The Woods Hole Assessment Model (WHAM): a general state-space assessment framework that incorporates time- and age-varying processes via random effects and 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; survival; 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), and typically ignore effects of the environment or interactions with other species, 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). 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 does have the potential to reduce uncertainty in reference points and improve projections of stock status (Miller et al., 2016).

An alternative approach is to allow biological parameters to vary stochastically over time, without explanation, i.e. the empirical approach *sensu* Punt et al. (2014). Statistical catch-at-age (SCAA) models typically only estimate year-specific recruitment (R_t) and F_t, often as deviations from a mean, 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., 2019). State-space models that treat parameters as unobserved states can avoid such subjectivity by estimating the penalty terms as variance parameters constraining random effects. 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, they generally have larger (more realistic) uncertainty and lower retrospective bias (**Miller et al. in prep**).

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). 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. 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 that include environmental effects on growth, M, and maturity in a similar state-space framework have also proven promising (Miller and Hyun, 2018; Miller et al., 2018; O'Leary et al., 2019; Xu, Timothy J. Miller, 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 survival, 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.

Finally, 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 four processes: survival (numbers at age transitions, NAA), natural mortality (M), selectivity (Sel), and environmental covariates ($Ecov$). Across processes, WHAM makes heavy use of first-order autoregression, AR1, to constrain the random effects, especially a 2D AR1 covariance structure which is AR1 over both age and year. 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.), $Sel_{a,y}$ (Xu, Thorson, et al., 2018), 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) / survival

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):

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

where $N_{a,y}$ are the numbers at age a in year y, Z is the total mortality rate ($F+M$), and 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):

$$\varepsilon_{1,y} \sim \mathcal{N}(0, \sigma_R^2)$$

The only difference between m1 and a SCAA is that the recruitment variance, σ_R^2, is estimated as a parameter within the model instead of being fixed or tuned externally.

m2 is the same as m1, except that the recruitment deviations are AR1 with autocorrelation parameter ρ_R:

$$\varepsilon_{1,y+1} \sim \mathcal{N}(\rho_R \varepsilon_{1,y}, \sigma_R^2)$$

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:

$$\varepsilon_{a,y} \sim \begin{cases} \mathcal{N}\left(0, \sigma_R^2\right), & \text{if } a = 1 \\ \mathcal{N}\left(0, \sigma_a^2\right), & \text{if } a > 1 \end{cases} \tag{2}$$

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 survival variations for young-of-the-year (recruitment) are typically larger than for other ages.

m4 treats the numbers at all ages as random effects, as in m3, but constrains the NAA deviations with a 2D

AR1 structure:

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

where $\mathbf{E} = \left(\varepsilon_{1,1}, \ldots, \varepsilon_{1,Y-1}, \varepsilon_{2,1}, \ldots, \varepsilon_{2,Y-1}, \ldots, \varepsilon_{A,1}, \ldots, \varepsilon_{A,Y-1}\right)'$ 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 ρ_a and ρ_y 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.

2.1.1.2 Natural mortality (*M*)

WHAM can estimate or fix μ_{M_a}, the mean natural mortality at age a, or estimate one M for all ages, μ_M, as fixed effect parameters. In addition, M can 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 estimate random effect deviations in M, $\delta_{a,y}$, similar to the NAA deviations (Cadigan, 2016; Stock et al., n.d.):

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

where σ_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 the 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 (*Sel*)

2.1.1.4 Environmental covariates (*Ecov*)

2.1.1.4.1 Time-series model

2.1.1.4.2 Observation model

2.1.1.4.3 Link to population

2.1.2 Data/observation model

2.1.2.1 Catch (agg, age comp)

2.1.2.2 Index (agg, age comp)

2.1.3 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?

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.

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.

2.1.4 Bias correction

- Analytical obs error. (Aldrin et al., 2020).
- Analytical process error.
- TMB epsilon. (Thorson, 2019; Thorson and Kristensen, 2016)

Should these all be used?

2.2 Simulation tests

Fit each model to original dataset. Use each model to simulate new data and random effects, keeping fixed effect parameters constant at values estimated in original fits. Re-fit each model to datasets simulated under each operating model. We used the stocks in Table 2.

We used R (R Core Team, 2020). WHAM is available as an R package (Miller and Stock, 2020). Table 1 summarizes the models fit for each process. Tutorials for how to specify additional random effect structures in WHAM are available at https://timjmiller.github.io/wham/.

3 Results

3.1 Original datasets

2D AR1 structure performed well across processes and stocks (Fig. 1).

3.1.1 Numbers-at-age

Fig. 2.

3.1.2 Natural mortality

Fig. 3.

3.1.3 Selectivity

Fig. 4.

3.1.4 Ecov-Recruitment

Fig. 5.

3.2 Simulation tests

Not all models converged. Fig. 6.

3.2.1 Numbers-at-age

Fig. 7.

Fig. 8.

3.2.2 Natural mortality

Fig. 9.

Fig. 10.

3.2.3 Selectivity

Fig. 11.

Fig. 12.

3.2.4 Ecov-Recruitment

Fig. 13.

4 Discussion

4.1 Overview

We described WHAM. Sim tests showed no bias in self-tests (when estimation model matched operating model). Some bias in cross-tests.

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. 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 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 survival 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?

4.3 Extensions

4.3.1 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).

4.3.2 Length/growth estimation

4.3.3 Ecov models

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

4.4 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

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

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

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

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

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

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

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.

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

Thorson, J.T., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research 175, 66–74. https://doi.org/10.1016/j.fishres.2015.11.016

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., 2019. Comparing the performance of three data-weighting methods when allowing for time-varying selectivity. Can. J. Fish. Aquat. Sci. 1–17. https://doi.org/10.1139/cjfas-2019-0107

Xu, H., Thorson, J.T., Methot, R.D., Taylor, I.G., 2018. 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	$\frac{B}{B_{40}}$	$\frac{F}{F_{40}}$
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

