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. Stock1\*, Haikun Xu2, Timothy J. Miller1, James T. Thorson3, and Janet A. Nye4

*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 attribute in fisheries stock assessment models and is typically treated as a deterministic process without process uncertainty. A recently developed state-space assessment model, however, treated the variation in survival as stochastic and independent of age and year. Several studies have shown that some population attributes such as recruitment can be autocorrelated, and that not accounting for autocorrelated population processes in stock assessment models may result in notably biased estimates of population attributes. These studies as well as the strong retrospective pattern found in the last assessment of Southern New England yellowtail flounder (*Limanda ferruginea*) motivated us to evaluate the two-dimensional (age and year) autocorrelations in survival. We found that survival deviations were first-order autocorrelated among both ages (autocorrelation = 0.56, standard error = 0.10) and years (autocorrelation = 0.33, standard error = 0.12). Moreover, the model which estimated the autocorrelations in survival deviations fitted data much better (AIC = 22) and had notably reduced retrospective patterns than the model which assumed independent survival. We also found that estimating the autocorrelations in survival deviations altered both the estimates and predictions of spawning biomass and fishing mortality by as much as 40%. Finally, by comparing the relative effect of an environmental covariate for recruitment and this 2-dimensional smoother for survival on model outcome, we showed that the survival smoother is notably more important to the assessment of Southern New England yellowtail flounder in terms of improving model fit according to AIC and providing reliable SSB prediction.

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

# 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. Population 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). To date, using a two-dimensional (2D) autocorrelation structure across both ages and years to model population processes is rare. One 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 () were simultaneously estimated. However, Cadigan (2016) did not compare the model estimates, predictions, goodness-of-fit, or retrospective patterns under alternative 2D autocorrelation structures for .

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 (Stone et al. 2004) except for the southern-most Southern New England-Mid Atlantic (SNEMA) stock. 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 SSBMSY 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 of spawning stock biomass (SSB) and fully-selected fishing mortality rate (*F*). The cause underlying the strong 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), stock assessment scientists sometimes impose a temporal trend to *M* in stock assessment models. *M* is an important parameter in stock assessment models because it directly influences stock productivity and 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). However, *M* is often difficult to estimate accurately in stock assessment models because it is confounded with some other parameters such as fishing mortality 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 . They found that misspecifying the temporal variation in likely induces larger biases than misspecifying the age-variation in does, underling the importance of correctly specifying the temporal trend in to stock assessment models. In fact, the impacts of misspecifying are positively related to , and was near the historical maximum for SNEMA yellowtail flounder in 2018, since was at a historic low (Legault and Palmer 2015; NEFSC 2020). Furthermore, the current SNEMA yellowtail flounder assessment specifies 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 declines with size (Lorenzen 1996; NEFSC 2012). Thus, there is reason to believe that misspecifying as constant could be a reason for the concerning retrospective patterns in recent assessments of the stock.

Instead of imposing a trend on by age or year, we explored a more flexible and objective method to address the retrospective problem: estimating a 2D smoother on survival or 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*, , is determined by *F*, *M*, and the numbers in the previous year: . Process errors, , can be included directly on as random effect deviations in survival (Gudmundsson and Gunnlaugsson 2012; Nielsen and Berg 2014; Miller et al. 2016) or on (Cadigan 2016), such that . We extend the model presented in Miller et al. (2016), in which annual survival was stochastic, but with the stochastic deviations in survival assumed to be uncorrelated by age or year, i.e. all . The can be equivalently called “random effects,” “deviations,” or “process errors” on numbers-at-age or survival. Henceforth, we refer to the as random effects or deviations.

In this study, we investigated 2D autocorrelation structures on both survival and 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 (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 , if a model could be estimated with 2D AR(1) smoothers on both survival and , 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.

# 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, , for age and year *y* () can be calculated by rewriting the stock equations:

where represents numbers at age and is the stock-recruit function in which an environmental time series () can be incorporated as a covariate. is the total number of observation and prediction years and 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 () 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., whenever ). 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 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 zero means. In other words, for all *a* and *y*:

(2)

where for all ages were assumed to be the same but different from age *a* = 1, i.e. recruitment, which we denote as . 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, the can be estimated internally in the model as fixed effect parameters. Survival deviations, however, are not necessarily independent. If survival deviations are autocorrelated among ages and years, they should follow a multivariate normal distribution:

(3)

where , is the  covariance matrix for the multivariate normal distribution and is calculated as the Kronecker product of the covariance matrix for the AR(1) process among ages () and the correlation matrix for the AR(1) process among years ():

where and 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 and are fixed at 0, there is no survival autocorrelation in either dimension and Eq. 3 collapses to Eq. 2. In fact, is the likelihood distribution function for this covariance structure:

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, for all ages were assumed equal but different from age *a* = 1, i.e. recruitment, denoted as .

