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

Brian C. Stock[1*], Haikun Xu[2], Timothy J. Miller[1], James T. Thorson[3], and Janet A. Nye[4]

[1] *NOAA Northeast Fisheries Science Center, Woods Hole, MA, USA*

[2] *Inter-American Tropical Tuna Commission, La Jolla, CA, USA*

[3] *NOAA Northwest Fisheries Science Center, Seattle, WA, USA*

[4] *Institute of Marine Science, University of North Carolina Chapel Hill, Morehead City, NC, USA*

**Corresponding author*:

Brian Stock, brian.stock@noaa.gov, +1 425-919-7879

NEFSC, 166 Water St, Woods Hole, MA, USA

Abstract

Survival is an important population process in fisheries stock assessment models and is typically treated as deterministic. Recently developed state-space assessment models can estimate stochastic deviations in survival, which represent variability in some ambiguous combination of natural mortality (*M*), fishing mortality (*F*), and migration. These survival deviations are generally treated as independent by age and year, despite our understanding that many population processes can be autocorrelated and that not accounting for autocorrelation can result in notable bias. We address these concerns, as well as the strong retrospective pattern found in the last assessment of Southern New England yellowtail flounder (*Limanda ferruginea*), by incorporating two-dimensional (2D, age and year) first-order autocorrelation in survival and *M*. We found that deviations were autocorrelated among both years (0.52 ± 0.10, 0.47 ± 0.21) and ages (0.34 ± 0.12, 0.39 ± 0.14) when estimated for survival or *M*, respectively. Models with the 2D AR(1) smoother on survival or *M* fitted the data better and had reduced retrospective pattern (lower AIC and Mohn's ρ). The lowest Mohn's ρ was achieved by simultaneously including 2D AR(1) *M* deviations and independent survival deviations. Including an environmental effect on recruitment further reduced Mohn's ρ for recruitment but did not affect Mohn's ρ for spawning stock biomass (SSB) or *F*. The 2D AR(1) smoother altered estimates of SSB and *F* by 10-15% in model years, whereas SSB changed by 30% in short-term projections with $F = 0$. We conclude that incorporating 2D autocorrelated variation in survival, *M*, or both could improve the assessment of Southern New England yellowtail flounder in terms of model fit and consistency of SSB projections.

Keywords: state-space model; stock assessment; random effects; survival; natural mortality; autocorrelation; yellowtail flounder

1. Introduction

Biological processes of a fish population usually, if not always, vary over time and age. For instance, a process such as recruitment can be autocorrelated in time if the environmental or ecological process by which it is driven is autocorrelated in time (Johnson et al. 2016; Thorson et al. 2014). Johnson et al. (2016) found that in cases where recruitment is highly autocorrelated, ignoring this autocorrelation in stock assessment models can lead to large biases in model predictions as well as the associated uncertainty intervals. Processes such as selectivity can also be autocorrelated among ages because adjacent age classes are often more similar in size, physiology, behavior, etc. than disparate age classes (Nielsen and Berg 2014; Berg and Nielsen 2016). When variations in selectivity are autocorrelated among ages, models that estimate this autocorrelation have been shown to fit data better than models that assume independent selectivity variations (Nielsen and Berg 2014; Xu et al. 2019). Using a two-dimensional (2D) autocorrelation structure across both ages and years to model population processes is rare, although Xu et al. (2019) recently implemented 2D AR(1) selectivity deviations in Stock Synthesis (Methot and Wetzel 2013). Another exception is Cadigan (2016), who developed a state-space, age-structured assessment model for Northern Cod (*Gadus morhua*) where the among-age and among-year autocorrelations in natural mortality rate (M) were simultaneously estimated. Cadigan (2016) did not, however, compare the model estimates, predictions, goodness-of-fit, or retrospective patterns under alternative 2D autocorrelation structures for M.

Yellowtail flounder (*Limanda ferruginea*) is a commercially important demersal flatfish in the Northwest Atlantic ranging from the Labrador Sea in the North to the Chesapeake Bay in the South (NEFSC 2012). There are four stocks of yellowtail flounder delineated by the following areas: Canadian Grand Banks, Cape Cod-Gulf of Maine, Georges Bank, and Southern

New England-Mid Atlantic. All four stocks experienced overfishing from the 1970s to the mid-1990s and since then all stocks have experienced some recovery except for the southern-most Southern New England-Mid Atlantic (SNEMA) stock (Stone et al. 2004). The SNEMA stock is currently assessed using a statistical catch-at-age model, the Age-Structured Assessment Program (ASAP; Legault and Restrepo 1999), and has declined in recent years to historic lows (5% of SSB_{MSY} proxy; NEFSC 2020). There are two major sources of uncertainty in recent stock assessments (NEFSC 2012, 2020). First, the assessment cannot explain the dramatic decrease in recruitment since the 1990s. Recent studies have linked poor recruitment to low spawning stock biomass (SSB) as well as unfavorable environmental conditions (more northward Gulf Stream and reduced cold pool; Miller et al. 2016; Xu et al. 2018). Second, there is a strong retrospective pattern for SSB and fully-selected fishing mortality rate (*F*). The cause underlying the retrospective patterns is unclear, but no doubt, strong retrospective patterns in SSB and *F* can induce bias and uncertainty in the determination of stock and harvest status (Brooks and Legault 2016; Miller and Legault 2017).

Retrospective pattern refers to the systematic inconsistency in estimates of fishery variables when additional years of data are included in the assessment model (Mohn 1999). It typically arises due to misspecifying temporal changes in input data or biological parameters, e.g. assuming a parameter is constant in the model when it varies in reality (Hurtado-Ferro et al. 2014; Legault 2009). To address retrospective issues in the assessments of some New England fish stocks such as Georges Bank yellowtail flounder (Legault et al. 2012) and Gulf of Maine Atlantic cod (NEFSC 2013), scientists sometimes impose a temporal trend on *M* in stock assessment models. *M* is an important parameter because it directly influences stock productivity. Moreover, misspecifying *M* can lead to biased estimation of population attributes

(Miller and Legault 2017; Thorson et al. 2015) and key reference points such as virgin biomass and maximum sustainable yield (Johnson et al. 2015). M is often difficult to estimate because it is confounded with other parameters such as F and recruitment, so M is usually pre-specified as a time-invariant constant to simplify model estimation (Deroba and Schueller 2013; Johnson et al. 2015; Legault and Palmer 2015). Deroba and Schueller (2013) conducted a simulation experiment to evaluate the estimation biases in SSB and recruitment that are induced by misspecifying M. Their results underline the importance of correctly specifying the temporal trend in M to stock assessment models, because misspecifying the temporal variation in M induced larger biases than misspecifying the age-variation in M. The impacts of misspecifying M are positively related to $M/(F+M)$, and $M/(F+M)$ was near the historical maximum for SNEMA yellowtail flounder in 2018 since F was at a historic low (Legault and Palmer 2015; NEFSC 2020). The current SNEMA yellowtail flounder assessment specifies M as a time-invariant, decreasing function of age, based on a time series average of weight-at-age data and the allometric relationship defining how M declines with size (Lorenzen 1996; NEFSC 2012). Thus, there is reason to believe that misspecifying M as constant could be a reason for the concerning retrospective patterns in recent assessments of the stock.

Instead of imposing a trend on M by age or year, we explored a more flexible and objective method to address the retrospective problem: estimating a 2D smoother that was first-order autoregressive, AR(1), over both age and year. We implemented this 2D AR(1) structure in the Woods Hole Assessment Model (WHAM), a state-space age-structured assessment framework developed at the NEFSC (Miller and Stock 2020; Stock and Miller, this issue). Whereas statistical catch-at-age models do not distinguish between observation and process errors, state-space models are able to simultaneously estimate process error (variance of

unobserved states, such as population numbers-at-age), and the observation errors in associated data (Nielsen and Berg 2014; Miller et al. 2016; Aeberhard et al. 2018). Statistical catch-at-age models assume that survival is deterministic, i.e. the number of age a fish in year y, $N_{a,y}$, is determined by F, M, and the numbers in the previous year: $N_{a,y} = N_{a-1,y-1}e^{-\left(F_{a-1,y-1}+M_{a-1,y-1}\right)}$. Process errors, ε, can be included directly on $N_{a,y}$ as random effect deviations in survival (Gudmundsson and Gunnlaugsson 2012; Nielsen and Berg 2014; Miller et al. 2016) or on $M_{a,y}$ (Cadigan 2016), such that $N_{a,y} = N_{a-1,y-1}e^{-\left(F_{a-1,y-1}+M_{a-1,y-1}\right)+\varepsilon_{a,y}}$. Here, we extend the model presented in Miller et al. (2016), which included stochastic deviations in annual survival assumed to be uncorrelated by age or year, i.e. all $\varepsilon_{a,y} \sim \mathrm{N}\,(0, \sigma_a^2)$. The $\varepsilon_{a,y}$ can be equivalently called "random effects," "deviations," or "process errors" on numbers-at-age or survival. Hereafter, we refer to the $\varepsilon_{a,y}$ as random effects or deviations.

