Estimating Population Trends with Stratified Random Sampling Under the Pressures of Climate Change

Benjamin A. Levy[1], Christopher M. Legault[2], Timothy J. Miller[2], Elizabeth N. Brooks[2]

[1]Ben's Institution, USA

[2]National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, MA, USA

Corresponding author: Ben Levy (benjamin.levy@noaa.gov)

Competing interests: The authors declare there are no competing interests.

Abstract

An Abstract

Keywords

keyword 1, keyword 2

Introduction

much of below is from https://apps-nefsc.fisheries.noaa.gov/nefsc/ecosystem-ecology/ or https://www.fisheries.noaa.gov/data-tools/fisheries-economics-united-states-data-and-visualizations

- The eastern continental shelf is ecologically diverse and economically important

The Northeast United States continental shelf spans from the Outer Banks of North Carolina to the Gulf of Maine. The region covers over 250,000 km^2 of ocean, extending over 200 km from shore in the largest areas in New England to just 30 km off shore in the southern regions. This ecologically diverse region contains approximately 18,000 vertebrate marine species. Commercial fisheries have been an important part of local economies for centuries. In 2019, New England fisheries produced $22 billion in sales, which sustained over 200,000 jobs. Maintaining a healthy ecosystem is therefore vital to sustained ecological health and economic prosperity of the region. [@nefmc20]

- Bottom trawl survey is important for monitoring population trends

Fish stocks in this highly productive and economically important region are managed by the National Oceanic and Atmospheric Administration's (NOAA) Northeast Fisheries Science Center (NEFSC) in Woods Hole, Massachusetts. Federal biologists assess the health and abundance of each commercial fish stock using fishery-independent bottom trawl survey data that has been collected by NOAA throughout the region since 1963 (cite survey paper). The survey uses a stratified random design where bottom trawl sampling takes place in predefined strata along the eastern continental shelf. The survey has created a rich time series data set with many uses including species-specific habitat identification, analysis of how environmental conditions influence species abundance, and estimating yearly species abundance trends to help inform stock assessments and ultimately quota limits **just listed a few uses of survey change/others?**.

The survey takes place twice each year- once in the spring and again in the fall. Since most spatial analyses and projections of future distributions typically assume a constant survey catchability and/or availability over time, NOAA's survey design includes sampling during approximately the same 2-3 week time period in each season.

- Climate change is happening Due to a combination of climate change and shifts in circulation, the Northeast United States continental shelf has experienced rapid warming in recent decades, resulting in a shift in spatial distributions of many species. Since stock assessment models rely on accurate descriptions of population dynamics and contemporary patterns of spatial abundance, there is concern that rapid undocumented changes in spatial distributions of species will bias future stock assessments. The implication of this is that the bottom trawl survey is actually sampling the population during a different life cycle stage than was originally assumed, which can lead to biased stock assessments. We are therefore interested in analyzing the impact of climate change on the accuracy of future stock assessment models as measured by NOAA's ongoing bottom-trawl survey along the East coast.

use more info from initial proposal

- Fish are changing spatial distribution and have altered life stages (?) because of climate change NYE paper
- Population indexing methods may be becoming biased as a result
- Briefly describe our study to test this

To test the ability of the bottom trawl survey to track population trends under shifting environmental conditions, we construct spatial models for fish where movement depend on temperature preferences. We can then consider the impact of climate change by simulating scenarios with repeating temperature patterns and those where temperature increases on average over time. In both cases we analyze the ability of stratified random sampling to track population trends.

Methods

- Describe simulation study

We construct spatial models for Yellowtail Flounder, Atlantic Cod, and Haddock on George's Bank, where movement of each species combine static species-specific habitat preferences with dynamic temperature preferences. Model dynamics are driven by dynamic temperature gradients estimated from data to create simulated data sets for each population where the true biomass is known. Using temperature gradients that repeat each year creates data sets with predictable, repeating spatial patterns, whereas using a temperature gradient that increases on average throughout the simulation leads to spatial distributions that shift over time. We conducting stratified random sampling on our simulation output to mimic the bottom trawl survey and compare the ability of contemporary indexing methods to track population trends.

