Implementing two-dimensional autocorrelation in 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 (0.53 ± 0.09, 0.63 ± 0.16) and ages (0.33 ± 0.12, 0.40 ± 0.16) when estimated for survival or *M*, respectively. Models with 2D autocorrelation on survival or *M* fit the data better and had reduced retrospective pattern than models without autocorrelation. The best fit model included 2D autocorrelated deviations in survival as well as independent deviations in *M*, and altered estimates of spawning stock biomass by 18% and *F* by 21% in model years. In short-term projections with $F = 0$, the model estimated 48% lower spawning stock biomass. We conclude that incorporating 2D autocorrelated variation in survival or *M* could improve the assessment of Southern New England yellowtail flounder in terms of model fit and consistency of biomass 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). Using a two-dimensional (2D) autocorrelation structure across both ages and years is relatively rare but has been used to model deviations in fishing mortality (Nielsen and Berg 2014; Kumar et al. 2020), natural mortality (Cadigan 2016), selectivity (Xu et al. 2019), and catch and survey index observations (Berg and Nielsen 2016).

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 stock, Southern New England-Mid Atlantic (SNEMA, 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 (NEFSC 2020). There are two major sources of uncertainty in recent SNEMA yellowtail

flounder assessments (NEFSC 2012, 2020). First, the assessment model cannot explain the dramatic decrease in recruitment since the 1990s. Recent studies link 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 (Mohn 1999) 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).

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 the natural mortality rate (M) in stock assessment models. This is because retrospective patterns typically arise 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).

In the search for possible misspecifications underlying a retrospective pattern, M is an important parameter to consider because it directly influences stock productivity. 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 usually specified as a time-invariant constant because it is often difficult to estimate (Deroba and Schueller 2013; Johnson et al. 2015; Legault and Palmer 2015). Deroba and Schueller (2013) found that misspecifying 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 in the latest SNEMA yellowtail flounder assessment $M/(F+M)$ is currently near the historical maximum since F is near its historic low (Legault and Palmer 2015; NEFSC 2020). The current 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 first-order autoregressive, AR(1), smoother over two dimensions, 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 Northeast Fisheries Science Center (NEFSC, Miller and Stock 2020; Stock and Miller, this issue). Using correlated process errors in state-space stock assessment models to reduce retrospective patterns is an emerging concept (ICES 2020). 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: $N_{a,y} = N_{a-1,y-1}e^{-(F_{a-1,y-1}+M_{a-1,y-1})}$. 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 $N_{a,y} = N_{a-1,y-1}e^{-(F_{a-1,y-1}+M_{a-1,y-1})+\varepsilon_{a,y}}$.

In this study, we extend the model presented in Miller et al. (2016) to include 2D AR(1) deviations in survival and M. We then apply it to the assessment of SNEMA yellowtail flounder. In particular, we assess whether it is better to place the 2D AR(1) smoother on survival versus M, attempt to estimate a model with 2D AR(1) smoothers on both survival and M, and measure the impact of including the 2D AR(1) smoother on estimates of SSB and *F*.

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 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 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 ($F + M$), A represents the plus-group, and g is the stock-recruit function in which an environmental time series (x) can be incorporated as a covariate. The $\varepsilon_{a,y}$ terms 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.

Strictly speaking, the survival deviation terms 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 $\varepsilon_{a,y} > Z_{a-1,y-1}$). 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 and directly modelled deviations in log(M). However, the survival deviations can also be caused by deviations in F 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:

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

where $\mathbf{E} = (\varepsilon_{1,1}, \dots, \varepsilon_{1,Y-1}, \varepsilon_{2,1}, \dots, \varepsilon_{2,Y-1}, \dots \dots, \varepsilon_{A,1}, \dots, \varepsilon_{A,Y-1})'$, $\boldsymbol{\Sigma_{total}}$ 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 ($\widetilde{\boldsymbol{\Sigma}}$):

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

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

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

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, $\hat{C}_{a,y} = N_{a,y}\frac{F_{a,y}}{Z_{a,y}}\left(1-e^{-Z_{a,y}}\right)$. Whether to estimate deviations in $\log(M_{a,y})$ or $\log(N_{a,y})$ also affects the calculation of reference points.

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

181 "Base" model was similar to a statistical catch-at-age model, e.g., the Age-Structured
182 Assessment Program (ASAP, Legault and Restrepo 1998; Miller and Legault 2015) or Stock
183 Synthesis (SS, Methot and Wetzel 2013), where recruitment deviations are typically estimated in
184 each year as independent fixed effects, $\varepsilon_{1,y} \sim \mathrm{N}\left(-\frac{\sigma_R^2}{2}, \sigma_R^2\right)$, and survival is deterministic.
185 However, while ASAP and SS do not estimate the recruitment variance, σ_R^2, in WHAM the $\varepsilon_{1,y}$
186 can either be treated as random effects with σ_R^2 estimated internally, or as fixed effect parameters
187 (Miller and Stock 2020; Stock and Miller, this issue). Here, we chose to model recruitment
188 deviations as random effects in the Base model, i.e., we estimate σ_R^2. The next model, NAA-1,
189 added recruitment autocorrelation, estimating one additional parameter from Eqn. 5, ρ_{year}.
190 NAA-2 through NAA-5 estimated "full state-space" models, with numbers at all ages treated as
191 random effects, but with different autocorrelation structures. NAA-2 estimated independent $\varepsilon_{a,y}$
192 as in Miller et al. (2016) and Nielsen and Berg (2014), NAA-3 and NAA-4 added autocorrelation
193 across ages and years, and NAA-5 estimated all parameters in the described 2D AR(1) smoother
194 (Eq. 5, Table 1). To isolate the effect of incorporating the 2D AR(1) smoother on survival, we
195 compared the model fit, retrospective pattern, and relative difference in SSB and *F* estimates
196 from NAA-5 versus NAA-2.
197 Next, we fit a series of models treating the numbers-at-age as in Base, but including
198 deviations in *M* as in Eq. 6 with the same set of autocorrelation structures: none, independent,
199 AR(1) by age, AR(1) by year, and 2D AR(1) (Table 2). As for the set of NAA models, we
200 isolated the effect of the 2D AR(1) smoother on *M* by comparing M-1 to M-4.
201 The last set of models tested the ability of WHAM to simultaneously estimate numbers-
202 at-age and *M* deviations as random effects, using only the independent and 2D AR(1)
203 autocorrelation structures for each (Table 3).