In this study, we investigated 2D autocorrelation structures on both survival and M in an assessment of SNEMA yellowtail flounder. In software packages such as AD Model Builder (ADMB; Fournier et al. 2012) and Template Model Builder (TMB; Kristensen et al. 2016), alternative models can be objectively compared based on metrics of retrospective pattern, e.g. Mohn's ρ (Mohn 1999), and model fit, e.g. Akaike's information criterion (AIC; Akaike 1973; Burnham and Anderson 2002). We assessed whether it was better to place the 2D AR(1) smoother on survival versus M, if a model could be estimated with 2D AR(1) smoothers on both survival and M, and the impact on estimates of SSB and F. Finally, we evaluated whether incorporating an environmental effect in the stock-recruit function could further improve retrospective patterns and model fit.

2. Material and Methods

2.1. 2D AR(1) smoother

We first compared various autocorrelation structures for survival deviations in WHAM. For simplicity, we only considered the first-order autocorrelation structure that has been widely used in previous studies (Cadigan 2016; Nielsen and Berg 2014). In WHAM, the stochastic survival deviations, $\varepsilon_{a,y}$, for age a and year y ($0 < y < Y$) can be calculated by rewriting the stock equations:

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

where N represents numbers at age, Z is the total mortality rate ($F + M$), and g is the stock-recruit function in which an environmental time series (x) can be incorporated as a covariate. Y is the total number of observation and prediction years and A represents the plus-group. This state-space model is unique in its ability of incorporating the stochastic change of the environmental covariate over time, uncertainty in associated observations, and its effect on recruitment as a covariate in the stock-recruit function (Miller et al. 2016; Stock and Miller, this issue). Strictly speaking, the survival deviation term (ε) stands for population migration into or out of the stock because it does not alter either the M or F in the Baranov catch equation (Gudmundsson and Gunnlaugsson 2012), and in fact realized "survival" can be greater than one (i.e., $N_{a,y} > N_{a-1,y-1}$ whenever $\varepsilon_{a,y} > Z_{a-1,y-1}$). Indeed, population mixing between adjacent yellowtail flounder stocks has been observed in tagging studies, but the extent of which was not large enough to significantly affect the population dynamics of each individual stock, including the depleted SNEMA stock (Cadrin 2003; Goethel et al. 2015). Gudmundsson and Gunnlaugsson

(2012) claimed that the survival deviation term can also be interpreted as "irregular natural mortality", because M impacts population dynamics primarily through stock equations. Cadigan (2016) and Aldrin et al. (2020) follow this interpretation. In fact, this survival deviation term can also be caused by deviations in fishing mortality or more generally deviations from the Baranov catch equation.

Miller et al. (2016) and Nielsen and Berg (2014) assumed that survival deviations are independent of age and time, and normally distributed with mean zero. In other words, for all a and y:

$$\varepsilon_{a,y} \sim \mathrm{N}\,(0, \sigma_a^2) \quad (2)$$

where σ_a for all ages $a > 1$ were assumed to be the same but different from age a = 1, i.e. recruitment, which we denote as σ_R. This assumption rests on the fact that survival variations for young-of-the-year (recruitment) are generally larger than for other ages. Unlike statistical catch-at-age models, σ_a and σ_R can be estimated internally as fixed effect parameters. Survival deviations, however, are not necessarily independent. If survival deviations are autocorrelated among ages and years, they follow a multivariate normal distribution:

$$\mathbf{E} \sim \mathrm{MVN}\,(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{total}}) \quad (3)$$

where $\mathbf{E} = (\varepsilon_{1,1}, \dots, \varepsilon_{1,Y-1}, \varepsilon_{2,1}, \dots, \varepsilon_{2,Y-1}, \dots \dots, \varepsilon_{A,1}, \dots, \varepsilon_{A,Y-1})'$, $\boldsymbol{\Sigma}_{\boldsymbol{total}}$ is the $A(Y-1) \times A(Y-1)$ covariance matrix for the multivariate normal distribution and is calculated as the Kronecker product of the $A \times A$ covariance matrix for the AR(1) process among ages ($\boldsymbol{\Sigma}$) and the $(Y-1) \times (Y-1)$ correlation matrix for the AR(1) process among years ($\widetilde{\boldsymbol{\Sigma}}$):

$$\boldsymbol{\Sigma}_{\boldsymbol{total}} = \boldsymbol{\Sigma} \otimes \widetilde{\boldsymbol{\Sigma}} \quad (4)$$

$$\boldsymbol{\Sigma}_{a,\tilde{a}} = \rho_{age}^{|a-\tilde{a}|} \sigma_a \sigma_{\tilde{a}}$$

$$\widetilde{\boldsymbol{\Sigma}}_{y,\tilde{y}} = \rho_{year}^{|y-\tilde{y}|}$$

where ρ_{age} and ρ_{year} are the two AR(1) coefficients in age and time, respectively. Either of them can be fixed at a constant between -1 and 1 or estimated in the state-space model as fixed effect parameters. Note that when both ρ_{age} and ρ_{year} are fixed at 0, there is no survival autocorrelation in either dimension and Eq. 3 collapses to Eq. 2. In fact, MVN $(\mathbf{0}, \boldsymbol{\Sigma_{total}})$ is the likelihood distribution function for this covariance structure:

$$\mathrm{Cov}\left(\varepsilon_{a,y}, \varepsilon_{\tilde{a},\tilde{y}}\right) = \frac{\sigma_a \sigma_{\tilde{a}} \rho_{age}^{|a-\tilde{a}|} \rho_{year}^{|y-\tilde{y}|}}{\left(1-\rho_{age}^2\right)\left(1-\rho_{year}^2\right)} \quad (5)$$

which means that the covariance between two survival deviations is positively related to how close the locations of the two survival deviations are on the age-time surface. As above, σ_a for all ages $a > 1$ were assumed equal but different from age a = 1, i.e. recruitment, denoted as σ_R.

Alternatively, we can apply this 2D AR(1) structure to random effect deviations in M, $\delta_{a,y}$, as in Cadigan (2016):

$$\log\left(M_{a,y}\right) = \mu_a + \delta_{a,y}$$

$$\mathrm{Cov}\left(\delta_{a,y}, \delta_{\tilde{a},\tilde{y}}\right) = \frac{\sigma_M^2 \varphi_{age}^{|a-\tilde{a}|} \varphi_{year}^{|y-\tilde{y}|}}{\left(1-\varphi_{age}^2\right)\left(1-\varphi_{year}^2\right)} \quad (6)$$

where μ_a is the mean $\log(M)$ at age a and can either be fixed or estimated. Applying deviations in $\log(M_{a,y})$ versus $\log(N_{a,y})$ determines where in the Baranov catch equation they affect the predicted catch, $\hat{C}_{a,y} = N_{a,y} \frac{F_{a,y}}{Z_{a,y}} (1 - e^{-Z_{a,y}})$. Another difference is that deviations in M affect the calculation of reference points, whereas deviations in $N_{a,y}$ do not.

2.2. Model descriptions

We first considered six models treating only the numbers-at-age (NAA) as random effects (Table 1). These models estimated deviations in survival by age and year, $\varepsilon_{a,y}$, assuming alternative autocorrelation structures formed by fixing or estimating the three parameters in Eq. 5. The

205 "Base" model was most similar to a statistical catch-at-age model, where recruitment is typically
206 estimated independently by year and survival is deterministic. NAA-1 added recruitment
207 autocorrelation. NAA-2 through NAA-5 estimated "full state-space" models, with numbers at all
208 ages treated as random effects, but with different autocorrelation structures. NAA-2 estimated
209 independent $\varepsilon_{a,y}$ as in Miller et al. (2016) and Nielsen and Berg (2014), NAA-3 and NAA-4
210 added autocorrelation across ages and years, and NAA-5 estimated all parameters in the
211 described 2D AR(1) smoother (Eq. 5, Table 1). To isolate the effect of incorporating the 2D
212 AR(1) smoother on survival, we compared the model fit, retrospective pattern, and relative
213 difference in SSB and *F* estimates from NAA-5 versus NAA-2.
214 Next, we fit a series of models treating the numbers-at-age as in Base, but including
215 deviations in *M* as in Eq. 6 with the same set of autocorrelation structures: none, independent,
216 AR(1) by age, AR(1) by year, and 2D AR(1) (Table 2). As for the set of NAA models, we
217 isolated the effect of the 2D AR(1) smoother on *M* by comparing M-1 to M-4. We then tested the
218 ability of WHAM to simultaneously estimate numbers-at-age and *M* as random effects, using
219 only the independent and 2D AR(1) autocorrelation structures for each (Table 3).
220 Finally, we tested whether linking an environmental covariate to recruitment could
221 further improve model fit and retrospective pattern (Table 4). Miller et al. (2016) and Xu et al.
222 (2018) found that incorporating the Cold Pool Index (CPI) or Gulf Stream Index (GSI) into the
223 stock-recruit function for SNEMA yellowtail flounder significantly improved model
224 performance as measured by AIC and Mohn's ρ. Here, we used the CPI instead of the GSI
225 because standard errors were straightforward to calculate for the CPI and these values could be
226 used directly in the assessment model as observation error. This avoided the need to estimate the
227 CPI observation error internally. We calculated the CPI as in Miller et al. (2016), updated it with

additional data through 2018, and incorporated it into the Beverton-Holt stock-recruit function as a limiting factor:

$$g\left(\theta, CPI_{y-1}, SSB_{y-1}\right) = \frac{SSB_{y-1}}{\beta e^{\gamma CPI_{y-1}} + \alpha SSB_{y-1}} \quad (7)$$

where g is from Eq. 1.