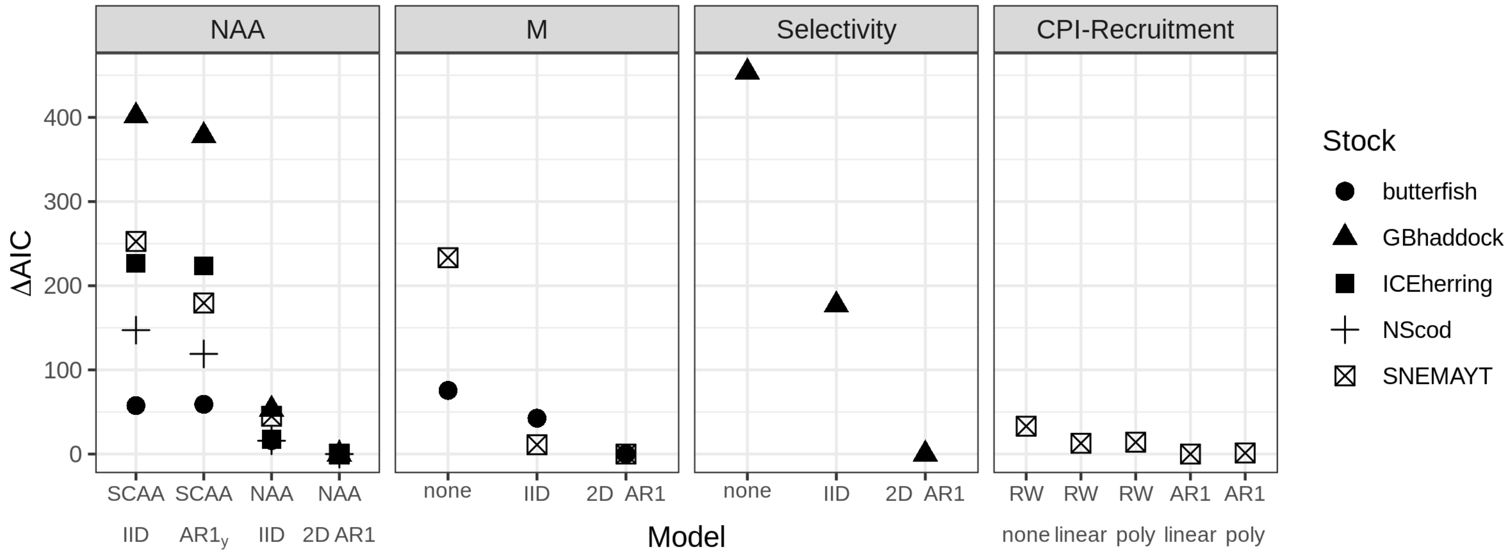

Figure 1: AIC differences by model and stock when fit to original datasets. Stock abbreviations: SNEMA yellowtail flounder (SNEMAYT), North Sea cod (NScod), Icelandic herring (ICEherring), and Georges Bank haddock (GBhaddock).

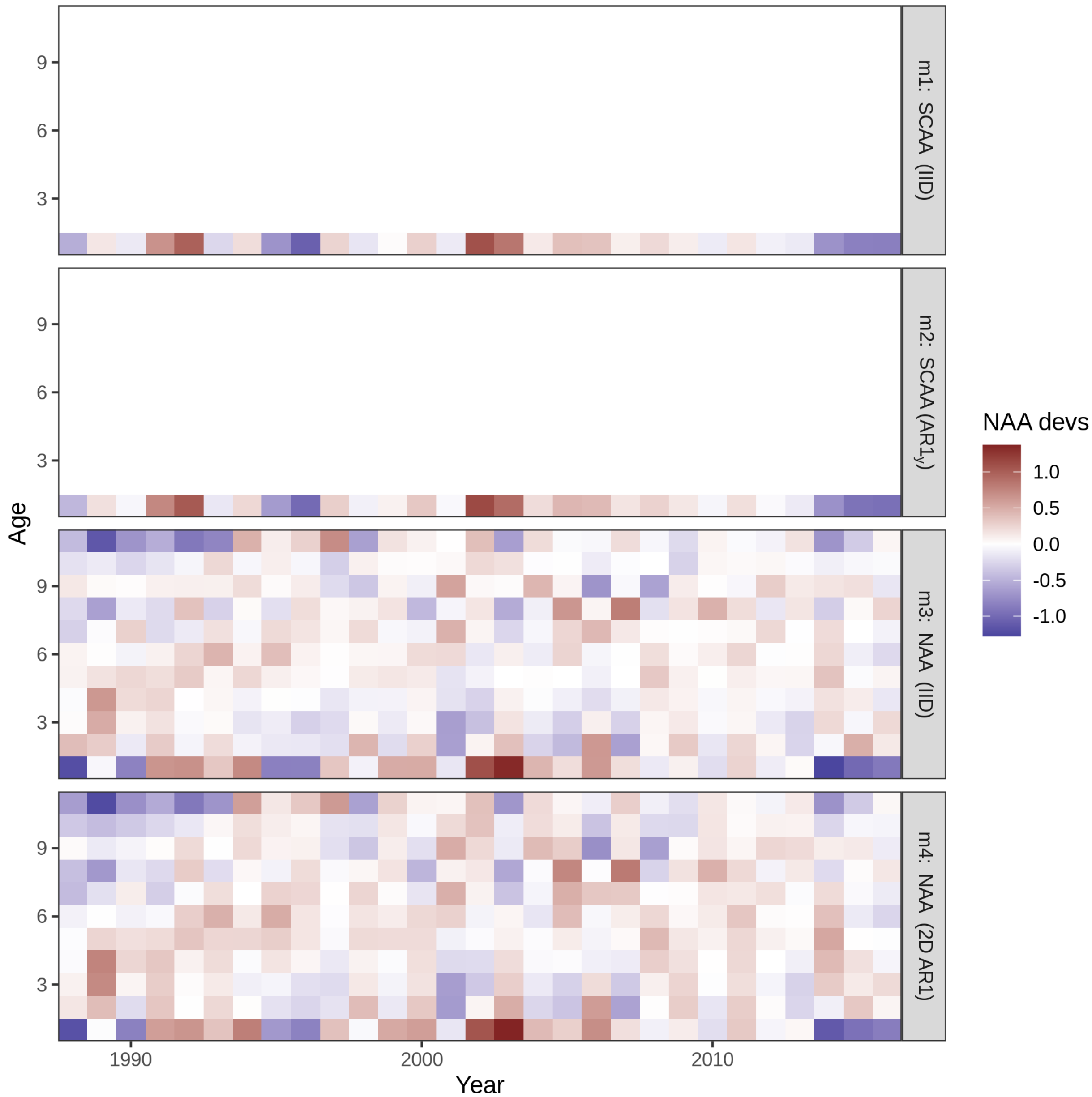

Figure 2: Survival deviations estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (IID). m2 = 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).