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. To conduct three-year projections of SSB, we fixed weight-at-age and maturity-at-age at the average values from the last five years of data, as is standard practice at the NEFSC (NEFSC 2020). In order to facilitate comparisons of short-term SSB projections between models, we fixed F at 0 in the projection years. We forecast variables treated as random effects, such as numbers-at-age, M, and the CPI, in the projection years by simply continuing the autoregressive processes. One final difference between this analysis and the current assessment was that we estimated σ_R^2, the variance of recruitment deviations, as a fixed effect parameter in all models. We did not estimate a stock-

recruitment relationship in any of the models. 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 Template Model Builder (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). Finally, we conducted simulation self- and cross-tests to estimate bias in parameters and derived quantities (Supplemental material). We tested sets of models without, with independent, and with 2D AR(1) random effect deviations in survival (Base, NAA-1, NAA-2, and NAA-5) and M (Base, M-1, and M-4).

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; compare Base and NAA-2 in 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 and years relative to the

models with independent deviations (Fig. 1). NAA-5, with the 2D AR(1) structure, had the best fit and reduced AIC by 44.9, $|\rho_R|$ by 0.30, $|\rho_{SSB}|$ by 0.11, and $|\rho_F|$ by 0.10 compared to NAA-2 with independent survival deviations (Table 1). Constraining the survival deviations with the 2D AR(1) structure reduced estimates of *F* by 9% and increased estimates of SSB by 6% in model years (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 53% when *F* was fixed at 0 (Fig. 2a-b).

In models that included autocorrelation by year, the survival deviations estimated in years near the end of the assessment impacted the projections of SSB. 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 with autocorrelation by year (NAA-1, NAA-4, and NAA-5 in Fig. 1). In the terminal year, NAA-4 and NAA-5 estimated negative survival deviations for ages 2-3 and near-zero deviations for older ages. The effect of the negative projected survival deviations resulted in the model with 2D AR(1) autocorrelation projecting lower SSB than the model with independent deviations, and this effect was more pronounced in 2020-2021 than in 2019 (Fig. 2b).

3.2. Deviations in *M* as random effects

Including deviations in *M*, instead of survival, also substantially improved model fit and the 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 11.1, $|\rho_R|$ by 0.02, $|\rho_{SSB}|$ by 0.11, and $|\rho_F|$ by 0.11 (Table 2). The estimated 2D

AR(1) *M* deviations had higher variance and higher autocorrelation than the 2D AR(1) survival deviations, and therefore appeared stronger and more smoothed (Figs. 1 and 3, $\sigma_M > \sigma_R > \sigma_a$, $\varphi_{year} > \rho_{year}$, and $\varphi_{age} > \rho_{age}$ in Tables 1-2). The effect of adding the 2D AR(1) structure on estimates of *F* and SSB was similar as for the NAA models: 13% lower *F* and 9% higher SSB during assessment years, similar terminal year status, and then 48% lower SSB in short-term projections (Fig. 2c-d). *M* deviations for younger ages in the model with 2D AR(1) autocorrelation were positive in the terminal year (Fig. 3). This was consistent with the NAA 2D AR(1) model estimating negative survival in this period (Fig. 1) and explained why adding 2D AR(1) autocorrelation on *M* deviations also led to lower SSB in short-term projections (Fig. 2).

3.3. Estimating deviations in both survival and *M*

The model that attempted to estimate deviations in both survival and *M* with 2D AR(1) autocorrelation failed to converge (Table 3). However, adding independent *M* deviations to 2D AR(1) on survival, NAA-M-3, and adding independent survival deviations to 2D AR(1) on *M*, NAA-M-2, substantially improved model fit (lower AIC by 29.9 and 13.5, respectively; Table 4). Both of these models had negligible Mohn's ρ_{SSB} and ρ_F (less than 0.05, Table 3 and Fig. 5). The two models estimated coherent survival and M deviations—years and ages with negative M deviations in NAA-M-2 had positive survival deviations in NAA-M-3 and vice-versa (e.g., ages 1-3 during the late 1970s-1980s in Fig. 4). All models with random effects on survival or M estimated substantially lower F and higher SSB than the Base model over the last 20 years of the assessment and including the 2D AR(1) structure further increased this difference (Figs. 2 and 6). While the state-space model with independent deviations projected SSB to increase at a similar rate to the Base model (NAA-2 in Table 4 and Fig. 6), models with the 2D AR(1) structure on survival or *M* predicted that SSB would increase at a reduced rate (Table 4, Figs. 2 and 6).

3.4. Simulation tests

All models had little-no bias in SSB, F, $B/B_{40\%}$, $F/F_{40\%}$, predicted catch, or recruitment in self-tests or in cross-tests when the operating model did not include survival or M deviations (Figs. S1-S9). Models without survival or M random effect deviations exhibited bias in all quantities when fit to data simulated with these random effects. The biases in SSB and F were always in opposite directions, as expected, and around 10-20%. All models estimated the recruitment variance, σ_R^2, and yearly autocorrelation parameters, ρ_{year} and φ_{year}, without bias (Fig. S3). The variance of age 2+ survival deviations, σ_a^2, the variance of M deviations, σ_M^2, and the autocorrelation by age, ρ_{age} and φ_{age}, were estimated with negative bias, which is the expected direction using maximum likelihood estimation instead of restricted maximum likelihood (REML). Last, the model with 2D AR(1) deviations on survival, NAA-5, had slightly lower bias in simulation self- and cross-tests than the model with 2D AR(1) deviations on M, M-4 (Figs. S1-S2).

