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

WHAM is great.

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.). *BCS: open to suggestions here... I added the retrospective bias bit back in since it links to the next paragraph. I know you wrote the lower rho was 'negligible' in the redfish paper, but we did show this in the 2D AR1 survival paper comparing Base to NAA-2. Others?*

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 commen 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). 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 rebuilding plans in the medium-term (3-10 years) for stocks that have not rebounded in response to dramatic decreases in F. For example, Miller et al. (2016) developed a state-space model for yellowtail flounder with an environmental effect on recruitment, which reduced retrospective patterns and residual variance. Additional applications using a similar state-space framework that include environmental effects on growth, M, and maturity have also proven promising (Miller and Hyun, 2018; Miller

et al., 2018; O'Leary et al., 2019; Xu et al., 2018).

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), which is in part a re-coding of the Age-Structured Assessment Program (ASAP, Legault and Restrepo, 1998; Miller and Legault, 2015) in TMB. WHAM retains characteristics of ASAP, such as its input file structure and reliance on empirical weight-at-age data, 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.

Either put core model equations here or in supplement

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.):

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

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:

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

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, $M_{a,y} = \mu_M W_{a,y}^b$, 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.):

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

where σ_M^2, 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 ($\sigma_M^2 = \varphi_a = \varphi_y = 0$ and not estimated). m2 allows IID M deviations, estimating σ_M^2 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 most 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:

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

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_s^2 is the AR1 variance, and ϕ_p, and ϕ_y are the AR1 correlation coefficients by parameter and year.

8

¹⁶⁷ Below, we demonstrate three models with random effect deviations on logistic selectivity, akin to those for M.

¹⁶⁸ m1 treats all numbers at age as independent random effects (i.e. NAA m3) but with no random effects on s

¹⁶⁹ ($\sigma_s^2 = \phi_p = \phi_y = 0$ and not estimated). m2 allows IID s deviations, estimating σ_s^2 but fixing $\phi_p = \phi_y = 0$. m3

¹⁷⁰ estimates the full 2D AR1 structure for s deviations.

2.1.1.4 Environmental covariates (*Ecov*)

¹⁷² WHAM models environmental covariate data using state-space models with process and observation com-

¹⁷³ ponents. The true, unobserved values (or "latent states", X_y) are then linked to the population dynamics

¹⁷⁴ equations with user-specified lag. For example, recruitment in year y may be influenced by X_{y-1} (lag 1),

¹⁷⁵ while natural mortality in year y may be influenced by X_y (lag 0). Multiple environmental covariates may be

¹⁷⁶ included, but only as independent processes. The *Ecov* and population model years do not need to match,

¹⁷⁷ and missing years are allowed. In particular, including *Ecov* data in the projection period can be useful.

2.1.1.4.1 Process model

¹⁷⁹ There are currently two options in WHAM for the *Ecov* process model: a normal random walk and AR1. We

¹⁸⁰ model the random walk as in Miller et al. (2016):

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

¹⁸¹ where σ_X^2 is the process variance and X_1 is estimated as a fixed effect parameter. One disadvantage of the

¹⁸² random walk is that its variance is nonstationary. In short-term projections, $\hat{X}_y$ will be equal to the last

¹⁸³ estimate with an observation and the uncertainty of $\hat{X}_y$ will increase over time. If $\hat{X}_y$ influences reference

¹⁸⁴ points, this leads to increasing uncertainty in stock status over time as well (Miller et al., 2016).

¹⁸⁵ For this reason, we generally prefer to model X_y as a stationary AR1 process as in Miller et al. (2018):

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

¹⁸⁶ where μ_X, σ_X^2, and $|\phi_X| < 1$ are the marginal mean, variance, and autocorrelation parameters. In addition

¹⁸⁷ to having stationary variance, another important difference between the random walk and AR1 in short-term

¹⁸⁸ projections is that the AR1 will gradually revert to the mean over time, unless environmental covariate

¹⁸⁹ 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 $\sigma_{x_y}^2$:

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

The observation variance in each year, $\sigma_{x_y}^2$, 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 $\sigma_{x_y}^2$ as random effects with parameters μ_{σ_x} and σ_{σ_x}:

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

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 $dR/dSSB$, 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

The population observation likelihood components are the same as in ASAP: aggregate catch and abundance index for each fleet and survey, and age composition for each fleet and survey.

2.1.2.1 Catch

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, $W_{a,y,i}$. Then,

$$
\begin{aligned}
\hat{C}_{y,i} &= \sum_a \hat{C}_{a,y,i} \\
\sigma^2_{C_{y,i}} &= e^{\eta_i} \sqrt{\log(\text{CV}^2_{y,i} + 1)} \\
\log(C_{y,i}) &\sim \mathcal{N}\left(\log(\hat{C}_{y,i}) - \frac{\sigma^2_{C_{y,i}}}{2}, \sigma^2_{C_{y,i}}\right)
\end{aligned}
\tag{6}
$$

where η_i are fleet-specific estimated parameters and $\text{CV}^2_{y,i}$ are user-specified. 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.

WHAM includes seven options for the catch and index age compositions: multinomial, Dirichlet, Dirichlet-multinomial, logistic normal treating zero observations as missing, logistic normal pooling zero observations with adjacent age classes, and two types of zero-one inflated logistic normal.

2.1.2.2 Indices

²³³ The survey indices of abundance are handled identically to the aggregate catch as in Eqn. 6 and catch age

²³⁴ compositions, except that the predicted index biomass at age i, $\hat{I}_{a,y,i}$ is a function of $N_{a,y}$, $Z_{a,y}$, $s_{a,y,i}$, $W_{a,y,i}$,

²³⁵ and catchability, q_i.

2.1.3 Projections

²³⁷ The default settings for short-term projections follow common practice for stock assessments in the U.S.

²³⁸ Northeast: the population is projected three years using the average selectivity, maturity, weight, and natural

²³⁹ mortality at age from the last five model years to calculate reference points. WHAM implements the same

²⁴⁰ options as ASAP4 for specifying F_y in the projection years: apply terminal year F_y, apply F averaged over

²⁴¹ specified years, apply $F_{X\%}$ (F at X% SPR, where X is specified), user-specified F_y, or user-specified catch.

²⁴² For models with random effects on *NAA*, *M*, or *Ecov*, the default is to continue the stochastic process into

²⁴³ the projection years. WHAM does not do this for selectivity because it is a function of management; instead,

²⁴⁴ selectivity is taken as the average of recent model years.

²⁴⁵ If the *Ecov* data extend beyond the population model years, WHAM will fit the *Ecov* model to all available

²⁴⁶ data and use the estimated X_y in projections. This may often be the case because lags can exist in both the

²⁴⁷ physical-biological mechanism and the assessment process. As an example, a model with an *Ecov* effect on

²⁴⁸ recruitment may link physical conditions in year t to recruitment in year $t + 1$, and the assessment may be

²⁴⁹ conducted in year t using population data through year $t - 1$. In this case, 3-year population projections only

²⁵⁰ need the *Ecov* model to be projected one year. While the default handling of *Ecov* projections is to continue

²⁵¹ the stochastic process, WHAM includes options to use terminal year x_y, x_y averaged over specified years, or

²⁵² specified x_y. The option to specify x_y allows users to investigate how alternative climate projections may

²⁵³ affect the stock.

2.2 Fits to original datasets

²⁵⁵ We fit the models described above to data from five stocks with a broad range of life history and status:

²⁵⁶ Southern New England-Mid Atlantic (SNEMA) yellowtail flounder, butterfish, North Sea cod, Icelandic

²⁵⁷ herring, and Georges Bank haddock (Tables 1 and 2). We fit the *NAA* random effects models to all five

²⁵⁸ stocks since these represent core WHAM functionality. We chose to highlight one stock each for the *M*, *s*,

²⁵⁹ and *Ecov* processes because all models did not converge for all stocks and processes: butterfish (*M*), Georges

²⁶⁰ Bank haddock (*s*), and SNEMA yellowtail flounder (*Ecov*). None of the original assessments estimated *M*.