Alternatively, we can apply this 2D AR(1) structure to random effect deviations in *M*, , as in Cadigan (2016):

(6)

where is the mean *M* and can either be fixed or estimated. Applying deviations in *M* versus determines where in the Baranov catch equation it affects the predicted catch, . Another difference is that deviations in *M* affect the calculation of reference points, whereas deviations in do not.

## 2.2. 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. 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. All the aforementioned data and the complete assessment report can be accessed at https://apps-nefsc.fisheries.noaa.gov/saw/sasi/sasi\_report\_options.php.

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, , assuming alternative autocorrelation structures formed by fixing or estimating the three parameters in Eq. 5. We refer to NAA-1 as the “base” model, because it was most similar to a statistical catch-at-age model, where recruitment is typically estimated independently by year and survival is deterministic. NAA-2 added recruitment autocorrelation. NAA-3 through NAA-6 estimated “full state-space” models, with numbers at all ages treated as random effects, but with different autocorrelation structures. NAA-3 estimated independently as in Miller et al. (2016) and Nielsen and Berg (2014), NAA-4 and NAA-5 added autocorrelation across ages and years, and NAA-6 estimated all parameters in the described 2D AR(1) smoother (Eq. 5, Table 1). To isolate the effect of incorporating the 2D AR(1) smoother on survival, we compared the model fit, retrospective pattern, and relative difference in SSB and F estimates from NAA-6 versus NAA-3.

Next, we fit a set of models treating the numbers-at-age as in the base model, NAA-1, but including deviations in *M* with the same set of autocorrelation structures: none, independent, AR(1) by age, AR(1) by year, and 2D AR(1). For each autocorrelation structure, we alternatively fixed or estimated the mean *M*, , resulting in a set of 10 nested models (Table 2). As for the set of NAA models, we isolated the effect of the 2D AR(1) smoother on *M* by comparing M-2 to M-5 (fixed ) and M-7 to M-10 (estimated ). We then tested the ability of WHAM to simultaneously estimate numbers-at-age and *M* as random effects, using only the independent and 2D AR(1) autocorrelation structures for each.

Finally, we tested whether linking an environmental covariate to recruitment could further improve model fit and retrospective pattern. Miller et al. (2016) and Xu et al. (2018) found that incorporating the Cold Pool Index (CPI) or Gulf Stream Index (GSI) into the stock-recruit function for SNEMA yellowtail flounder significantly improved model performance as measured by AIC and Mohn’s . Here, we used the CPI instead of the GSI because standard errors were straightforward to calculate for the CPI and these values could be used directly in the assessment model as observation error. This avoided the need to estimate the CPI observation error internally. We calculated the CPI as in Miller et al. (2016) and incorporated it into the Beverton-Holt stock-recruit function as a limiting factor:

where is from Eq. 1.

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. 2015), 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. 2015). We compared model fit and retrospective pattern using AIC and Mohn’s (Mohn 1999), using seven retrospective peels as in the latest assessment (NEFSC 2020).

# 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 ; Table 3). Estimating survival deviations with autocorrelation by age, year, or both further reduced AIC and Mohn’s . According to both AIC and the magnitude of estimated and , 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 3). The survival deviations estimated by models with autocorrelation were smoothed across ages and years relative to the models with independent deviations (Fig. 1). NAA-6, with the 2D AR(1) structure, had the best fit and retrospective pattern, reducing AIC by 36.2, by 0.41, by 0.18, and by 0.14 compared to NAA-3 with independent survival deviations (Table 3). Simply constraining the survival deviations with the 2D AR(1) structure reduced estimates of F and increased estimates of SSB by about 15% (relative difference between NAA-6 and NAA-3; Fig. 2). NAA-3 and NAA-6 estimated similar SSB in the terminal year of the assessment, 2018, but then differed in their SSB estimates in the projection years by about 30%, even though F was held constant at similar values (Fig. 2).