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. These results imply that including survival and M deviations in the SNEMA yellowtail flounder assessment would provide more consistent estimates of stock and harvest status. Different from previous assessments in the region which have 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 evaluate performance against models with independent or 2D AR(1) deviations. Specific to SNEMA yellowtail flounder, we found that the 2D AR(1) smoother impacted SSB and *F* estimates by 6-13% in model years, and this increased to 14-21% when random effect deviations on both survival and M were included. Relative to models with independent or no deviations in survival or *M*, all models with the 2D AR(1) smoother estimated higher SSB in the last two decades but lower SSB in near-term projections (Fig. 6). Thus, the decision whether to implement the 2D AR(1) smoother in the assessment of SNEMA yellowtail flounder may be consequential.

Although placing the 2D AR(1) structure on survival or *M* deviations is clearly supported, we suggest putting it on survival for three reasons. First, the model with 2D AR(1) survival deviations and independent *M* deviations had lower AIC by a wide margin (35.7, Table 4). Second, models with the 2D AR(1) smoother on survival had greatly reduced uncertainty in SSB projections compared to all other models, especially in the second and third projection years (Table 4). Last, the model with 2D AR(1) deviations on survival performed slightly better in simulation self- and cross-tests than the model with the 2D AR(1) structure on *M* deviations (Figs. S1-S2). Nevertheless, all models with the 2D AR(1) structure on survival or *M* estimated consistent effects compared to the models without: lower *F* and higher SSB in assessment years, and lower SSB in projection years (Fig. 6).

Near-term SSB forecasts changed substantially, by around 50%, when the 2D AR(1) smoother was included to constrain deviations in survival or *M*. In models where the survival or *M* deviations are independent (e.g., Base, NAA-2, and M-1), they do not affect projections of

SSB unless they are linked to an environmental covariate that is also projected. In contrast, including autocorrelation by year (ρ_{year} or φ_{year}) propagates non-zero survival or M deviations into short-term projections, with the trend near the assessment terminal year becoming important. In the case of SNEMA yellowtail flounder, in recent years models with φ_{year} estimated positive *M* deviations and models with ρ_{year} estimated negative survival deviations. This clearly resulted in lower projected SSB. Note that although the projected deviations asymptotically approach zero over time (Figs. 1, 3, and 4), SSB in a given projection year is the result of *cumulative* survival deviations, which means that the influence of the survival smoother on SSB is not necessarily weaker over time (Fig. 2b,d).

We estimated 2D autocorrelated deviations in *M* as Cadigan (2016), although there were noteworthy differences between the studies. Cadigan (2016) developed a state-space, age-structured assessment model for Northern Cod (*Gadus morhua*) and estimated 2D AR(1) deviations in *M*. Cadigan (2016) did not, however, compare the model estimates, predictions, goodness-of-fit, or retrospective patterns under alternative 2D autocorrelation structures for *M*. Cadigan (2016) included extensive tagging data to inform *M* in his model estimation and when he fit the model without tagging data, he found that the process and measurement error variance parameters were highly confounded. Cadigan (2016) also dealt with other issues such as uncertain catches and time-varying survey catchability (*q*), and these may be reasons why his model with 2D AR(1) *M* deviations did not converge without tagging observations. Finally, Cadigan (2016) treated *F*, *q*, and selectivity as in Nielsen and Berg (2014), which is more flexible and less constrained than in our analysis using WHAM and may be another factor in the ability to estimate 2D AR(1) deviations in *M*. Other than these differences between our study and Cadigan (2016), we do not know when modeling *M* deviations as random effects will be more or

less likely to succeed in general. A study that fit models to simulated data arising from alternative life history, selectivity, catchability, and data availability could shed light on this question.

The 2D AR(1) structure could be 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 residual 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 could also be applied to other potentially autocorrelated biological processes, 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). 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).

One potential drawback to allowing survival or *M* to vary in time is that calculating biological reference points (BRPs) becomes more complex. 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, estimated survival deviations were predominantly negative and *M* deviations were mostly positive since 2000. 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 to compare the performance of deterministic versus time-varying reference points when coupled with management procedures and assessment models with time-variation in survival and *M*.

Care should be taken when interpreting the main findings found in this study. First, population dynamics in the 3-year projection time period were predicted by fixing $F = 0$, and non-zero catches will almost certainly occur. Second, all the conclusions made in this study are stock-specific. For example, the relative importance of the deviations in survival versus *M* and the 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. Including an environment-linked stock-recruitment relationship is another way to account for time-varying productivity and directly impacts near-term predictions of recruitment (e.g., Miller et al. 2016; Xu et al. 2018). However, for fish stocks with low selectivity at ages 1-3, changes in predicted recruitment will not propagate to the fished age classes in 3-year projections and therefore will not appreciably impact SSB predictions. A key difference is that the 2D AR(1) smoother impacts the predicted numbers at all ages, not just recruitment, and we expect it to be relatively more important to near-term SSB predictions for late-maturing and

long-lived fish stocks. As demonstrated here, deviations in survival or *M* of older ages near the end of an assessment can propagate through the entire age structure in near-term projections and substantially modify SSB predictions (Figs. 4a and 6). Finally, movement is another process that may affect the relative importance of deviations in survival versus *M* and clearly depends on the stock in question. For yellowtail flounder, population mixing between adjacent stocks has been observed in tagging studies but not too a large enough extent to significantly affect the population dynamics of individual stocks (Cadrin 2003; Goethel et al. 2015).

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 more complex models with deviations in survival and *M* (Table 4). Using the model with independent survival deviations as a baseline, adding the 2D AR(1) structure increased run time by 3x and additionally including independent M deviations increased run time by 5x (Table 4). While the run times are short, on the order of one minute, these results are limited to the dimension of the SNEMA yellowtail flounder assessment (e.g., the number of age classes and time steps) and it is possible that run time could be an issue for assessments with many more years and ages, especially if run time increases worse than linearly. In addition, standard practice when introducing new

assessment models is to evaluate them using simulation testing, which involves thousands of model fits. Still, the most complex model took only 1.28 minutes to run on a laptop computer, which suggests that computation speed is unlikely to be a hurdle to incorporating additional complexity into stock assessments via random effects in TMB.