²⁶¹ Except for the haddock *s* demonstration, we used the same selectivity parameterization and time blocks as in

²⁶² the original assessments. We used the m3 *NAA* model (all NAA deviations are IID random effects) in the *s*

²⁶³ and *Ecov* demonstrations, but the m1 *NAA* model (only recruitment deviations are IID random effects) for

²⁶⁴ the *M* demonstrations, because the NAA transitions can be interpreted as survival and including random

²⁶⁵ effect deviations on both NAA and *M* can lead to model non-convergence (Stock et al., n.d.). Within each

²⁶⁶ process, we compared the performance of the different random effects models using AIC.

²⁶⁷ We fit all models using the open-source statistical software R (R Core Team, 2020) and TMB (Kristensen et al.,

²⁶⁸ 2016), as implemented in the R package WHAM (Miller and Stock, 2020). Documentation and tutorials for how

²⁶⁹ to specify additional random effect structures in WHAM are available at https://timjmiller.github.io/wham/.

²⁷⁰ 2.3 Simulation tests

²⁷¹ After fitting each model to the original datasets, we used the simulation feature of TMB to conduct self- and

²⁷² cross-tests (*sensu* Deroba et al., 2015). For each model, we generated 100 sets of new data and random effects,

²⁷³ keeping the fixed effect parameters constant at values estimated in original fits. We then re-fit all models

²⁷⁴ to datasets simulated under each as an operating model. We calculated the relative error in parameters

²⁷⁵ constraining random effects (Table 1) and quantities of interest, such as spawning stock biomass (SSB), F,

²⁷⁶ $\frac{\text{B}}{\text{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R. We calculated relative error as $\frac{\hat{\theta}_i}{\theta} - 1$, where θ is the true value used to simulate data

²⁷⁷ and θ_i is the value estimated in simulation i. To estimate bias of each estimation model for a given operating

²⁷⁸ model, we calculated 95% confidence intervals of the median relative error using the binomial distribution

²⁷⁹ (Thompson, 1936). To summarize the bias across simulations and years for each model, we calculated the

²⁸⁰ median quantity across years and took the mean of the medians across simulations. Finally, for each operating

²⁸¹ model we calculated the proportion of simulations in which each estimation model converged and had the

²⁸² lowest AIC, aggregated across stocks.

²⁸³ 3 Results

²⁸⁴ 3.1 Original datasets

²⁸⁵ The 2D AR1 covariance structure for random effects on *NAA*, *M*, and selectivity performed well across stocks

²⁸⁶ and processes, evidenced by lower AIC than the IID random effects models (Fig. 1). This AIC difference was

²⁸⁷ larger for selectivity than *NAA* or *M*, but the differences for *NAA* and *M* were also non-trivial, ranging from

²⁸⁸ 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, $\frac{\text{B}}{\text{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, R, and the variance and correlation parameters controlling random effects was generally small and not significant based on confidence intervals. 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; when SSB was biased high, F was biased low, and vice-versa. Predicted catch was never biased. Bias of variance parameters was either negligible or negative, as expected. 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. 6). 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

For four of the five stocks, σ_R^2 was estimated without bias in self-tests (Fig. 7). 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.

Across the five stocks, all NAA models estimated SSB, F, $\frac{\text{B}}{\text{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, 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); Fig. 8). The convergence rate for most models and stocks was above 95%, again with the exception of the SCAA models for Icelandic herring (ref supp figure). 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

15

25% (ref supp figures). The more complex NAA models estimated all quantities without bias regardless of operating model.

3.2.2 Natural mortality

As for the NAA models, σ_R^2 was estimated without bias in the M model self-tests (Fig. 9). σ_M^2 was biased low, φ_y had no bias, and the direction of bias in φ_a was inconsistent.

SSB, F, $\frac{\text{B}}{\text{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, 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 (e.g. for butterfish, Fig. 10).

3.2.3 Selectivity

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).

Models with IID or 2D AR1 s random effects estimated SSB, F, $\frac{\text{B}}{\text{B}_{40\%}}$, $\frac{F}{F_{40\%}}$, and R with little to no bias across operating models (Fig. 12). The model without s random effects showed substantial bias when fit to data simulated from m2 or m3.

3.2.4 Ecov-Recruitment

The random walk CPI models estimated σ_X^2 with less bias than the AR1 CPI models (Fig. 13). 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 natural mortality, variation in selectivity or natural mortality, or stochastic changes in the numbers at age are of interest. The simulation tests showed negligible or no bias in estimation of important assessment outputs (SSB, F, stock and harvest status).

We described WHAM. Sim tests showed no bias in self-tests (when estimation model matched operating model). Some bias in cross-tests. Bias in SSB and F opposite such that no bias in predicted catch, Deroba 2015.

(Miller et al., 2018) The biased maximum likelihood estimation of the autoregres- sive process parameters we found in the simulation study was expected (Wilson 1989). The reduction in bias for most of these parameters using REML estimation was also expected (Cheang and Reinsel 2000), but substantial bias still remained,

Contrast mechanistic vs. empirical approaches. Empirical approaches allows time-varying productivity and with AR1 can propagate effect of changing productivity in short-term projections (Stock et al., n.d.). However, the empirical approach cannot predict values beyond extremes in observed time-series (*TJM: This is always true? I suspect certain types of time series models might predict outside of time series. E.g., if there is an increase with the last observation being the greatest, some higher order AR models might predict the next few observations higher than the last*). This is an issue because many ecosystems are changing to such extent that conditions in the near future have never been observed. E.g. last time it was close to this warm on the U.S. Northeast Shelf was the 1940s, and few fisheries or oceanographic time-series stretch back that far. Thus, longer term projections likely require mechanistic approach. Think about other sources of data? Fish catch, not just surveys + age data? Continuous Plankton Recorder (other side of Atlantic, but correlated?). 1951 was almost as warm (annual mean), but warm in winter instead of summer (diff seasonal change).

Mechanistic approach: more likely to find environmental effects in cases with:

- history of overfishing (Free et al., 2019)
- more rapid environmental change (Free et al., 2019)
- stocks at edge of species' range (Free et al., 2019)
- opportunistic species (short-lived) vs. longer-lived (Free et al., 2019)
- longer time series
- stronger signals (wider ranges of observed stock status and environmental conditions)
- periodic signal where we've recorded more than one cycle (e.g. PDO/sst sardine)
- lower trophic level (e.g. generally tighter relationship between PO and zooplankton, then small pelagics TL2, then TL3+). PO forecasting group look at zooplankton indices? See Perretti 2017 (recruitment regime shifts)

Reducing uncertainty may be just as valuable as improving accuracy / reducing bias.

4.2 Differences from other assessment model frameworks

How is WHAM different from SAM?

Not sure where to put this... some could be in Intro or Discussion. Definitely will be a question in readers' minds so may be good to introduce early?

WHAM estimates annual full F as fixed effects and selectivity of surveys or fisheries as constant or time-varying random effects with various autoregressive assumptions possible. SAM estimates FAA as a multivariate log-normal random walk. WHAM can estimate interannual transitions as random walk processes like SAM or as a function of autoregressive deviations. Alternatively (or simultaneously) WHAM can estimate deviations in natural mortality as autoregressive processes like NCAM (Cadigan, 2016). Uniquely, WHAM can model multiple covariate time series as state-space processes and include their effects (nonlinearly) in various ways on recruitment or natural mortality.

Like SS, ASAP, BAM, WHAM treats aggregate and composition observations separately for fisheries and indices. This differs from SAM where observations of numbers at age are treated as multivariate log-normal.

Most assessments in the U.S. assume separability in $F_{a,t}$, estimate F_t and Sel_a. WHAM does this. SAM estimates $F_{a,t}$ directly. WHAM and SAM also make different separability assumptions for the catch/index data (aggregate total + age comps vs. $C_{a,t}$ directly). Should be similar (?) but could test.

To move from traditional SCAA models to state-space models with random effects in the Northeast US, WHAM can be configured to fit models very similar to ASAP and therefore bridge between these frameworks. Goal is to replicate ASAP assessments in the U.S. Northeast. Can easily turn on/off random effects.

Observation model is natural for landings data that are measured as total weight plus age composition sampling. Age composition sampling often done separately with survey data.

Treating F and Sel separately can be useful for projections. Oftentimes we want to specify F in projections to calculate a reference point, as opposed to continuing a F time-series process.

4.3 Future work

WHAM will be used in upcoming research track assessments. Could transition to operational. Potential to improve several NEFSC assessments.