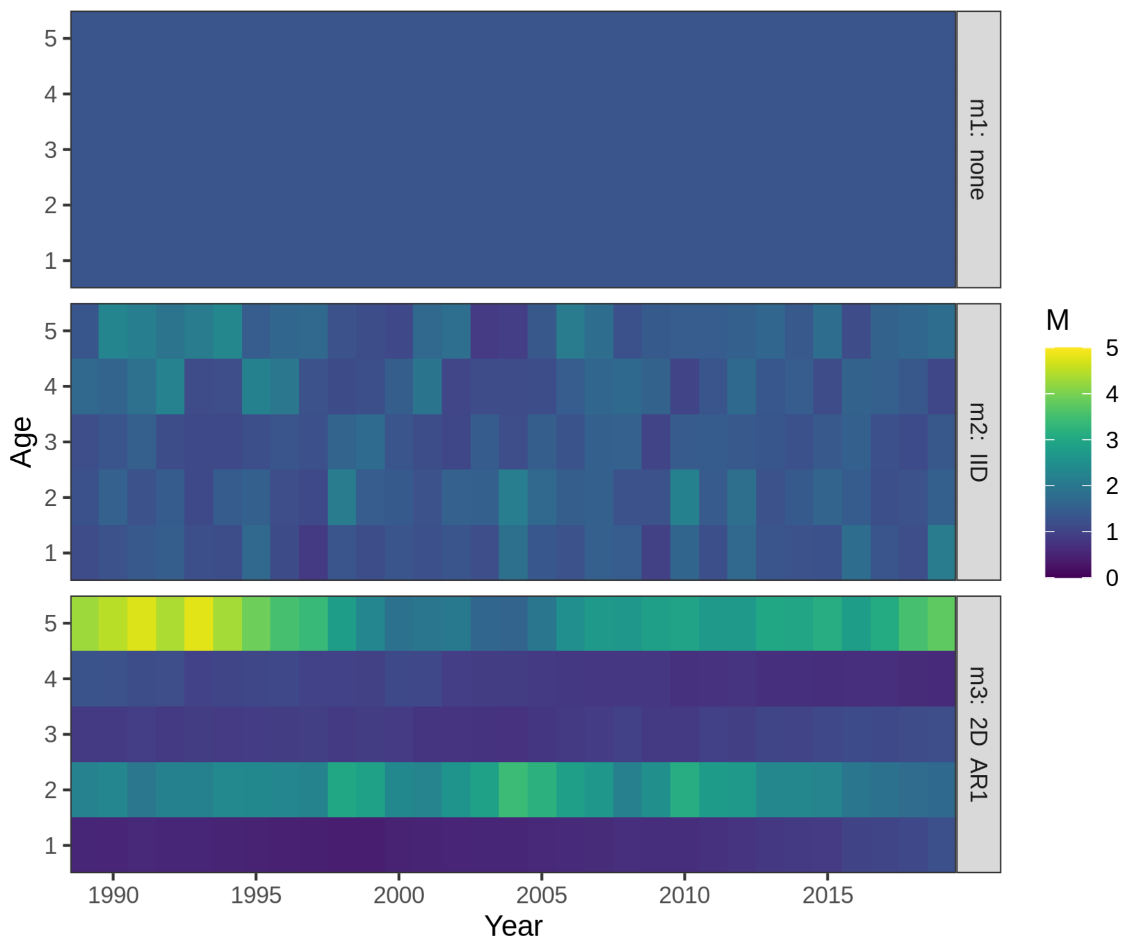

Figure 3: Natural mortality (M) estimated for butterfish using three random effects models. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).

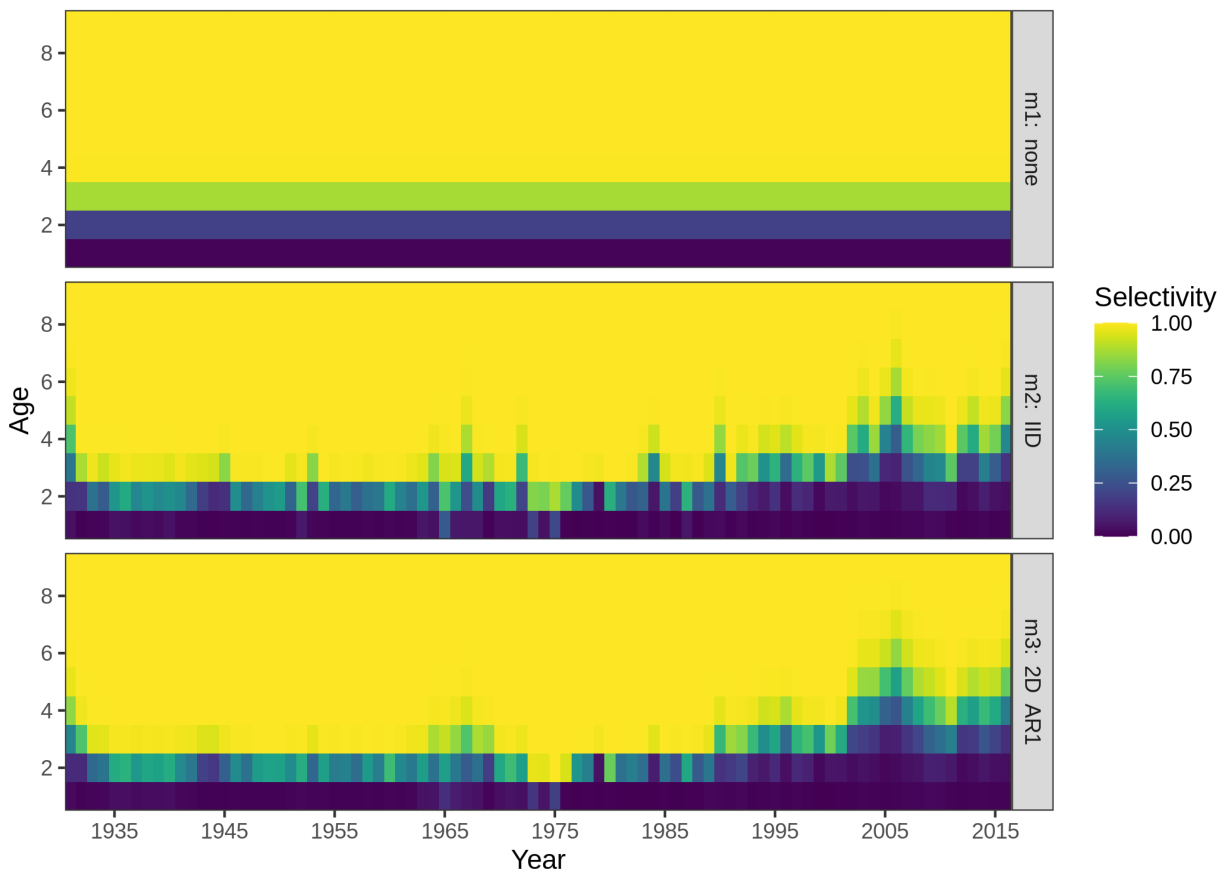

Figure 4: Selectivity estimated for Georges Bank haddock using three random effects models. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).