Population Model Formulation

– Used MixFishSim. Describe edits made to package

We use the R package *MixFishSim* (MFS) to model our populations [@dolder2020highly]. MFS is a discrete spatiotemporal simulation tool where users can model multiple species under varying environmental conditions. The package uses a delay-difference population model with discrete processes for growth, death, and recruitment of the population. We formulate the following inputs for the MFS package to address our research question:

Study Area

We obtained a shapefile for the 15 strata that comprise George's Bank to use as our modeling environment. We discritized the region into a raster with 88 rows and 144 columns. Haddock inhabit all 15 strata in the domain Cod inhabit 13 strata, and yellowtail exist in 9 strata. Figure 1 shows the regions used in our models.

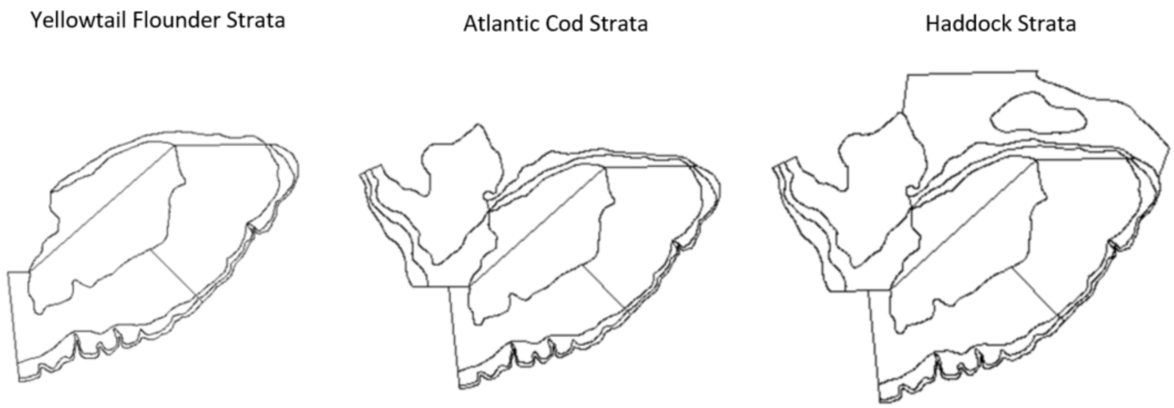

Figure 1: Strata inhabited by each species in our population models.

Population Dynamics and Recruitment

The time step for our models is one week. MFS uses a modified two-stage Deriso-Schnute delay difference equation that models the biomass of in each cell in our study area [@dolder2020highly]. Individual terms in the formulation account for growth of mature adults, natural and fishing mortality, and the addition of new recruits. We chose to represent recruitment in the model using a Beverton-Holt formulation. Recruitment is a function of the adult biomass that existed in the previous year and is added to the population incrementally throughout each species' predefined spawning period. Parameter inputs were either obtained from the literature or chosen to produce desired model dynamics. A full list of parameters used in our model can be seen below in Tables **??** and 5. **missing spawning weeks, recruitment weeks, lambda**

Movement

The package was designed to generate theoretical habitat preferences using Gaussian Random Fields that combine with hypothetical temperature gradients to drive the probability of movement from cell I to cell J using the formulation

$$Pr(C_{wk+1} = J|C_{wk} = I) = \frac{e^{-\lambda \cdot d_{I,J}} \cdot (Hab^2_{J,p} \cdot Tol_{J,p,wk})}{\sum_{c=1}^{C} e^{-\lambda \cdot d} \cdot (Hab^2_{c,p} \cdot Tol_{c,p,wk})}, \tag{1}$$

where

$e^{-\lambda \cdot d_{I,J}}$ accounts for distance between cells I and J,

$Hab^2_{J,p}$ is the static habitat value for species p in cell J, and

$Tol_{c,p,wk}$ is the value from normally distributed temperature tolerance for species p in cell c in week wk.