- 2D AR(1) selectivity. Most assessments in the U.S. assume separability in $F_{a,t}$, i.e. estimate F_t and Sel_a. WHAM does this. SAM estimates $F_{a,t}$ directly. WHAM and SAM make different separability

assumptions for the catch/index data as well (aggregate total + age comps vs. $C_{a,t}$ directly). Should be similar (?) but could test.

- How many time/age-varying random effects can be estimated simultaneously? Stock et al. (n.d.) estimated random effect deviations in abundance at age and M, as well as an environmental covariate effect on recruitment.

- Ecov-Recruitment simulation study. How much information does Ecov need to have to be useful?

4.4 Extensions

4.4.1 Multivariate spatiotemporal environmental data

Most examples that include environment-recruitment effects are univariate, but in many cases it is likely that multiple factors have to align for successful recruitment (e.g. temperature and currents for Nassau Grouper).

4.4.2 Length/growth estimation

4.4.3 Ecov models

- AR(k)
- splines
- Gaussian process/EDM/Munch/Sugihara

4.5 Conclusion

Development of TMB has facilitated significant advancement in fisheries assessment, allowing us to treat population processes as random effects. A grand challenge in fisheries is to assess and manage stocks in a changing environment. Increasingly have the environmental data. Population time-series are lengthening. WHAM is a step in this direction.

Acknowledgements

Supplementary material

More figures.

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

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

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

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.

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

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

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).

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

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

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

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

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

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

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)	$\sigma_R, \sigma_a, \rho_y, \rho_a$
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)	$\sigma_R, \sigma_M, \varphi_y, \varphi_a$
Selectivity (Sel)		
m1: none	No random effects on selectivity	σ_R, σ_a
m2: IID	Selectivity deviations are IID random effects	$\sigma_R, \sigma_a, \sigma_{Sel}$
m3: 2D AR1	Selectivity deviations are random effects with correlation by year and age (2D AR1)	$\sigma_R, \sigma_a, \sigma_{Sel}, \phi_y, \phi_a$
Ecov-Recruitment (Ecov)		
m1: RW-none	Ecov: random walk (RW), effect on β: none	$\sigma_R, \sigma_a, \sigma_x$
m2: RW-linear	Ecov: random walk (RW), effect on β: linear	$\sigma_R, \sigma_a, \sigma_x, \beta_1$
m3: RW-poly	Ecov: random walk (RW), effect on β: 2nd order polynomial (poly)	$\sigma_R, \sigma_a, \sigma_x, \beta_1, \beta_2$
m4: AR1-linear	Ecov: autocorrelated (AR1), effect on β: linear	$\sigma_R, \sigma_a, \sigma_x, \phi_x, \beta_1$
m5: AR1-poly	Ecov: autocorrelated (AR1), effect on β: 2nd order polynomial (poly)	$\sigma_R, \sigma_a, \sigma_x, \phi_x, \beta_1, \beta_2$

Table 2: Stocks used in simulation tests.