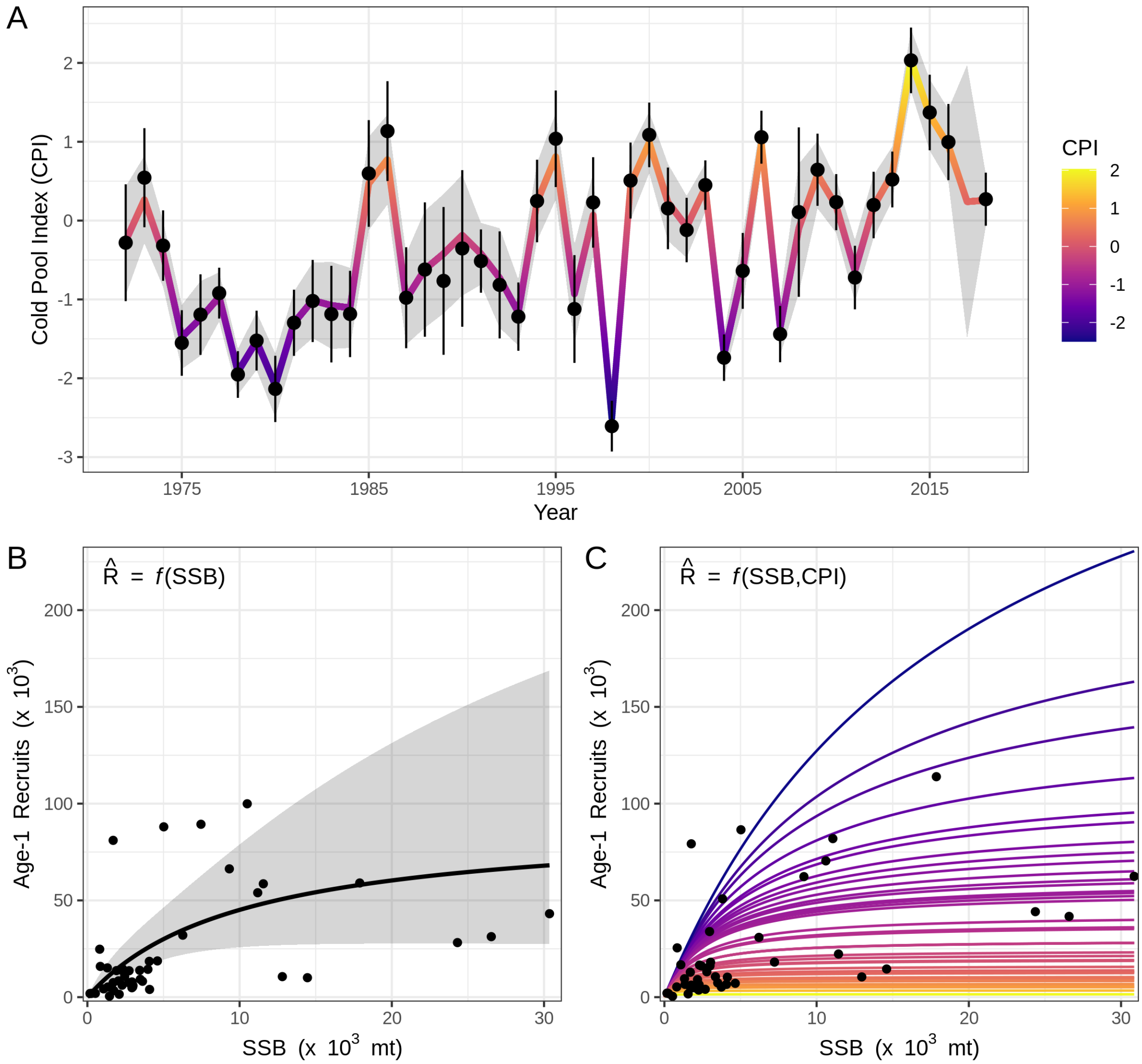

Figure 5: Beverton-Holt stock-recruit relationships fit for Southern New England-Mid Atlantic yellowtail flounder, with and without effects of the Cold Pool Index (CPI). A) CPI estimated from the model with lowest AIC (m4, AR1-linear). Points are observations with 95% CI, and the line with shading is 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).