Owing to that negative survival deviations were predicted for the oldest age classes, which are the primary SSB contributors, the survival smoother (S0(age, year)) immediately lower the first SSB prediction compared to S0(base) (Fig. 5). Note that the survival deviation at age was predicted to a[symptot](javascript:void(0);)ically approach zero over time (see Fig. 2), so the direct influence of the survival smoother on SSB predictions are expected to be increasingly weaker over time. However, the predicted negative survival deviations in the first prediction year still propagate to older age classes over time and indirectly influence SSB predictions in the second and third prediction years. In this case, the direct and indirect impacts of the survival smoother on SSB predictions both lead to increasingly lower SSB predictions from S0(age, year) compared to S S0(base) and even S(base).

## 3.2. Deviations in *M* as random effects

Including deviations in *M*, instead of survival, also substantially improved model fit and retrospective pattern (Table 4). This was true whether or not was estimated. 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. Again, the 2D AR(1) structure had the best fit, regardless of whether was estimated (M-5 and M-10; Table 4). Compared to the models with independent M deviations, including the the 2D AR(1) structure reduced AIC by 5.9, by 0.05, by 0.10, and by 0.06 when was not estimated (Table 4). When was estimated, however, adding the 2D AR(1) structure resulted in worse retrospective pattern ( by 0.18, by 0.14, and by 0.18; Table 4). The estimated 2D AR(1) *M* deviations were smoothed less than the 2D AR(1) survival deviations, as evidenced by and (Figs. 1 and 3, Tables 3-4). When was not estimated, the effect of adding the 2D AR(1) structure on estimates of F and SSB was similar as for the NAA models: lower F and higher SSB by about 10-15% during assessment years, similar terminal year status, and then lower SSB in short-term projections by a greater amount (Fig. 2). Models that estimated increased *M* compared to the value fixed in the assessment, and adding the 2D AR(1) structure on *M* deviations resulted in larger differences in F and SSB.

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

Models that attempted to estimate deviations in both survival and *M* with 2D AR(1) autocorrelation failed to converge. However, adding independent *M* deviations to NAA-6 (2D AR(1) on survival) and adding independent survival deviations to M-5 and M-10 (2D AR(1) on *M*) substantially improved model fit (lower AIC by 25.9, 9.5, and 3.8, respectively; Tables 3-5). Differences in Mohn’s between these models were not large or consistent. In the model with lowest Mohn’s , deviations in both survival and M reinforced each other—positive survival deviations were estimated for the same years and ages as negative M deviations (e.g. ages 1-3 during the late 1970s-1980s in Fig. 4). Thus, the magnitude of the deviations reduced from approximately -2 to 2, as when only survival or *M* deviations were estimated (Figs. 1 and 3), to the range -1 to 1 (Fig. 4). Including independent deviations in survival and *M* did not have much impact on F or SSB estimates relative to including only 2D AR(1) deviations on either, except for when was estimated (Fig. 2).

## 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 which converged (Tables 5-6). Across models with different random effects on survival and *M*, including a CPI effect on recruitment reduced Mohn’s by about 0.1 on average and did not have much influence on or (Fig. 5). The NAA-M-CPI model with lowest Mohn’s had the lowest Mohn’s of all models considered (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 were already included (Fig. 7).

# Discussion

Using a state-space, age-structured assessment model developed for SNEMA yellowtail flounder, we showed that implementing the 2D AR(1) survival smoother considerably improved model fit and reduced the retrospective patterns for SSB and . Different from some previous studies which addressed the retrospective problem in New England stock assessments by applying the temporal trend in , this paper provides a more objective, flexible, and generic approach to reduce retrospective pattern in stock assessments. Particularly, both the two AR(1) coefficients for the survival smoother can be pre-specified as a constant or estimated in the state-space model as fixed effect parameters. By comparing the AIC of four alternative models with differing autocorrelation assumptions for survival deviations, we showed that it is more important to estimate the among-year than among-age autocorrelation in survival deviations of SNEMA yellowtail flounder. The among-year autocorrelation in survival deviations was estimated to be high (0.56) in the best fitting model, indicating that biological processes for ages larger than 1 can also be autocorrelated across time. Moreover, the estimated survival deviations for age 5&6 were quantitatively comparable to those for age 1, informing us that the variation in adult survival is at least as influential as that in recruitment to the population dynamics of SNEMA yellowtail flounder.