Acknowledgements

We thank Larry Alade, Kelli Johnson, and one anonymous reviewer for helpful comments. This research was performed while BCS held an NRC Research Associateship award at the Northeast Fisheries Science Center. HX and JAN were funded by NOAA FATE Grant #NA12OAR4320071.

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.

Kumar, R., Cadigan, N. G., Zheng, N., Varkey, D. A., and Morgan, M. J. 2020. A state-space spatial survey-based stock assessment (SSURBA) model to inform spatial variation in relative stock trends. Canadian Journal of Fisheries and Aquatic Sciences, 77: 1638–1658.

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 assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. *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.

Zhang, F., Regular, P. M., Wheeland, L., Rideout, R. M., and Morgan, M. J. 2020. Accounting for non-stationary stock–recruitment relationships in the development of MSY-based reference points. ICES Journal of Marine Science. https://doi.org/10.1093/icesjms/fsaa176.

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 estimates all parameters in Eqn. 5, which constrains survival deviations according to a 2D autoregressive, AR(1), process across years and ages. NAA-5 had the lowest negative log likelihood (-log$\mathcal{L}$) and Akaike's information criterion (AIC). Mohn's ρ were averaged over seven retrospective peels for three quantities: recruitment (*R*), spawning stock biomass (*SSB*), and fishing mortality averaged over ages 4-5 (*F*). Maximum likelihood estimates of the parameters constraining random effects are listed with standard error in parentheses. Dashes indicate parameters which are not included in a given model.

			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.	1.67 (0.18)	—	—	—	-919.603	-1685.2	252.6	6.42	1.02	-0.43
NAA-1	Age-1	$AR(1)_{year}$	0.66 (0.08)	—	0.92 (0.05)	—	-957.164	-1758.3	179.5	4.56	0.90	-0.38
NAA-2	All ages	Indep.	1.22 (0.15)	0.61 (0.05)	—	—	-1024.455	-1892.9	44.9	0.81	0.17	-0.10
NAA-3	All ages	$AR(1)_{age}$	0.93 (0.13)	0.54 (0.05)	—	0.50 (0.09)	-1036.632	-1915.3	22.5	0.40	0.04	0.02
NAA-4	All ages	$AR(1)_{year}$	0.78 (0.11)	0.48 (0.05)	0.61 (0.08)	—	-1045.568	-1933.1	4.7	0.67	0.10	-0.04
NAA-5	All ages	2D AR(1)	0.73 (0.11)	0.47 (0.05)	0.53 (0.09)	0.33 (0.12)	-1048.883	-1937.8	0.0	0.51	0.06	0.00

564 Table 2. Model descriptions and results where only recruitment and natural mortality (*M*) deviations were estimated as random effects.
565 All models treated the numbers-at-age as in Base, with independent recruitment deviations and deterministic survival. M-4 estimated
566 all parameters in Eqn. 6 as in Cadigan (2016), which constrains log *M* deviations according to a 2D autoregressive, AR(1), process
567 across years and ages. M-4 had the lowest negative log likelihood (-log$\mathcal{L}$) and Akaike's information criterion (AIC). Mohn's ρ were
568 averaged over seven retrospective peels for three quantities: recruitment (*R*), spawning stock biomass (*SSB*), and fishing mortality
569 averaged over ages 4-5 (*F*). Maximum likelihood estimates of the parameters constraining *M* random effects are listed with standard
570 error in parentheses. Dashes indicate parameters which are not included in a given model.

		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	—	—	—	—	-919.603	-1685.2	233.3	6.42	1.02	-0.43
M-1	Indep.	1.21 (0.10)	—	—	-1031.676	-1907.4	11.1	0.12	0.18	-0.11
M-2	AR(1) age	1.15 (0.43)	—	0.26 (0.48)	-968.364	-1778.7	139.8	2.88	0.12	-0.10
M-3	AR(1) year	0.17 (0.08)	0.98 (0.02)	—	-981.554	-1805.1	113.4	1.50	-0.14	0.32
M-4	2D AR(1)	0.79 (0.14)	0.63 (0.16)	0.40 (0.16)	-1039.268	-1918.5	0.0	-0.10	0.07	-0.00

571

572 Table 3. Model results where deviations in both numbers-at-age (NAA) and natural mortality (*M*) were estimated as random effects.
573 NAA-M-3 had the lowest negative log likelihood (-log$\mathcal{L}$) and Akaike's information criterion (AIC). Mohn's ρ were averaged over
574 seven retrospective peels for three quantities: recruitment (*R*), spawning stock biomass (*SSB*), and fishing mortality averaged over
575 ages 4-5 (*F*). Parameters are described in Eqns. 5 and 6. The model that included all parameters, 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	-1047.749	-1937.5	30.2	0.21	0.12	-0.05
NAA-M-2	σ_R, σ_a	$\sigma_M, \varphi_{year}, \varphi_{age}$	-1046.981	-1932.0	35.7	-0.14	0.03	0.04
NAA-M-3	$\sigma_R, \sigma_a, \rho_{year}, \rho_{age}$	σ_M	-1064.853	-1967.7	0.0	0.45	0.05	0.01
NAA-M-4	$\sigma_R, \sigma_a, \rho_{year}, \rho_{age}$	$\sigma_M, \varphi_{year}, \varphi_{age}$						

576

Table 4. Performance metrics and short-term projections of spawning stock biomass (SSB) for models with independent (indep.) or 2D autoregressive, AR(1), random effect deviations in numbers-at-age (NAA) and natural mortality (M). NAA-M-3 had the lowest Akaike's information criterion (AIC). Mohn's ρ were averaged over seven retrospective peels for three quantities: recruitment (*R*), spawning stock biomass (*SSB*), and fishing mortality averaged over ages 4-5 (*F*). *F* was set to 0 in projection years. Model size is the number of random effects (dimension of the Hessian matrix with respect to random effects). Model run times were a function of model size and whether TMB detected sparseness of the Hessian matrix.

