmental terms" a top priority (Hare et al., 2016). Applications of state-space models with environmental effects on recruitment, growth, M, and maturity have proven promising (Miller and Hyun, 2018; Miller et al., 2018; O'Leary et al., 2019; Xu et al., 2018). In addition to providing short-term (1-3 years) catch advice with reduced retrospective bias, it is hoped that environment-linked assessments will help create realistic evaluation of sustainable stock and harvest levels in the medium-term (3-10 years) for stocks that have not rebounded in response to dramatic decreases in F.

To address the needs of fisheries management in a changing climate, we seek an assessment framework that combines both the empirical and mechanistic approaches. Namely, it should be able to 1) estimate time-varying parameters as random effects (i.e. a state-space model), and 2) include environmental effects

directly on biological parameters. The framework should also allow for easy testing against status quo SCAA models to ease gradual adoption through the "research track" or "benchmark" assessment process (Lynch et al., 2018). The objectives of this manuscript are to introduce the Woods Hole Assessment Model (WHAM) framework and demonstrate its ability to:

1. estimate time- and age-varying random effects on annual changes in abundance at age, M, and selectivity;
2. fit environmental time-series with process and observation error, missing data, and a link to a population process; and
3. simulate new data and random effects to conduct self- and cross-tests (*sensu* Deroba et al., 2015) to estimate bias in parameters and derived quantities.

Throughout, we describe how the above are implemented using the open-source WHAM software package (Miller and Stock, 2020).

2 Methods

2.1 Model description

WHAM is a generalization and extension of Miller et al. (2016) in TMB. It is in many respects similar to the Age-Structured Assessment Program (ASAP, Legault and Restrepo, 1998; Miller and Legault, 2015) and can be configured to fit statistical catch-at-age models nearly identically. There is functionality built into WHAM to migrate ASAP input files to R inputs needed for WHAM, and WHAM uses many of the same types of data inputs, such as empirical weight-at-age, so that existing assessments in the U.S. Northeast can be easily replicated and tested against models with state-space and environmental effects in a single framework.

2.1.1 Processes with random effects

WHAM primarily diverges from ASAP through the implementation of random effects on three processes: inter-annual transitions in numbers at age (NAA), natural mortality (M), and selectivity (s), as well as allowing effects of environmental covariates ($Ecov$) on recruitment and natural mortality (but see ASAP4; Miller and Legault, 2015). The environmental covariates and their observations are treated using state-space models with true, unobserved values treated as random effects and observation on them having error. Other than environmental covariates, the processes are assumed to have a two dimensional (2D) first-order autoregressive (AR1) covariance structure over age and year, although correlation in either or both dimensions

can be turned off. The 2D AR1 structure has been widely used to model deviations by age and year in the parameters $F_{a,y}$ (Nielsen and Berg, 2014), $M_{a,y}$ (Cadigan, 2016; Stock et al., n.d.), $s_{a,y}$ (Xu et al., 2019), and $N_{a,y}$ (Stock et al., n.d.), as well as in the catch ($C_{a,y}$) and survey index ($I_{a,y}$) observations (Berg and Nielsen, 2016).

2.1.1.1 Numbers at age (NAA)

The stock equations in WHAM that describe the transitions between numbers at age are identical to Miller et al. (2016) and Nielsen and Berg (2014):

$$\log N_{a,y} = \begin{cases} \log\left(f(SSB_{y-1})\right) + \varepsilon_{1,y}, & \text{if } a = 1 \\ \log\left(N_{a-1,y-1}\right) - Z_{a-1,y-1} + \varepsilon_{a,y}, & \text{if } 1 < a < A \\ \log\left(N_{A-1,y-1}e^{-Z_{A-1,y-1}} + N_{A,y-1}e^{-Z_{A,y-1}}\right) + \varepsilon_{A,y}, & \text{if } a = A \end{cases} \quad (1)$$

where $N_{a,y}$ are the numbers at age a in year y, Z is the total mortality rate ($F + M$), f is the stock-recruit function, Y is the total number of observation and prediction years, and A represents the plus-group. In this analysis we demonstrate four possible models for the NAA deviations, $\varepsilon_{a,y}$.

m1 is most similar to a SCAA model, where only recruitment deviations, $\varepsilon_{1,y}$, are estimated (i.e. $\varepsilon_{a,y} = 0$ for $a > 1$ in Eqn 1). In m1, the recruitment deviations are assumed to be independent and identically distributed (IID):

$$\varepsilon_{1,y} \sim \mathcal{N}\left(-\frac{\sigma_R^2}{2}, \sigma_R^2\right)$$

The $-\frac{\sigma_R^2}{2}$ bias correction term is included by default so that the expected recruitment, $E(N_{1,y} = R_y)$ equals the expected R_y from the stock-recruit function (Methot and Taylor, 2011; Thorson, 2019). This bias correction adjustment can also be turned off. The only difference between m1 and a SCAA is that annual recruitments are random effects and σ_R^2 is an estimated parameter within the model.

m2 is the same as m1, except that the recruitment deviations are stationary AR1 with autocorrelation parameter $-1 < \rho_y < 1$:

$$\varepsilon_{1,y+1} \sim \mathcal{N}\left(\rho_y \varepsilon_{1,y} - \frac{\sigma_R^2}{2(1-\rho_y^2)}, \sigma_R^2\right)$$

m3 is the "full state-space" model from Nielsen and Berg (2014) and Miller et al. (2016), where all numbers

at age are independent random effects and:

$$\varepsilon_{a,y} \sim \begin{cases} \mathcal{N}\left(-\frac{\sigma_R^2}{2}, \sigma_R^2\right), & \text{if } a = 1 \\ \mathcal{N}\left(-\frac{\sigma_a^2}{2}, \sigma_a^2\right), & \text{if } a > 1 \end{cases} \tag{2}$$

where σ_a^2 for all ages $a > 1$ are assumed to be the same but different from age $a = 1$, i.e. recruitment. This assumption is sensible because variability of deviations between expected and realized recruitment are typically larger than deviations from expected abundance at older ages.

m4 treats the numbers at all ages as random effects, as in m3, but the *NAA* deviations have a 2D stationary AR1 structure as in Stock et al. (n.d.):

$$\mathbf{E} \sim \mathcal{MVN}\left(0, \Sigma\right)$$

where $\mathbf{E} = (\varepsilon_{1,1}, \ldots, \varepsilon_{1,Y-1}, \varepsilon_{2,1}, \ldots, \varepsilon_{2,Y-1}, \ldots, \varepsilon_{A,1}, \ldots, \varepsilon_{A,Y-1})'$ is a vector of all *NAA* deviations, $\mathbf{\Sigma}$ is the covariance matrix of $\mathbf{E}$ defined by:

$$\mathrm{Cov}\left(\varepsilon_{a,y}, \varepsilon_{\tilde{a},\tilde{y}}\right) = \frac{\sigma_a \sigma_{\tilde{a}} \rho_a^{|a-\tilde{a}|} \rho_y^{|y-\tilde{y}|}}{\left(1-\rho_a^2\right)\left(1-\rho_y^2\right)}$$

and $-1 < \rho_a < 1$ and $-1 < \rho_y < 1$ are the AR1 coefficients in age and year, respectively. As in m3, σ_a^2 for all ages $a > 1$ are assumed to be the same but different from age $a = 1$, σ_R^2. The bias correction term for age $a > 1$ in m4 is $-\frac{\sigma_a^2}{2(1-\rho_y^2)(1-\rho_a^2)}$.

2.1.1.2 Natural mortality (M)

For natural mortality, there are mean parameters for each age, μ_{M_a}, each of which may be estimated freely or fixed at the initial values. The μ_{M_a} may also be estimated in sets of ages, e.g. estimate one mean M shared across ages 3-5, $\mu_{M_3} = \mu_{M_4} = \mu_{M_5}$. There is also an option for M to be specified as a function of weight-at-age, $M_{a,y} = \mu_M \mathrm{W}_{a,y}^b$, as in Lorenzen (1996) and Miller and Hyun (2018). Regardless of whether μ_{M_a} are fixed or estimated, WHAM can also be configured to estimate deviations in M, $\delta_{a,y}$, as random effects analogous to the *NAA* deviations (Cadigan, 2016; Stock et al., n.d.):

$$\begin{aligned} \log\left(M_{a,y}\right) &= \mu_{M_a} + \delta_{a,y} \\ \mathrm{Cov}\left(\delta_{a,y}, \delta_{\tilde{a},\tilde{y}}\right) &= \frac{\sigma_M^2 \varphi_a^{|a-\tilde{a}|} \varphi_y^{|y-\tilde{y}|}}{\left(1-\varphi_a^2\right)\left(1-\varphi_y^2\right)} \end{aligned} \tag{3}$$

where σ^2_M, φ_a, and φ_y are the AR1 variance and correlation coefficients in age and year, respectively.

In this analysis, we demonstrate three alternative M random effects models. For simplicity, all models treat μ_{M_a} as known, as in each of the original assessments. m1 is identical to the base NAA model, with no random effects on M ($\sigma^2_M = \varphi_a = \varphi_y = 0$ and not estimated). m2 allows IID M deviations, estimating σ^2_M but fixing $\varphi_a = \varphi_y = 0$. m3 estimates the full 2D AR1 structure for M deviations.

2.1.1.3 Selectivity (s)

As in ASAP and many other SCAA assessment frameworks, WHAM assumes separability in the fishing mortality rate by age and year, e.g. $F_{a,y} = F_y s_a$, where F_y is the "fully selected" fishing mortality rate in year y and s_a is the selectivity at age. We note that this differs from the approach in SAM (Nielsen and Berg, 2014), where the $F_{a,y}$ are estimated directly as multivariate random effects without the separability assumption. Three parametric forms are available (logistic, double-logistic, and decreasing-logistic), as well as a non-parametric option to estimate each s_a individually ("age-specific"). To allow for temporal changes in selectivity as in ASAP, WHAM can estimate selectivity in user-specified time blocks. WHAM estimates selectivity parameters on the logit scale to avoid boundary problems during estimation.

WHAM estimates annual full F_y and mean selectivity parameters as fixed effects. Deviations in selectivity parameters can be estimated as random effects, $\zeta_{p,y}$, with autocorrelation by parameter (p), year (y), both, or neither. This is done similarly to Xu et al. (2019), except that the deviations are placed on the parameters instead of the mean $s_{a,y}$ in order to guarantee that $0 < s_{a,y} < 1$. For example, logistic selectivity with two parameters a_{50} and k is estimated as:

$$\begin{aligned} s_{a,y} &= \tfrac{1}{1+e^{-(a-a_{50_y})/k_y}} \\ a_{50_y} &= l_{a_{50}} + \tfrac{u_{a_{50}} - l_{a_{50}}}{1+e^{-(\nu_1+\zeta_{1,y})}} \\ k_y &= l_k + \tfrac{u_k - l_k}{1+e^{-(\nu_2+\zeta_{2,y})}} \\ \mathrm{Cov}\left(\zeta_{1,y}, \zeta_{2,\tilde{y}}\right) &= \tfrac{\sigma_s^2 \phi_p \phi_y^{|y-\tilde{y}|}}{\left(1-\phi_p^2\right)\left(1-\phi_y^2\right)} \end{aligned} \tag{4}$$

where ν_1 is the logit-scale mean a_{50} parameter with lower and upper bounds $l_{a_{50}}$ and $u_{a_{50}}$, ν_2 is the logit-scale mean k parameter with lower and upper bounds l_k and u_k, σ^2_s is the AR1 variance, and ϕ_p, and ϕ_y are the AR1 correlation coefficients by parameter and year.

Below, we demonstrate three models with random effect deviations on logistic selectivity, akin to those for M. m1 treats all numbers at age as independent random effects (i.e. NAA m3) but with no random effects on s ($\sigma^2_s = \phi_p = \phi_y = 0$ and not estimated). m2 allows IID s deviations, estimating σ^2_s but fixing $\phi_p = \phi_y = 0$. m3

estimates the full 2D AR1 structure for s deviations.

2.1.1.4 Environmental covariates (*Ecov*)

WHAM models environmental covariate data using state-space models with process and observation components. The true, unobserved values (or "latent states", X_y) are then linked to the population dynamics equations with user-specified lag. For example, recruitment in year y may be influenced by X_{y-1} (lag 1), while natural mortality in year y may be influenced by X_y (lag 0). Multiple environmental covariates may be included, but only as independent processes. The *Ecov* and population model years do not need to match, and missing years are allowed. In particular, including *Ecov* data in the projection period can be useful.

2.1.1.4.1 Process model

There are currently two options in WHAM for the *Ecov* process model: a normal random walk and AR1. We model the random walk as in Miller et al. (2016):

$$X_{y+1}|X_y \sim \mathcal{N}\left(X_y, \sigma_X^2\right)$$

where σ_X^2 is the process variance and X_1 is estimated as a fixed effect parameter. One disadvantage of the random walk is that its variance is nonstationary. In short-term projections, $\hat{X}_y$ will be equal to the last estimate with an observation and the uncertainty of $\hat{X}_y$ will increase over time. If $\hat{X}_y$ influences reference points, this leads to increasing uncertainty in stock status over time as well (Miller et al., 2016).