We found that the improvement in model fit that was induced by the survival smoother alone was notably larger than that was induced by the recruitment covariate alone. We also found that the survival smoother had profound impacts on SSB and estimates as well as near-term SSB predictions, while the recruitment covariate negligibly impacted the estimates of SSB and . These two findings further underscored the importance of implementing the 2D AR(1) survival smoother to the assessment of SNEMA yellowtail flounder. Indeed, this 2D AR(1) smoother for survival deviations can also extended to three dimensions if, for example, survival is modelled to be sex-specific. Also, the generic 2D AR(1) smoother we introduce in this paper can be applied to other potentially autocorrelated biological processes (e.g., selectivity and growth) as well, which we predict would be an important research topic for continuously developing mixed-effect stock assessment models.

For SNEMA yellowtail flounder, implementing the 2D AR(1) survival smoother in the state-space model considerably reduced the severity of retrospective pattern and modified near-term SSB prediction as well as the associated uncertainty interval. Indeed, both the Mohn’s rhos for SSB and *F* became negligibly different from zero when the survival smoother was implemented, which means that the state-space model with the survival smoother is more reliable in terms of determining the stock and harvest status of SNEMA yellowtail flounder. Because the among-year autocorrelation in survival deviations was estimated to be high, the best predictions for survival deviation at age from the model with the survival smoother (S0(age, year)) have the same sign as those estimated for the terminal year and decline exponentially towards zero with slope determined by the autocorrelation coefficient. In contrast, the best predictions for survival deviation at age from the model without the survival smoother (S0(base)) are zero as a result of the independent assumption for survival deviations. Therefore, S0(age, year) provided lower SSB predictions than S0(base) for 2012-2014. In addition to the median SSB, the associated uncertainty predicted by the model with and without the survival smoother was also notably different. The SSB predictions from the model with the survival smoother have wider uncertainty interval, namely, not accounting for the positive autocorrelation in survival deviations corresponds to under-estimating the uncertainty in near-term SSB predictions.

Near-term SSB predictions changed substantially when the GSI was incorporated into the stock-recruit function or the 2D AR(1) smoother was implemented for survival. The reason that the recruitment covariate (GSI) only impacted the second- and third-year SSB predictions can be found in the maturity pattern. In 2012, the GSI-induced recruitment signal does not affect SSB prediction because the maturity at age 1 is zero. In 2013 and 2014, the GSI-induced recruitment signal propagates to age 2 and 3 at which maturity is 0.47 and 1, respectively, so the recruitment covariate made a significant difference in the second (-8%) and third (-24%) SSB predictions. It is worth noting that while the change in maturity from age 1 to older ages is dramatic, the increasingly amplified effect of the recruitment covariate on SSB predictions over time is also contributed by much larger weights at older ages than age 1. The survival smoother, in contrast, has notably (-10%) altered the first SSB prediction in 2012 already because the change in predicted survival it induces are large and even comparable to the specified in the model (0.23-0.41, corresponding to age 6-1). Later on, the survival smoother induces even larger changes in the second (-24%) and third (-30%) SSB predictions, both of which are larger than the corresponding ones due to the recruitment covariate. In terms of the accuracy of near-term SSB predictions, implementing the 2D AR(1) survival smoother is likely to be more important to SNEMA yellowtail flounder than incorporating an environmental effect on recruitment.

The 2D AR(1) survival smoother, however, does not make calculating biological reference points (BRPs) easier. We envision two different ways to calculate BRPs when using this survival smoother:

1. *Deterministic BRP*: The simplest calculation for BRPs is to ignore the stochastic variation in survival and calculate the yield curve given average survival. This procedure ignores the process error in biological processes and is typically used to calculate BRPs.
2. *Dynamic BRP*: Regime shifts in recruitment have been shown to trigger rebuilding rules that target biomass levels that are essentially impossible under new recruitment dynamics (Szuwalski et al. 2015). Therefore, BRPs can vary across time and be calculated annually based upon the survival deviation estimated for each year.

In the case study, we found that the estimated survival deviations were predominantly negative since the 1990s, so ignoring this trend in survival may result in largely biased BRPs estimation for the recent three decades. Whereas dynamic BRPs seems to be the more appropriate way of calculating BRPs, it is more challenging because an assumption regarding how survival deviations are attributed to fishing mortality, natural mortality, and migration is required before calculating the dynamic BRPs. These three attributors 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 given different forms of time-variation in survival.