2.3. Application to SNEMA yellowtail flounder

We evaluated the performance of our proposed 2D AR(1) survival smoother by using data from the 2019 SNEMA yellowtail flounder stock assessment as a case study (NEFSC 2020). We included likelihood components for the following observations through 2018: (1) three indices of abundance from the spring, fall, and winter NEFSC bottom trawl surveys; (2) aggregate catch from one commercial fleet; and (3) age composition from the three bottom trawl surveys and the commercial catch. As in Miller et al. (2016), age composition data were assumed to follow a logistic-normal distribution with pooling of zero observations (Atchison and Shen 1980). Empirical weight-at-age, natural mortality-at-age, and maturity-at-age were treated as known. Maturity was fixed at 0.0052, 0.6836, 0.9854, 0.9970, 0.9963, and 1 for ages 1-6, while natural mortality was specified as 0.405, 0.336, 0.296, 0.275, 0.256, and 0.2311 yr^{-1} for ages 1-6 (NEFSC 2020). Selectivity of the fleet was divided into six time blocks as in NEFSC (2020). We estimated logistic selectivity for the fleet and indices, except in three time blocks where age-specific, flat-topped selectivity facilitated convergence, i.e. we fixed selectivity at 1 for older ages and estimated selectivity at younger ages as free parameters. During three prediction years, weight and maturity were fixed at the values in the terminal year of the data, and *F* was fixed at 0 to facilitate comparisons of SSB projections between models. Variables treated as random effects, such as numbers-at-age, *M*, and the CPI, were forecast in the prediction years by simply

continuing the stochastic process, e.g. AR(1) for the CPI. The data and assessment report can be accessed at https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi_report_options.php.

We fit the models using WHAM, an R package that utilizes TMB to fit age-structured, state-space stock assessments (Miller and Stock 2020). TMB calculates the marginal likelihood of fixed effect parameters using the Laplace approximation to integrate across random effect parameters (Kristensen et al. 2016), and fixed effect parameters are then estimated by maximizing the marginal likelihood within R (R Core Team 2020). After the fixed effect parameters are estimated, TMB predicts the random effect coefficients using empirical Bayes (Kristensen et al. 2016). We compared model fit and retrospective pattern using AIC and Mohn's ρ (Mohn 1999), using seven retrospective peels as in the latest assessment (NEFSC 2020).

3. Results

3.1. Numbers-at-age (survival) as random effects

Treating numbers at all ages as random effects resulted in markedly better model fit (lower AIC) and reduced retrospective pattern (lower Mohn's ρ; Table 1). Estimating survival deviations with autocorrelation by age, year, or both further reduced AIC and Mohn's ρ. According to both AIC and the magnitude of estimated ρ_{age} and ρ_{year}, the among-year autocorrelation in survival deviations was higher and had larger impact on model fit than the among-age autocorrelation in survival deviations (Table 1). The survival deviations estimated by models with autocorrelation were smoothed across ages or/and years relative to the models with independent deviations (Fig. 1). NAA-5, with the 2D AR(1) structure, had the best fit and retrospective pattern, reducing AIC by 36.2, ρ_R by 0.41, ρ_{SSB} by 0.18, and ρ_F by 0.14 compared to NAA-2 with independent survival deviations (Table 1). Simply constraining the survival deviations with the 2D AR(1) structure reduced estimates of *F* by 12% and increased estimates of SSB by 9% (mean relative

difference between NAA-5 and NAA-2; Fig. 2a-b). NAA-2 and NAA-5 estimated similar SSB in the terminal year of the assessment, but then differed in their SSB estimates in the projection years by 22% when *F* was fixed at 0 (Fig. 2a-b).

The survival deviations estimated in years near the end of the assessment impacted the projections of SSB. For instance, all NAA models estimated very low recruitment in 2015, i.e. strong negative survival of age-1 fish, and because ρ_{year} was > 0 this propagated through the end of the assessment and into the projection years for models that included autocorrelation by year (Fig. 1). In the terminal year, these models also estimated negative survival deviations for ages 2-4, but slightly positive deviations for the plus-group. The combined effect of the survival deviations across ages resulted in the model with 2D AR(1) autocorrelation projecting lower SSB in 2019, but higher SSB in 2020-2021, than the model with independent deviations (Fig. 2b). Note that although the projected survival deviations asymptotically approached zero over time (Fig. 1), SSB in a given projection year is the result of *cumulative* survival deviations, and the influence of the survival smoother on SSB is not necessarily weaker over time (Fig. 2b).

3.2. Deviations in *M* as random effects

Including deviations in *M*, instead of survival, also substantially improved model fit and retrospective pattern (Table 2). In contrast to treating numbers-at-age as random effects, including 1D autocorrelation by age or year led to worse fit and retrospective pattern than estimating independent *M* deviations. The 2D AR(1) structure again had the best fit (M-4; Table 2). Compared to the models with independent *M* deviations, including the 2D AR(1) structure reduced AIC by 5.9, $|\rho_R|$ by 0.05, $|\rho_{SSB}|$ by 0.10, and $|\rho_F|$ by 0.06 (Table 2). The estimated 2D AR(1) *M* deviations had higher variance and lower autocorrelation in time than the 2D AR(1) survival deviations, and therefore appeared less smoothed (Figs. 1 and 3, $\sigma_M > \sigma_a$ and $\varphi_{year} <$

296 ρ_{year} in Tables 1-2). The effect of adding the 2D AR(1) structure on estimates of *F* and SSB was
297 similar as for the NAA models: 10% lower *F* and 7% higher SSB during assessment years,
298 similar terminal year status, and then 22% lower SSB in short-term projections (Fig. 2). *M*
299 deviations for younger ages in models with 2D AR(1) autocorrelation were positive in the
300 terminal year (Fig. 3). This was consistent with the NAA 2D AR(1) models estimating negative
301 survival in this period (Fig. 1), and explained why adding 2D AR(1) autocorrelation on *M*
302 deviations led to lower SSB in short-term projections (Fig. 2).

303 **3.3. Estimating deviations in both survival and *M***

304 The model that attempted to estimate deviations in both survival and *M* with 2D AR(1)
305 autocorrelation failed to converge. However, adding independent *M* deviations to NAA-5 (2D
306 AR(1) on survival) and adding independent survival deviations to M-4 (2D AR(1) on *M*)
307 substantially improved model fit (lower AIC by 25.9 and 9.5, respectively; Tables 1-3).
308 Differences in Mohn's ρ between these models were not large or consistent. In the model with
309 lowest Mohn's ρ, deviations in both survival and *M* reinforced each other—positive survival
310 deviations were estimated for the same years and ages as negative *M* deviations (e.g. ages 1-3
311 during the late 1970s-1980s in Fig. 4). Thus, the magnitude of the deviations reduced from
312 approximately -2 to 2, as when only survival or *M* deviations were estimated (Figs. 1 and 3), to
313 the range -1 to 1 (Fig. 4). Adding independent deviations in survival to the model with 2D AR(1)
314 deviations in *M* resulted in higher φ_{year} (increase from 0.47 to 0.76), which propagated the
315 elevated *M* in the terminal year stronger and further into the projection years (M-4 in Fig. 3; top
316 panel in Fig. 4). The result was a more pronounced decrease in projected SSB (Fig. 2d).

3.4. Incorporating the CPI-recruitment effect

Finally, we investigated the effect of additionally incorporating an environmental covariate, the CPI, into the stock-recruit function for the NAA-M models (Table 4). Across models with different random effects on survival and *M*, including a CPI effect on recruitment reduced Mohn's ρ_R by about 0.1 on average and did not have much influence on ρ_{SSB} or ρ_F (Fig. 5). NAA-M-CPI-2 had the lowest Mohn's ρ of all models considered (Table 5, Fig. 6). Including the CPI-recruitment effect did not have much impact on the estimates of SSB and *F*, given that deviations in survival and *M* for all ages were already included (Table 5, Fig. 7).

4. Discussion

Using a state-space, age-structured assessment model developed for SNEMA yellowtail flounder, we showed that implementing a 2D AR(1) smoother on survival or *M* considerably improved model fit and reduced the retrospective patterns for SSB, *F*, and recruitment. Different from previous assessments in the region which addressed a retrospective problem by *a priori* specifying a temporal trend in M (e.g. Georges Bank yellowtail flounder and Gulf of Maine Atlantic Cod; Legault et al. 2012; NEFSC 2013, 2020), this paper provides a more objective, flexible, and generic approach to reduce retrospective pattern in stock assessments. In WHAM, the two autocorrelation coefficients in the 2D AR(1) smoother can either be specified at fixed values or estimated as parameters in the assessment model. This makes it easy to specify or estimate a temporal trend in *M* or survival, and then compare performance against models with constant or 2D AR(1) deviations. Specific to SNEMA yellowtail flounder, we showed that it is more important to estimate the among-year than among-age autocorrelation in survival deviations by comparing the AIC of several alternative models with differing autocorrelation assumptions. However, including both dimensions as a 2D AR(1) process consistently resulted in

the best model fit and lowest retrospective pattern. Moreover, the estimated survival and *M* deviations for older ages were of comparable magnitude to those for age 1 (Fig. 4), which means that the variation in processes at age 2+ may be more important than recruitment variation to the population dynamics of SNEMA yellowtail flounder.