For this reason, we generally prefer to model X_y as a stationary AR1 process as in Miller et al. (2018):

$$\begin{aligned} X_1 &\sim \mathcal{N}\left(\mu_X, \tfrac{\sigma_X^2}{1-\phi_X^2}\right) \\ X_y &\sim \mathcal{N}\left(\mu_X(1-\phi_X) + \phi_X X_{y-1}, \sigma_X^2\right) \end{aligned} \tag{5}$$

where μ_X, σ_X^2, and $|\phi_X| < 1$ are the marginal mean, variance, and autocorrelation parameters. In addition to having stationary variance, another important difference between the random walk and AR1 in short-term projections is that the AR1 will gradually revert to the mean over time, unless environmental covariate observations are included in the projection period.

2.1.1.4.2 Observation model

The environmental covariate observations, x_y, are assumed to be normally distributed with mean X_y and variance $\sigma^2_{x_y}$:

$$x_y | X_y \sim \mathcal{N}\left(X_y, \sigma^2_{x_y}\right)$$

The observation variance in each year, $\sigma^2_{x_y}$, can be treated as known with year-specific values (as in Miller et al., 2016) or one overall value shared among years. They can also be estimated as parameters, likewise either as yearly values or one overall value. If included, WHAM estimates yearly $\sigma^2_{x_y}$ as random effects with parameters μ_{σ_x} and σ_{σ_x}:

$$\sigma^2_{x_y} \sim \mathcal{N}\left(\mu_{\sigma_x}, \sigma^2_{\sigma_x}\right)$$

2.1.1.4.3 Link to population

WHAM currently provides options to link the modeled environmental covariate, X_y, to the population dynamics via recruitment or natural mortality. It is also sometimes useful to fit the *Ecov* model without a link to the population dynamics so that models with and without environmental effects have the same data in the likelihood and can be compared via AIC.

In the case of recruitment, the options follow Fry (1971) and Iles and Beverton (1998): "controlling" (density-independent mortality), "limiting" (carrying capacity effect, e.g. X_y determines the amount of suitable habitat), "lethal" (threshold effect, i.e. R_t goes to 0 at some X_y value), "masking" (X_y decreases $\mathrm{d}R/\mathrm{d}SSB$, as expected if X_y affects metabolism or growth), and "directive" (e.g. behavioral). Of these, WHAM currently allows controlling, limiting, or masking effects in the Beverton-Holt stock-recruit function, and controlling or masking effects in the Ricker function. For natural mortality, environmental effects are placed on μ_M, shared across ages.

Regardless of where the environment-population link is, the effect can be either linear or polynomial. Nonlinear effects of environmental covariates are common in ecology, and quadratic effects are to be expected in cases where intermediate values are optimal (Agostini et al., 2008; Brett, 1971). WHAM includes a function to calculate orthogonal polynomials in TMB, akin to the `poly()` function in `R`.

In this analysis, we compare five models with limiting effects on Beverton-Holt recruitment:

$$\hat{R}_{y+1} = \frac{\alpha \text{SSB}_y}{1 + e^{\beta_0 + \beta_1 X_y + \beta_2 X_y^2} \text{SSB}_y}$$

where SSB_y is spawning stock biomass in year y, α and β_0 are the standard parameters of the Beverton-Holt function, and β_1 and β_2 are polynomial effect terms that modify β_0 based on the value of the estimated environmental covariate, X_y.

m1 treats X_y as a random walk ($\phi_X = 1$) but does not include an effect on recruitment ($\beta_1 = \beta_2 = 0$). We include m1 in order to compare AIC of the original model to those with effects on recruitment. m2 and m3 also treat X_y as a random walk, but m2 estimates β_1 and m3 estimates both β_1 and β_2. m4 and m5 estimate X_y as an AR1 process, and m4 estimates β_1 and m5 estimates both β_1 and β_2.

2.1.2 Population observation model

Like ASAP, there are observation likelihood components for aggregate catch and abundance index for each fleet and index, and age composition for each fleet and index.

2.1.2.1 Aggregate catch and indices

The predicted catch at age for fleet i, $\hat{C}_{a,y,i}$, is a function of $N_{a,y}$, $M_{a,y}$, $F_{y,i}$, $s_{a,y,i}$, and empirical weight at age, $W_{a,y,i}$:

$$\hat{C}_{a,y,i} = N_{a,y} W_{a,y,i} \left(1 - e^{-Z_{a,y}}\right) \frac{F_{a,y,i}}{Z_{a,y}}.$$

The log-aggregate catch $\hat{C}_{y,i} = \sum_a \hat{C}_{a,y,i}$ observation is assumed to have a normal distribution

$$\log(C_{y,i}) \sim \mathcal{N}\left(\log(\hat{C}_{y,i}) - \tfrac{\sigma^2_{C_{y,i}}}{2}, \sigma^2_{C_{y,i}}\right). \tag{6}$$

where the standard deviation

$$\sigma_{C_{y,i}} = e^{\eta_i} \sigma_{\tilde{C}_{y,i}}$$

is a function of an input standard deviation $\sigma_{\tilde{C}_{y,i}}$ and a fleet-specific parameter η_i that is fixed at 0 by default, but may be estimated. The bias correction term, $-\frac{\sigma^2_{C_{y,i}}}{2}$, is included by default based on Aldrin et al. (2020) but can be turned off.

Observations of aggregate indices of abundance are handled identically to the aggregate catch as in Eqn. 6

except that for index i the predicted index at age is

$$\hat{I}_{a,y,i} = q_i s_{a,y,i} N_{a,y} W_{a,y,i} e^{-Z_{a,y} f_{y,i}}$$

where q_i is the catchability and $f_{y,i}$ is the fraction of the annual time step elapsed when the index is observed. There are options for indices to be in terms of abundance (numbers) or biomass and $W_{a,y,i} = 1$ for the former.

2.1.2.2 Catch and index age composition

WHAM includes several options for the catch and index age compositions including multinomial (default), Dirichlet, Dirichlet-multinomial, logistic normal and a zero-one inflated logistic normal. In all of the applications and simulation studies here, we assumed age composition observations were multinational distributed.

2.1.3 Projections

The default settings for short-term projections follow common practice for stock assessments in the U.S. Northeast: the population is projected three years using the average selectivity, maturity, weight, and natural mortality at age from the last five model years to calculate reference points (NEFSC, 2020a). WHAM implements similar options as ASAP for specifying F_y in the projection years: terminal year F_y, average F over specified years, $F_{X\%}$ (F at X% SPR, where X is specified and 40 by default), user-specified F_y, or F derived from user-specified catch. For all options except user-specified F_y, the uncertainty in projected F is propagated into the uncertainty of projected population attributes. For models with random effects on *NAA*, *M*, or *Ecov*, the default is to continue the stochastic process into the projection years. WHAM does not currently do this for selectivity because, like F, it is a function of management. Instead, selectivity is taken as the average of recent model years.

If the *Ecov* data extend beyond the population model years, WHAM will fit the *Ecov* model to all available data and use the estimated X_y in projections. This may often be the case because lags can exist in both the physical-biological mechanism and the assessment process. As an example, a model with an *Ecov* effect on recruitment may link physical conditions in year t to recruitment in year $t+1$, and the assessment conducted in year t may only use population data through year $t-1$. In this case, 3-year population projections only need the *Ecov* model to be projected one year. While the default handling of *Ecov* projections is to continue the stochastic process, WHAM includes options to use terminal year x_y, x_y averaged over specified years, or specified x_y. The option to specify x_y allows users to investigate how alternative climate projections may

affect the stock.

2.2 Fits to original datasets

We fit the models described above to data from five stocks with a broad range of life history, status, and model dimension (number of ages and years): Southern New England-Mid Atlantic (SNEMA) yellowtail flounder, butterfish, North Sea cod, Icelandic herring, and Georges Bank (GB) haddock (Tables 1 and 2). We fit the *NAA* random effects models to all five stocks since these represent core WHAM functionality. We chose to highlight one stock each for the *M*, *s*, and *Ecov* processes because all models did not converge for all stocks and processes: butterfish (M), GB haddock (s), and SNEMA yellowtail flounder (*Ecov*). We fixed μ_{M_a} at the values used or estimated in the original assessments. Except for the GB haddock s models, we used the same selectivity parameterization and time blocks as in the original assessments. We used the m3 *NAA* model (all NAA deviations are IID random effects) in the s and *Ecov* demonstrations, but the m1 *NAA* model (only recruitment deviations are IID random effects) for the M demonstrations, because the NAA transitions can be interpreted as survival and including random effect deviations on both NAA and M can lead to model non-convergence (Stock et al., n.d.). Within each process, we compared the performance of the different random effects models using AIC.

We fit all models using the open-source statistical software R (R Core Team, 2020) and TMB (Kristensen et al., 2016), as implemented in the R package WHAM (Miller and Stock, 2020). Documentation and tutorials for how to specify additional random effect structures in WHAM are available at https://timjmiller.github.io/wham/. Code and data files to run the analysis presented here are available at https://github.com/brianstock-NOAA/wham-sim.

2.3 Simulation tests

After fitting each model to the original datasets, we used the simulation feature of TMB to conduct self- and cross-tests (*sensu* Deroba et al., 2015). For each model, we generated 100 sets of new data and random effects, keeping the fixed effect parameters constant at values estimated in original fits. We then re-fit all models to datasets simulated under each as an operating model. We calculated the relative error in parameters constraining random effects (Table 1) and quantities of interest, such as spawning stock biomass (SSB), F, $\frac{\mathrm{B}}{\mathrm{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R. We calculated relative error as $\frac{\hat{\theta}_i}{\theta_i} - 1$, where θ_i is the true value for simulated data set i and $\hat{\theta}_i$ is the value estimated from fitting the model to the simulated data. To estimate bias of each estimation model for a given operating model, we calculated 95% confidence intervals of the median relative

error using the binomial distribution (Thompson, 1936). To summarize the bias across simulations and years for each model, we calculated the median quantity across years and took the mean of the medians across simulations. Finally, for each operating model we calculated the proportion of simulations in which each estimation model converged and had the lowest AIC, aggregated across stocks.

3 Results

3.1 Original datasets

The 2D AR1 covariance structure for random effects on NAA, M, and selectivity performed well across stocks and processes, evidenced by lower AIC than the IID random effects models (Fig. 1). This AIC difference was larger for selectivity than NAA or M, but the differences for NAA and M were also non-trivial, ranging from 11.1–53.2.

3.1.1 Numbers-at-age (NAA)

Treating all ages as random effects was strongly supported by AIC, compared to treating only recruitment deviations as random effects (Fig. 1). Including autoregressive NAA random effects was also generally supported by AIC, either the AR1 when only recruitment deviations were random effects or the 2D AR1 when all ages were random effects. The estimated AR and IID random effects were similar, with the AR random effects slightly smoothed compared to the IID random effects (e.g. for Icelandic herring in Fig. 2).

3.1.2 Natural mortality (M)

As for the NAA models, including random effect deviations on M was supported by AIC (Fig. 1). Although the 2D AR1 M model did not converge for North Sea cod, it had the lowest AIC for butterfish and SNEMA yellowtail flounder. In contrast to the NAA models, the patterns in estimated 2D AR1 and IID M random effects differed noticeably (e.g. for butterfish in Fig. 3). The butterfish IID M model estimated elevated M for age 5+ fish early and late in the time-series, with lower M in intervening years. The butterfish 2D AR1 M model estimated a similar, but much exaggerated, pattern for age 5+ fish and reduced M for ages 1 and 4. Since μ_M was held constant, the M random effects models had to decrease M in some ages and years in order to increase M in others.

3.1.3 Selectivity (s)

For Georges Bank haddock, including 2D AR1 random effect deviations on selectivity was strongly supported by AIC (Fig. 1). The 2D AR1 s model estimated similar patterns in selectivity compared to the IID s model, except with smoother variations by age and year (Fig. 4). Compared to the model with constant selectivity, the IID and 2D AR1 models estimated lower s for age 3 in recent years and higher s for age 3 before 1990. They also estimated higher s for age 2 before 1990, especially from 1973 to 1976.

3.1.4 Environmental covariate effect on recruitment (*Ecov*)

As in previous analyses (Miller et al., 2016; Xu et al., 2018), including an effect of the CPI on recruitment for SNEMA yellowtail flounder was clearly supported by AIC (ΔAIC of 20.3-33.0 between m1 and m2-m4, Fig. 1). The AR1-linear model (m4) had the lowest AIC—fitting the CPI using an AR1 model was preferred over the random walk (ΔAIC of 12.7 between m2 and m4), and the quadratic term, β_2 was deemed unnecessary (ΔAIC of 1.3 between m5 and m4). As expected, the AR1 process model estimated higher uncertainty in years with higher observation error (e.g. 1982–1992) and missing observations (2017, Fig. 5A). The CPI negatively influenced recruitment, i.e. $\hat{R}$ was higher following years with lower CPI (lower fall bottom temperature, Fig. 5C). Including the CPI-recruitment link noticeably changed the estimates of SSB and R in some years (compare locations of points between Fig. 5B-C).