	Description			Performance metrics				SSB projections (mt)			Characteristics		
Model	NAA random effects		M random effects	ΔAIC	ρ_R	ρ_{SSB}	ρ_F	2019	2020	2021	Size	Sparse Hessian	Run time (min)
Base	Age-1	Indep.	—	282.5	6.42	1.02	-0.43	312 (194, 500)	1151 (117, 11282)	2324 (294, 18377)	45	No	1.22
NAA-2	All ages	Indep.	—	74.8	0.81	0.17	-0.10	482 (201, 1153)	1556 (233, 10384)	2793 (466, 16722)	270	Yes	0.24
NAA-5	All ages	2D AR(1)	—	29.9	0.51	0.06	0.00	298 (119, 745)	602 (119, 3042)	1133 (154, 8342)	270	Yes	0.72
M-1	Age-1	Indep.	Indep.	60.3	0.12	0.18	-0.11	250 (71, 884)	1972 (231, 16873)	4512 (764, 26638)	321	No	7.65
M-4	Age-1	Indep.	2D AR(1)	49.2	-0.10	0.07	-0.00	203 (79, 518)	704 (45, 11091)	1782 (127, 24955)	321	No	8.10
NAA-M-2	All ages	Indep.	2D AR(1)	35.7	-0.14	0.03	0.04	185 (67, 510)	383 (28, 5163)	793 (39, 16020)	546	Yes	1.06
NAA-M-3	All ages	2D AR(1)	Indep.	0.0	0.45	0.05	0.01	318 (120, 844)	680 (138, 3352)	1279 (164, 9981)	546	Yes	1.28

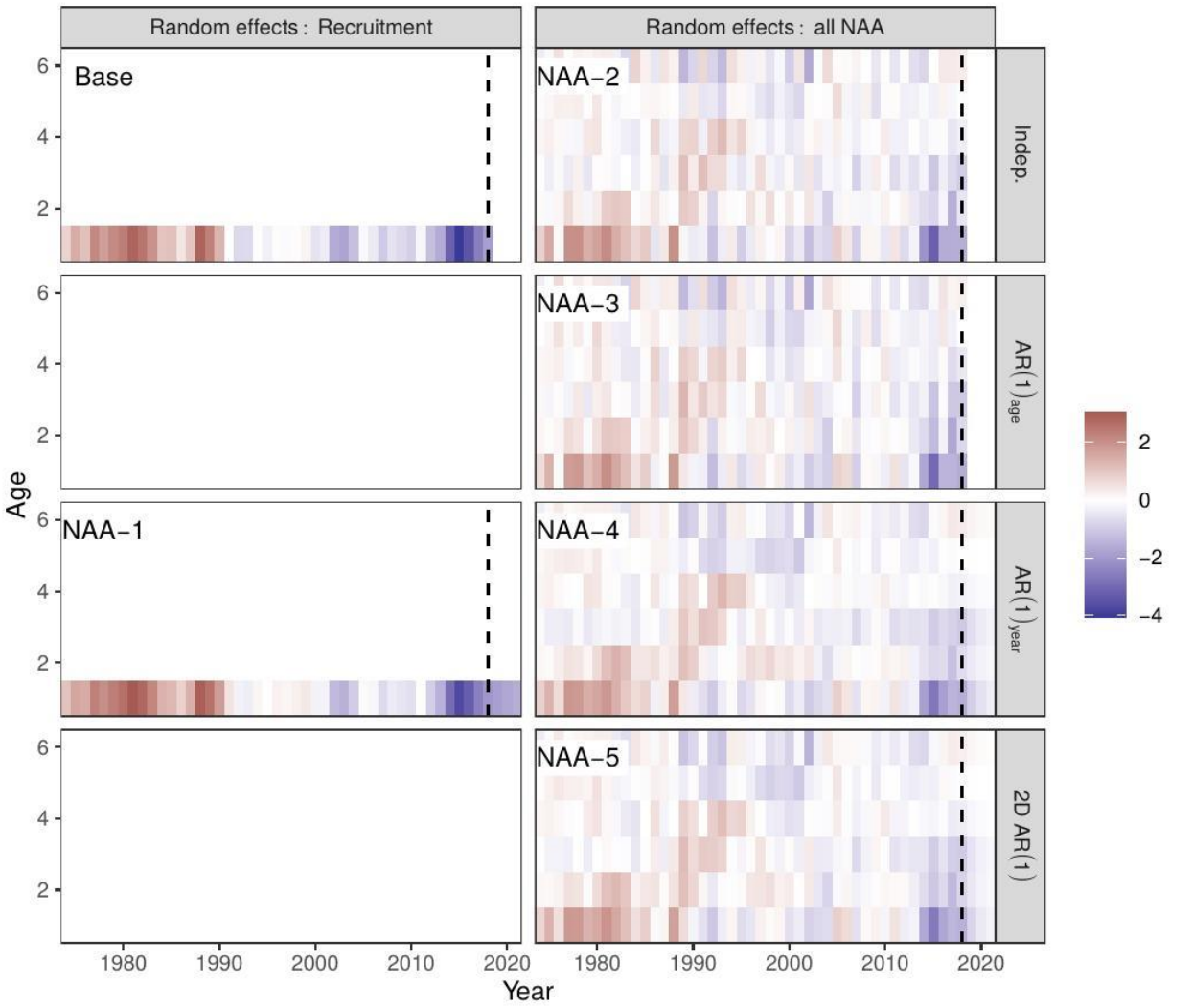


Figure 1. Deviations in log numbers-at-age (NAA) by year and age estimated by models in which only NAA are random effects. Models in the left column treat only age-1, i.e., recruitment, deviations as random effects, whereas models in the right column treat all NAA deviations as random effects. Models are grouped into rows by correlation structure: Indep. = independent (no correlation), $AR(1)_{age}$ = autoregressive by age, $AR(1)_{year}$ = autoregressive by year, and 2D AR(1) = autoregressive by age and year. The vertical dashed line denotes the terminal year in the assessment, 2018. Model descriptions are listed in Table 1.