Care should be taken when interpreting the main findings found in this study. First of all, the survival for the plus group is not strictly survival. It can be a caveat in this study considering that the predicted survival deviations for the plus group changed notably as the survival smoother implemented in the state-space model. Secondly, population dynamics in 2012-2014 was predicted in the absence of fishing (i.e., for ) to remove the potential impacts of differing F scenarios on near-term SSB predictions. This assumption likely results in an exaggerated shift of the predicted age composition towards older age classes. However, a parallel SSB prediction run assuming more realistic fishing mortality patterns (status quo fishing mortality) for 2012-2014 provided similar age structure predictions (not shown), because *F* reached the historically low level (~0.2) in 2011 due to the low stock productivity in recent years. Thus, the conclusions made under zero *F* for 2012-2014 are robust considering the current SNEMA yellowtail flounder harvest status. Lastly, all the conclusions made in this study are species-specific. For example, the relative importance of the recruitment covariate and survival 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 survival smoother can directly impact the predicted survival at all ages, and thereby, is expected to be more important to near-term SSB prediction for late-matured and long-lived fish stocks. While a recruitment covariate can only directly impact near-term recruitment prediction. For late-matured and long-lived fish stocks, the covariate-induced change in recruitment prediction is not able to propagate to the majority of spawning age classes and notably impact SSB prediction in the near-term.

Lastly, it is important to note that implementing the 2D AR(1) survival smoother reduced the computation speed as it broke the sparseness of the Hessian matrix. In this case study with 69 fixed effects parameters and 287 random effect parameters, the implementation increased model convergence time from 10 seconds to 200 seconds in a laptop computer. The extent to which the convergence time increases in other cases due to the implementation is expected to be positively associated with the case-specific number of age classes and time steps (e.g., years) considered in the model. The implementation can actually be even more computational expensive when a better representation of the among-age survival autocorrelation structure is included, e.g., the irregular lattice correlation structure investigated in Berg and Nielsen (2016). However, with continuing improvement in computation ability, the consequence of reduced convergence speed is unlikely to be a hurdle in implementing this method in the future.

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

Berg, C.W.; Nielsen, A. Accounting for correlated observations in an age-based state-space stock assessment model. ICES Journal of Marine Science: Journal du Conseil:fsw046; 2016

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

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. Stock structure of yellowtail flounder off the northeastern United States. 2003

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

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

Goethel, D.R.; Legault, C.M.; Cadrin, S.X. 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: Journal du Conseil. 72:164-177; 2015

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

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.; Muradian, M.L. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science: Journal du Conseil:fsu198; 2014

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

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.; Ono, K. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES Journal of Marine Science: Journal du Conseil. 72:137-150; 2015

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

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

Legault, C.M.; Palmer, M.C. 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; 2015

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

Lorenzen, K. 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; 1996

Miller, T.J.; Hare, J.A.; Alade, L.A. 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; 2016

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

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: Journal du Conseil. 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; 600 p.

NEFSC. 2013. 57th Northeast Regional Stock Assessment Workshop (57th SAW) Assessment Report. US Deptartment of Commerce, Northeast Fisheries Science Center Reference Document.

NEFSC. 2020. Operational Assessment of 14 Northeast Groundfish Stocks, Updated Through 2018. \*will update citation when officially published.

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

Ono, K.; Punt, A.E.; Rivot, E. Model performance analysis for Bayesian biomass dynamics models using bias, precision and reliability metrics. Fisheries Research. 125:173-183; 2012

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. WHAM description and simulation test.

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

Szuwalski, C.S.; Vert‐Pre, K.A.; Punt, A.E.; Branch, T.A.; Hilborn, R. Examining common assumptions about recruitment: a meta‐analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries. 16:633-648; 2015

Thorson, J.T.; Jensen, O.P.; Zipkin, E.F. 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; 2014

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

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.