3.2 Simulation tests

Several findings were consistent across processes and stocks. When the estimation and operating model were consistent, bias of SSB, F, $\frac{\mathrm{B}}{\mathrm{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R was generally small and not significant based on confidence intervals (Figs. 6–8 and S1–S6). Bias was also generally small when more complex models were fitted to less complicated operating model simulations. In contrast, the bias was often non-trivial when the estimation model was less complex than the operating model, especially when models without random effects were fit to data simulated from models with random effects. Bias in SSB and F were always opposite, i.e. when SSB was biased high, F was biased low, and vice versa. Predicted catch was never biased. Bias of the variance and correlation parameters controlling random effects was generally negligible or negative, as expected (Figs. 9–12). Restricted maximum likelihood (REML) should be used if more accurate estimation of these parameters is a priority.

In cross-tests, the percentage of simulations in which AIC selected the correct model varied between 62–99%

by process, stock, and operating model (Fig. 13). Estimation models more complex than the operating model were more likely to be chosen for NAA models than for M or selectivity.

3.2.1 Numbers-at-age

Across the five stocks, all NAA models estimated SSB, F, $\frac{\mathrm{B}}{\mathrm{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R with very minimal bias in self-tests. An exception was the estimation of F for Icelandic herring, particularly for the SCAA models (median relative error with 95% CI for m1: -0.060 (-0.073, -0.047), m2: -0.060 (-0.073, -0.047), m3: 0.015 (-0.004, 0.034), and m4: 0.049 (0.017, 0.080); Figs. 6 and S7–S9). The convergence rate for most models and stocks was above 95%, again with the exception of the SCAA models for Icelandic herring (Fig. S10a). In cross-tests, SCAA models exhibited non-trivial bias when estimating data simulated from models that treated numbers at all ages as random effects. The degree of bias varied by quantity and stock between -25% and 25% (Figs. 6 and S1–S4). The more complex NAA models estimated all quantities without bias regardless of operating model.

For four of the five stocks, σ_R^2 was estimated without bias in self-tests (Fig. 9). The exception was butterfish, for which σ_R^2 was negatively biased in m1, m2, and m4 (but not m3). In contrast, both m3 and m4 estimated σ_a^2 with negative bias for all five stocks. There was no consistent pattern in bias for the correlation parameters ρ_a and ρ_y.

3.2.2 Natural mortality

SSB, F, $\frac{\mathrm{B}}{\mathrm{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R were estimated without significant bias in self-tests, although m1 and m2 exhibited bias in SSB and $\frac{F}{F_{40\%}}$ when fit to data simulated from m3 (Figs. 7 and S5–S6).

As for the NAA models, σ_R^2 was estimated without bias in the M model self-tests (Fig. 10). σ_M^2 was biased low, φ_y had no bias, and the direction of bias in φ_a was inconsistent.

3.2.3 Selectivity

Models with IID or 2D AR1 s random effects estimated SSB, F, $\frac{\mathrm{B}}{\mathrm{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R with little to no bias across operating models (Fig. 8). The model without s random effects showed substantial bias when fit to data simulated from m2 or m3.

For Georges Bank haddock, the variance and correlation parameters σ_a^2, σ_s^2, ϕ_y, and ϕ_a were all estimated with slight negative bias in self-tests of models with s random effects (Fig. 11).

3.2.4 Ecov-Recruitment

The random walk CPI models estimated σ_X^2 with less bias than the AR1 CPI models (Fig. 12). β_0 was biased high in all models, although this was not significant for the model with lowest AIC (m4, AR1-linear). All models estimated the parameters ϕ_X, α, β_1, and β_2 without significant bias.

4 Discussion

4.1 Overview

Our results suggest that the WHAM package can be a useful tool for stock assessment when environmental effects on recruitment, or stochastic changes in the numbers at age transitions, selectivity, or natural mortality are of interest. The simulation tests showed negligible or no bias in estimation of important assessment outputs (SSB, F, stock and harvest status) when the operating and estimation models matched. The less complex models, without random effects or autoregressive structure, exhibited some bias in cross-tests, while the more complex models did not. In these cases, bias in SSB and F were opposite such that predicted catch was unbiased. The WHAM models with IID or 2D AR1 random effect deviations performed well across stocks and processes, which suggests that they warrant consideration in future stock-specific studies.

4.2 Relationships to other existing assessment model frameworks

WHAM assumes separability in $F_{a,y}$, i.e. $F_{a,y} = F_y s_a$, and estimates annual full F_y as fixed effects and selectivity of surveys or fisheries as constant or time-varying random effects with various autoregressive assumptions possible. Most other assessment frameworks in the U.S, e.g. ASAP, Stock Synthesis (SS; Methot and Wetzel, 2013), an Assessment Model for Alaska (AMAK; Anon., 2015), and the Beaufort Assessment Model (BAM; Williams and Shertzer, 2015), make this same separability assumption which is useful for specifying a fully-selected F in projections to calculate reference points or generating catch advice. Estimating selectivity in time blocks is common practice in these frameworks. In contrast, SAM estimates $F_{a,y}$ directly as a multivariate random walk process (Nielsen and Berg, 2014). The autoregressive models for selectivity parameters under certain configurations of WHAM should allow for similar F at age patterns as in SAM. Stock Synthesis allows 2D AR1 random effects on $s_{a,y}$ instead of the parameters, and then the variance parameter is estimated by an iterative tuning algorithm (Xu et al., 2019). It is unclear whether estimating time-varying selectivity as random effects produces better results than assuming time blocks, or if there are

advantages to using any of the three approaches for estimating time-varying selectivity used by WHAM, SAM, or SS.

Likewise, WHAM and the other U.S.-based assessment models treat catch, index, and composition observations differently than SAM. The U.S-based assessment models treat aggregate and composition observations separately for fisheries and indices whereas SAM treats observations of catch and indices at age $C_{a,y}$ and $I_{a,y}$ as multivariate log-normal. Separate observation models for aggregate and composition observations is natural when sampling for total catch differs from that for length and age composition.

Like SAM, WHAM can estimate interannual transitions in NAA as a random walk processes, but WHAM can also be configured to treat these deviations as stationary autoregressive processes. Alternatively (or simultaneously, Stock et al., n.d.) WHAM can estimate deviations in natural mortality as autoregressive processes like NCAM (Cadigan, 2016). Uniquely, WHAM can model multiple environmental covariate time series as state-space processes and include their effects, possibly nonlinearly, in various ways on recruitment or natural mortality. Although most applications thusfar investigate effects of physical processes on demographic parameters, indices of predation might also be considered (Marshall et al., 2019).

4.3 Future use of WHAM

The productivity of several groundfish stocks in the U.S. Northeast has declined in recent decades, and conducting assessments in WHAM with time-varying processes via random effects or environment-productivity links could account for these trends and potentially reduce retrospective bias (Perretti et al., 2017; Stock et al., n.d.; Tableau et al., 2018). WHAM can be configured to fit SCAA models very similar to ASAP and therefore bridge between the two frameworks. Including random effects or environmental covariates in an assessment model is a large structural change, and simulation cross-tests such as we have demonstrated here should be conducted through the research track process (Lynch et al., 2018). If comparisons against status quo SCAA models prove favorable, WHAM could transition to being used in operational assessments. However, because the details of how to include time-varying processes, as well as the effect on the assessment output, will vary by stock, this evaluation may need to be conducted on a stock-by-stock basis.

Random effects allow for changing productivity and, if they are considered as an autoregressive process, can propagate the effect of these changes on assessment output in short-term projections (Stock et al., n.d.). However, the AR1 process demonstrated here trends to the mean in projections and will not predict values beyond extremes in observed time-series. This is an issue worth examining because many marine ecosystems are changing to such extent that recent conditions are time-series extremes, and conditions in the

near future may continue to expand the range of observations (e.g., sea surface temperatures over the U.S. Northeast Shelf in the last decade; Chen et al., 2020). More complex nonstationary time-series models such as autoregressive integrated moving average (ARIMA) that can forecast beyond observed values could be implemented. Nonlinear (e.g. splines) or double linear integration (Di Lorenzo and Ohman, 2013) techniques could also be worth pursuing, as well as time-delay embedding methods (Munch et al., 2018, 2017), which not only allow for nonstationary dynamics but also do not require a specified functional form. The best approach to making stochastic projections of productivity responses to environmental conditions that are rapidly changing and at time-series extremes merits further research.

In addition to the empirical (i.e. random effects) approach, it is worth considering the mechanistic approach (i.e. explicitly modeled links to environmental covariates) for a given stock whose recruitment or M is suspected to have shifted due to a clear, plausible hypothesized external influence. Several factors may result in a higher likelihood of determining that the mechanistic approach is useful in an assessment: a history of overfishing (Free et al., 2019; but high F can also swamp the signal of an environmental influence, Haltuch and Punt, 2011), more rapid environmental change, stocks at the edge of species' range, opportunistic (short-lived) species (*sensu* Winemiller and Rose, 1992), lower trophic level species, longer time series, periodic signals in which more than once cycle has been recorded (e.g. Pacific Decadal Oscillation and sardine), and stronger signals (wider ranges of observed stock status and environmental conditions) (Free et al., 2019; Haltuch et al., 2019; Haltuch and Punt, 2011; Marshall et al., 2019).

The perception of precision in model output such as projected population biomass or catch advice is affected by whether productivity componenents are treated as either constant or as autoregressive processes, much like whether certain parameters are either fixed or estimated in traditional assessment models. Modeling frameworks such as WHAM allow the performance of these alternative models to be compared and when allowing temporal variation in productivity components is justified, the greater uncertainty in the assessment output is more realistic. Moreover, model uncertainty could be included in our preception of output uncertainty by fitting ensembles of WHAM models that represent alternative states of nature (Anderson et al., 2017; Möllmann et al., 2014). Finally, the simulation capabilities of WHAM can also be useful in situations where variation in productivity is hypothesized, but information to estimate this variation is unavialable. WHAM can be configured as an operating model to simulate plausible environmentally-driven changes in recruitment or M or plausible empirical variation in order to estimate the sensitivity of status quo models.

4.4 Conclusion

A major present-day challenge in fisheries is to assess and manage stocks in a changing environment. We foresee environment-linked stock assessments becoming increasingly feasible and realistic as fisheries and oceanographic time-series lengthen and our ecological understanding deepens. We have developed WHAM with this in mind. Finally, we note that the development of TMB has been a critical advance for fisheries assessment modeling frameworks such as WHAM, allowing us to rapidly fit models that treat population and environmental processes as time-varying random effects in a state-space framework.

Acknowledgements

This research was performed while BCS held an NRC Research Associateship award at the NEFSC. NOAA Fish & Climate grant number ??.

References

Aeberhard, W.H., Mills Flemming, J., Nielsen, A., 2018. Review of State-Space Models for Fisheries Science. Annu. Rev. Stat. Appl. 5, 215–235. https://doi.org/10.1146/annurev-statistics-031017-100427

Agostini, V., Hendrix, A., Hollowed, A., Wilson, C., Pierce, S., Francis, R., 2008. Climate-ocean variability and Pacific hake: A geostatistical modeling approach. Journal of Marine Systems 71, 237–248. https://doi.org/10.1016/j.jmarsys.2007.01.010

Aldrin, M., Tvete, I., Aanes, S., Subbey, S., 2020. The specification of the data model part in the SAM model matters. Fisheries Research 229, 105585. https://doi.org/10.1016/j.fishres.2020.105585

Anderson, S.C., Cooper, A.B., Jensen, O.P., Minto, C., Thorson, J.T., Walsh, J.C., Afflerbach, J., Dickey-Collas, M., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Mosqueira, I., Rosenberg, A.A., Selig, E.R., 2017. Improving estimates of population status and trend with superensemble models. Fish and Fisheries 18, 732–741. https://doi.org/10.1111/faf.12200

Anon., 2015. Assessment Model for Alaska Description of GUI and Instructions.