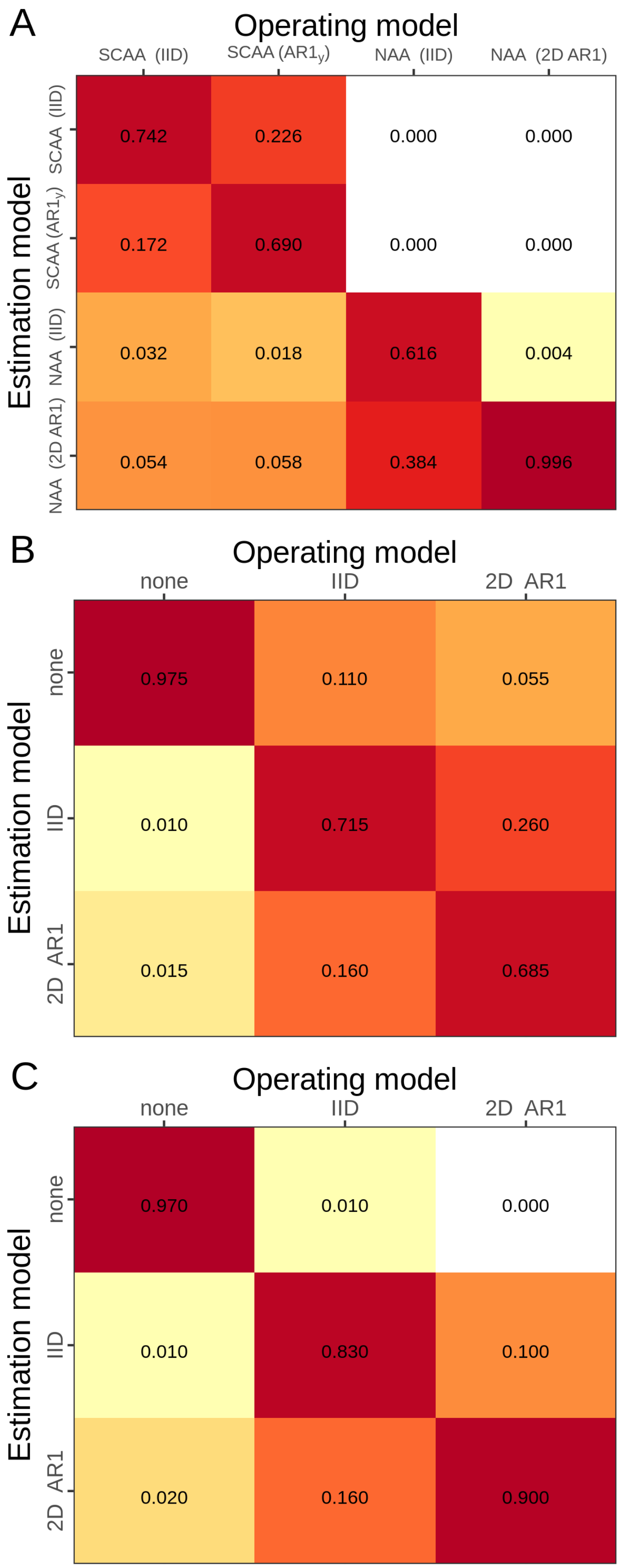

Figure 6: Proportion of simulations in which each model had the lowest AIC. A) Numbers-at-age (NAA), aggregated across all five stocks. B) Natural mortality (M), aggregated over two stocks (SNEMAYT and NScod). C) Selectivity (GBhaddock). Not all estimation models converged for each simulation, even when the operating model matched.

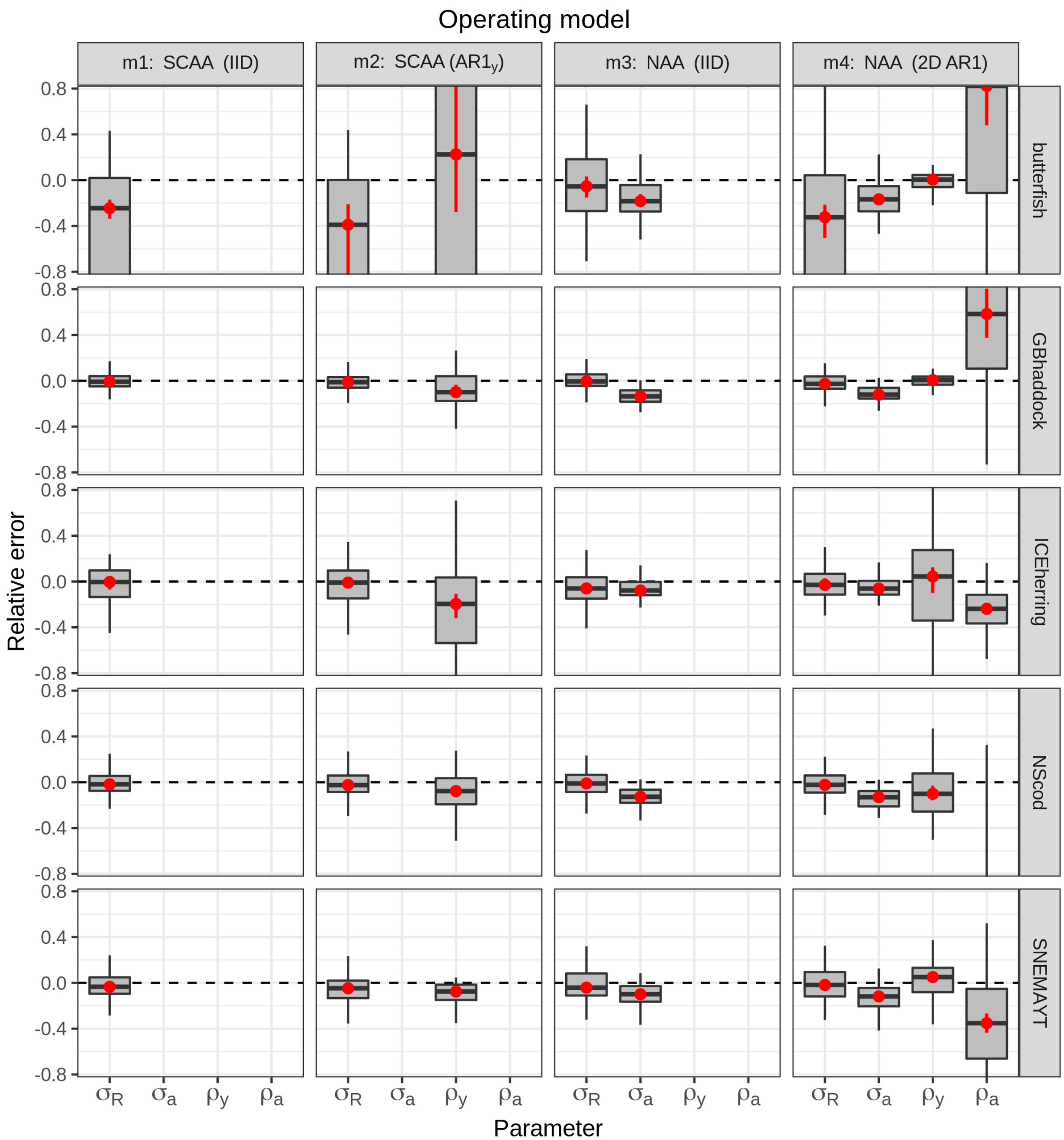

Figure 7: Relative error of parameters constraining numbers-at-age (NAA) random effects. Four models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = only recruitment deviations are random effects (most similar to traditional 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.