Table 1. Descriptions of models with only numbers-at-age (NAA) estimated as random effects. In other words, these models estimated deviations in survival by year and age, , with the covariance given by Eq. 5. NAA-1 is most similar to a statistical catch-at-age model, where recruitment is typically estimated independently by year and survival is deterministic.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Ages treated as random effects | Correlation structure | Estimated parameters | Notes |
| NAA-1 | Age-1 (recruits) | Indep. |  | base |
| NAA-2 | Age-1 (recruits) | AR(1) year | , |  |
| NAA-3 | All ages | Indep. | , | Miller et al. (2016) |
| NAA-4 | All ages | AR(1) age | , , |  |
| NAA-5 | All ages | AR(1) year | , , |  |
| NAA-6 | All ages | 2D AR(1) | , , , |  |

Table 2. Descriptions of models with only natural mortality (*M*) estimated as random effects. These models estimated deviations in *M* by year and age, , with the covariance given by Eq. 6. All models treated the numbers-at-age as in the base model, NAA-1 (recruitment independent by year, deterministic survival). M-1 was also equivalent to the base model and did not estimate any *M* parameters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Estimate ,  mean *M* | Correlation structure | Estimated parameters | Notes |
| M-1 | --- | --- | --- | base |
| M-2 | --- | Indep. |  |  |
| M-3 | --- | AR(1) age |  |  |
| M-4 | --- | AR(1) year | , |  |
| M-5 | --- | 2D AR(1) | , , | Cadigan (2016) |
| M-6 | Yes | --- |  |  |
| M-7 | Yes | Indep. | , |  |
| M-7 | Yes | AR(1) age | , |  |
| M-9 | Yes | AR(1) year | , , |  |
| M-10 | Yes | 2D AR(1) | , , , |  |

Table 3. Model results where only numbers-at-age (NAA) were estimated as random effects. NAA-6, where survival is a 2D AR1 process across years and ages, had the lowest AIC and Mohn’s . Mohn’s 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 | | |
| Model |  |  |  |  |  | -log | AIC |  |  |  |  |  |
| NAA-1 | 1.18 (0.13) | — | — | — |  | -934.024 | -1712.0 | 221.6 |  | 5.86 | 0.99 | -0.42 |
| NAA-2 | 0.66 (0.08) | — | 0.92 (0.05) | — |  | -957.164 | -1756.3 | 177.3 |  | 4.56 | 0.90 | -0.38 |
| NAA-3 | 1.01 (0.12) | 0.58 (0.05) | — | — |  | -1027.684 | -1897.4 | 36.2 |  | 0.97 | 0.27 | -0.17 |
| NAA-4 | 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-5 | 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-6 | 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 4. Model results where only recruitment and natural mortality (M) were estimated as random effects. M-10, which estimated mean M, , as well as 2D AR1 random effects on M, had the lowest AIC but not Mohn’s . M-5 and M-7 had lower Mohn’s than M-10, but higher AIC. Numbers-at-age (NAA) were treated as in model NAA-1. Only estimating , without any random effects on NAA or M, substantially reduced AIC and Mohn’s (compare M-1 to M-6). Mohn’s 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 | | |
| Model |  |  |  |  |  | -log | AIC |  |  |  |  |  |
| M-1 | — | — | — | — |  | -934.024 | -1712.0 | 227.5 |  | 5.86 | 0.99 | -0.42 |
| M-2 | — | 1.18 (0.09) | — | — |  | -1037.981 | -1918.0 | 21.5 |  | 0.12 | 0.20 | -0.13 |
| M-3 | — | 1.15 (0.43) | — | 0.25 (0.49) |  | -979.398 | -1798.8 | 140.7 |  | 2.06 | 0.10 | -0.08 |
| M-4 | — | 0.13 (0.06) | 0.98 (0.02) | — |  | -988.290 | -1816.6 | 122.9 |  | 1.19 | -0.12 | 0.30 |
| M-5 | — | 0.89 (0.12) | 0.47 (0.21) | 0.39 (0.14) |  | -1042.965 | -1923.9 | 15.6 |  | -0.07 | 0.10 | -0.06 |
| M-6 | 0.79 (0.04) | — | — | — |  | -976.715 | -1795.4 | 144.1 |  | 2.36 | 0.15 | -0.09 |
| M-7 | 0.70 (0.12) | 0.71 (0.10) | — | — |  | -1045.295 | -1930.6 | 8.9 |  | -0.13 | -0.03 | 0.09 |
| M-8 | 0.81 (0.11) | 0.42 (0.16) | — | -0.86 (0.19) |  | -983.717 | -1805.4 | 134.1 |  | 1.98 | 0.12 | -0.07 |
| M-9 | 0.36 (0.22) | 0.12 (0.07) | 0.98 (0.03) | — |  | -983.959 | -1805.9 | 133.6 |  | 1.55 | -0.10 | 0.26 |
| M-10 | 0.83 (0.16) | 0.56 (0.08) | 0.37 (0.14) | 0.30 (0.16) |  | -1051.753 | -1939.5 | 0.0 |  | -0.31 | -0.17 | 0.27 |