We found that the 2D AR(1) smoother impacted SSB and F estimates in model years as well as near-term SSB predictions, while including an effect of the CPI on recruitment had less impact on the estimates of SSB and F. This underscores the importance of implementing survival or *M* deviations with the 2D AR(1) smoother to the assessment of SNEMA yellowtail flounder. The 2D AR(1) structure could also extended to three dimensions if, for example, survival is modelled to also be sex- or cohort-specific. A 3D AR(1) process across year, age, and cohort could be appropriate since cohort effects are visible in the estimated 2D AR(1) survival and *M* deviations (Figs. 1, 3, and 4). The generic 2D AR(1) random effects structure described here can also be applied to other potentially autocorrelated biological processes, e.g. selectivity as in Xu et al. (2019), and we imagine this will be an important research topic in the development of next-generation stock assessment models with mixed effects (Punt et al. 2020). Indeed, WHAM makes heavy use of 2D AR(1) random effects, currently allowing users to specify them on numbers-at-age, *M*, and selectivity (Stock and Miller, this issue).

For SNEMA yellowtail flounder, implementing the 2D AR(1) smoother on survival or *M* considerably reduced the severity of retrospective pattern and modified near-term predictions of SSB (Figs. 2 and 6). Incorporating 2D AR(1) deviations in survival led to lower AIC, while the model with 2D AR(1) deviations in *M* had lower Mohn's ρ (Tables 1-2, Fig. 6). The model with 2D AR(1) deviations in *M*, independent deviations in survival, and the CPI-recruitment effect had the lowest Mohn's ρ, with $|\rho_{SSB}|$, $|\rho_F|$, and $|\rho_R|$ all less than 0.11 (Fig. 6). This implies that

363 including these factors in the SNEMA yellowtail flounder assessment would provide more
364 consistent estimates of stock and harvest status.

365 Near-term SSB forecasts changed substantially, by 22%, when the 2D AR(1) smoother
366 was included to constrain deviations in survival or *M*. In models where the survival or *M*
367 deviations are independent (e.g. Base, NAA-2, and M-1), they do not affect projections of SSB
368 unless they are linked to an environmental covariate that is also projected. In contrast, including
369 autocorrelation by year (ρ_{year} or φ_{year}) propagates non-zero survival or M deviations into short-
370 term projections, with the trend near the assessment terminal year becoming important. In the
371 case of SNEMA yellowtail flounder, models with φ_{year} estimated positive *M* deviations in
372 recent years, and this clearly resulted in lower projected SSB. The effect of including the 2D
373 AR(1) on survival deviations was less straightforward, because projected survival deviations
374 were negative for younger ages and positive for older ages. The combined effect in the first
375 projection year was a reduction in forecasted SSB relative to the model with independent
376 deviations. In subsequent years, the forecasted SSB increased. The explanation for this
377 counterintuitive result is that the older ages contribute more to SSB than younger ages, and the
378 cumulative effect of consecutive years with slightly elevated survival of the plus-group trumped
379 the reduced survival of younger ages. Therefore, implementing 2D AR(1) autocorrelation in
380 survival or *M* is likely to have greater impact on near-term SSB predictions in the SNEMA
381 yellowtail flounder assessment than incorporating an environmental effect on recruitment.
382 Unfortunately, whether to include the 2D AR(1) smoother on survival or *M* was not clear—the
383 model with 2D AR(1) survival deviations (NAA-5) had lower AIC, while the model with 2D
384 AR(1) *M* deviations (M-4) had lower Mohn's ρ. However, both models estimated similar
385 impacts to SSB projections: increased uncertainty in all projection years and increased mean

SSB in 2019-2020. Taking an ensemble, or model averaging, approach is one sensible path forward, especially since allowing correlated time- and age-variation in survival versus *M* makes similar biological assumptions and affected estimates of SSB and F in the same direction (Table 5; Figs. 2 and 7).

Whether 2D AR(1) autocorrelation is included or not, allowing survival or *M* to vary in time does not make calculating biological reference points (BRPs) easier. We envision two different ways to calculate BRPs for models with deviations in survival or *M*:

1. *Deterministic BRPs*: The simplest calculation for BRPs is to ignore the stochastic variation in survival and calculate the yield curve given average survival. This procedure ignores variation in biological processes and is typically used to calculate BRPs.

2. *Dynamic BRPs*: BRPs can vary across time and be calculated annually based upon the survival deviation estimated for each year.

For SNEMA yellowtail flounder, since the 1990s estimated survival deviations were predominantly negative, and *M* deviations were mostly positive. Ignoring these trends in productivity may result in biased estimates of BRPs in recent decades. Furthermore, treating survival or *M* deviations as uncorrelated in time neglects to propagate productivity changes in short-term projections. Whereas calculating dynamic BRPs seems to be more appropriate, it poses a challenge to management because an assumption regarding how survival or *M* deviations are attributed to fishing mortality, natural mortality, and migration is required. These three processes are generally confounded in stock assessment models, which means that it is difficult, if possible, to partition their influences on the estimates of population attributes including

survival. We recommend future research exploring their relative performance when coupled with management procedures and different parameterizations of time-variation in survival and *M*.

Care should be taken when interpreting the main findings found in this study. First, survival of the plus-group is not strictly survival. This matters because including the 2D AR(1) smoother on survival estimated increased survival of the plus-group, which resulted in higher short-term forecasts of SSB relative to models with independent survival deviations or 2D AR(1) deviations on *M*. Second, population dynamics in the 3-year projection time period were predicted by fixing $F = 0$, and non-zero catches will almost certainly occur. Last, all the conclusions made in this study are species-specific. For example, the relative importance of the recruitment covariate, deviations in survival versus *M*, and 2D AR(1) smoother to SSB prediction is highly dependent upon the parameters (e.g., age at maturity, longevity, selectivity and weight-at-age) that influence the age structure and life history of the stock. Generally speaking, the 2D AR(1) smoother can directly impact the predicted numbers at all ages, and thereby, is expected to be more important to near-term SSB prediction for late-matured and long-lived fish stocks. On the other hand, a recruitment covariate can only directly impact near-term predictions of recruitment. For late-matured and long-lived fish stocks, the covariate-induced change in recruitment prediction is not able to propagate to most of the mature age classes, and therefore will not appreciably impact SSB predictions in the near-term.

One final note is that the run time required to fit the models varied substantially and in unintuitive patterns. Much of TMB's advantage over ADMB in computational speed depends on its algorithm for automatically detecting sparseness of the Hessian matrix (Kristensen et al. 2016), and we found that this sparseness detection was the most important determinant of model run time. Directly specifying the survival deviations, $\varepsilon_{a,y}$, as random effect parameters did not

result in a sparse Hessian, but parameterizing the log numbers at age, $\log(N_{a,y})$, and then calculating the $\varepsilon_{a,y}$ as derived quantities, did. When only numbers at age 1 (i.e. recruits) were random effects, the Hessian was not detected as sparse. Thus, some of the least complex models we considered (e.g. Base, NAA-1, and M-1) took longer to run than the most complex model, NAA-M-CPI-2 (Table 5). With or without a sparse Hessian, adding 2D AR(1) autocorrelation increased run time by about 3x. Increasing model complexity had little impact on run time as long as the Hessian was detected as sparse, and the most complex model took 1.35 minutes to run on a laptop computer (Table 5). Therefore, while these results are limited to the dimension of the SNEMA yellowtail flounder application (e.g. the number of age classes and time steps), they suggest that computation speed is unlikely to be a hurdle to incorporating additional complexity into stock assessments via random effects in TMB.

Acknowledgements

This research was performed while BCS held an NRC Research Associateship award at the Northeast Fisheries Science Center. We thank Chris Melrose for updating the CPI with data through 2018, using code originally written by Jon Hare.

References

Aeberhard, W. H., Mills Flemming, J., and Nielsen, A. 2018. Review of State-Space Models for Fisheries Science. Annual Review of Statistics and Its Application 5: 215–235.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In Proceedings of the Second International Symposium on Information Theory, ed. B. N. Petrov and F. Csaki, 267–281. Budapest: Akademiai Kiado. Reprinted in Breakthroughs in Statistics, ed. S. Kotz, 610–624. New York: Springer (1992).

Aldrin, M., Tvete, I. F., Aanes, S., and Subbey, S. 2020. The specification of the data model part in the SAM model matters. Fisheries Research 229: 105585.

Atchison, J., and Shen, S.M. 1980. Logistic-normal distributions: Some properties and uses. Biometrika. 67: 261-272.

Berg, C. W., and Nielsen, A. 2016. Accounting for correlated observations in an age-based state-space stock assessment model. ICES Journal of Marine Science 73: 1788–1797.

Brooks, E. N., and Legault, C. M. 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Canadian Journal of Fisheries and Aquatic Sciences 73: 935–950.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach: Springer Science & Business Media.

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.

Cadrin, S.X. 2003. Stock structure of yellowtail flounder off the northeastern United States. Dissertations and Master's Theses (Campus Access). Paper AAI3103697.

Deroba, J.J., and Schueller, A.M. 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. Fisheries Research. 146:27-40.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software. 27: 233-249.

Goethel, D.R., Legault, C.M., and Cadrin, S.X. 2015. Demonstration of a spatially explicit, tag-integrated stock assessment model with application to three interconnected stocks of yellowtail flounder off of New England. ICES Journal of Marine Science 72: 164-177.

Gudmundsson, G., and Gunnlaugsson, T. 2012. Selection and estimation of sequential catch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences 69: 1760-1772.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., and Muradian, M.L. 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science 72: 99-110.

Johnson, K.F., Councill, E., Thorson, J.T., Brooks, E., Methot, R.D., and Punt, A.E. 2016. Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? Fisheries Research 183:222-232.

Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-pre, K.A., Anderson, S.C., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R.R., Muradian, M.L., and Ono, K. 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES Journal of Marine Science 72: 137-150.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., and Bell, B. M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software 70: 1-21.

Legault, C.M. 2009. Report of the retrospective working group. NOAA NMFS Northeast Fisheries Science Center Reference Document:09-01.

Legault, C.M., Alade, L., Stone, H.H., and Gross, W.E. 2012. Stock assessment of Georges Bank yellowtail flounder for 2012. TRAC Ref Doc. 2:133.

Legault, C.M., and Palmer, M.C. 2015. In what direction should the fishing mortality target change when natural mortality increases within an assessment? Canadian Journal of Fisheries and Aquatic Sciences 73: 1-9.

Legault, C.M., and Restrepo, V.R. 1999. A flexible forward age-structured assessment program. ICCAT Col Vol Sci Pap. 49: 246-253.

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.

Methot Jr, R. D., and Wetzel, C. R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142: 86-99.

Miller, T.J., Hare, J.A., and 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.

Miller, T.J. and Legault, C.M. 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fisheries Research 186: 109-120.

Miller, T.J. and Stock, B.C. 2020. The Woods Hole Assessment Model (WHAM). Version 1.0. https://timjmiller.github.io/wham/.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science 56: 473-488.

NEFSC. 2012. 54th Northeast Regional Stock Assessment Workshop (54th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc 12-18, 600p. https://repository.library.noaa.gov/view/noaa/4193

NEFSC. 2013. 55th Northeast Regional Stock Assessment Workshop (55th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc 13-01. 41p. https://repository.library.noaa.gov/view/noaa/4330.

NEFSC. 2020. Operational Assessment of 14 Northeast Groundfish Stocks, Updated Through 2018. https://s3.amazonaws.com/nefmc.org/9_Prepublication-NE-Grndfsh-10-3-2019_191202_105733.pdf. *pre-print, will update citation when officially published.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research 158: 96-101.

Punt, A. E., Dunn, A., Elvarsson, B. Þ., Hampton, J., Hoyle, S. D., Maunder, M. N., Methot, R. D., et al. 2020. Essential features of the next-generation integrated fisheries stock assessment package: A perspective. Fisheries Research 229: 105617.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Stock, B.C. and Miller, T.J. this issue. The Woods Hole Assessment Model (WHAM): a general state-space age-structured stock assessment framework designed to incorporate environmental effects. *final ref TBD.

Stone, H.H., Gavaris, S., Legault, C.M., Neilson, J.D., and Cadrin, S.X. 2004. Collapse and recovery of the yellowtail flounder (Limanda ferruginea) fishery on Georges Bank. Journal of Sea Research. 51: 261-270.

Thorson, J.T., Jensen, O.P., and Zipkin, E.F. 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. Canadian Journal of Fisheries and Aquatic Sciences 71: 973-983.

Thorson, J.T., Monnahan, C.C., and Cope, J.M. 2015. The potential impact of time-variation in vital rates on fisheries management targets for marine fishes. Fisheries Research 169: 8-17.

Xu, H., Miller, T. J., Hameed, S., Alade, L. A., and 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.

Xu, H., Thorson, J. T., Methot, R. D., and Taylor, I. G. 2019. A new semi-parametric method for autocorrelated age-and time-varying selectivity in age-structured assessment models. Canadian Journal of Fisheries and Aquatic Sciences 76: 268-285.

Table 1. Model descriptions and results where only numbers-at-age (NAA) were estimated as random effects. "Base" is most similar to a statistical catch-at-age model, with independent recruitment deviations and deterministic survival. NAA-2 is the state-space model with independent survival deviations, as in Miller et al. (2016). NAA-5, where survival is a 2D AR(1) process across years and ages, had the lowest AIC and Mohn's ρ. Mohn's ρ abbreviations: R = recruitment, SSB = spawning stock biomass, and F = fishing mortality averaged over ages 4-5. Maximum likelihood estimates of parameters constraining random effects are listed with standard error in parentheses.

			Estimated parameters				Model fit			Mohn's ρ		
Model	Ages treated as random effects	Correlation structure	σ_R	σ_a	ρ_{year}	ρ_{age}	-log($\mathcal{L}$)	AIC	ΔAIC	ρ_R	ρ_{SSB}	ρ_F
Base	Age-1	Indep.	1.18 (0.13)	—	—	—	-934.024	-1712.0	221.6	5.86	0.99	-0.42
NAA-1	Age-1	AR(1) year	0.66 (0.08)	—	0.92 (0.05)	—	-957.164	-1756.3	177.3	4.56	0.90	-0.38
NAA-2	All ages	Indep.	1.01 (0.12)	0.58 (0.05)	—	—	-1027.684	-1897.4	36.2	0.97	0.27	-0.17
NAA-3	All ages	AR(1) age	0.86 (0.12)	0.53 (0.05)	—	0.46 (0.10)	-1036.859	-1913.7	19.9	0.55	0.12	-0.06
NAA-4	All ages	AR(1) year	0.76 (0.11)	0.48 (0.05)	0.60 (0.09)	—	-1044.252	-1928.5	5.1	0.73	0.14	-0.08
NAA-5	All ages	2D AR(1)	0.72 (0.10)	0.47 (0.05)	0.52 (0.10)	0.34 (0.12)	-1047.803	-1933.6	0.0	0.56	0.09	-0.03

Table 2. Model descriptions and results where only recruitment and natural mortality (*M*) were estimated as random effects. All models treated the numbers-at-age as in Base: recruitment independent by year and deterministic survival. M-4 estimated 2D AR(1) random effects on *M* as in Cadigan (2016) and had the lowest AIC and Mohn's ρ. Mohn's ρ abbreviations: *R* = recruitment, *SSB* = spawning stock biomass, and *F* = fishing mortality averaged over ages 4-5. Maximum likelihood estimates of parameters constraining random effects are listed with standard error in parentheses.

		Estimated parameters			Model fit			Mohn's ρ		
Model	Correlation structure	σ_M	φ_{year}	φ_{age}	-log($\mathcal{L}$)	AIC	ΔAIC	ρ_R	ρ_{SSB}	ρ_F
Base	—	—	—	—	-934.024	-1712.0	211.9	5.86	0.99	-0.42
M-1	Indep.	1.18 (0.09)	—	—	-1037.981	-1918.0	5.9	0.12	0.20	-0.13
M-2	AR(1) age	1.15 (0.43)	—	0.25 (0.49)	-979.398	-1798.8	125.1	2.06	0.10	-0.08
M-3	AR(1) year	0.13 (0.06)	0.98 (0.02)	—	-988.290	-1816.6	107.3	1.19	-0.12	0.30
M-4	2D AR(1)	0.89 (0.12)	0.47 (0.21)	0.39 (0.14)	-1042.965	-1923.9	0.0	-0.07	0.10	-0.06

564 Table 3. Model results where both numbers-at-age (NAA) and natural mortality (*M*) were estimated as random effects. NAA-M-3 had
565 the lowest AIC but not Mohn's ρ. NAA-M-2 had higher AIC but the lowest Mohn's ρ. Mohn's ρ abbreviations: *R* = recruitment, *SSB*
566 = spawning stock biomass, and *F* = fishing mortality averaged over ages 4-5. NAA-M-4 did not converge.

	Estimated parameters		Model fit			Mohn's ρ		
Model	NAA	M	-log($\mathcal{L}$)	AIC	ΔAIC	ρ_R	ρ_{SSB}	ρ_F
NAA-M-1	σ_R, σ_a	σ_M	-1051.090	-1942.2	17.3	0.20	0.17	-0.09
NAA-M-2	σ_R, σ_a	$\sigma_M, \varphi_{year}, \varphi_{age}$	-1048.691	-1933.4	26.1	-0.22	0.02	0.04
NAA-M-3	$\sigma_R, \sigma_a, \rho_{year}, \rho_{age}$	σ_M	-1061.727	-1959.5	0.0	0.52	0.10	-0.04
NAA-M-4	$\sigma_R, \sigma_a, \rho_{year}, \rho_{age}$	$\sigma_M, \varphi_{year}, \varphi_{age}$						

567

Table 4. Results from models that estimated an effect of the Cold Pool Index (CPI) on recruitment, in addition to random effects on numbers-at-age (NAA) and natural mortality (*M*). NAA-M-CPI-1 had the lowest AIC but not Mohn's ρ. NAA-M-CPI-2 had higher AIC but the lowest Mohn's ρ. Mohn's ρ abbreviations: *R* = recruitment, *SSB* = spawning stock biomass, and *F* = fishing mortality averaged over ages 4-5. NAA-M-CPI-3 and NAA-M-CPI-4 did not converge.

	Estimated parameters		Model fit			Mohn's ρ		
Model	NAA	M	-log($\mathcal{L}$)*	AIC	ΔAIC	ρ_R	ρ_{SSB}	ρ_F
NAA-M-CPI-1	σ_R, σ_a	σ_M	-996.345	-1824.7	0.0	0.16	0.17	-0.09
NAA-M-CPI-2	σ_R, σ_a	σ_M, φ_{year}, φ_{age}	-994.979	-1818.0	6.7	-0.11	0.02	0.06
NAA-M-CPI-3	σ_R, σ_a, ρ_{year}, ρ_{age}	σ_M						
NAA-M-CPI-4	σ_R, σ_a, ρ_{year}, ρ_{age}	σ_M, φ_{year}, φ_{age}						

*Note that the likelihoods and AIC in Table 4 are not comparable to those in Tables 1-3 because additional data (the CPI) have been included.