Since we are modeling real species on the northeast continental shelf, we formulate the habitat and temperature components as follows.

Habitat Input

Species-specific habitat preferences were derived using the *lrren* tool from the R package *envi* **cite** to create a niche model for each species. The *lrren* tool estimates an ecological niche using the relative risk function by relating presence/absence data to two covariate predictors. We used bottom trawl point data from 2009-2021 as our presence/absence input by using a value of 0 for any tow that failed to catch the given species and weighting a successful catch by the biomass of the given tow **cite trawl data?**. Depth and mean sediment size were used as our covariate predictors **cite**. Since the values in $Hab^2_{J,p}$ are required to be between 0 and 1, we transform the spatial estimates from *lrren* to fall between these bounds. See Figure ??? for a visual representation of this process being applied to Cod. Figure 2 depicts habitat preferences $Hab^2_{J,p}$ for each species.

Temperature Input

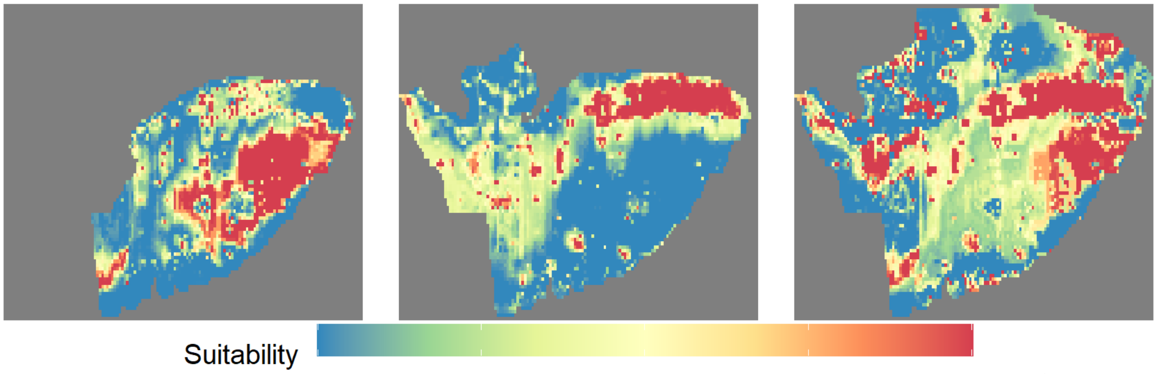

Figure 2: Static habitat preferences for each species in our population models (Yellotwtail, Cod, Haddock).

Each species is assumed to have normally distributed temperature preferences ($N(\mu, \sigma)$). We assume Yellowtail Flounder's preferences are $N(8.75, 4.25)$, while Haddock and Cod have preferences $N(9, 4)$. We chose these values by combining information in the literature with temperatures recorded in the bottom trawl survey. Weekly estimated temperature data for the region for 2012 was obtained from FVCOM **cite**. We chose 2012 because the data displayed an average temperature pattern that consistently oscillated between maximum and minimum temperature values. This data was also transform to create an oscillating pattern that increases 5 degrees Celsius on average over the duration of the simulation. **show images of temperature and/or and/or temp videos??? and/or average temperature oscillations?**

–Describe difference between increasing and constant temperature scenarios (images?)

In equation (1), $Hab^2_{J,p}$ is constant for the duration of the simulation, while $Tol_{c,p,wk}$ changes each week. Using a temperature gradient that repeats each year produces the same spatial preferences in a given week each year, which results in consistent spatial biomass patterns. Scenarios where the temperature increases over time creates spatial preferences that evolve as the water warms, which creates spatial biomass patterns that shift in a given week over the duration of the simulation.