Stock	Modules tested				Model dim		Biol. par.		Stock status	
	NAA	M	Sel	Ecov	# Ages	# Years	M	σ_R	$\frac{B}{B_{40}}$	$\frac{F}{F_{40}}$
SNEMA yellowtail flounder	x	x		x	6	49	0.2-0.4	1.67	0.01	0.44
Butterfish	x	x			5	31	1.3	0.23	2.57	0.03
North Sea cod	x	x			6	54	0.2-1.2	0.87	0.14	2.00
Icelandic herring	x				11	30	0.1	0.55	0.40	1.81
Georges Bank haddock	x		x		9	86	0.2	1.65	5.16	0.12
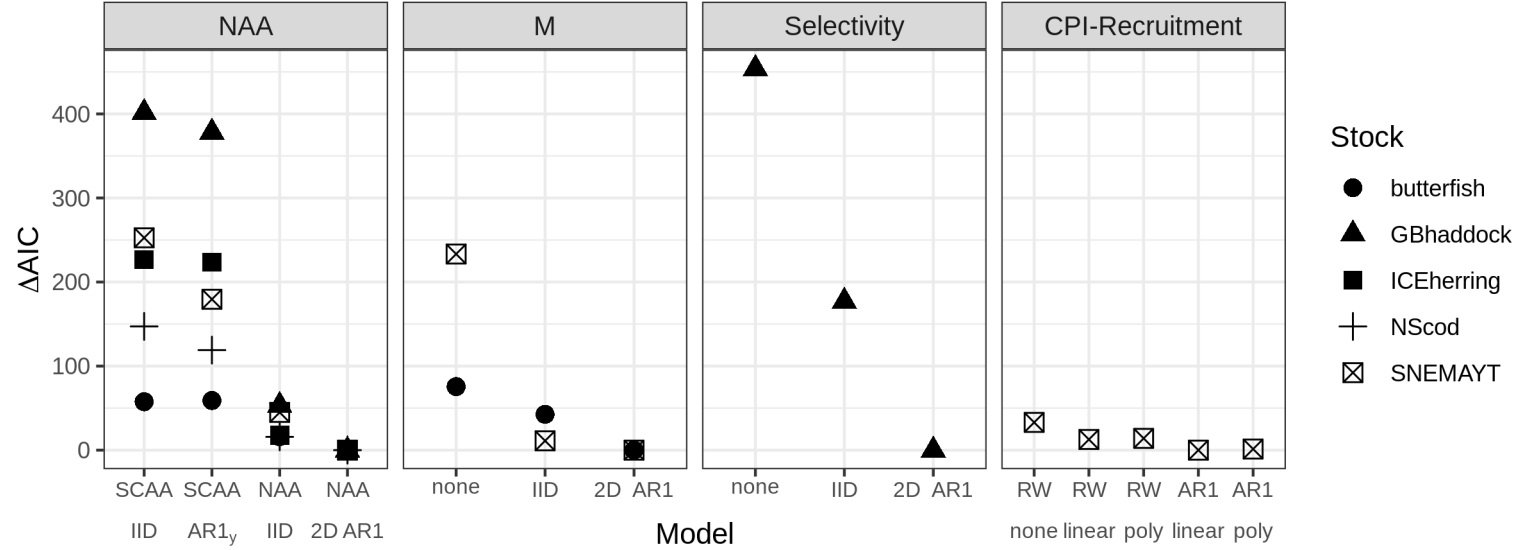
Figure 1: AIC differences by model and stock when fit to original datasets. Stock abbreviations: SNEMA yellowtail flounder (SNEMAYT), North Sea cod (NScod), Icelandic herring (ICEherring), and Georges Bank haddock (GBhaddock).
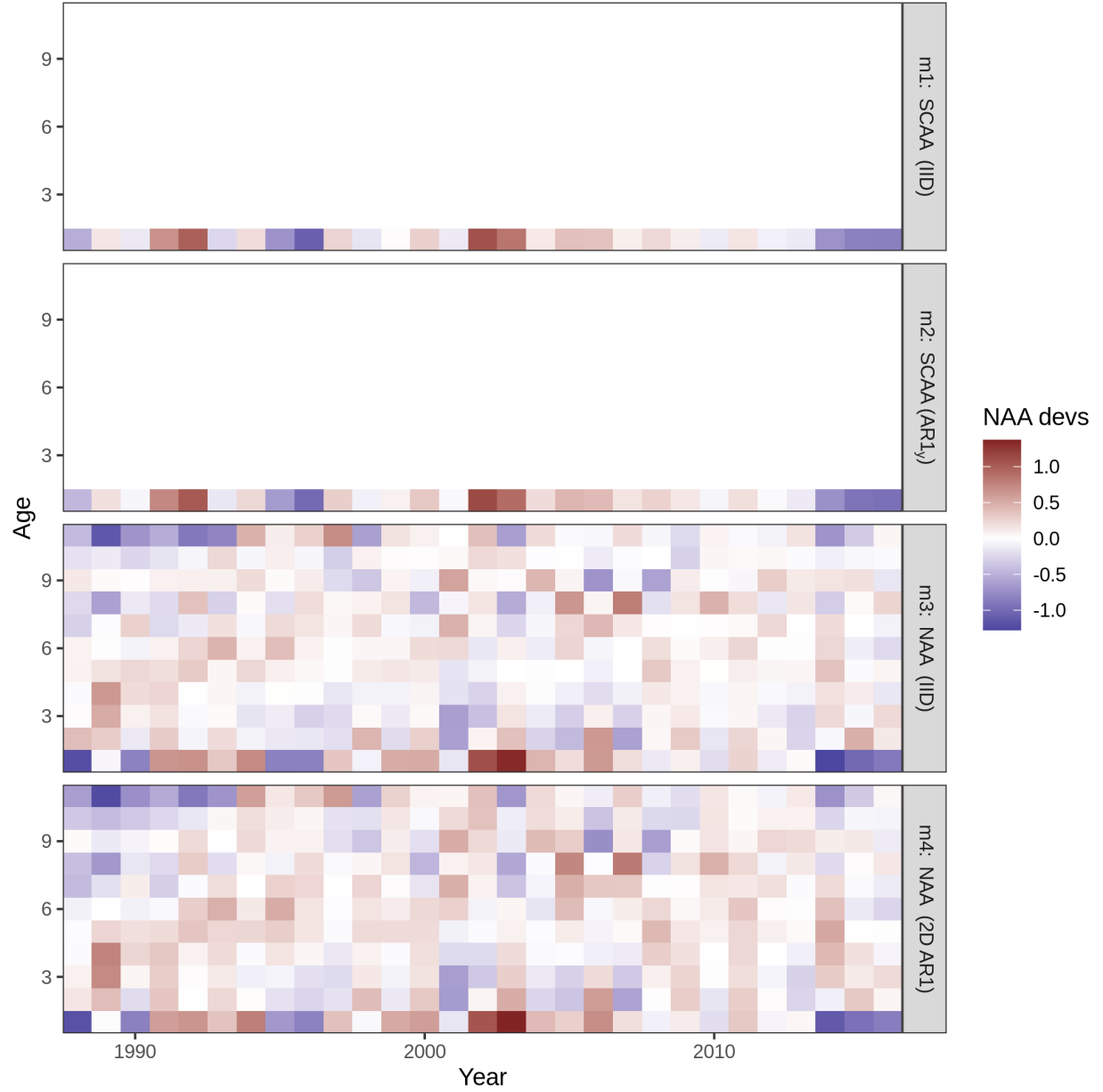
Figure 2: Abundanc-at-age deviations estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (IID). 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).
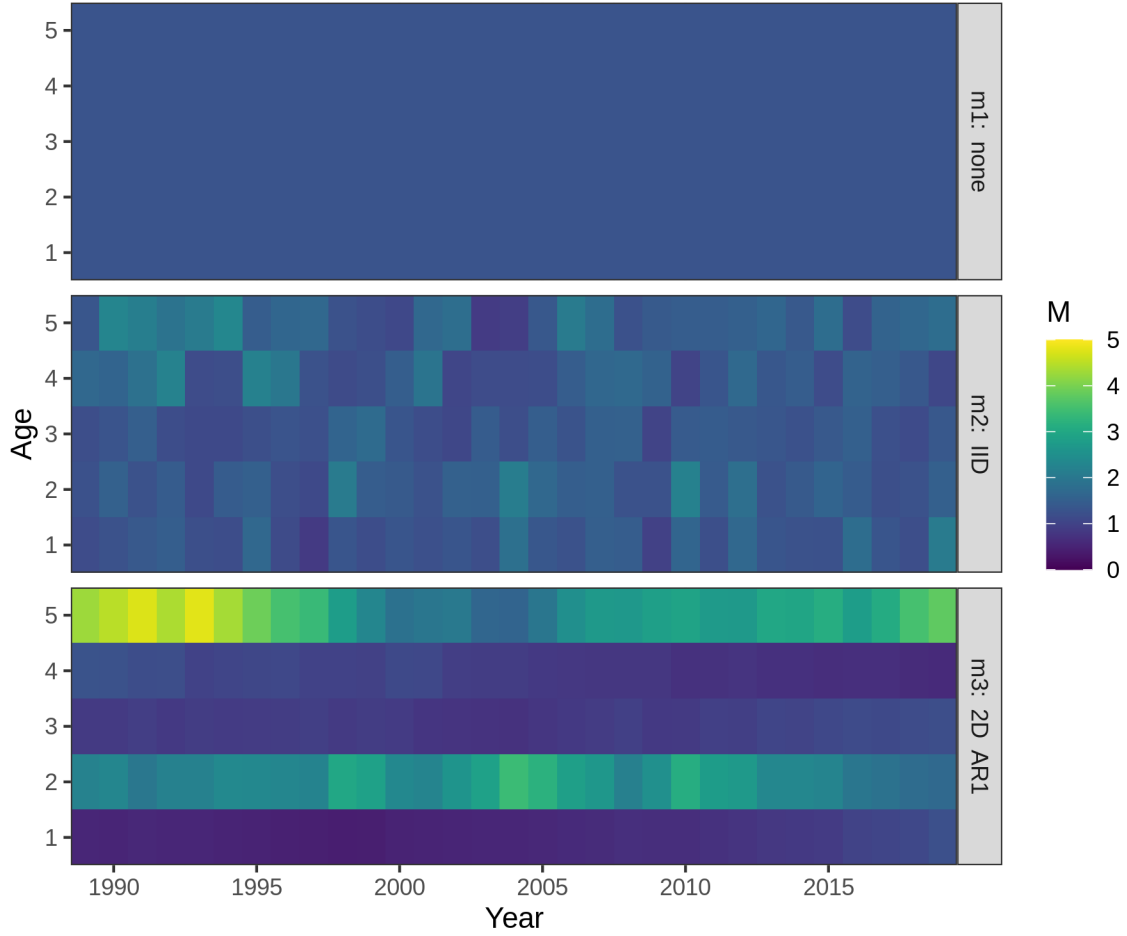