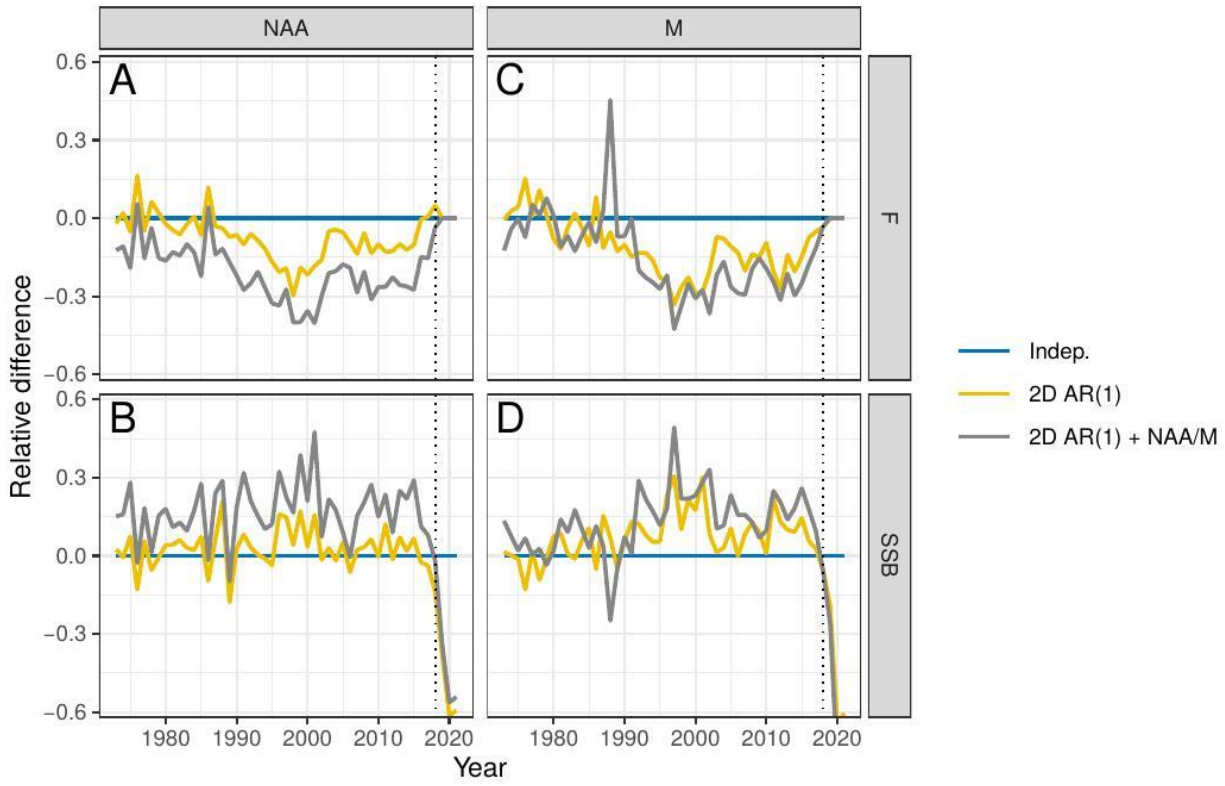


Figure 2. Relative difference in estimates of fishing mortality (*F*, top row) and spawning stock biomass (SSB, bottom row) from constraining deviations in numbers-at-age (NAA, left) and natural mortality (*M*, right column) to follow a 2D autoregressive correlation structure over ages and years, 2D AR(1). Relative difference was calculated using the model with independent (Indep.) deviations in the process listed in column heading as the baseline, i.e., $\theta_{2D\ AR(1)}/\theta_{Indep.} - 1$, where θ is either *F* or SSB. Results labeled as "2D AR1 + NAA/M" are from models with 2D AR(1) deviations in the process by column as well as independent 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.