– Describe each scenario that is considered

We consider three population parameter scenarios for each of our three species- a scenario where parameters result in each population increasing over time, one where the populations are relatively constant over time, and a scenario where the parameter combination results in each population decreasing over time. Each of these three scenarios is paired with a temperature gradient that repeats as well as one that increasing roughly 5 degrees Celsius over the duration of the 20 year simulation.

Simulating Bottom Trawl Survey and Population Indexing

-Describe post hoc sampling process and how data is used

After each simulation is complete, we mimic the bottom trawl survey by conducting stratified random sampling in each inhabited strata twice each year. We sample in the same weeks that the Spring and Fall surveys take place and the number of the samples taken in each strata reflect true values. Most strata contain enough cells to sample a unique location in each survey over the duration of the simulation. For smaller strata we must repeat some sample locations. We then use the biomass collected from our samples in contemporary population indexing methods to estimate population trends. Knowing the true population values in our simulations allows us to compare the error calculated from each estimation method.

–Stratified mean vs VAST with and without covariates

We compare the stratified mean estimate of abundance to estimates obtained from the Vector-Autoregressive Spatio-temporal (VAST) model. The stratified mean is a typical survey-based approach that scales individual samples to the strata-level by considering the area of each strata, before scaling to the region-level based on the relative size of each strata. VAST is a spatio-temporal statistical framework that models both abundance (biomass) and probability of occurrence (presence/absence). If desired, VAST also allows users to include covariate data to better inform the model.

Results

Discussion

A range of data-limited methods for setting catch advice were evaluated for stocks where assessment models may be rejected due to strong, positive retrospective patterns. A method was considered to perform well if it limited overfishing without resulting in light exploitation rates ($F << F_{MSY}$), thereby allowing depleted stocks to recover to SSB_{MSY} (or for healthy stocks to remain there), and for high and stable catches (close to MSY).

Overall, none of the methods evaluated performed best across the scenarios exploring the different sources of the retrospective pattern (unreported catch or increasing M) and different levels of historical fishing intensity. A number of methods did perform well in many cases, however, while others performed consistently poorly, resulting in frequent and intense overfishing ($F >> F_{MSY}$). We performed simulations for a couple of scenarios with no source of retrospective patterns and found the expected result that all DLMs and the SCAA performed better (SSB, F, and catch were all closer to the MSY reference points) than when either source of retrospective patterns was present. Due to the focus of this study, we did not examine the no retrospective source in detail and do not comment on it further.

Currently, in the Northeast U.S., if an assessment model is rejected due to a large rho value in SSB, the catch advice from that model is ignored and some data-limited approach is used. However, the rho-adjusted SCAA model performed better than a number of the alternatives explored here. Therefore, there should not necessarily be an expectation that a data-limited method will perform better than the rejected assessment model. The SCAA only resulted in high exploitation rates ($F >> F_{MSY}$) when unreported catch was the source of the retrospective pattern and for the scenario where $F = F_{MSY}$ at the end of the base period that left the stock in relatively good condition ($SSB \sim SSB_{MSY}$). In contrast, this method was particularly effective when the stock was depleted and there was unreported