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

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimated parameters | | |  | Model fit | | |  | Mohn’s | | |
| Model | NAA | M |  |  | -log | AIC |  |  |  |  |  |
| NAA-M-1 | , |  |  |  | -1051.090 | -1942.2 | 17.3 |  | 0.20 | 0.17 | -0.09 |
| NAA-M-2 | , | , , |  |  | -1048.691 | -1933.4 | 26.1 |  | -0.22 | 0.02 | 0.04 |
| NAA-M-3 | , |  | Yes |  |  |  |  |  |  |  |  |
| NAA-M-4 | , | , , | Yes |  | -1054.627 | -1943.3 | 16.2 |  | -0.52 | -0.28 | 0.54 |
| NAA-M-5 | , , , |  |  |  | -1061.727 | -1959.5 | 0.0 |  | 0.52 | 0.10 | -0.04 |
| NAA-M-6 | , , , | , , |  |  |  |  |  |  |  |  |  |
| NAA-M-7 | , , , |  | Yes |  |  |  |  |  |  |  |  |
| NAA-M-8 | , , , | , , | Yes |  |  |  |  |  |  |  |  |

Table 6. 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-4 had the lowest AIC but not Mohn’s . NAA-M-CPI-2 had higher AIC but the lowest Mohn’s . NAA-M-CPI-5 did not converge. Mohn’s abbreviations: *R* = recruitment, *SSB* = spawning stock biomass, and *F* = fishing mortality averaged over ages 4-5.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimated parameters | | |  | Model fit | | |  | Mohn’s | | |
| Model | NAA | M |  |  | -log\* | AIC |  |  |  |  |  |
| NAA-M-CPI-1 | , |  |  |  | -996.345 | -1824.7 | 3.7 |  | 0.16 | 0.17 | -0.09 |
| NAA-M-CPI-2 | , | , , |  |  | -994.979 | -1818.0 | 10.4 |  | -0.11 | 0.02 | 0.06 |
| NAA-M-CPI-4 | , | , , | Yes |  | -1001.182 | -1828.4 | 0.0 |  | -0.42 | -0.25 | 0.47 |
| NAA-M-CPI-5 | , , , |  |  |  | -1000.819 |  |  |  |  |  |  |

\*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.

A screenshot of a cell phone

Description automatically generated

Figure 1. Survival deviations by year and age estimated by models in which only numbers-at-age (NAA) were random effects.

A close up of a map

Description automatically generated

Figure 2. Relative difference in estimates of F (top panels) and SSB (lower panels) from constraining deviations in survival (left) and *M* (center, right) to follow a 2D AR(1) correlation structure over ages and years. Relative difference was calculated as , where is either F or SSB. In the center column, was fixed, whereas was estimated in the right column. IID = independent deviations by age and year (as per 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.

A screenshot of a social media post

Description automatically generated

Figure 3. Deviations in natural mortality (*M*) by year and age estimated by models without numbers-at-age (NAA) random effects. Estimating (right column) clearly reduced the magnitude of *M* deviations for AR(1) models (M-8 and M-9 vs. M-3 and M-4), but did not reduce the *M* deviations in models where the deviations varied by year and age (M-7 and M-10 vs. M-2 and M-5).

A picture containing brick

Description automatically generated

Figure 4. Deviations in natural mortality (*M*, top panel) and numbers-at-age (NAA, bottom panel) from the final model with lowest Mohn’s , NAA-M-CPI-2.

A picture containing bird

Description automatically generated

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 between otherwise equivalent models. Across the various NAA and M models, including the CPI-recruitment link minimally impacted Mohn’s or , and slightly lowered Mohn’s by about 0.1 on average.