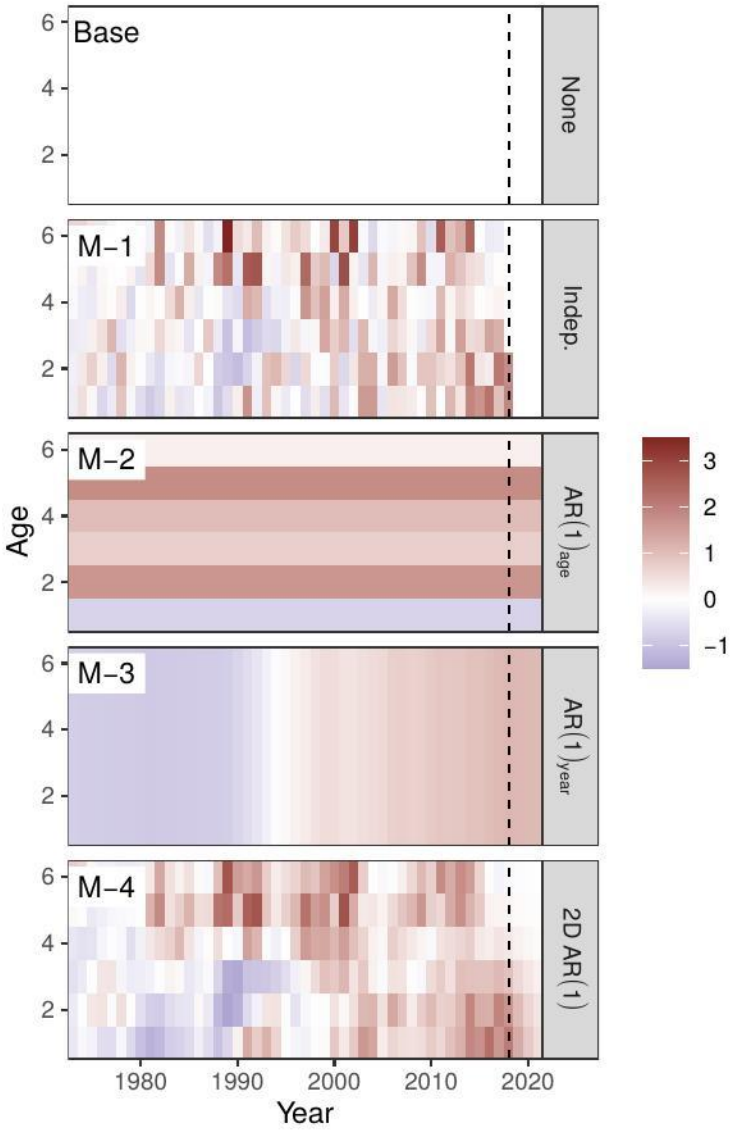


Figure 3. Deviations in log natural mortality (*M*) by year and age estimated by models without numbers-at-age (NAA) random effects. Models are grouped into rows by the correlation structure of the *M* deviations: none = no deviations from the *M* values specified in the assessment (Base model), indep. = independent (no correlation, M-1), $AR(1)_{age}$ = autoregressive by age (M-2), $AR(1)_{year}$ = autoregressive by year (M-3), and 2D AR(1) = autoregressive by age and year (M-4). The vertical dashed line marks the terminal year in the assessment, 2018. Model descriptions are listed in Table 2.

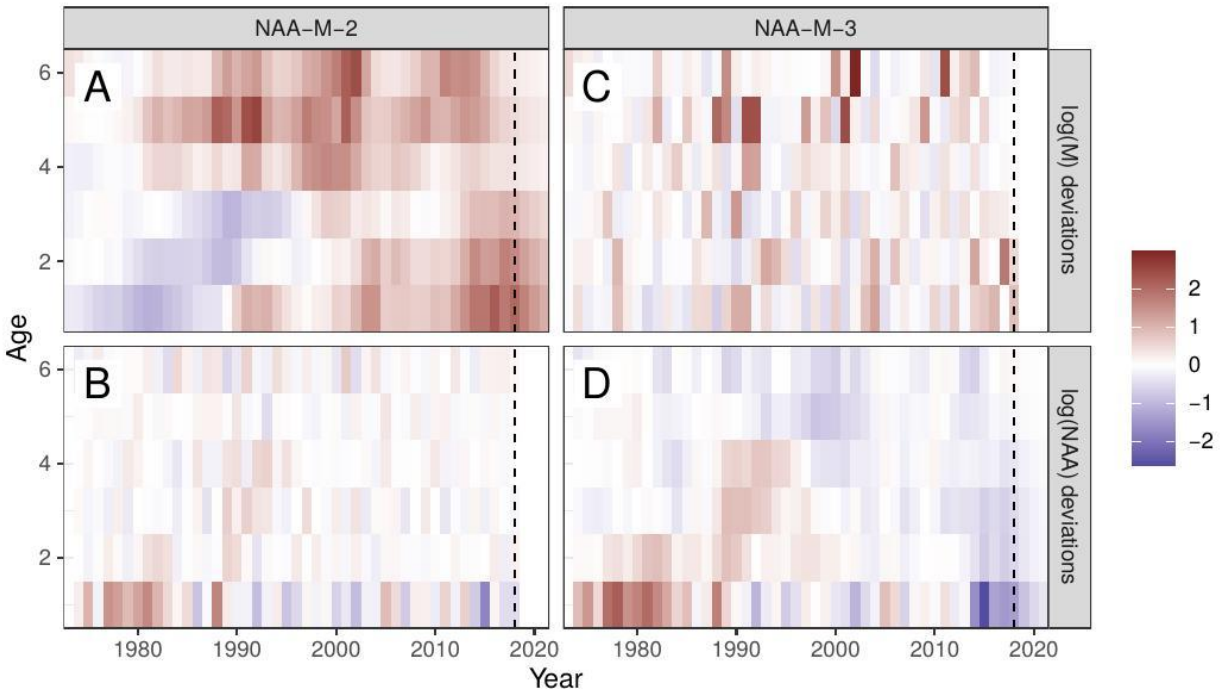


Figure 4. Deviations in natural mortality (log *M*, top row) and numbers-at-age (log NAA, bottom row) from models NAA-M-2 (left column) and NAA-M-3 (right column). The vertical dashed line marks the terminal year in the assessment, 2018. Deviations are zero in the projection years when they are treated as independent (B, C), and non-zero when they are autoregressive by year and age, i.e., 2D AR(1) (A, D).