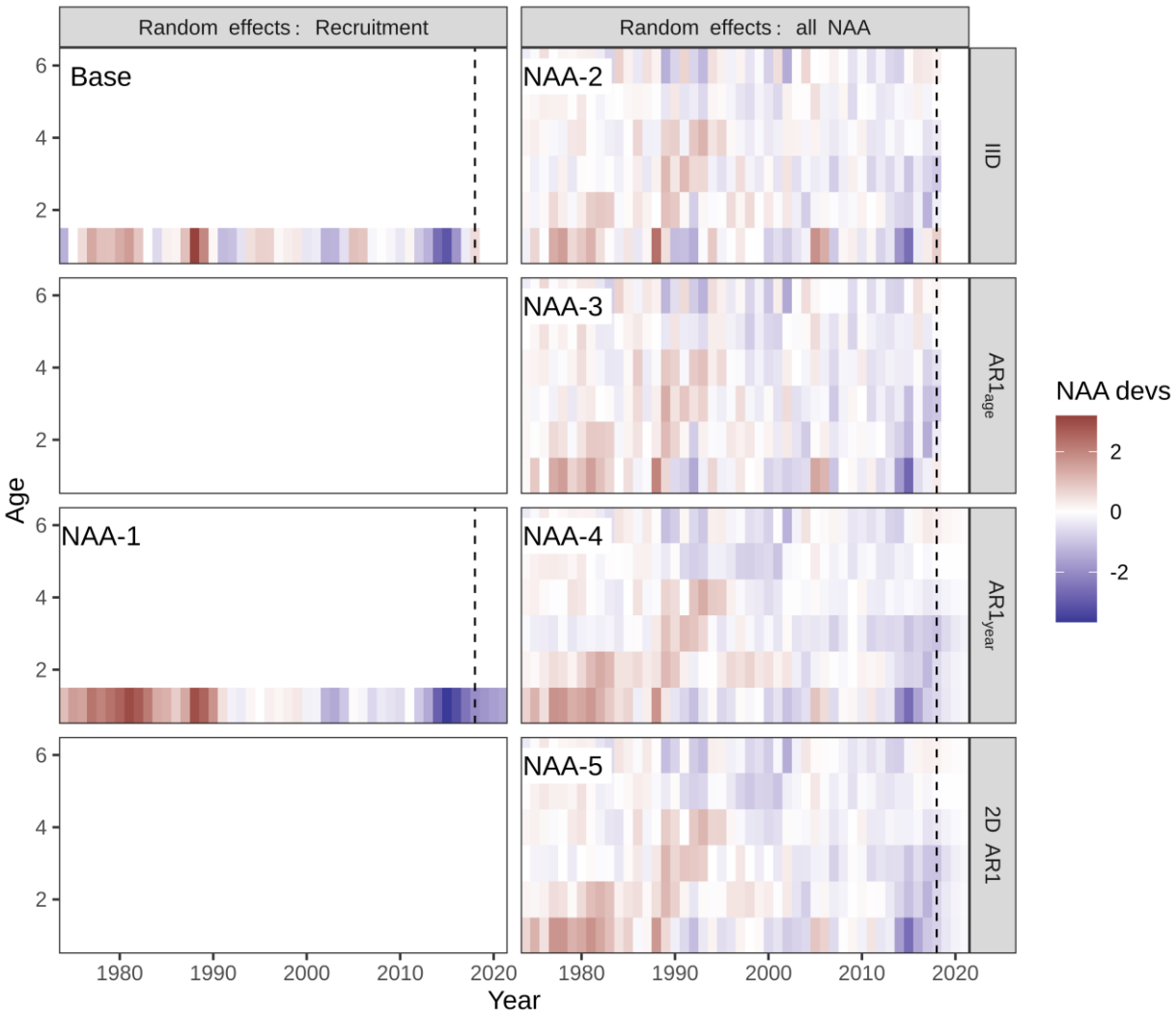
Table 5. Short-term projections of spawning stock biomass (SSB) for the models from each class with lowest Mohn's ρ. Base = recruitment independent by year, deterministic survival, and no random effects on *M*. *F* was set to 0 in projection years. Although only NAA-M-CPI-2 included an effect of the CPI on recruitment, all models fit an AR(1) process to CPI observations in order to allow direct comparison using AIC.

	Description				Model performance				SSB projections		
Model	NAA random effects		M random effects	CPI	ΔAIC	ρ_R	ρ_{SSB}	ρ_F	2019	2020	2021
Base	Age-1	Indep.	—	—	247.4	5.86	0.99	-0.42	273 (173, 430)	362 (197, 665)	462 (202, 1055)
NAA-2	All ages	Indep.	—	—	62.1	0.97	0.27	-0.17	356 (159, 800)	488 (172, 1386)	650 (191, 2209)
NAA-5	All ages	2D AR(1)	—	—	25.9	0.56	0.09	-0.03	260 (106, 638)	466 (102, 2135)	867 (116, 6493)
M-4	Age-1	Indep.	2D AR(1)	—	35.5	-0.07	0.09	-0.06	278 (102, 754)	414 (95, 1794)	659 (112, 3876)
NAA-M-2	All ages	Indep.	2D AR(1)	—	26.1	-0.21	0.04	0.03	213 (75, 605)	260 (50, 1343)	368 (43, 3131)
NAA-M-3	All ages	2D AR(1)	Indep.	—	0.0	0.52	0.10	-0.04	280 (116, 678)	536 (117, 2459)	1023 (138, 7561)
NAA-M-CPI-2	All ages	Indep.	2D AR(1)	Yes	5.9	-0.10	0.04	0.04	207 (71, 603)	310 (55, 1750)	526 (57, 4873)

580 Table 6. Model run times as a function of model size and whether TMB detected sparseness of
581 the Hessian matrix. Model size = number of random effects (dimension of the Hessian matrix
582 with respect to random effects).

Model	NAA random effects		M random effects	Model size	Sparse Hessian	Run time (min)
Base	Age-1	Indep.	—	45	No	1.51
NAA-2	All ages	Indep.	—	270	Yes	0.37
NAA-5	All ages	2D AR(1)	—	270	Yes	1.12
M-1	Age-1	Indep.	Indep.	321	No	2.67
M-4	Age-1	Indep.	2D AR(1)	321	No	10.30
NAA-M-2	All ages	Indep.	2D AR(1)	546	Yes	1.46
NAA-M-3	All ages	2D AR(1)	Indep.	546	Yes	1.99
NAA-M-CPI-2	All ages	Indep.	2D AR(1)	593	Yes	1.35

583



584

585 Figure 1. Survival deviations by year and age estimated by models in which only numbers-at-age
586 (NAA) were random effects. Base = the statistical catch-at-age model with recruitment as
587 random effects without autocorrelation. NAA-1 = autocorrelated recruitment. The models in the
588 right column treat all NAA as random effects with different correlation structures. IID =
589 independent and identically distributed, i.e. no correlation. The vertical dashed line denotes the
590 terminal year in the assessment, 2018.