Figure 3: Natural mortality (M) estimated for butterfish using three random effects models. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).
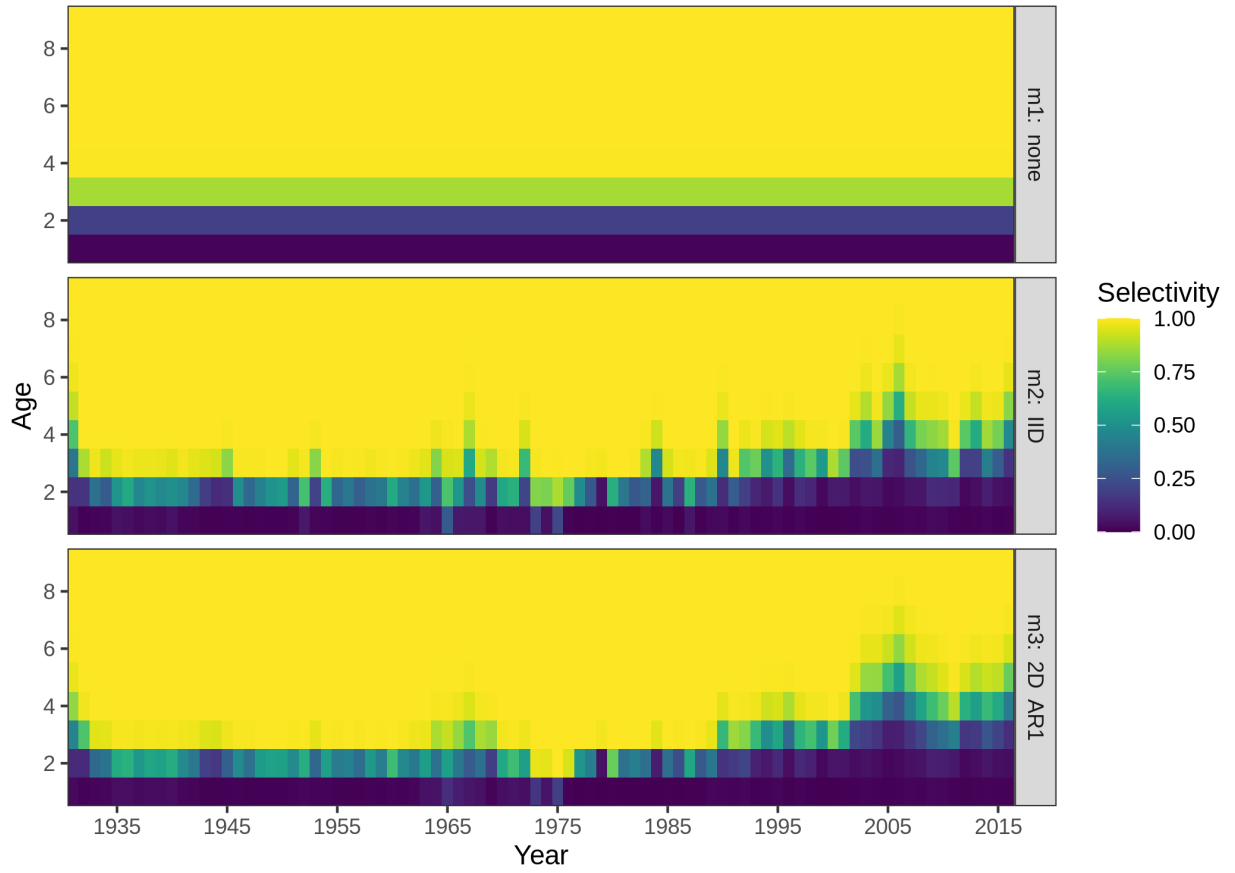
Figure 4: Selectivity estimated for Georges Bank haddock using three random effects models. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).
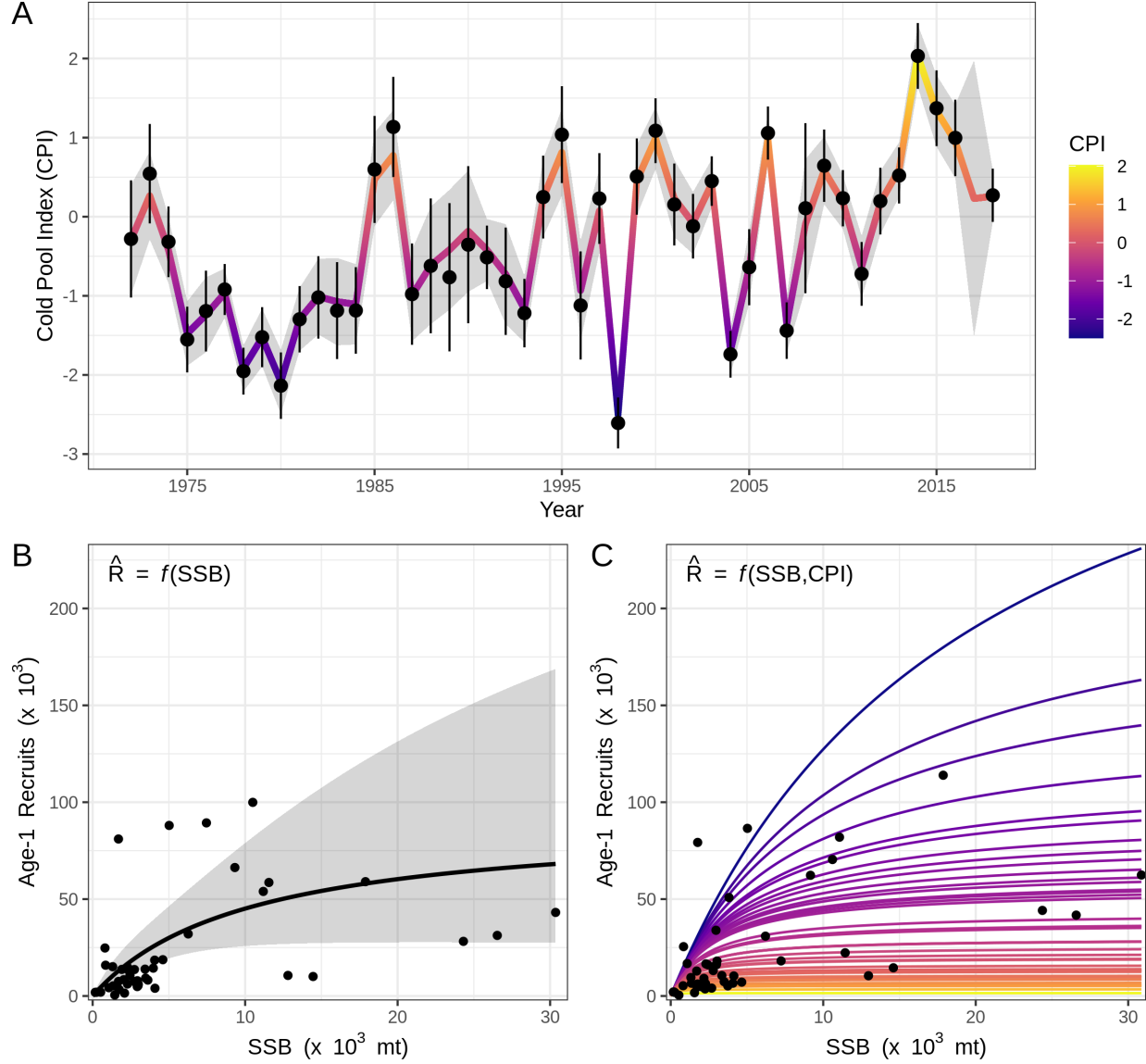
Figure 5: Beverton-Holt stock-recruit relationships fit for Southern New England-Mid Atlantic yellowtail flounder, with and without effects of the Cold Pool Index (CPI). A) CPI estimated from the model with lowest AIC (m4, AR1-linear). Points are observations with 95% CI, and the line with shading is 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).
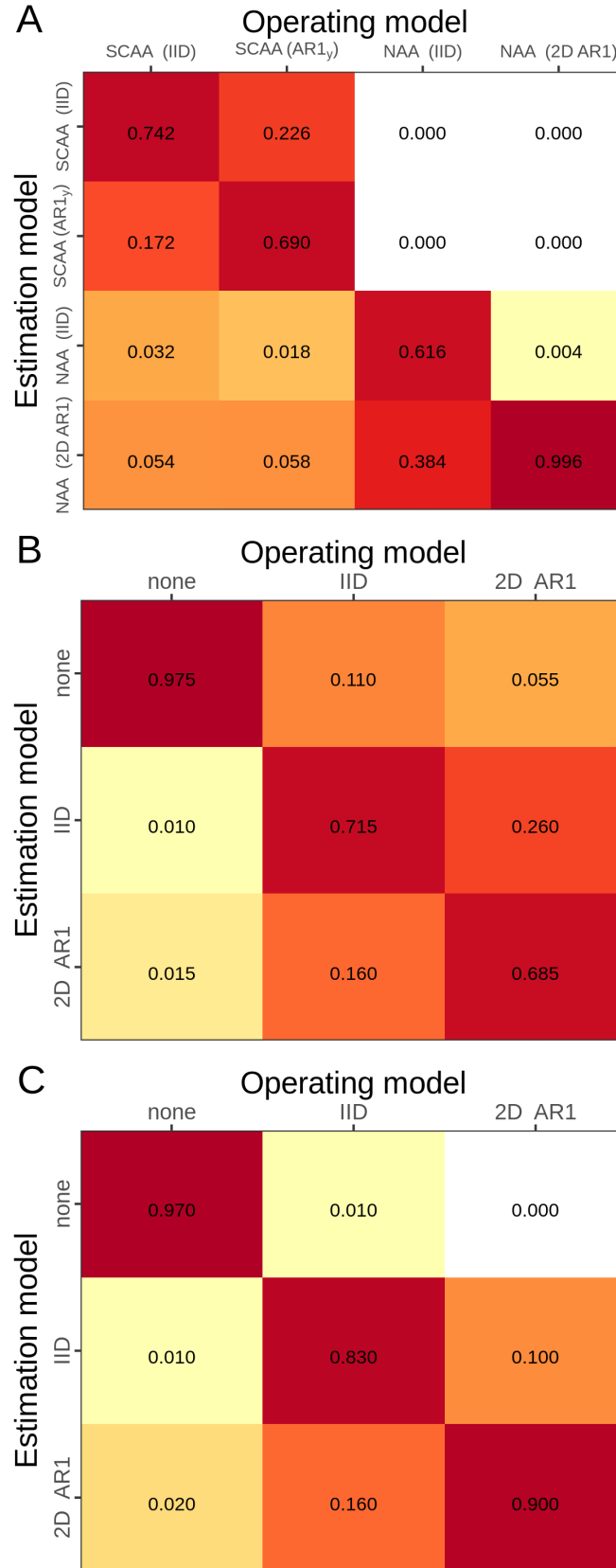