Berg, C.W., Nielsen, A., 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES J Mar Sci 73, 1788–1797. https://doi.org/10.1093/icesjms/fsw046

Brett, J.R., 1971. Energetic responses of salmon to temperature. A study of some thermal relations in the physiology and freshwater ecology of sockeye salmon (*Oncorhynchus Nerka*). Am Zool 11, 99–113. https://doi.org/10.1093/icb/11.1.99

Brooks, E.N., Legault, C.M., 2016. Retrospective forecasting evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. 73, 935–950. https://doi.org/10.1139/cjfas-2015-0163

Cadigan, N.G., 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences 73, 296–308. https://doi.org/10.1139/cjfas-2015-0047

Chen, Z., Kwon, Y.-O., Chen, K., Fratantoni, P., Gawarkiewicz, G., Joyce, T.M., 2020. Long-Term SST Variability on the Northwest Atlantic Continental Shelf and Slope. Geophys. Res. Lett. 47. https://doi.org/10.1029/2019GL085455

De Oliveira, J., Butterworth, D., 2005. Limits to the use of environmental indices to reduce risk and/or increase yield in the South African anchovy fishery. African Journal of Marine Science 27, 191–203. https:

//doi.org/10.2989/18142320509504078

Deroba, J.J., Butterworth, D.S., Methot, R.D., De Oliveira, J.a.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y.-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.-J.F., 2015. Simulation testing the robustness of stock assessment models to error: Some results from the ICES strategic initiative on stock assessment methods. ICES J Mar Sci 72, 19–30. https://doi.org/10.1093/icesjms/fst237

Di Lorenzo, E., Ohman, M.D., 2013. A double-integration hypothesis to explain ocean ecosystem response to climate forcing. Proceedings of the National Academy of Sciences 110, 2496–2499. https://doi.org/10.1073/pnas.1218022110

Free, C.M., Thorson, J.T., Pinsky, M.L., Oken, K.L., Wiedenmann, J., Jensen, O.P., 2019. Impacts of historical warming on marine fisheries production. Science 363, 979–983. https://doi.org/10.1126/science.aau1758

Fry, F., 1971. The effect of environmental factors on the physiology of fish, in: Fish Physiology. Elsevier, pp. 1–98. https://doi.org/10.1016/S1546-5098(08)60146-6

Garstang, W., 1900. The Impoverishment of the Sea. A Critical Summary of the Experimental and Statistical Evidence bearing upon the Alleged Depletion of the Trawling Grounds. Journal of the Marine Biological Association of the United Kingdom 6, 1–69. https://doi.org/10.1017/S0025315400072374

Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. Applied Statistics 43, 117–126.

Haltuch, M.A., Brooks, E., Brodziak, J., Devine, J., Johnson, K., Klibansky, N., Nash, R., Payne, M., Shertzer, K., Subbey, S., Wells, B., 2019. Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. Fisheries Research 217, 198–216. https://doi.org/10.1016/j.fishres.2018.12.016

Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Canadian Journal of Fisheries and Aquatic Sciences 68, 912–926. https://doi.org/10.1139/f2011-030

Hare, J.A., Borggaard, D.L., Friedland, K.D., Anderson, J., Burns, P., Chu, K., Clay, P.M., Collins, M.J., Cooper, P., Fratantoni, P.S., Johnson, M.R., Manderson, J.P., Milke, L., Miller, T.J., Orphanides, C.D., Saba, V.S., 2016. Northeast Regional Action Plan - NOAA Fisheries Climate Science Strategy (No. NMFS-NE-239). NOAA Fisheries, Northeast Fisheries Science Center, Woods Hole, MA.

Hill, K.T., Crone, P.R., Zwolinski, J.P., 2018. Assessment of the Pacific sardine resource in 2018 for U.S. Management in 2018-2019 (No. NOAA Technical Memorandum NMFS-SWFSC-600). US Department of Commerce.

Hjort, J., 1914. Fluctuations in the great fisheries of Northern Europe viewed in the light of biological research. Rapports et Procès-Verbaux des Réunions du Conseil Permanent International Pour L'Exploration de la Mer 20, 1–228.

ICES, 2020. Workshop on the review and future of state space stock assessment models in ICES (WKRFSAM). ICES Scientific Reports 2, 23p. https://doi.org/10.17895/ices.pub.6004

ICES, 2017a. Report of the working group on assessment of demersal stocks in the North Sea and Skagerrak (2017). ICES CM 2017/ACOM:21, 26 April-5 May 2017, ICES HQ.

ICES, 2017b. Report of the North Western Working Group (NWWG). ICES CM 2017/ACOM:08, 27 April-4 May 2017, Copenhagen, Denmark.

Iles, T.C., Beverton, R.J.H., 1998. Stock, recruitment and moderating processes in flatfish. Journal of Sea Research, Proceedings of the Third International Symposium on Flatfish Ecology, Part II 39, 41–55. https://doi.org/10.1016/S1385-1101(97)00022-1

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., Bell, B.M., 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software 70, 1–21. https://doi.org/10.18637/jss.v070.i05

Larkin, P., 1996. Concepts and issues in marine ecosystem management. Reviews in Fish Biology and Fisheries 6, 139–164. https://doi.org/10.1007/BF00182341

Legault, C.M., Restrepo, V.R., 1998. A Flexible Forward Age-Structured Assessment Program (No. 49).

Link, J.S., 2002. What Does Ecosystem-Based Fisheries Management Mean? Fisheries 27, 5.

Lorenzen, K., 1996. The relationship between body weight and natural mortality in juvenile and adult fish: A comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49, 627–642. https://doi.org/10.1111/j.1095-8649.1996.tb00060.x

Lynch, P.D., Methot, R.D., Link, J.S. (Eds.), 2018. Implementing a Next Generation Stock Assessment Enterprise. An Update to the NOAA Fisheries Stock Assessment Improvement Plan, in:. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/ SPO-183, p. 127. https://doi.org/10.7755/TMSPO.183

Marshall, K.N., Koehn, L.E., Levin, P.S., Essington, T.E., Jensen, O.P., 2019. Inclusion of ecosystem information in US fish stock assessments suggests progress toward ecosystem-based fisheries management.

ICES J Mar Sci 76, 1–9. https://doi.org/10.1093/icesjms/fsy152

McClatchie, S., Goericke, R., Auad, G., Hill, K., 2010. Re-assessment of the stockRecruit and temperatureRecruit relationships for Pacific sardine (Sardinops sagax). Can. J. Fish. Aquat. Sci. 67, 1782–1790. https://doi.org/10.1139/F10-101

Mendelssohn, R., 1988. Some problems in estimating population sizes from catch-at-age data. Fishery Bulletin 86, 617–630.

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68, 1744–1760. https://doi.org/10.1139/f2011-092

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142, 86–99. https://doi.org/10.1016/j.fishres.2012.10.012

Miller, T.J., Hare, J.A., Alade, L.A., 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences 73, 1261–1270. https://doi.org/10.1139/cjfas-2015-0339

Miller, T.J., Hyun, S.-Y., 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Canadian Journal of Fisheries and Aquatic Sciences 75, 691–703. https://doi.org/10.1139/cjfas-2017-0035

Miller, T.J., Legault, C.M., 2015. Technical details for ASAP version 4 (No. Ref Doc. 15-17). US Dept Commer, Northeast Fish Sci Cent.

Miller, T.J., O'Brien, L., Fratantoni, P.S., 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. Can. J. Fish. Aquat. Sci. 1–13. https://doi.org/10.1139/cjfas-2017-0124

Miller, T.J., Stock, B.C., 2020. The Woods Hole Assessment Model (WHAM).

Möllmann, C., Lindegren, M., Blenckner, T., Bergström, L., Casini, M., Diekmann, R., Flinkman, J., Müller-Karulis, B., Neuenfeldt, S., Schmidt, J.O., Tomczak, M., Voss, R., Gårdmark, A., 2014. Implementing ecosystem-based fisheries management: From single-species to integrated ecosystem assessment and advice for Baltic Sea fish stocks. ICES J Mar Sci 71, 1187–1197. https://doi.org/10.1093/icesjms/fst123

Munch, S.B., Giron-Nava, A., Sugihara, G., 2018. Nonlinear dynamics and noise in fisheries recruitment: A

602 global meta-analysis. Fish and Fisheries 19, 964–973. https://doi.org/10.1111/faf.12304

603 Munch, S.B., Poynor, V., Arriaza, J.L., 2017. Circumventing structural uncertainty: A Bayesian perspective
604 on nonlinear forecasting for ecology. Ecological Complexity 32, 134–143. https://doi.org/10.1016/j.ecocom.
605 2016.08.006

606 Myers, R.A., 1998. When do environment-recruitment correlations work? Reviews in Fish Biology and
607 Fisheries 8, 285–305.

608 NEFSC, 2020a. Operational assessment of 14 Northeast groundfish stocks, updated through 2018. U.S. Dept.
609 Commer., NOAA, NMFS, NEFSC, Woods Hole, MA.

610 NEFSC, 2020b. Butterfish 2020 assessment update report. U.S. Dept. Commer., NOAA, NMFS, NEFSC,
611 Woods Hole, MA.

612 Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space
613 models. Fisheries Research 158, 96–101. https://doi.org/10.1016/j.fishres.2014.01.014

614 O'Leary, C.A., Miller, T.J., Thorson, J.T., Nye, J.A., 2019. Understanding historical summer flounder (
615 *Paralichthys Dentatus*) abundance patterns through the incorporation of oceanography-dependent vital rates
616 in Bayesian hierarchical models. Can. J. Fish. Aquat. Sci. 76, 1275–1294. https://doi.org/10.1139/cjfas-
617 2018-0092

618 Patrick, W.S., Link, J.S., 2015. Myths that Continue to Impede Progress in Ecosystem-Based Fisheries
619 Management. Fisheries 40, 155–160. https://doi.org/10.1080/03632415.2015.1024308

620 Perretti, C.T., Fogarty, M.J., Friedland, K.D., Hare, J.A., Lucey, S.M., McBride, R.S., Miller, T.J., Morse,
621 R.E., O'Brien, L., Pereira, J.J., Smith, L.A., Wuenschel, M.J., 2017. Regime shifts in fish recruitment on
622 the Northeast US Continental Shelf. Marine Ecology Progress Series 574, 1–11. https://doi.org/10.3354/
623 meps12183

624 Pershing, A.J., Alexander, M.A., Hernandez, C.M., Kerr, L.A., Bris, A.L., Mills, K.E., Nye, J.A., Record,
625 N.R., Scannell, H.A., Scott, J.D., Sherwood, G.D., Thomas, A.C., 2015. Slow adaptation in the face of rapid
626 warming leads to collapse of the Gulf of Maine cod fishery. Science 350, 809–812. https://doi.org/10.1126/
627 science.aac9819

628 Punt, A.E., A'mar, T., Bond, N.A., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haltuch, M.A.,
629 Hollowed, A.B., Szuwalski, C., 2014. Fisheries management under climate and environmental uncertainty:
630 Control rules and performance simulation. ICES J Mar Sci 71, 2208–2220. https://doi.org/10.1093/icesjms/
631 fst057

R Core Team, 2020. R: A Language and Environment for Statistical Computing.

Rose, G.A., Rowe, S., 2015. Northern cod comeback. Can. J. Fish. Aquat. Sci. 72, 1789–1798. https://doi.org/10.1139/cjfas-2015-0346

Shelton, P.A., Sinclair, A.F., Chouinard, G.A., Mohn, R., Duplisea, D.E., 2006. Fishing under low productivity conditions is further delaying recovery of Northwest Atlantic cod (Gadus morhua). Can. J. Fish. Aquat. Sci. 63, 235–238. https://doi.org/10.1139/f05-253

Shotwell, S.K., Hanselman, D.H., Belkin, I.M., 2014. Toward biophysical synergy: Investigating advection along the Polar Front to identify factors influencing Alaska sablefish recruitment. Deep Sea Research Part II: Topical Studies in Oceanography 107, 40–53. https://doi.org/10.1016/j.dsr2.2012.08.024

Stock, B.C., Xu, H., Miller, T.J., Thorson, J.T., Nye, J.A., n.d. Implementing a 2-dimensional smoother on either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder.

Stock, C.A., Alexander, M.A., Bond, N.A., Brander, K.M., Cheung, W.W., Curchitser, E.N., Delworth, T.L., Dunne, J.P., Griffies, S.M., Haltuch, M.A., Hare, J.A., Hollowed, A.B., Lehodey, P., Levin, S.A., Link, J.S., Rose, K.A., Rykaczewski, R.R., Sarmiento, J.L., Stouffer, R.J., Schwing, F.B., Vecchi, G.A., Werner, F.E., 2011. On the use of IPCC-class models to assess the impact of climate on Living Marine Resources. Progress in Oceanography 88, 1–27. https://doi.org/10.1016/j.pocean.2010.09.001

Sullivan, P.J., 1992. A Kalman filter approach to catch-at-length analysis. Biometrics 48, 237–257.

Tableau, A., Collie, J.S., Bell, R.J., Minto, C., 2018. Decadal changes in the productivity of New England fish populations. Can. J. Fish. Aquat. Sci. 76, 1528–1540. https://doi.org/10.1139/cjfas-2018-0255

Thompson, W.R., 1936. On Confidence Ranges for the Median and Other Expectation Distributions for Populations of Unknown Distribution Form. Ann. Math. Statist. 7, 122–128. https://doi.org/10.1214/aoms/1177732502

Thorson, J.T., 2019. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. Fisheries Research 217, 133–139. https://doi.org/10.1016/j.fishres.2018.02.005

Walters, C.J., Collie, J.S., 1988. Is Research on Environmental Factors Useful to Fisheries Management? Can. J. Fish. Aquat. Sci. 45, 1848–1854. https://doi.org/10.1139/f88-217

Williams, E.H., Shertzer, K.W., 2015. Technical documentation of the Beaufort Assessment Model (BAM).