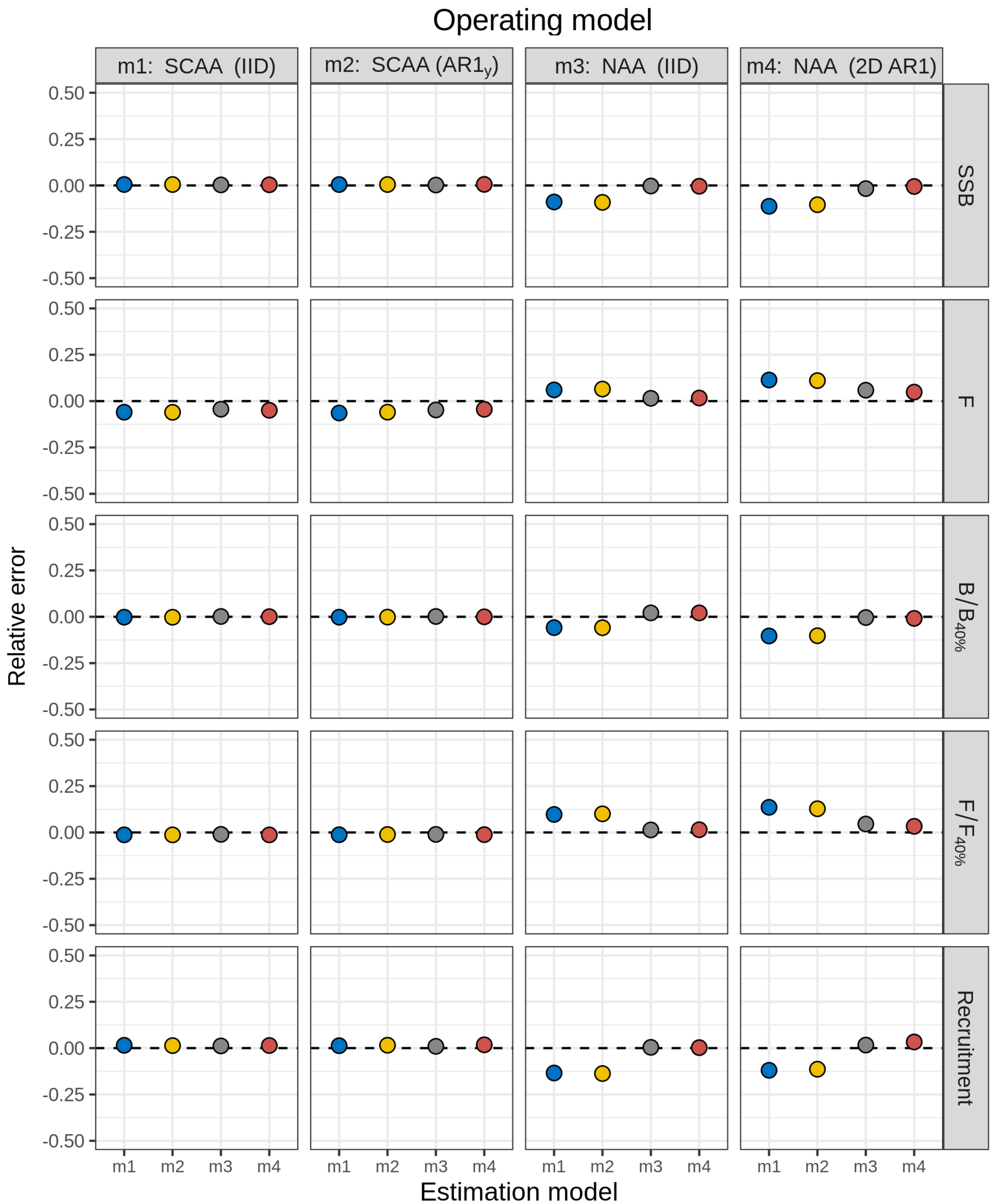

Figure 8: Relative error of key quantities estimated for Icelandic herring using four models of numbers-at-age (NAA) random effects. m1 = only recruitment deviations are random effects (most similar to traditional statistical catch-at-age, SCAA), and deviations are independent and identically distributed (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).

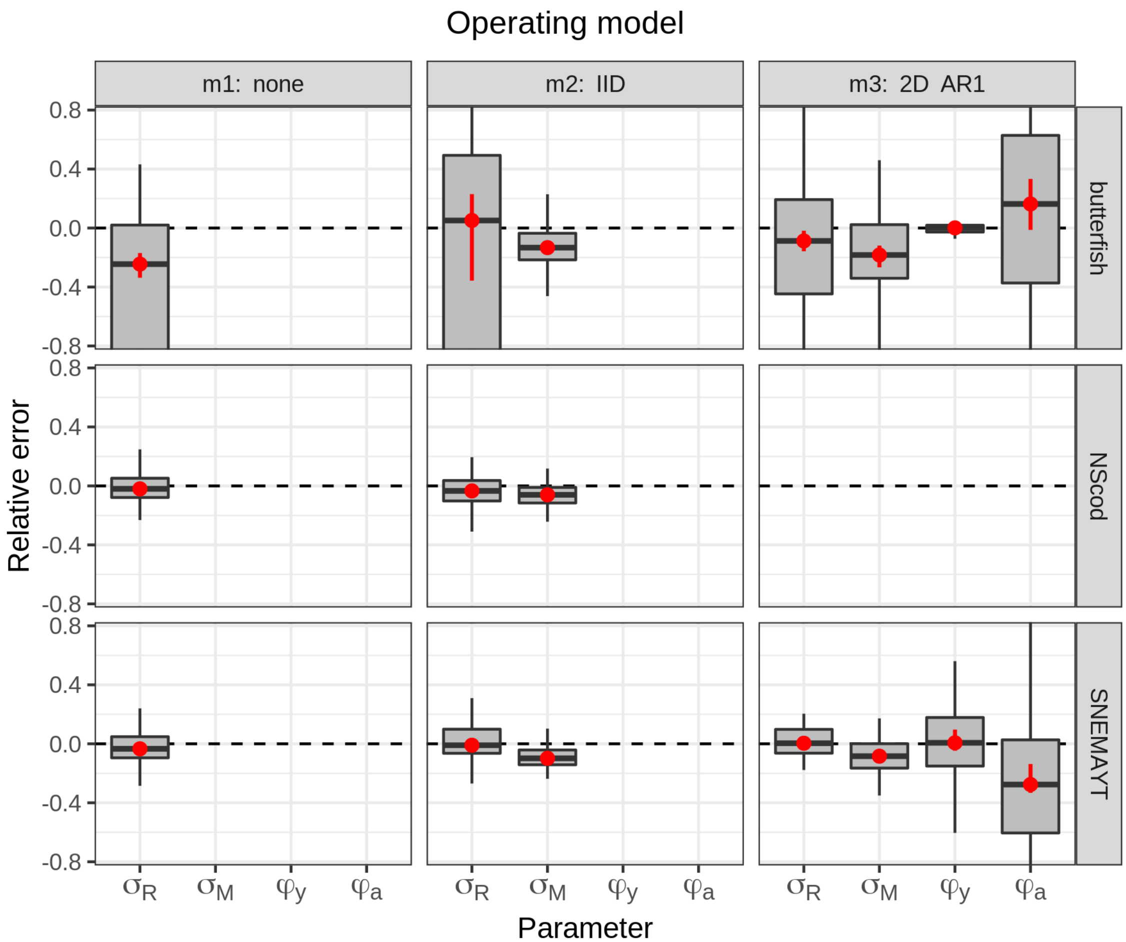

Figure 9: Relative error of parameters constraining natural mortality (M) random effects. Three models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = no random effects on M. m2 = M deviations were independent and identically 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).