591

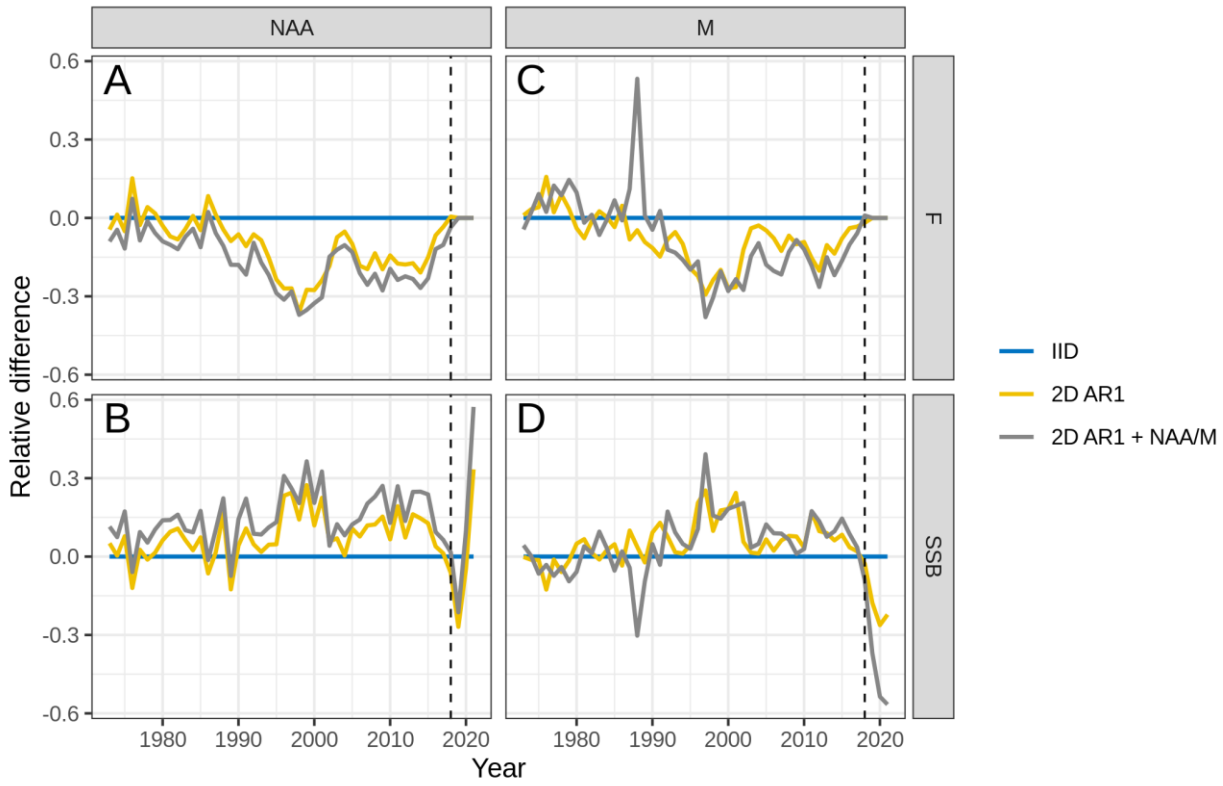


Figure 2. Relative difference in estimates of *F* (top panels) and SSB (lower panels) from constraining deviations in survival (left) and *M* (right) to follow a 2D AR(1) correlation structure over ages and years. Relative difference was calculated as $\theta_{2D\ AR(1)}/\theta_{IID} - 1$, where θ is either *F* or SSB. IID = independent deviations by age and year (for process listed in column heading), 2D AR1 = 2D AR(1) deviations, and 2D AR1 + NAA/M = 2D AR(1) deviations as well as IID deviations in the off-column heading. The vertical dashed line marks the terminal year in the assessment, 2018. *F* was fixed at 0 in projection years.

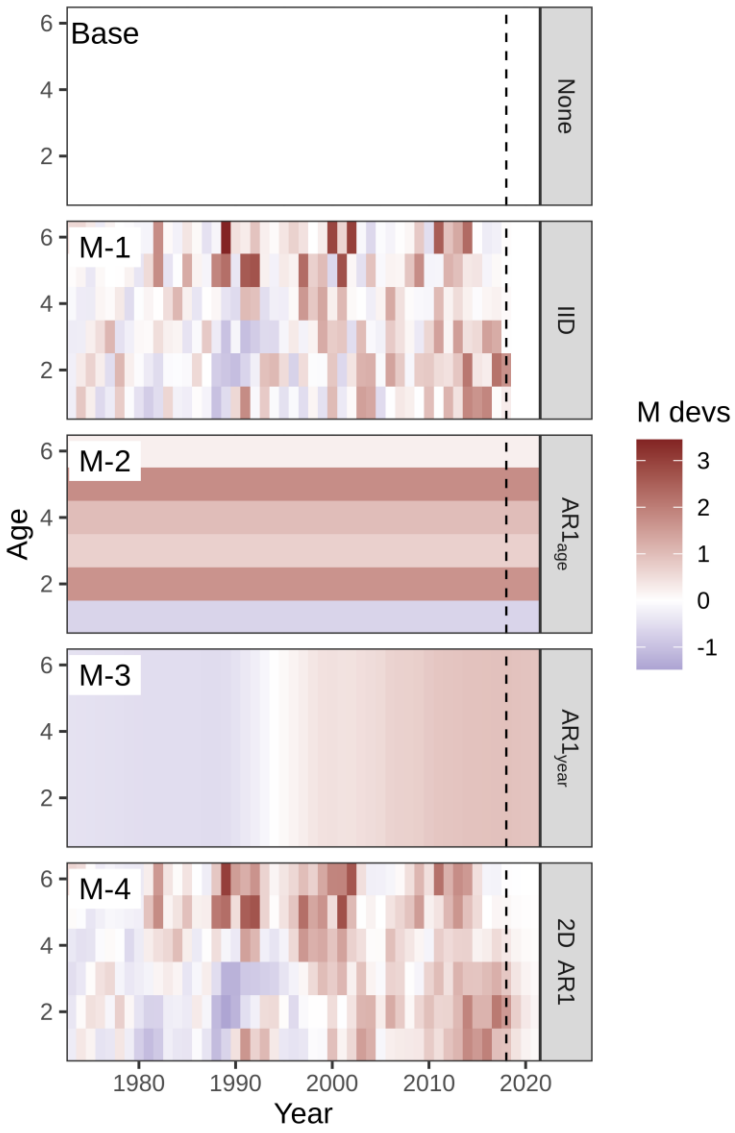


Figure 3. Deviations in log natural mortality (M) by year and age estimated by models without numbers-at-age (NAA) random effects. Base = no deviations from the M_a values specified in the assessment. The vertical dashed line marks the terminal year in the assessment, 2018.

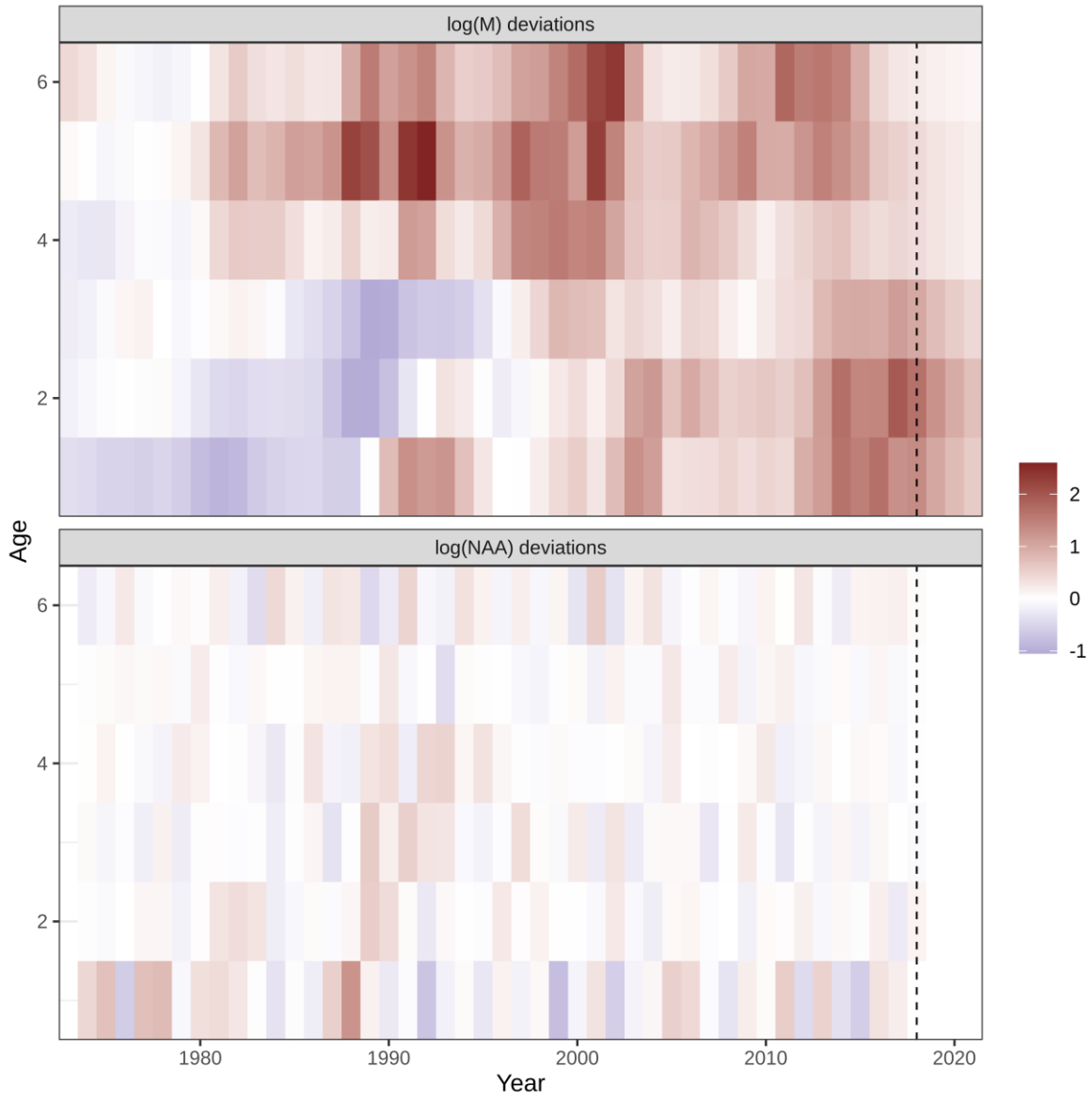


Figure 4. Deviations in natural mortality (log *M*, top panel) and numbers-at-age (log NAA, bottom panel) from the final model with lowest Mohn's ρ, NAA-M-CPI-2. The vertical dashed line marks the terminal year in the assessment, 2018. In the projection years, NAA deviations are zero and *M* deviations are non-zero because NAA-M-CPI-2 includes the 2D AR1 structure on *M* and not NAA.