Winemiller, K.O., Rose, K.A., 1992. Patterns of Life-History Diversification in North American Fishes:

Implications for Population Regulation. Canadian Journal of Fisheries and Aquatic Sciences 49, 2196–2218. https://doi.org/10.1139/f92-242

Xu, H., Miller, T.J., Hameed, S., Alade, L.A., Nye, J.A., 2018. Evaluating the utility of the Gulf Stream Index for predicting recruitment of Southern New England-Mid Atlantic yellowtail flounder. Fisheries Oceanography 27, 85–95. https://doi.org/10.1111/fog.12236

Xu, H., Thorson, J.T., Methot, R.D., 2020. Comparing the performance of three data weighting methods when allowing for time-varying selectivity. Can. J. Fish. Aquat. Sci. 77, 247–263. https://doi.org/10.1139/cjfas-2019-0107

Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2019. A new semi-parametric method for autocorrelated age- and time-varying selectivity in age-structured assessment models. Can. J. Fish. Aquat. Sci. 76, 268–285. https://doi.org/10.1139/cjfas-2017-0446

Zwolinski, J.P., Demer, D.A., 2012. A cold oceanographic regime with high exploitation rates in the Northeast Pacific forecasts a collapse of the sardine stock. Proceedings of the National Academy of Sciences 109, 4175–4180. https://doi.org/10.1073/pnas.1113806109

Table 1: Model descriptions and estimated parameters. Parameter descriptions and equations are given in text. Note that the base model in the M module is NAA m1, and the base model in the Selectivity and Ecov-Recruitment modules is NAA m3. Ecov m1 fits the Cold Pool Index data and estimates σ_x in order to allow comparison to m2-m5 using AIC (same data needed in likelihood).

Model	Description	Estimated parameters
Numbers-at-age (NAA)		
m1: SCAA (IID)	Recruitment deviations are IID random effects	σ_R
m2: SCAA (AR1)	Recruitment deviations are autocorrelated (AR1) random effects	σ_R, ρ_y
m3: NAA (IID)	All NAA deviations are IID random effects	σ_R, σ_a
m4: NAA (2D AR1)	All NAA deviations are random effects with correlation by year and age (2D AR1)	σ_R, σ_a, ρ_y, ρ_a
Natural mortality (M)		
m1: none	No random effects on M	σ_R
m2: IID	M deviations are IID random effects	σ_R, σ_M
m3: 2D AR1	M deviations are random effects with correlation by year and age (2D AR1)	σ_R, σ_M, φ_y, φ_a
Selectivity (Sel)		
m1: none	No random effects on selectivity	σ_R, σ_a
m2: IID	Selectivity deviations are IID random effects	σ_R, σ_a, σ_{Sel}
m3: 2D AR1	Selectivity deviations are random effects with correlation by year and age (2D AR1)	σ_R, σ_a, σ_{Sel}, ϕ_y, ϕ_a
Ecov-Recruitment (Ecov)		
m1: RW-none	Ecov: random walk (RW), effect on β: none	σ_R, σ_a, σ_x
m2: RW-linear	Ecov: random walk (RW), effect on β: linear	σ_R, σ_a, σ_x, β_1
m3: RW-poly	Ecov: random walk (RW), effect on β: 2nd order polynomial (poly)	σ_R, σ_a, σ_x, β_1, β_2
m4: AR1-linear	Ecov: autocorrelated (AR1), effect on β: linear	σ_R, σ_a, σ_x, ϕ_x, β_1
m5: AR1-poly	Ecov: autocorrelated (AR1), effect on β: 2nd order polynomial (poly)	σ_R, σ_a, σ_x, ϕ_x, β_1, β_2

Table 2: Stocks used in simulation tests.

	Modules tested				Model dim		Biol. par.		Stock status		
Stock	NAA	M	Sel	Ecov	# Ages	# Years	M	σ_R	$\frac{B}{B_{40}}$	$\frac{F}{F_{40}}$	Source
SNEMA yellowtail flounder	x	x		x	6	49	0.2-0.4	1.67	0.01	0.44	NEFSC (2020a)
Butterfish	x	x			5	31	1.3	0.23	2.57	0.03	NEFSC (2020b)
North Sea cod	x	x			6	54	0.2-1.2	0.87	0.14	2.00	ICES (2017a)
Icelandic herring	x				11	30	0.1	0.55	0.40	1.81	ICES (2017b)
Georges Bank haddock	x		x		9	86	0.2	1.65	5.16	0.12	NEFSC (2020a)

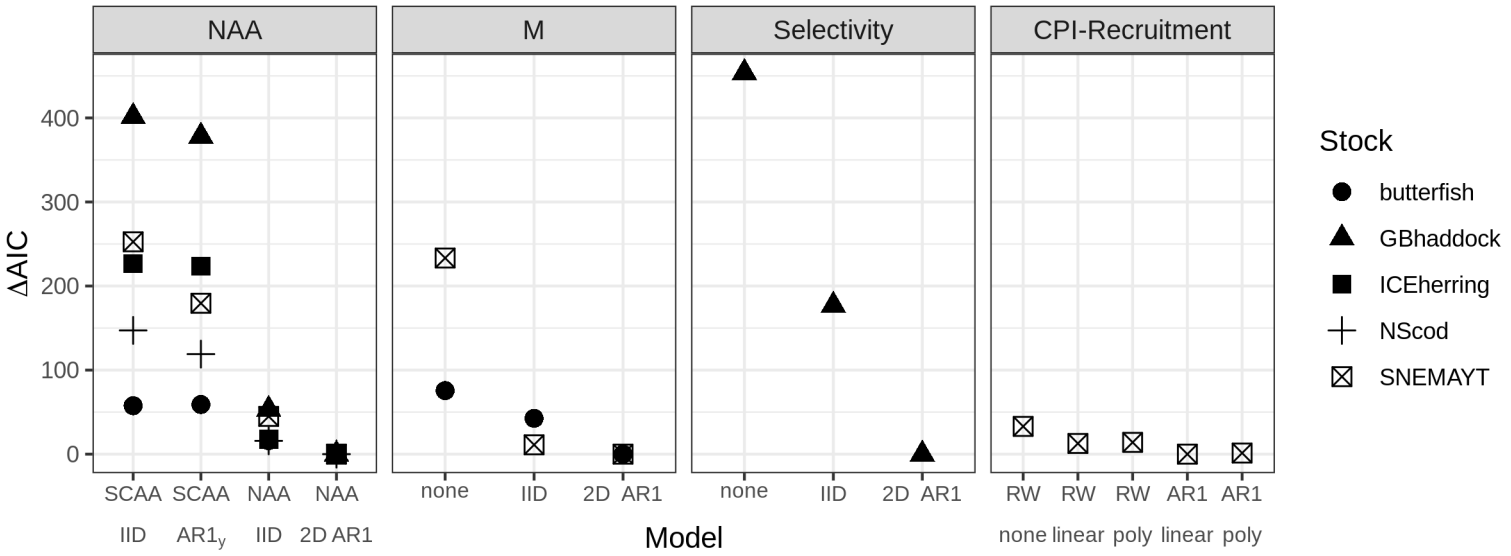

Figure 1: AIC differences by model and stock when fit to original datasets. Stock abbreviations: SNEMA yellowtail flounder (SNEMAYT), North Sea cod (NScod), Icelandic herring (ICEherring), and Georges Bank haddock (GBhaddock).

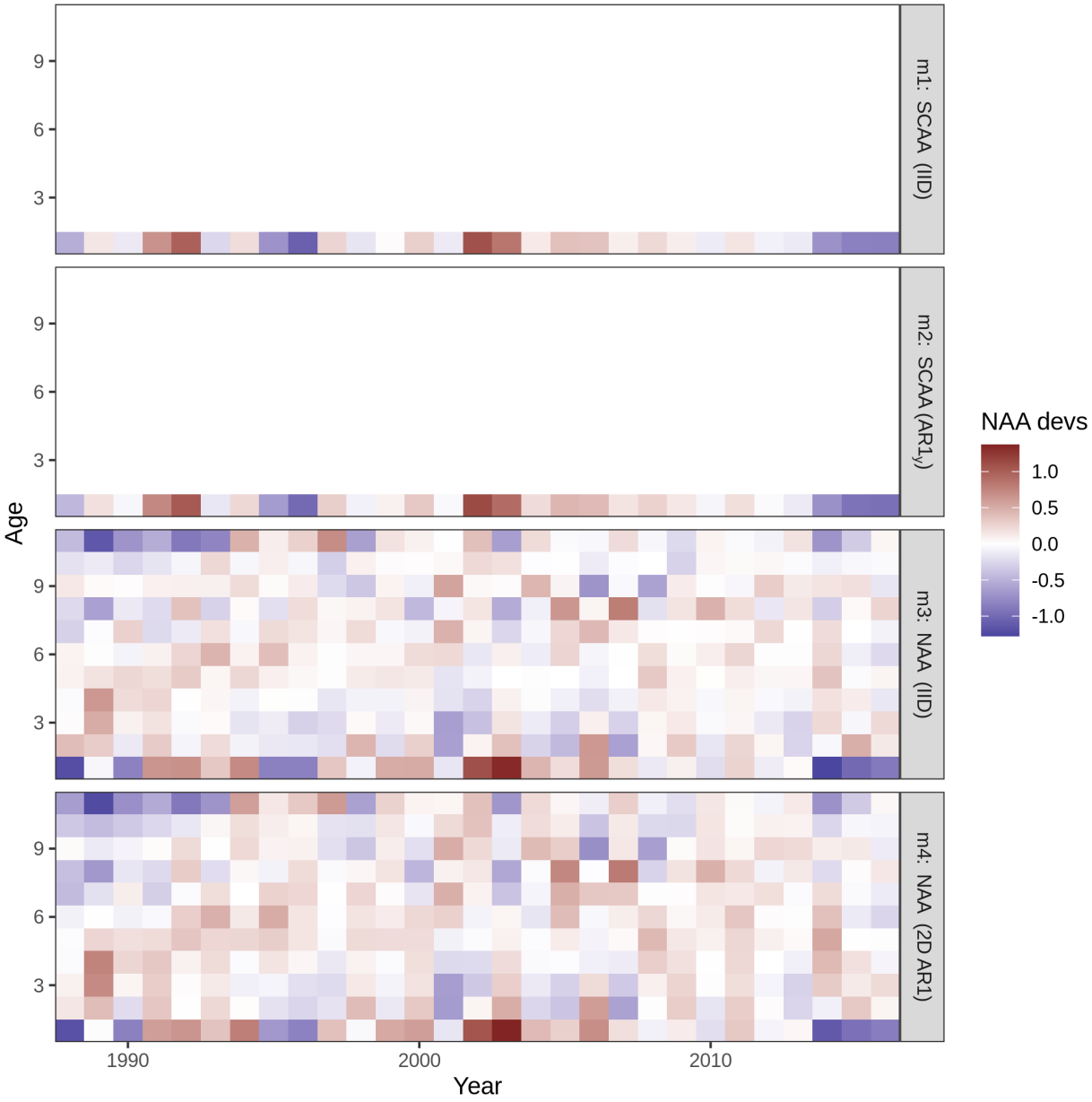

Figure 2: Abundanc-at-age deviations estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (IID). m2 = as m1, but with autocorrelated recruitment deviations ($AR1_y$). m3 = all NAA deviations are IID random effects. m4 = as m3, but deviations are correlated by age and year (2D AR1).

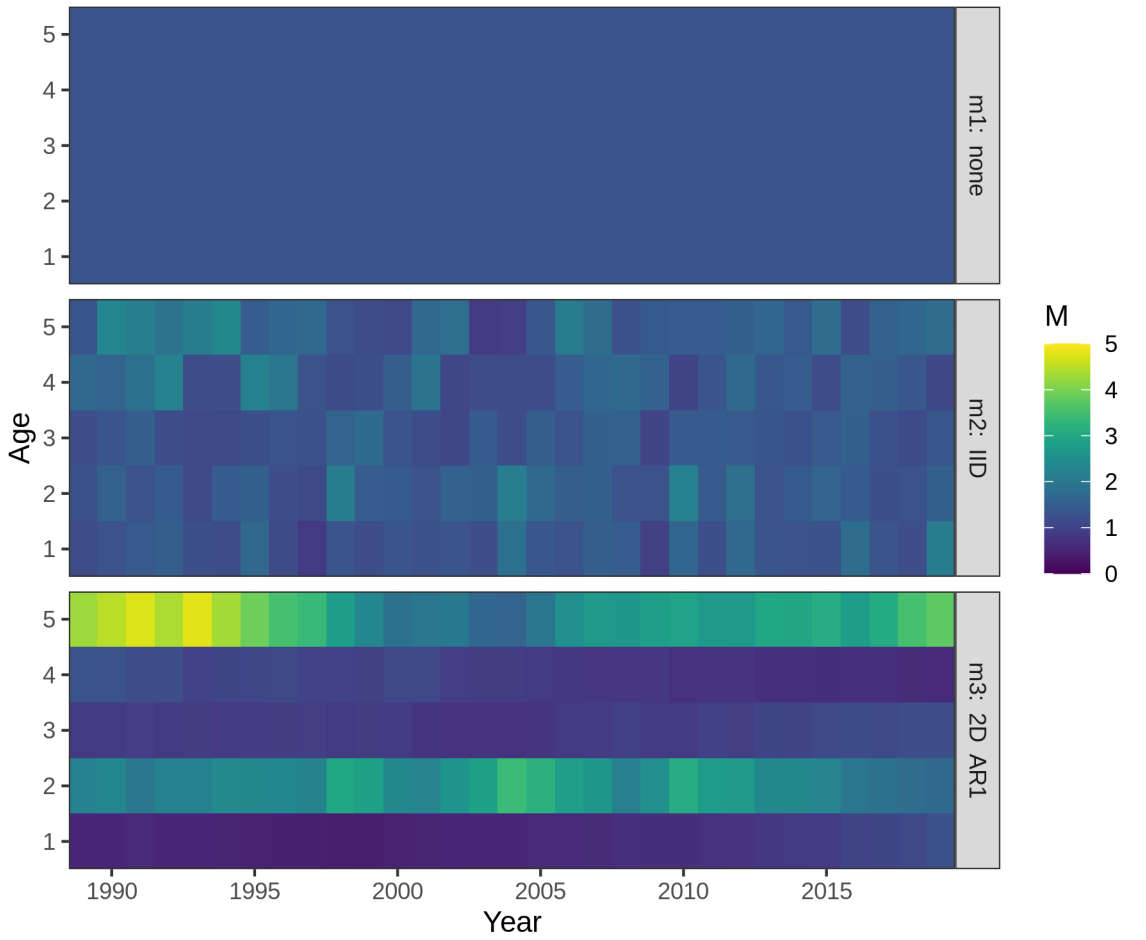

Figure 3: Natural mortality (M) estimated for butterfish using three random effects models. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).