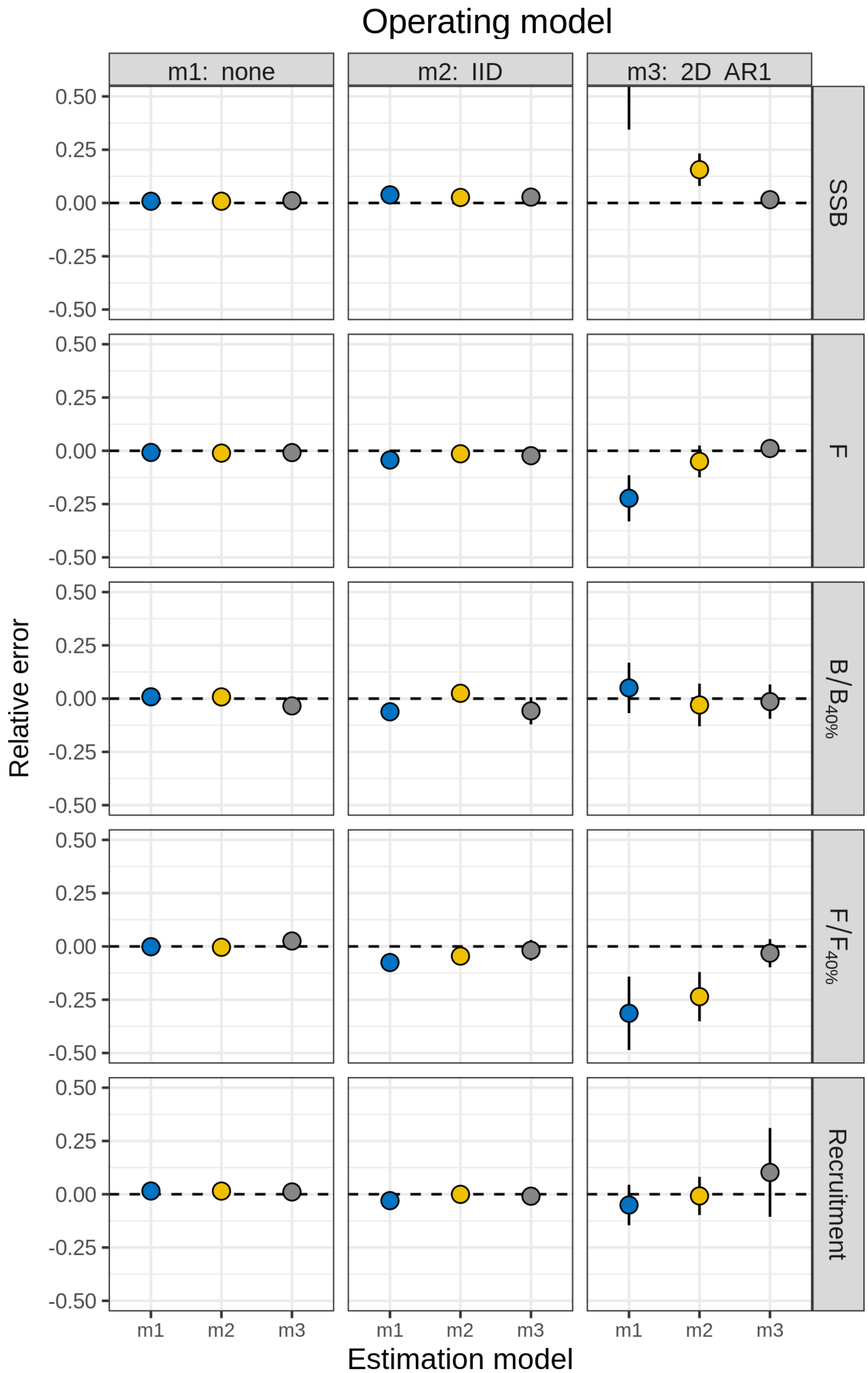

Figure 10: Relative error of key quantities estimated for butterfish using three models of natural mortality (M) random effects. m1 = no random effects on M. m2 = M deviations are independent and identically distributed (IID). m3 = M deviations are correlated by age and year (2D AR1).

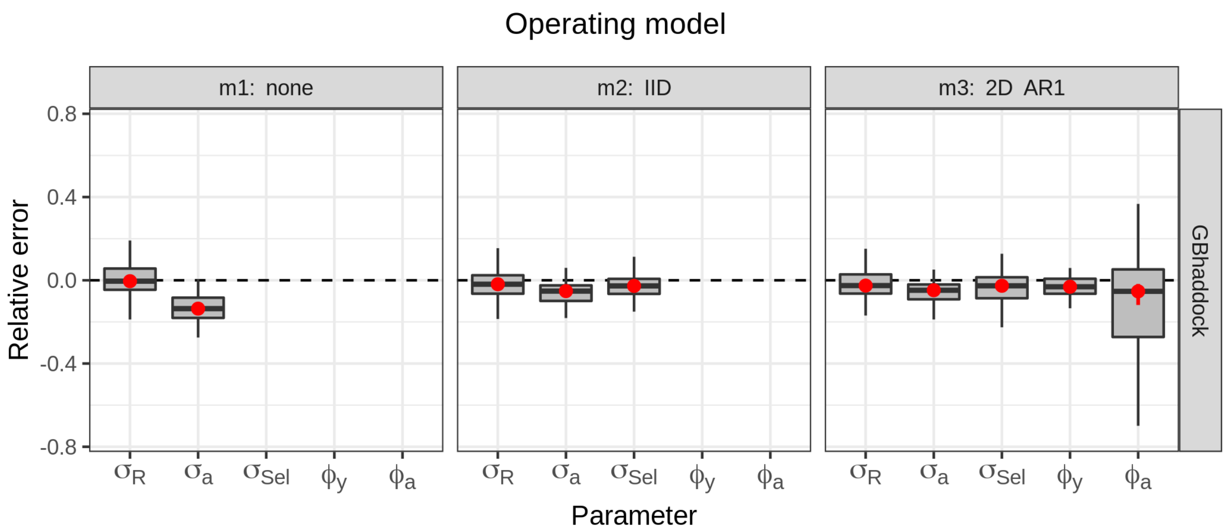

Figure 11: Relative error of parameters constraining selectivity random effects for Georges Bank haddock (GBhaddock). Three models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. m1 = no random effects (constant selectivity). m2 = 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.