Table 1: Yellowtail error results

X	X.1	X.2	X.3	X.4	Constant.Population	X.5	Increasing.P
Temp	Covariate	Strata	Noise	season	Stratified Mean	VAST Estimate	Stratified M
const	no cov	all	no	spring	0.21	0.11	0.16
const	no cov	all	yes	spring	0.25	0.16	0.22
const	w/ cov	all	no	spring	0.21	0.07	0.16
const	w/ cov	all	yes	spring	0.25	0.08	0.22
const	no cov	all	no	fall	0.32	0.68	0.34
const	no cov	all	yes	fall	0.31	0.77	0.46
const	w/ cov	all	no	fall	0.32	0.08	0.34
const	w/ cov	all	yes	fall	0.31	0.11	0.46
const	no cov	reduced	no	spring	0.27	0.19	0.2
const	no cov	reduced	yes	spring	0.26	0.15	0.22
const	w/ cov	reduced	no	spring	0.27	0.19	0.2
const	w/ cov	reduced	yes	spring	0.26	0.14	0.22
const	no cov	reduced	no	fall	0.47	0.25	0.41
const	no cov	reduced	yes	fall	0.49	0.36	0.46
const	w/ cov	reduced	no	fall	0.47	0.19	0.41
const	w/ cov	reduced	yes	fall	0.49	0.17	0.46
increasing	no cov	all	no	spring	0.28	0.11	0.32
increasing	no cov	all	yes	spring	0.28	0.15	0.34
increasing	w/ cov	all	no	spring	0.28	0.06	0.32
increasing	w/ cov	all	yes	spring	0.28	0.12	0.34
increasing	no cov	all	no	fall	0.51	1.26	0.3
increasing	no cov	all	yes	fall	0.5	1.38	0.39
increasing	w/ cov	all	no	fall	0.51	0.23	0.3
increasing	w/ cov	all	yes	fall	0.5	0.28	0.3
increasing	no cov	reduced	no	spring	0.31	0.17	0.4
increasing	no cov	reduced	yes	spring	0.29	0.19	0.41
increasing	w/ cov	reduced	no	spring	0.31	0.17	0.4
increasing	w/ cov	reduced	yes	spring	0.29	0.15	0.41
increasing	no cov	reduced	no	fall	0.64	0.75	0.7
increasing	no cov	reduced	yes	fall	0.66	0.89	0.69
increasing	w/ cov	reduced	no	fall	0.64	0.2	0.7
increasing	w/ cov	reduced	yes	fall	0.66	0.13	0.69

Table 2: Cod error results

Temp	Strata	Noise	season	VAST.No.Cov	VAST.w..Cov	Stratified.Mean	X	X.1
const	all	no	spring	0.11	0.12	0.36	NA	
const	all	yes	spring	0.12	0.15	0.35	NA	Cod
const	all	no	fall	0.19	0.05	0.49	NA	Decreasin
const	all	yes	fall	0.30	0.23	0.41	NA	
const	reduced	no	spring	0.17	0.24	0.41	NA	
const	reduced	yes	spring	0.20	0.23	0.46	NA	
const	reduced	no	fall	0.21	0.33	0.60	NA	
const	reduced	yes	fall	0.18	0.31	0.58	NA	
increasing	all	no	spring	0.12	0.15	0.25	NA	
increasing	all	yes	spring	0.19	0.19	0.27	NA	
increasing	all	no	fall	0.76	0.13	0.45	NA	
increasing	all	yes	fall	0.89	0.33	0.44	NA	
increasing	reduced	no	spring	0.14	0.22	0.32	NA	
increasing	reduced	yes	spring	0.15	0.21	0.29	NA	
increasing	reduced	no	fall	0.60	0.31	0.54	NA	
increasing	reduced	yes	fall	0.62	0.32	0.53	NA	

Table 3: Haddock error results

Temp	Strata	Noise	season	VAST.No.Cov	VAST.w..Cov	Stratified.Mean	X	X.1
const	all	no	spring	0.49	0.18	0.18	NA	
const	all	yes	spring	0.73	0.43	0.21	NA	Haddock
const	all	no	fall	0.28	0.05	0.26	NA	Increasin
const	all	yes	fall	0.41	0.06	0.27	NA	
const	reduced	no	spring	0.34	0.35	0.45	NA	
const	reduced	yes	spring	0.30	0.33	0.46	NA	
const	reduced	no	fall	0.36	0.48	0.54	NA	
const	reduced	yes	fall	0.33	0.46	0.52	NA	
increasing	all	no	spring	0.25	0.05	0.26	NA	
increasing	all	yes	spring	0.30	0.06	0.31	NA	
increasing	all	no	fall	0.89	0.23	0.40	NA	
increasing	all	yes	fall	1.04	0.35	0.42	NA	
increasing	reduced	no	spring	0.32	0.40	0.44	NA	
increasing	reduced	yes	spring	0.38	0.37	0.37	NA	
increasing	reduced	no	fall	0.44	0.64	0.72	NA	
increasing	reduced	yes	fall	0.42	0.62	0.70	NA	