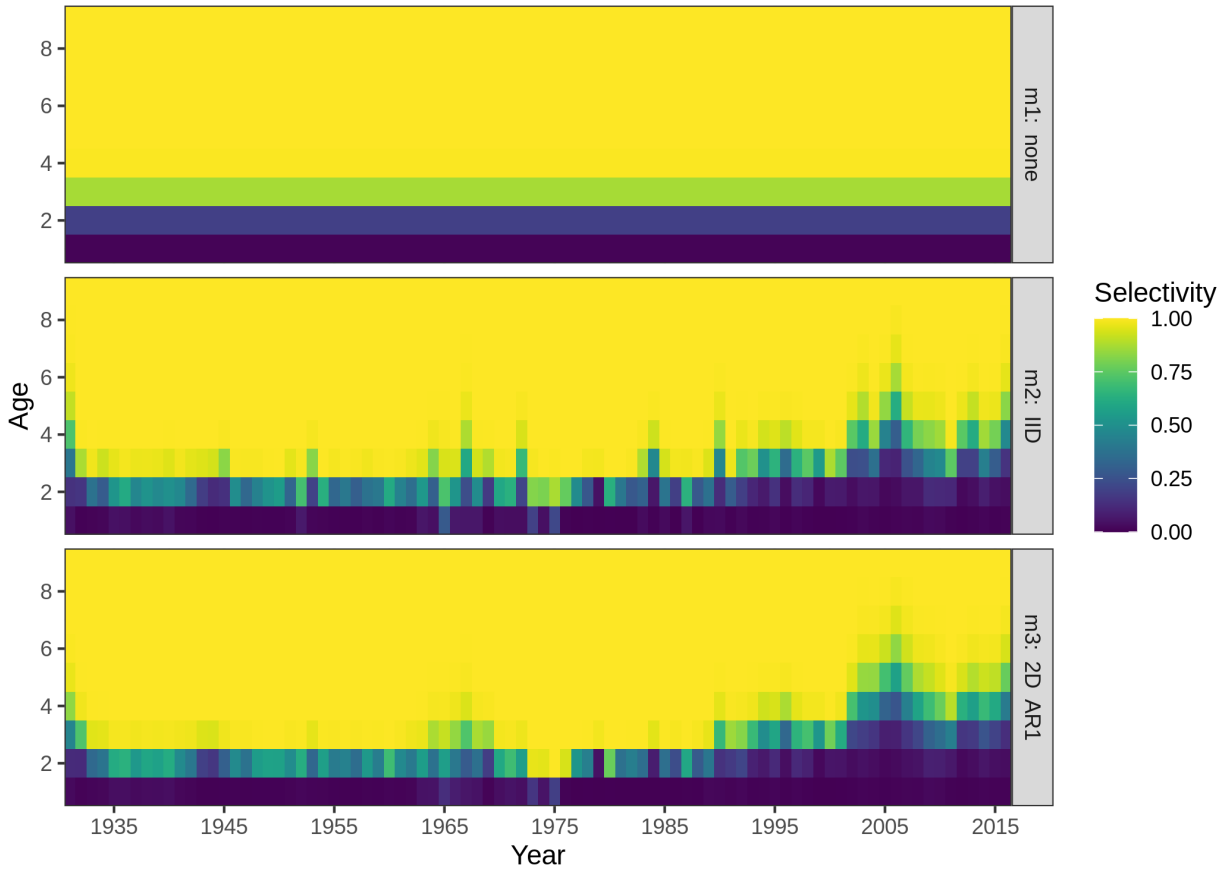

Figure 4: Selectivity estimated for Georges Bank haddock using three random effects models. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).

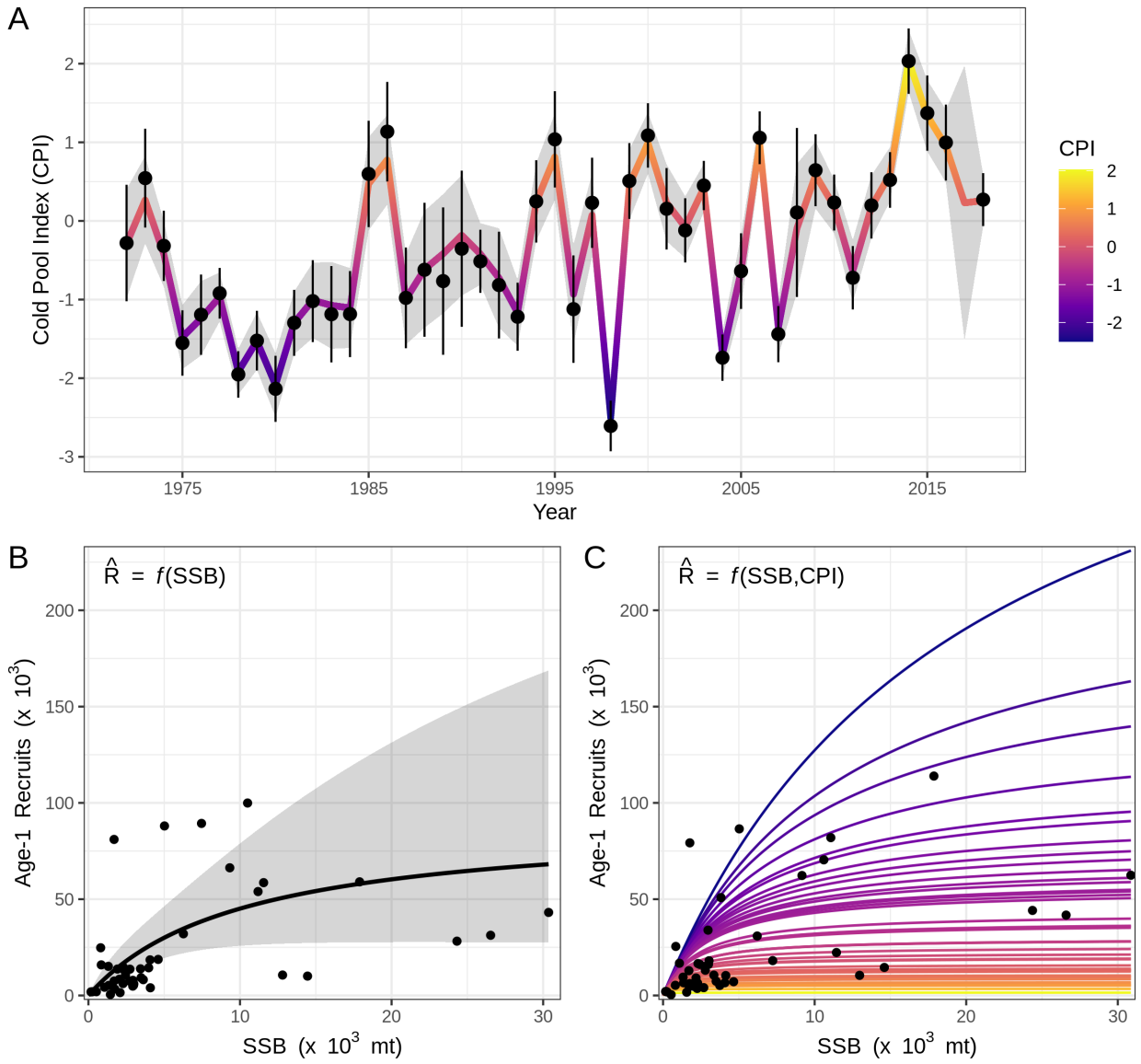

Figure 5: Beverton-Holt stock-recruit relationships fit for Southern New England-Mid Atlantic yellowtail flounder, with and without effects of the Cold Pool Index (CPI). A) CPI estimated from the model with lowest AIC (m4, AR1-linear). Points are observations with 95% CI, and the line with shading is the model-estimated CPI with 95% CI. Note the increased uncertainty surrounding the CPI estimate in 2017 (no observation). B) Estimates of spawning stock biomass (SSB), recruitment, and the stock-recruit function from the model without a CPI effect, m1. C) Estimates of SSB and recruitment from m4, with an effect of the CPI on β. Lines depict the expected stock-recruit relationship in each year t, given the CPI in year $t-1$ (color).

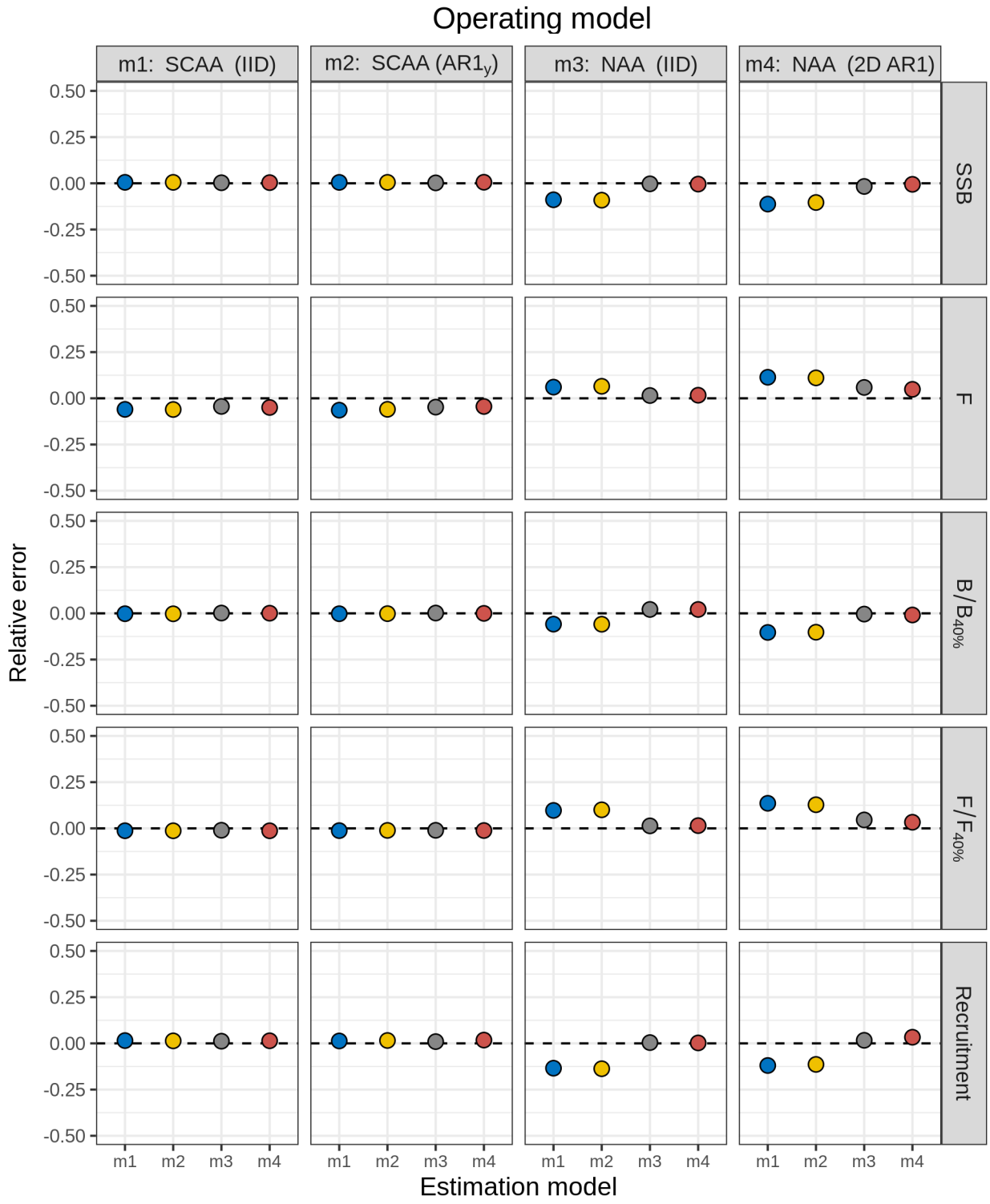

Figure 6: Relative error of key quantities estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (IID). m2 = as m1, but with autocorrelated recruitment deviations ($AR1_y$). m3 = all NAA deviations are IID random effects. m4 = as m3, but deviations are correlated by age and year (2D AR1).