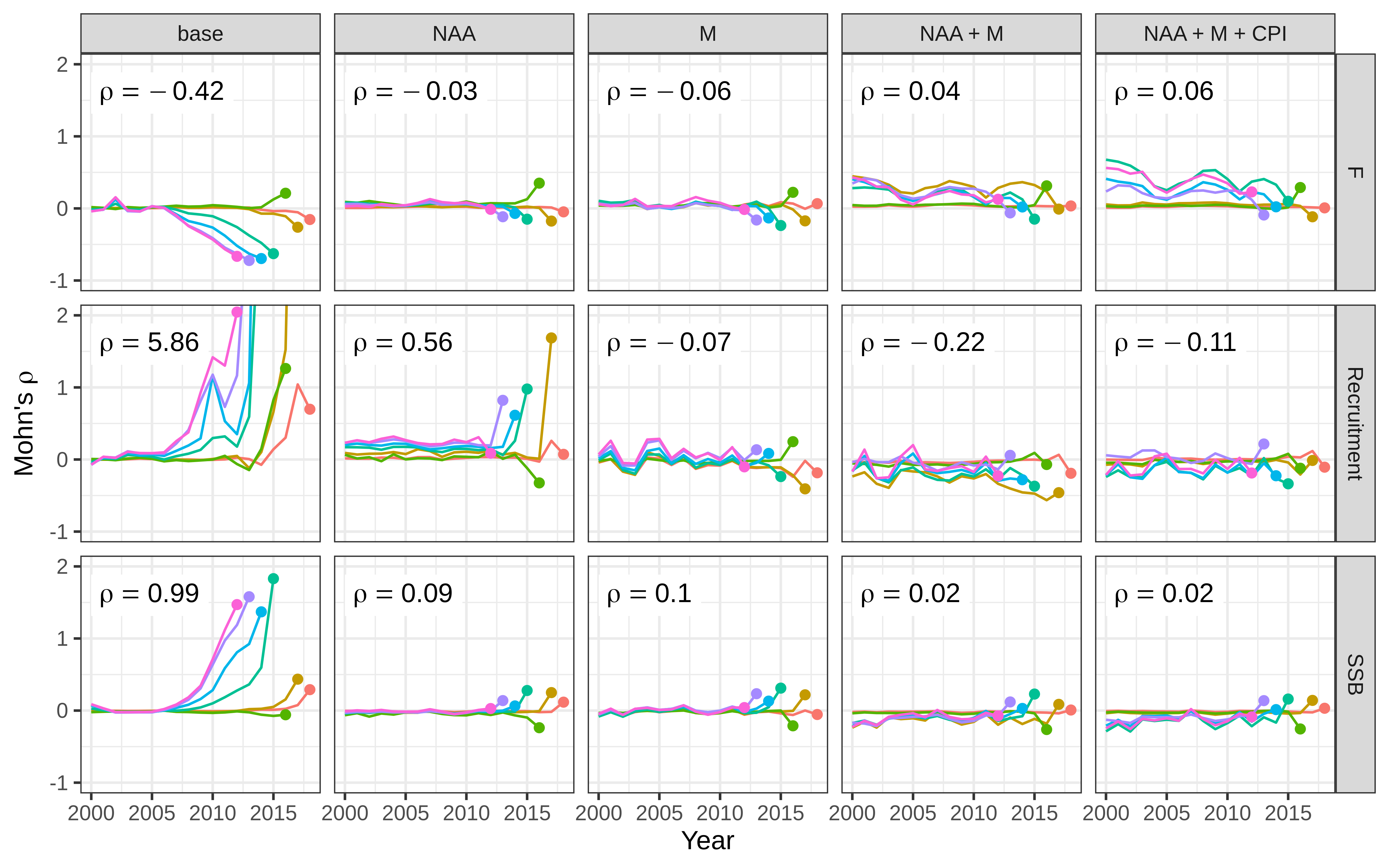


Figure 6. Retrospective patterns in fishing mortality (F, top row), recruitment (middle row), and spawning stock biomass (SSB, bottom row). Lines and points depict Mohn’s from seven peels, and the average Mohn’s is given in each panel. Columns show results by model: base = the statistical catch-at-age model, NAA-1, and the others are the models from Tables 3-6 with lowest Mohn’s (NAA-6, M-5, NAA-M-2, and NAA-M-CPI-2).

A close up of a map

Description automatically generated

Figure 7. Trends in fishing mortality (F) and spawning stock biomass (SSB). base = the statistical catch-at-age model, NAA-1. Within each model class, the model with lowest Mohn’s is shown (NAA-6, M-5, NAA-M-2, and NAA-M-CPI-2; Tables 3-6). The dashed line denotes the terminal year in the assessment, 2018. Note that the assessment years (1973-2018) extend beyond the x-axis limit.