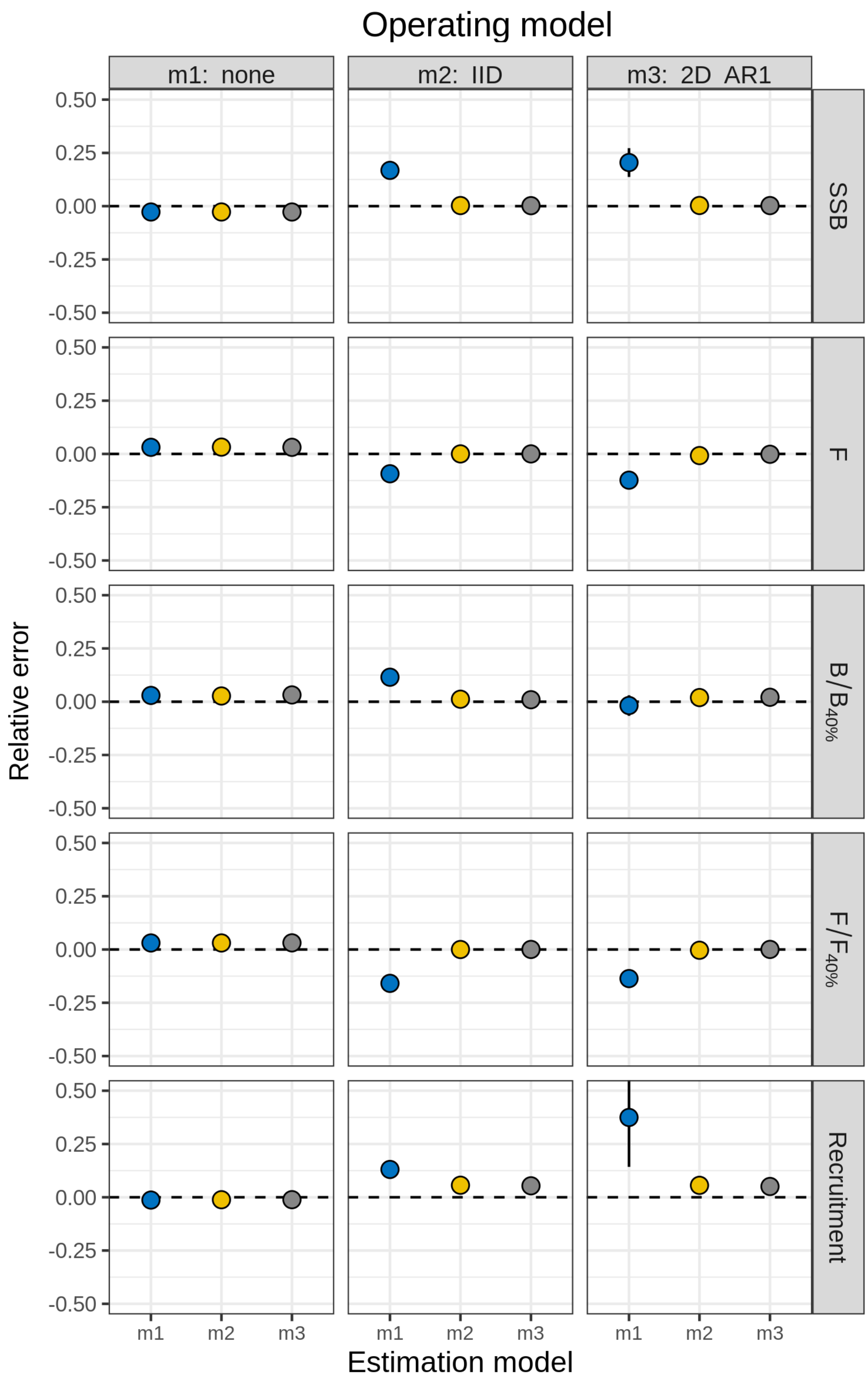

Figure 12: Relative error of key quantities estimated for Georges Bank haddock using three models of selectivity random effects. m1 = no random effects (constant logistic selectivity). m2 = selectivity deviations are independent and identically distributed (IID). m3 = selectivity deviations are correlated by parameter and year (2D AR1).

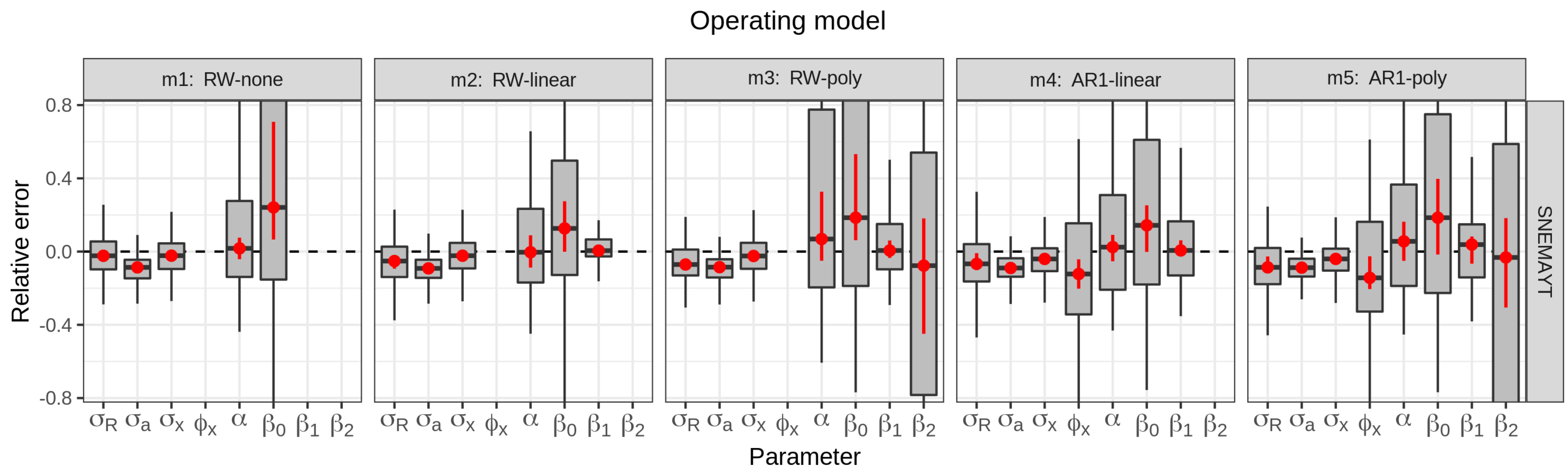

Figure 13: Relative error of parameters constraining variation in recruitment for Southern New England-Mid Atlantic yellowtail flounder (SNEMAYT). Five models were used to simulate 100 datasets keeping fixed effect parameters constant, and then re-fit to each simulated dataset. All models estimated recruitment 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.