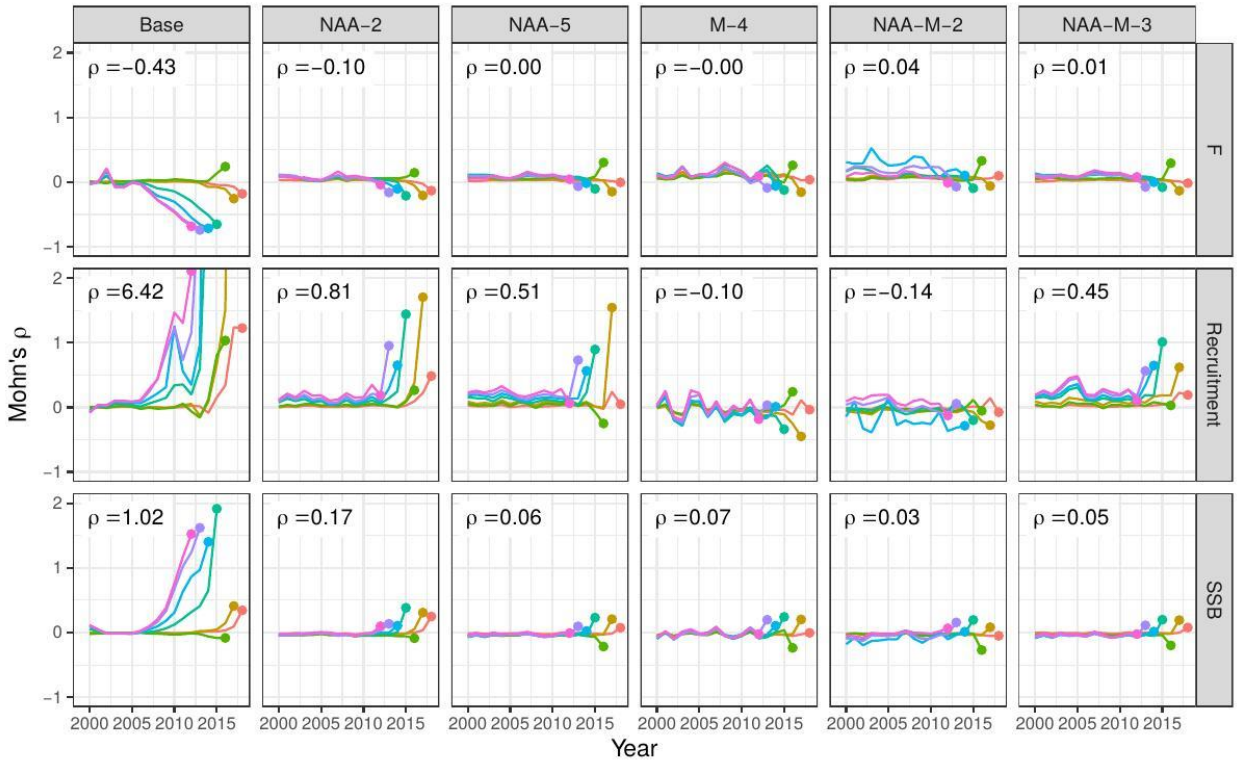
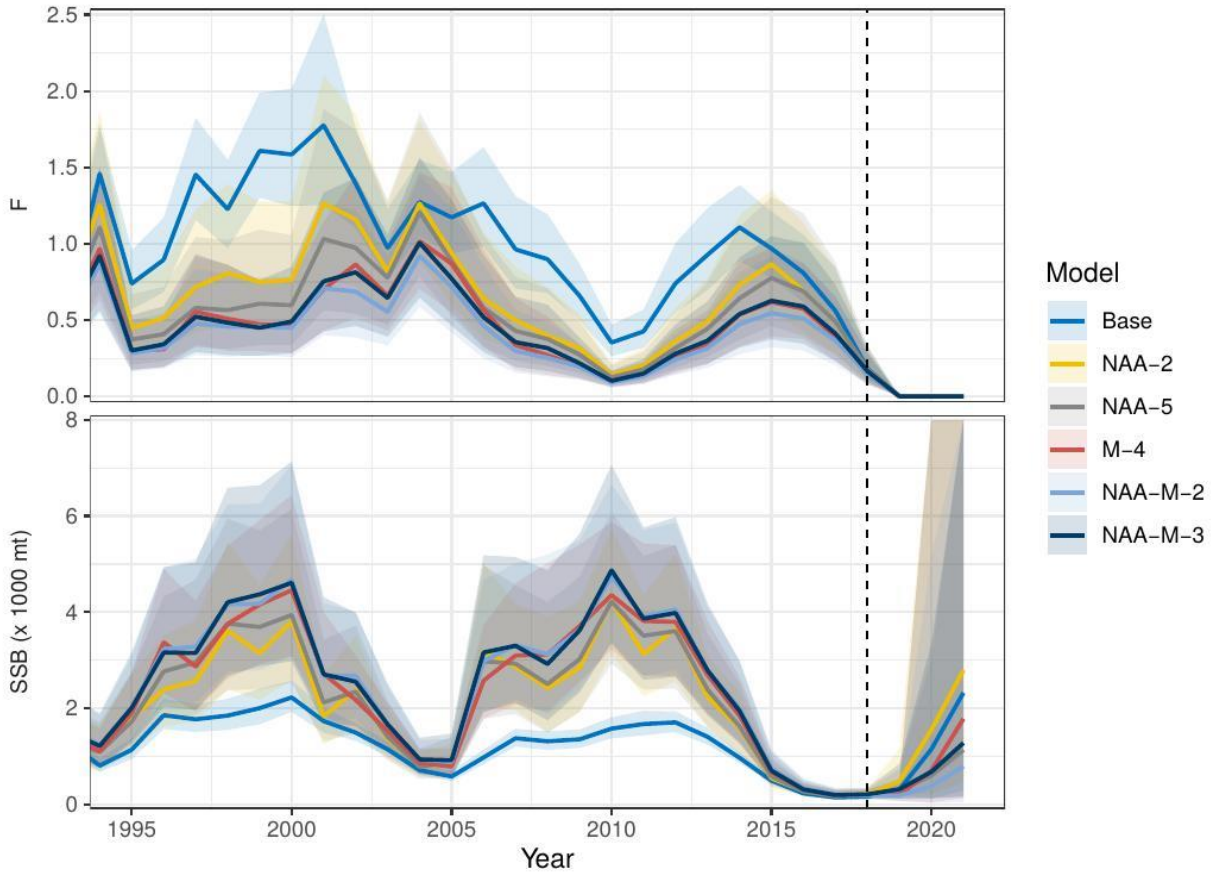


Figure 5. Retrospective patterns in fishing mortality (*F*, top row), recruitment (middle row), and spawning stock biomass (SSB, bottom row). Lines and points depict Mohn's ρ from seven peels, and the average Mohn's ρ is given in each panel. Columns show results by model, and model descriptions are listed in Table 4.



627

628 Figure 6. Trends in fishing mortality (*F*) and spawning stock biomass (SSB) from models with
629 independent or 2D autoregressive deviations in numbers-at-age (NAA), natural mortality (*M*), or
630 both. Model descriptions are given in Table 4. The dashed line denotes the terminal year in the
631 assessment, 2018. *F* is fixed at 0 for all models in projection years, 2019-2021. The first year in
632 the assessments is 1973, beyond the left x-axis limit.