Figure 6: Proportion of simulations in which each model had the lowest AIC. A) Numbers-at-age (NAA), aggregated across all five stocks. B) Natural mortality (M), aggregated over two stocks (SNEMAYT and NScod). C) Selectivity (GBhaddock). Not all estimation models converged for each simulation, even when the operating model matched.
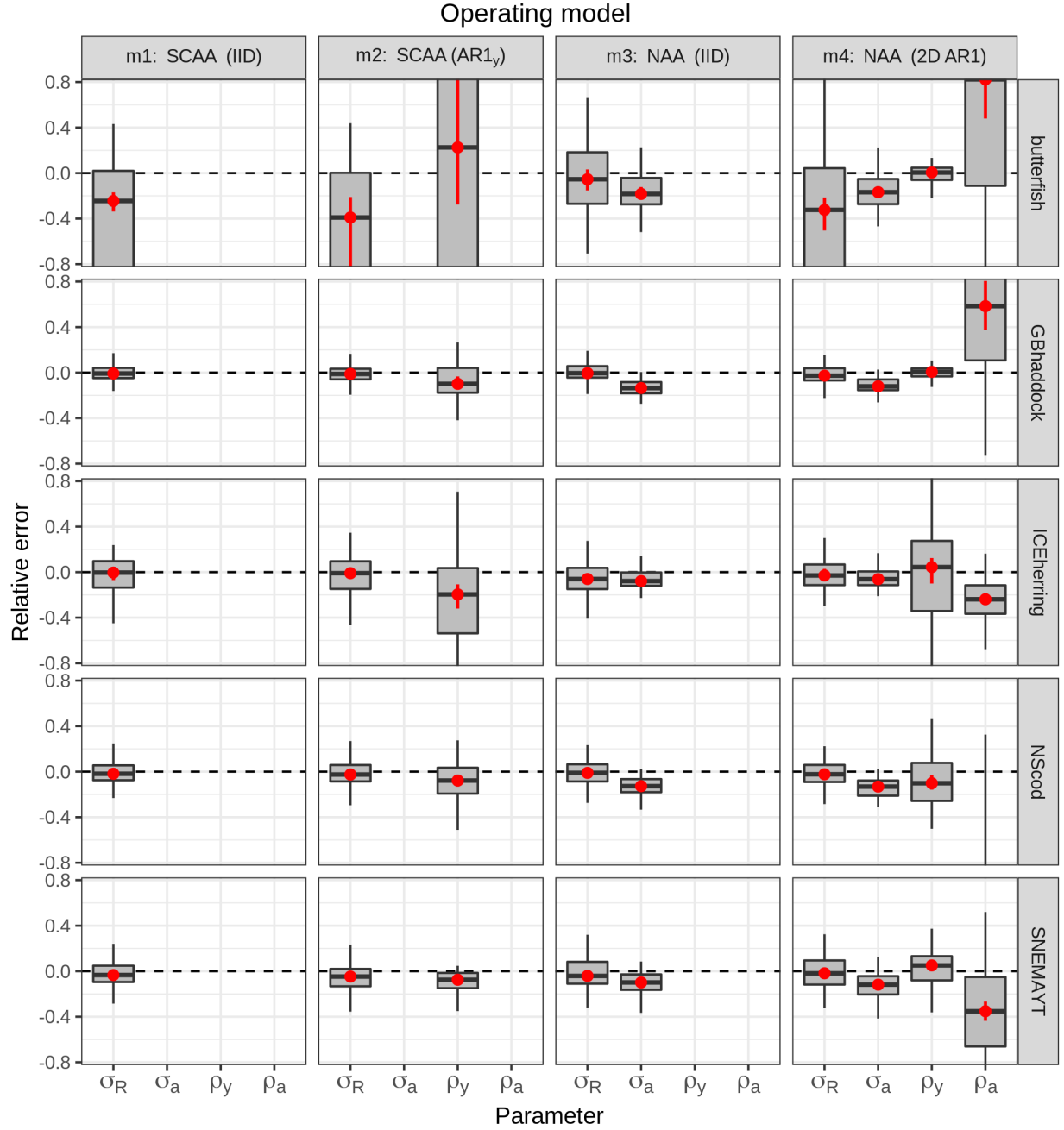
Figure 7: Relative error of parameters constraining numbers-at-age (NAA) random effects. Four models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = only recruitment deviations are random effects (most similar to traditional 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,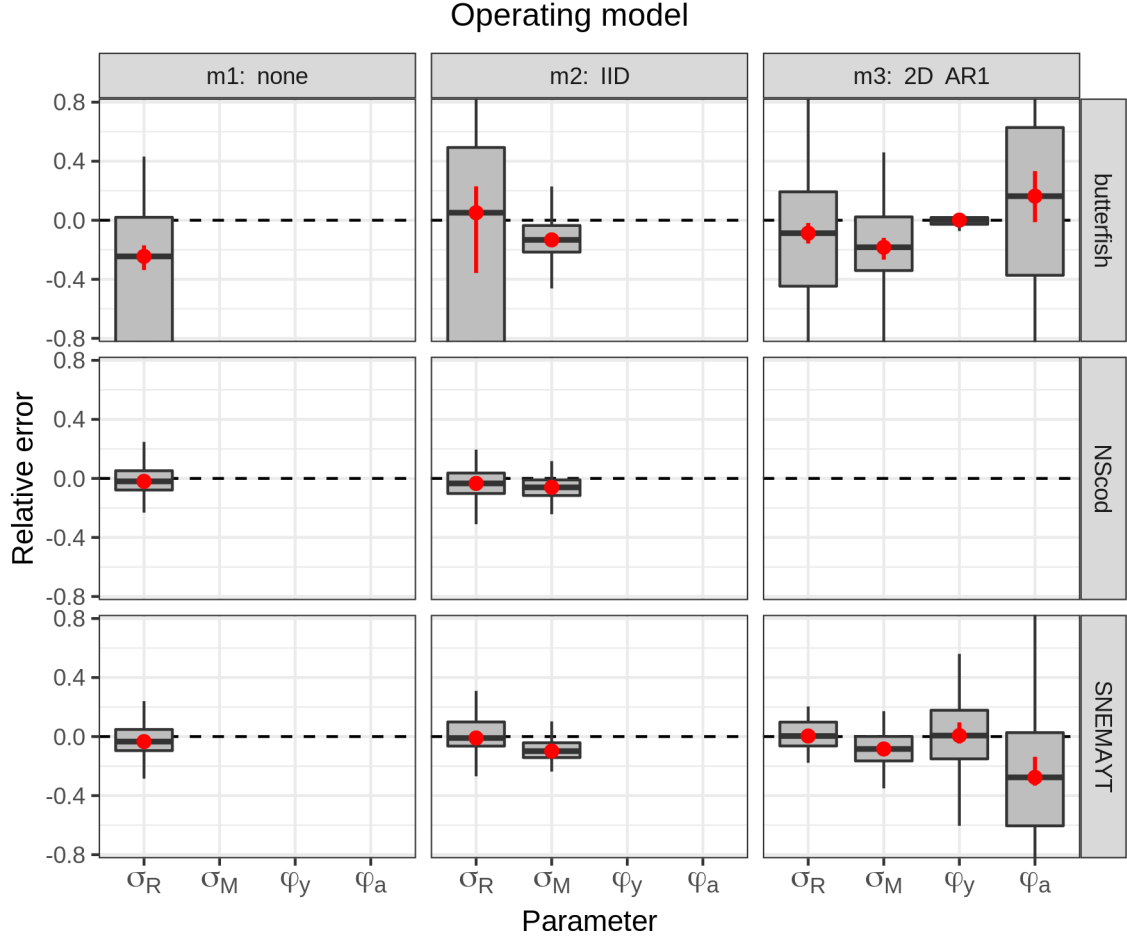 but deviations are correlated by age and year (2D AR1). Relative error was calculated as $\frac{\hat{\theta}_i}{\theta} - 1$, where $\hat{\theta}_i$ was the estimate in simulation i for parameter θ, and θ was the true value (estimate from original dataset). Red points and lines show median relative error with 95% CI.