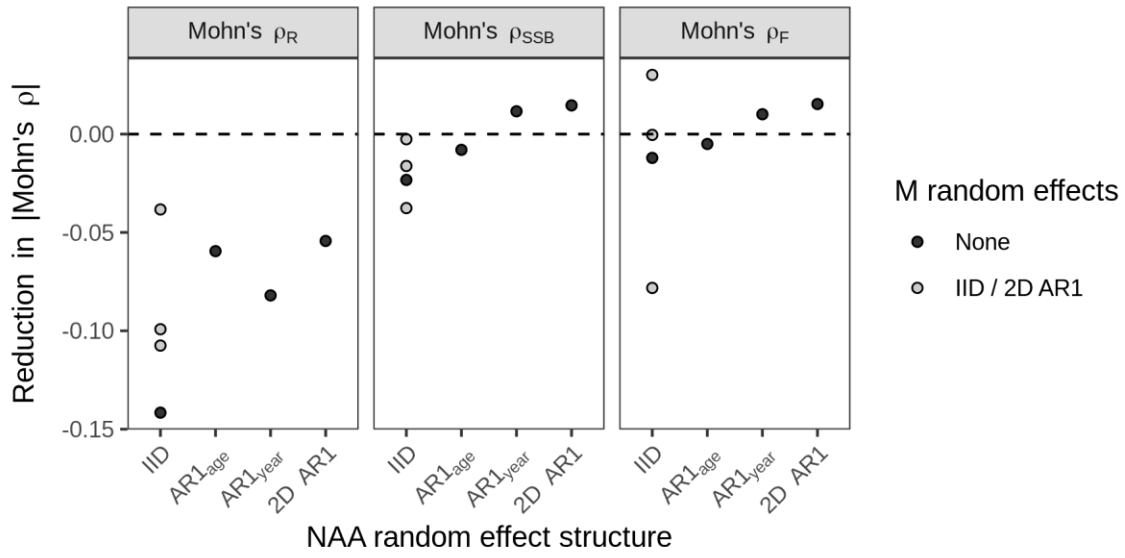


Figure 5. Reduction in retrospective patterns by including an effect of the Cold Pool Index (CPI) on recruitment, measured as the difference in Mohn's ρ between otherwise equivalent models. Across the various NAA and *M* models, including the CPI-recruitment link minimally impacted Mohn's ρ_{SSB} or ρ_F, and slightly lowered Mohn's ρ_R by about 0.1 on average.

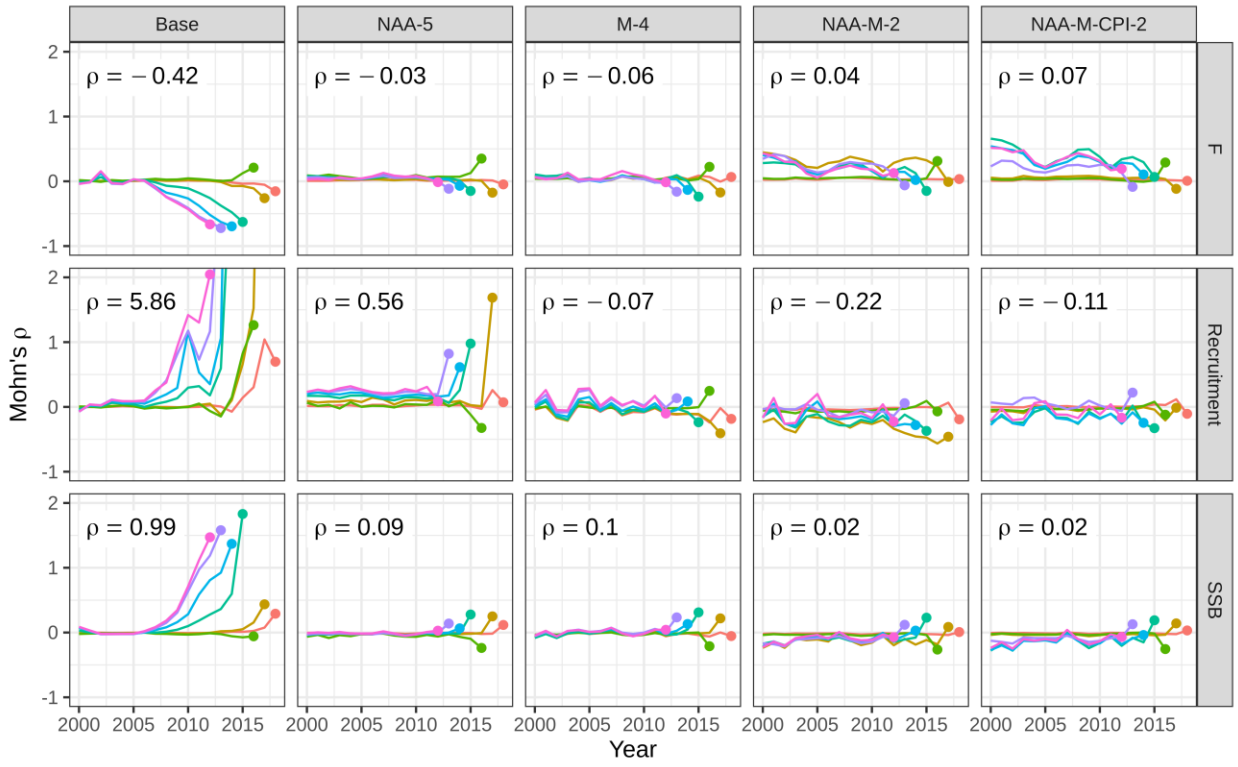
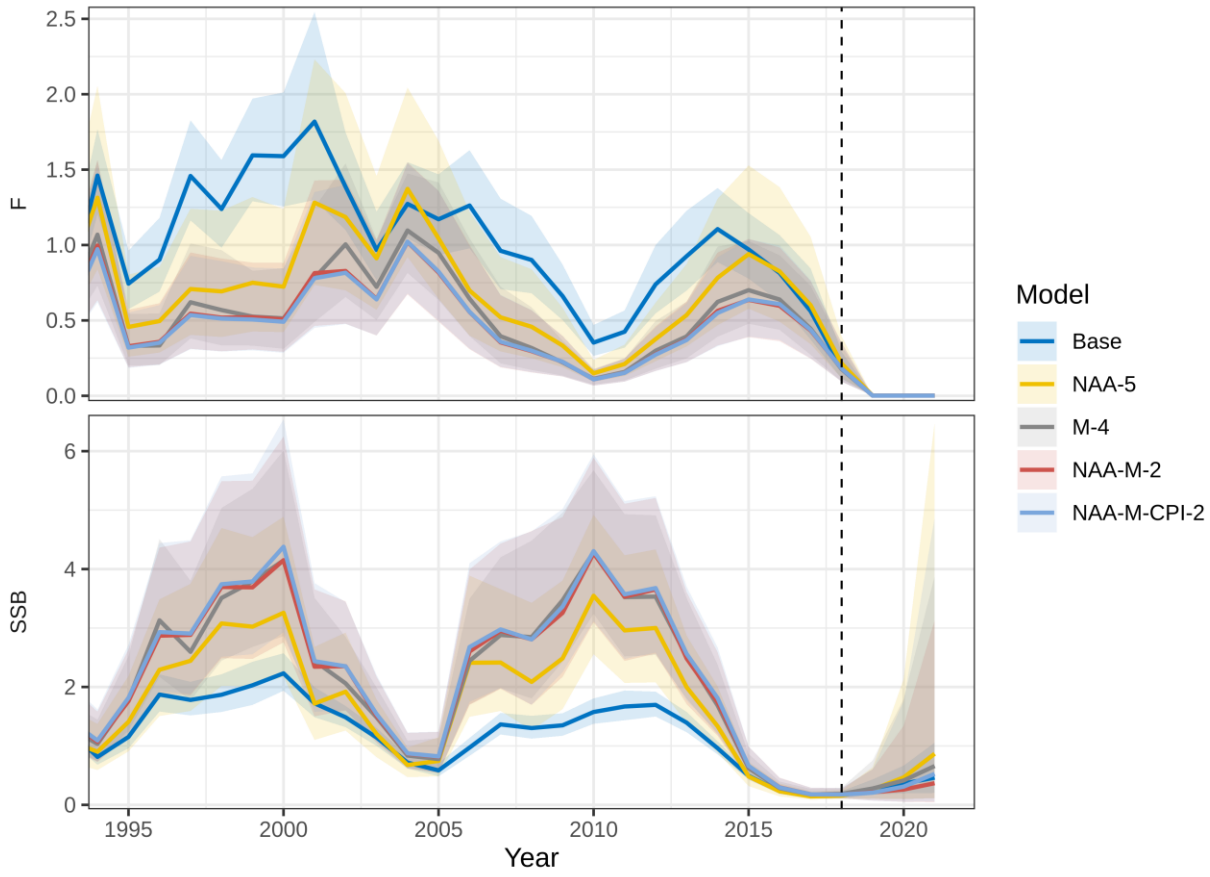


Figure 6. Retrospective patterns in fishing mortality (*F*, top row), recruitment (middle row), and spawning stock biomass (SSB, bottom row). Lines and points depict Mohn's ρ from seven peels, and the average Mohn's ρ is given in each panel. Columns show results by model. Base = the statistical catch-at-age model. Within each model class, the model with lowest Mohn's ρ is shown (NAA-5, M-4, NAA-M-2, and NAA-M-CPI-2; Tables 1-4).



627

628 Figure 7. Trends in fishing mortality (*F*) and spawning stock biomass (SSB). Base = the
629 statistical catch-at-age model. Within each model group, the model with lowest Mohn's ρ is
630 shown (NAA-5, M-4, NAA-M-2, and NAA-M-CPI-2; Tables 1-4). The dashed line denotes the
631 terminal year in the assessment, 2018. *F* is fixed at 0 for all models in projection years, 2019-
632 2021. The first year in the assessments is 1973, beyond the left x-axis limit.