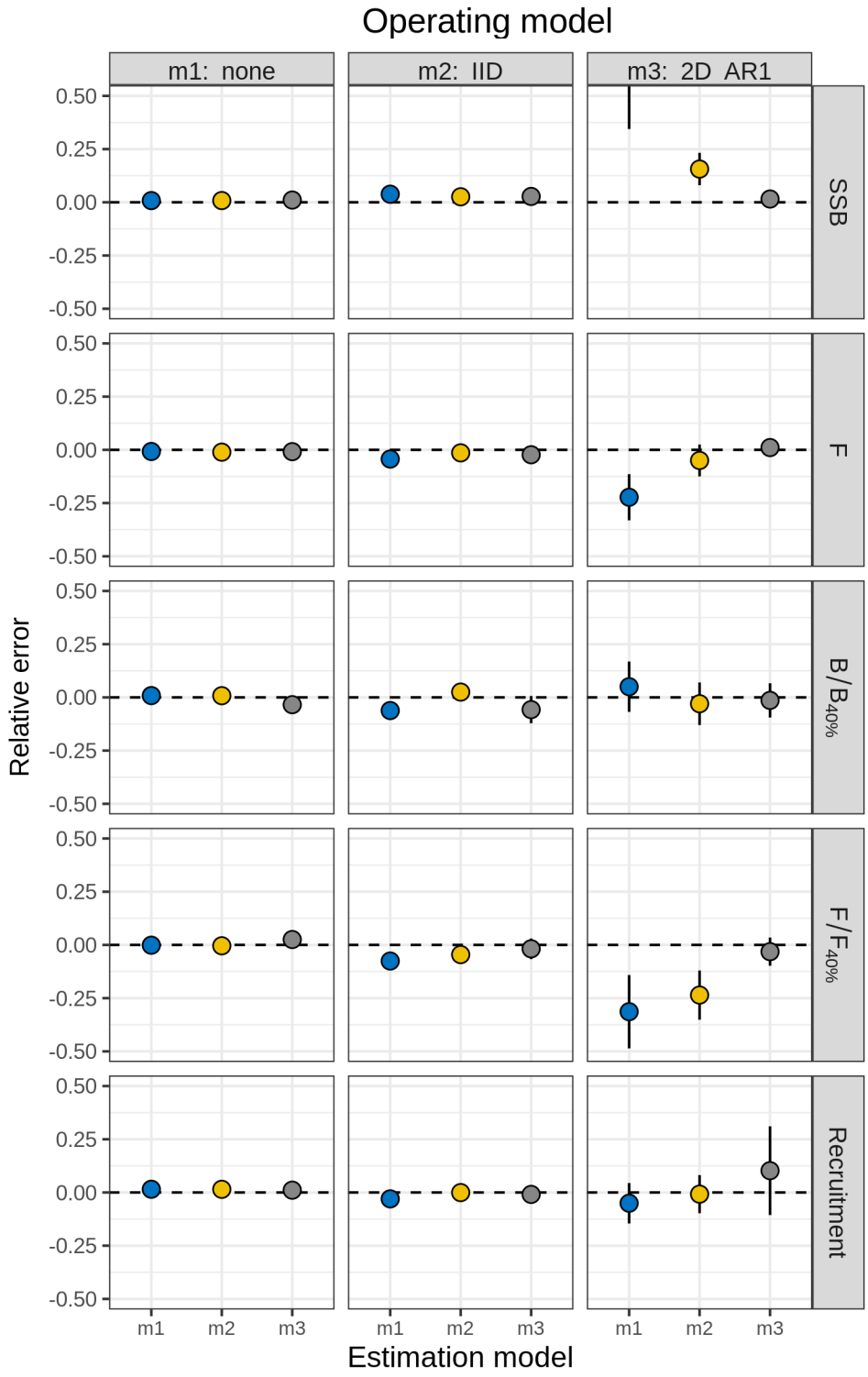

Figure 7: Relative error of key quantities estimated for butterfish using three models of natural mortality (M) random effects. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).

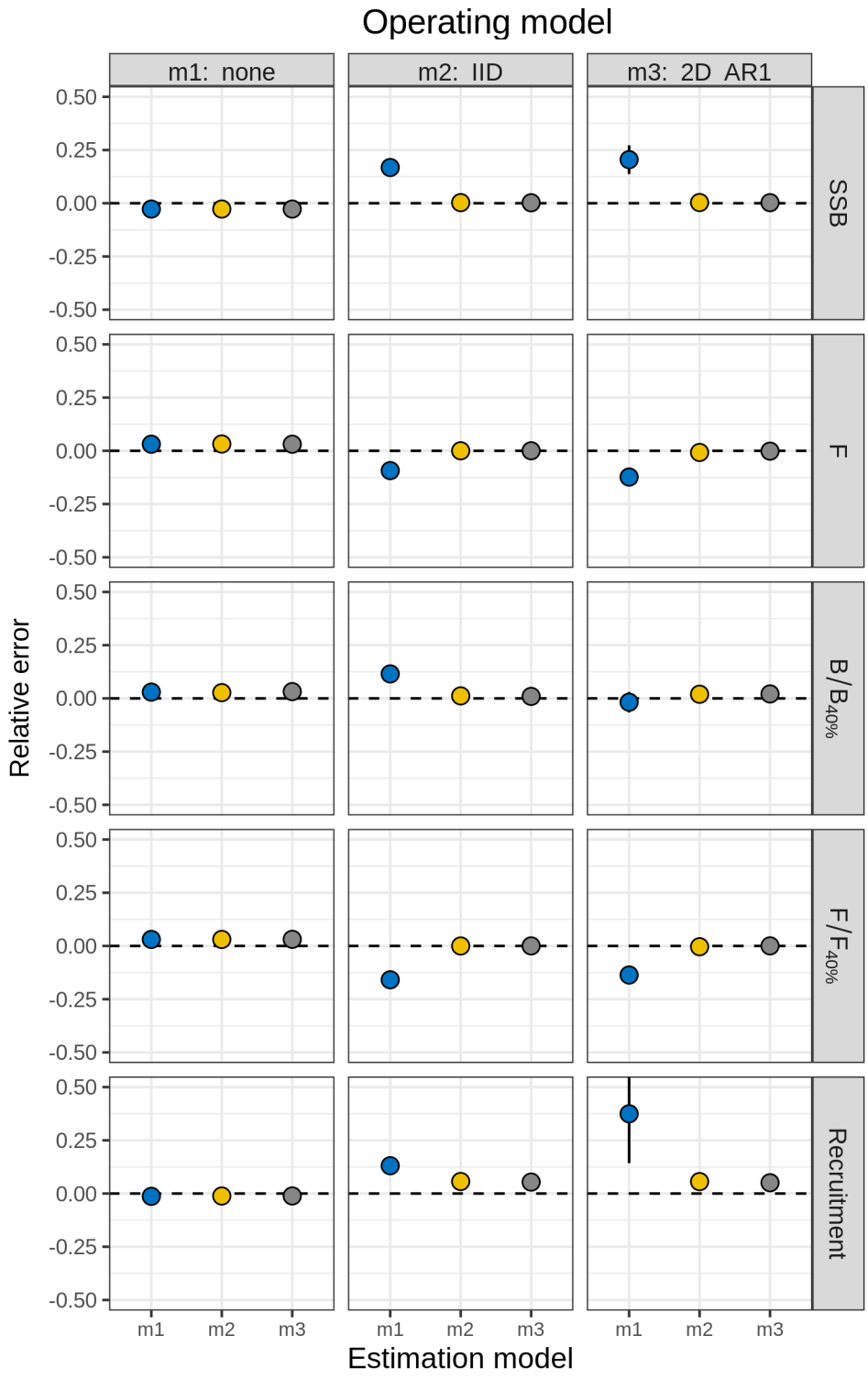

Figure 8: Relative error of key quantities estimated for Georges Bank haddock using three models of selectivity random effects. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).

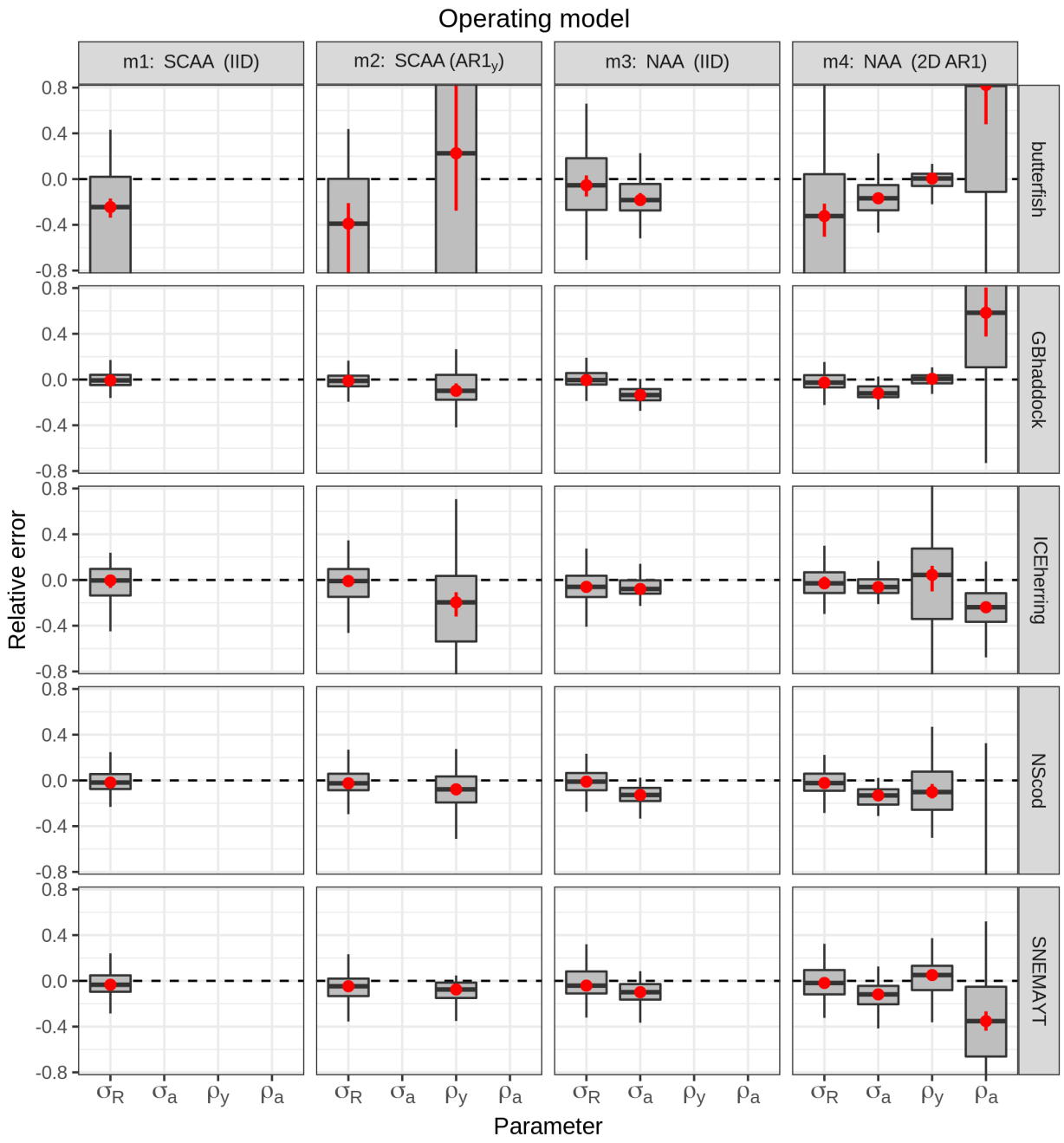

Figure 9: Relative error of parameters constraining numbers-at-age (NAA) random effects. Four models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (IID). m2 = as m1, but with autocorrelated recruitment deviations ($AR1_y$). m3 = all NAA deviations are IID random effects. m4 = as m3, but deviations are correlated by age and year (2D AR1). Relative error was calculated as $\frac{\hat{\theta}_i}{\theta} - 1$, where $\hat{\theta}_i$ was the estimate in simulation i for parameter θ, and θ was the true value (estimate from original dataset). Red points and lines show median relative error with 95% CI.