34
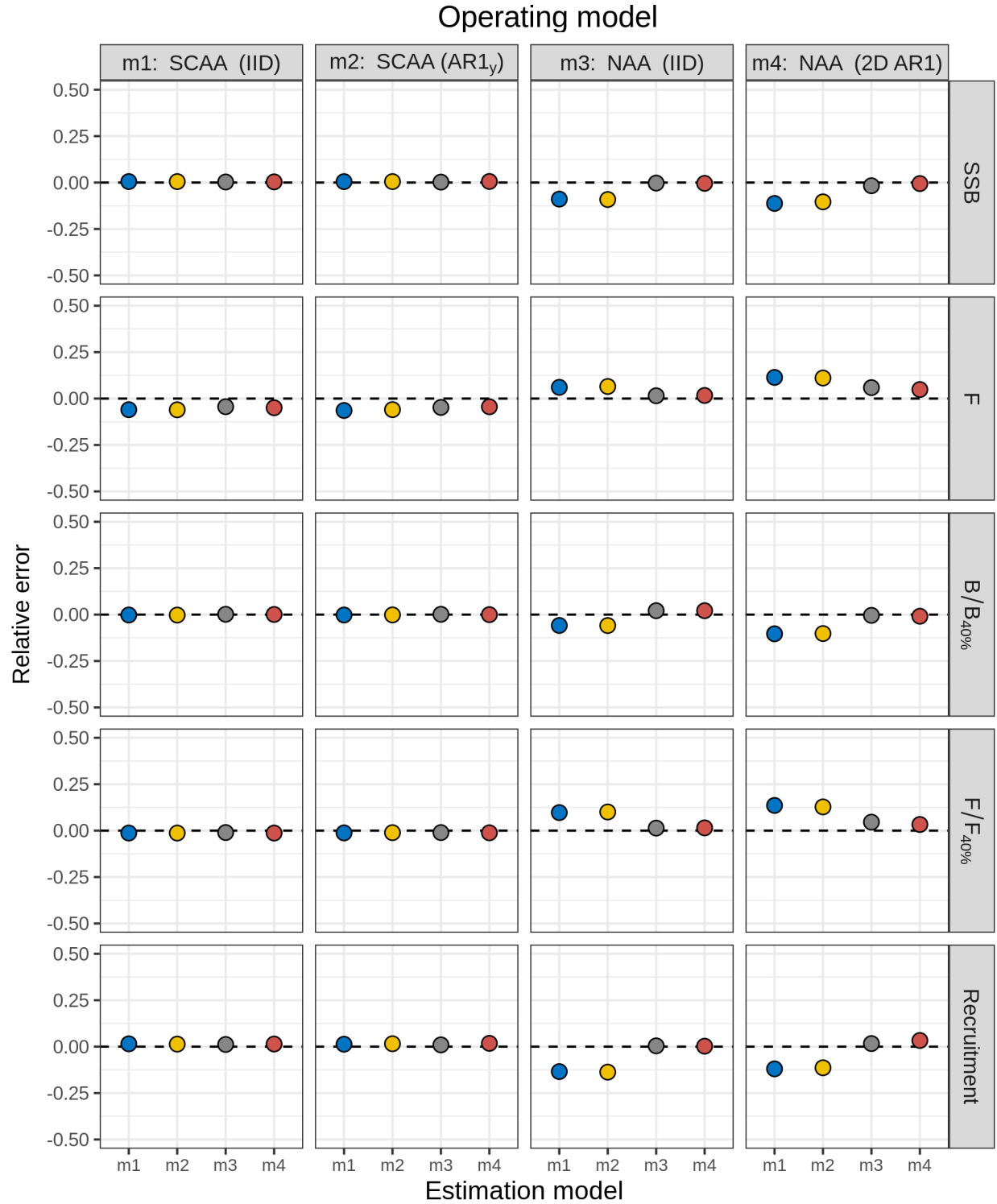
Figure 8: Relative error of key quantities estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (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 9: 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).
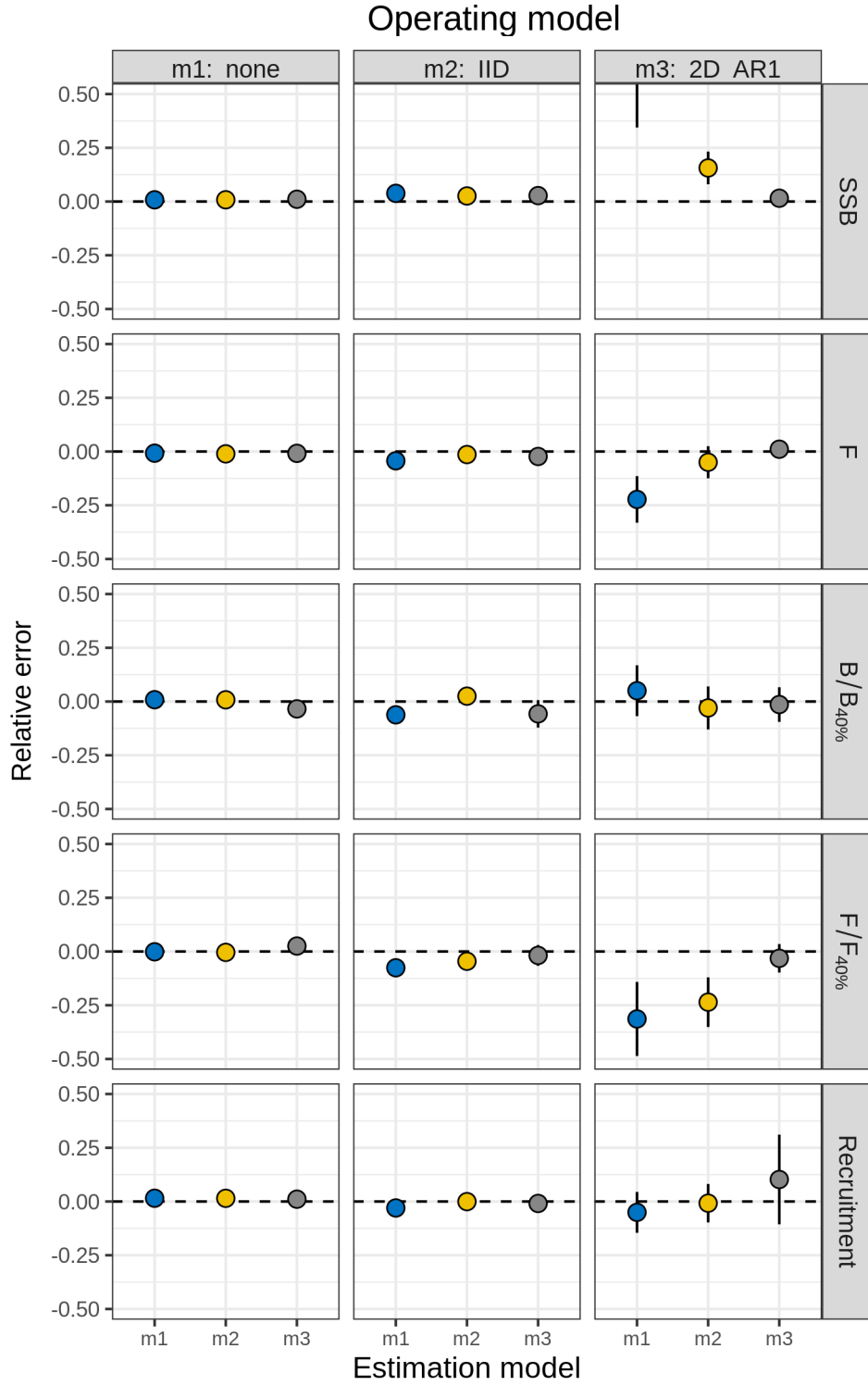
Figure 10: Relative error of key quantities estimated for butterfish using three models of natural mortality (M) random effects. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).
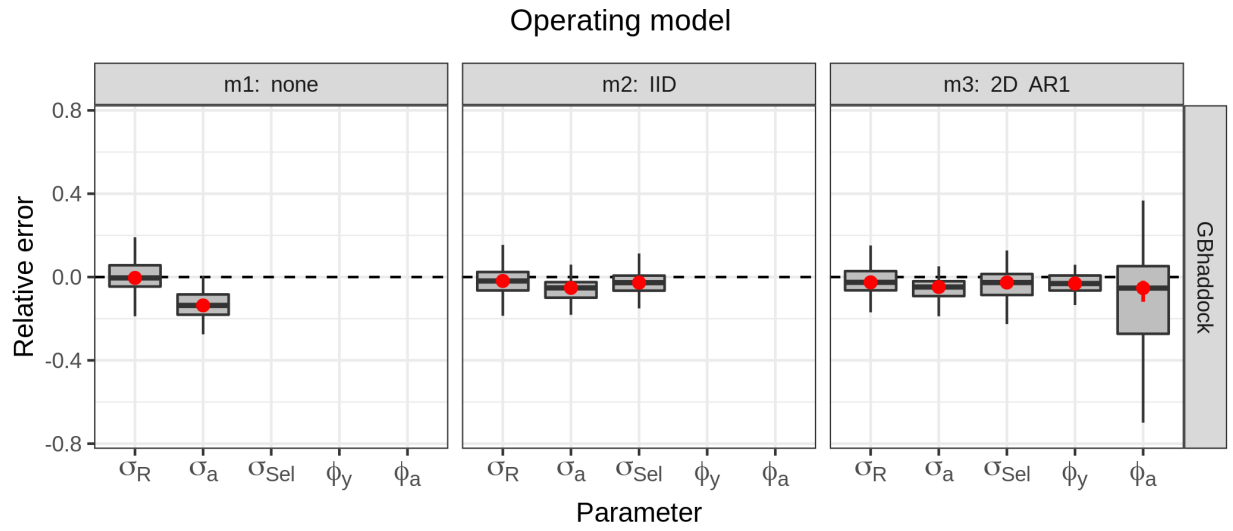
Figure 11: Relative error of parameters constraining selectivity random effects for Georges Bank haddock (GBhaddock). Three models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = no random effects (constant selectivity). m2 = 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.
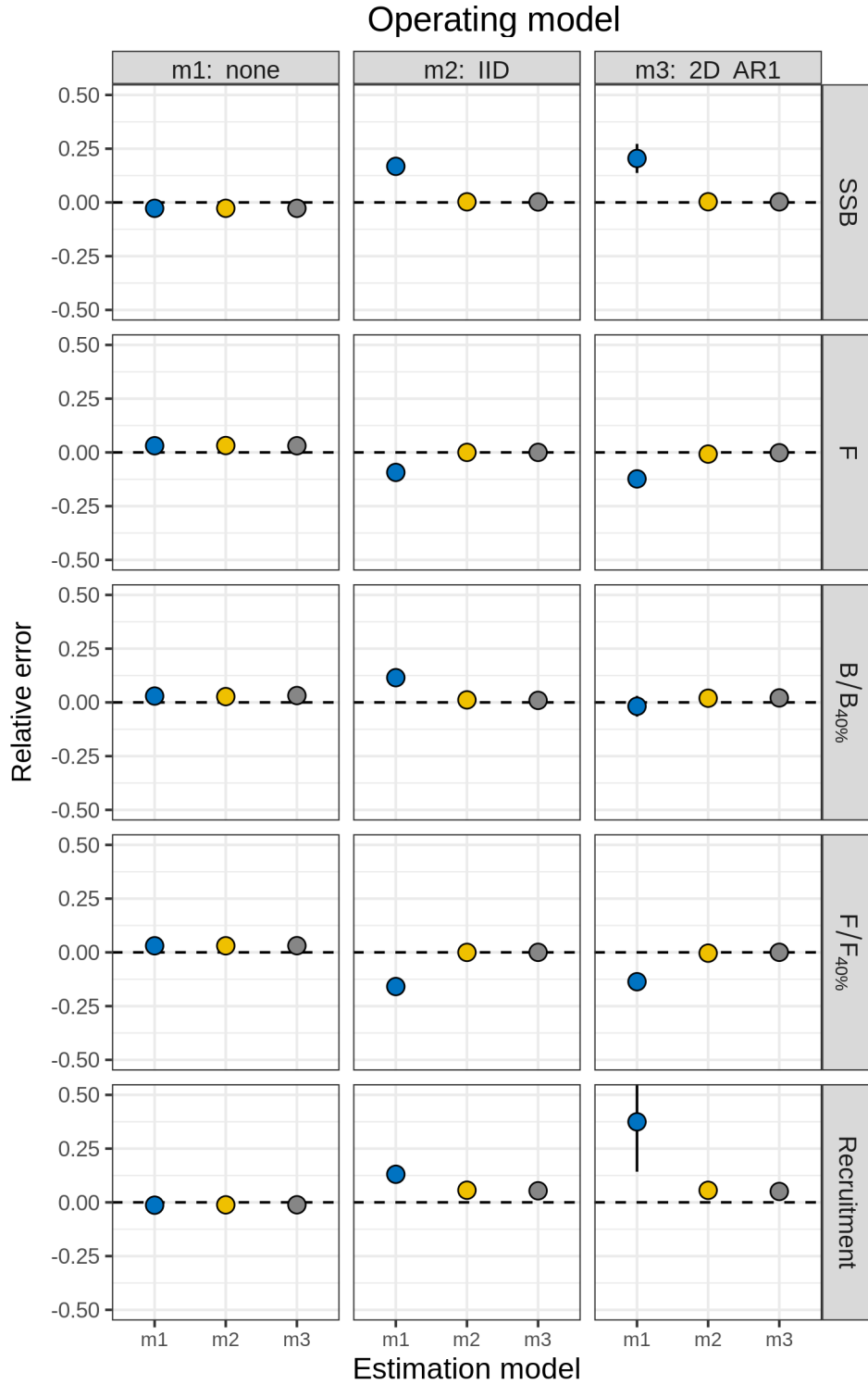
Figure 12: Relative error of key quantities estimated for Georges Bank haddock using three models of selectivity random effects. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).
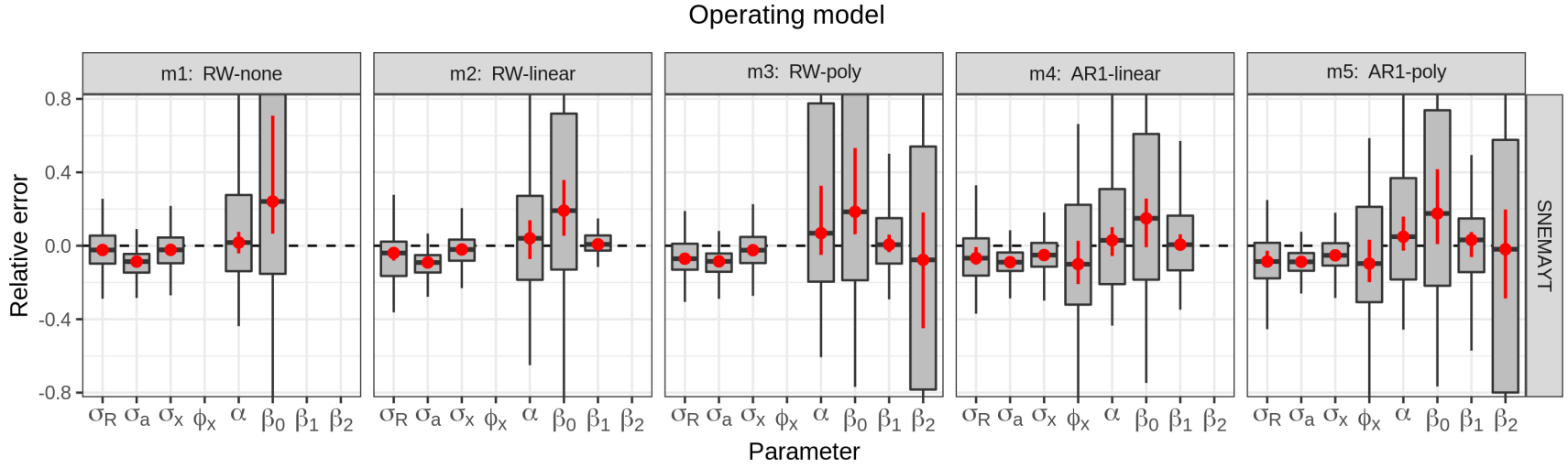
Figure 13: Relative error of parameters constraining variation in recruitment for Southern New England-Mid Atlantic yellowtail flounder (SNEMAYT). Five models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. All models estimated recruitment 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.