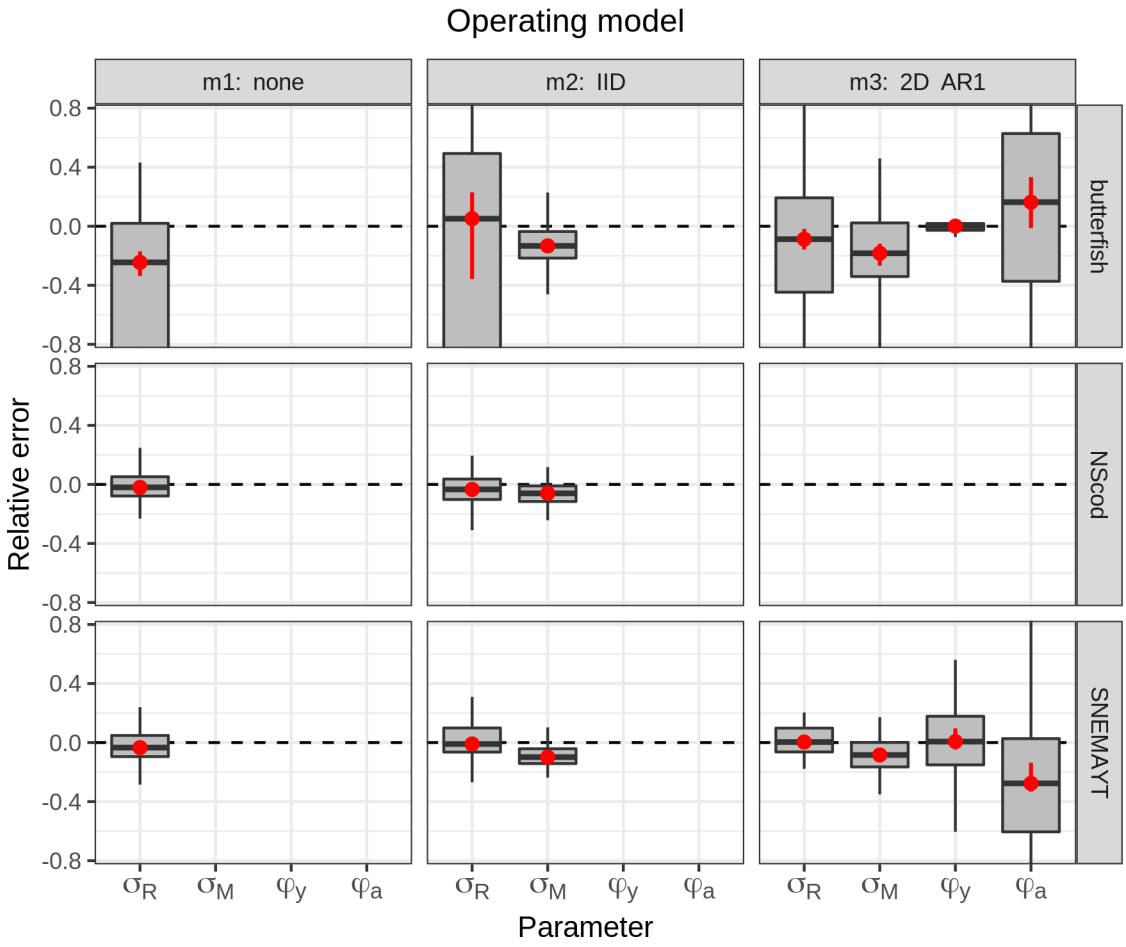

Figure 10: Relative error of parameters constraining natural mortality (M) random effects. Three models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = no random effects on M. m2 = M deviations were independent and identically distributed (IID). m3 = M deviations were correlated by age and year (2D AR1). Relative error was calculated as $\frac{\hat{\theta}_i}{\theta} - 1$, where $\hat{\theta}_i$ was the estimate in simulation i for parameter θ, and θ was the true value (estimate from original dataset). Red points and lines show median relative error with 95% CI. Stock abbreviations: SNEMA yellowtail flounder (SNEMAYT) and North Sea cod (NScod, m3 did not converge).

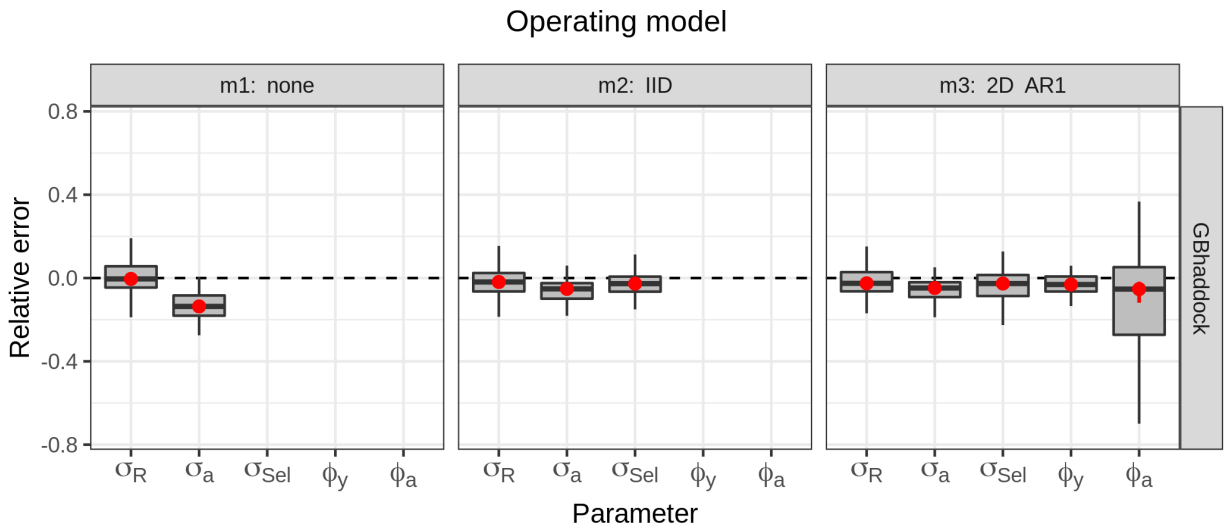

Figure 11: Relative error of parameters constraining selectivity random effects for Georges Bank haddock (GBhaddock). Three models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = no random effects (constant selectivity). m2 = selectivity deviations were independent and identically distributed (IID). m3 = selectivity deviations were correlated by parameter and year (2D AR1). Relative error was calculated as $\frac{\hat{\theta}_i}{\theta} - 1$, where $\hat{\theta}_i$ was the estimate in simulation i for parameter θ, and θ was the true value (estimate from original dataset). Red points and lines show median relative error with 95% CI.

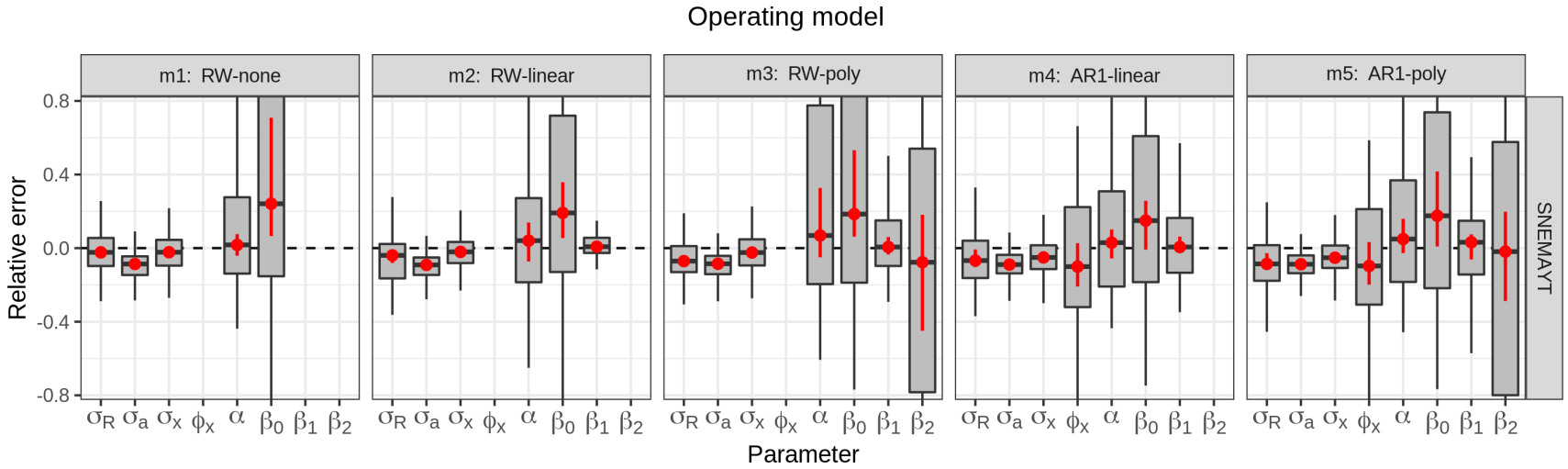

Figure 12: Relative error of parameters constraining variation in recruitment for Southern New England-Mid Atlantic yellowtail flounder (SNEMAYT). Five models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. All models estimated recruitment using the Beverton-Holt function and included CPI effects on β: $\hat{R}_{t+1} = \frac{\alpha S_t}{1+e^{\beta_0+\beta_1 x_t+\beta_2 x_t^2} S_t}$. m1 = Cold Pool Index (CPI) modeled as a random walk (RW) with no effect on recruitment ($\beta_1 = \beta_2 = 0$). m2 = CPI as RW, linear effect on β. m3 = CPI as RW, 2nd order polynomial effect on β. m4 = CPI as AR1, linear effect. m5 = CPI as AR1, polynomial effect. Relative error was calculated as $\frac{\hat{\theta}_i}{\theta} - 1$, where $\hat{\theta}_i$ was the estimate in simulation i for parameter θ, and θ was the true value (estimate from original dataset). Red points and lines show median relative error with 95% CI.

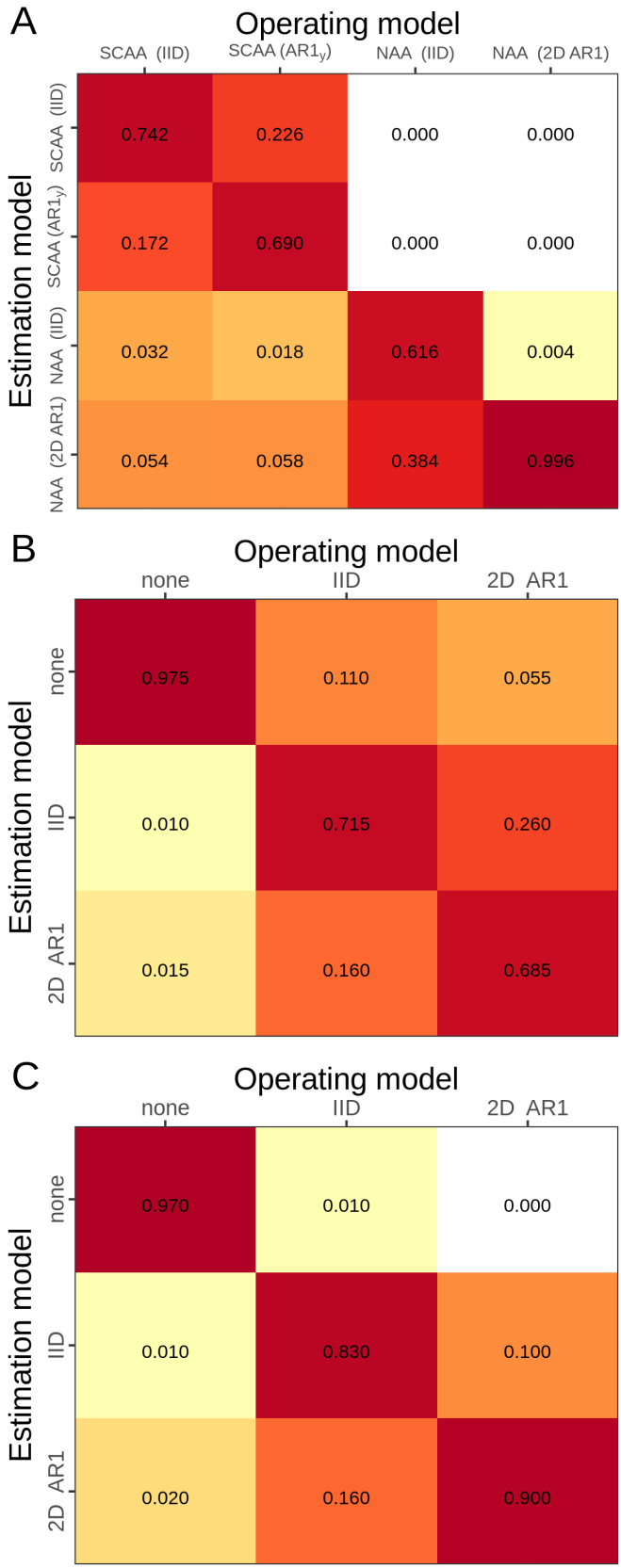

Figure 13: Proportion of simulations in which each model had the lowest AIC. A) Numbers-at-age (NAA), aggregated across all five stocks. B) Natural mortality (M), aggregated over two stocks (SNEMAYT and NScod). C) Selectivity (GBhaddock). Not all estimation models converged for each simulation, even when the operating model matched.