catch. When M was the source of the retrospective pattern, the rho-adjusted SCAA method typically resulted in light exploitation rates, on average. The light exploitation rates in these cases were likely driven by the combination of using a rho-adjustment, but also using the lower M from the beginning of the base period rather than the higher M that occurred during the feedback period. Using an M value that is too low in a stock assessment will typically bias estimates of biomass and reference points too low, resulting in catch advice that is below target levels [@Johnsonetal2014; @Puntetal2021M]. The consequences of using a value for M that is too low versus too high is also asymmetrical [@Johnsonetal2014], with negative consequences being more severe when M is assumed too high than low, and the results here are consistent with these previous conclusions.

The methods that adjusted recent average catches based on trends in the survey (Ismooth and Islope) performed well overall in terms of catch, stock status, and variation in catch. The method using the expanded survey biomass with the recent exploitation rate (ES-Frecent) also performed well and similarly to Ismooth. The performance of these methods was also generally robust among scenarios, with the exception of when there were unreported catches and the stock was depleted (see below). The generally positive performance of these methods was consistent with @Hilbornetal2002 and @CoxKronlund2008, both of which evaluated a variant of a "hold-steady" DLM. In the case of @Hilbornetal2002, the "hold-steady" DLM policy was designed to adjust catches in order to keep rockfish (*Sebastes spp.*) populations at recently observed index levels, and did so by functioning as a constant escapement harvest control rule where target catches were set to zero below some pre-specified index level. In the variant used by @CoxKronlund2008, catches were adjusted to maintain a sablefish (*Anoplopoma fimbria*) population at a pre-specified index level thought to be sustainable and desirable in terms of meeting fishery objectives (e.g., high catch), but never permitted target catches of zero and so functioned as a constant exploitation rate control rule. The "hold-steady" DLM of @CoxKronlund2008 performed similarly in terms of catch, stock depletion, and variation in catch, as a constant exploitation rate policy where target catch was specified as the product of desired exploitation rate and an estimate of biomass from a SCAA model. This result was robust to uncertainty in initial stock status and steepness [@CoxKronlund2008]. The SCAA model was always correctly specified (i.e., expected to produce unbiased estimates on average), however, and no comparison to the results of this research in the presence of retrospective patterns is possible [@CoxKronlund2008]. The "hold-steady" policy of @Hilbornetal2002 performed similarly to or better in terms of catch and stock status than other harvest control rules that relied on assessment estimates of biomass (i.e., 40:10 and constant F). The performance of the "hold-steady" DLM was also more robust to uncertainty in steepness and to the presence of unreported catch [@Hilbornetal2002]. The performance of the two harvest policies that relied on assessment estimates of biomass (i.e., constant exploitation rate and a "40:10" biomass-based policy) also degraded when the estimates of biomass were biased, which is an issue that does not effect the "hold-steady" DLM [@Hilbornetal2002]. The bias in the assessment estimates considered in @Hilbornetal2002 were not necessarily induced by a retrospective pattern, however, and no consideration of making a rho-adjustment was possible in that study.

The Ismooth method is currently used to set catches for Georges Bank cod [@nefsc19] and red hake (*Urophycis chuss*; @nefsc20). Variations of the ES-Frecent have been used for witch flounder and Georges Bank yellowtail flounder. While the findings here generally support the continued use of the Ismooth and ES-Frecent methods, they may not be well suited for depleted stocks where unreported catches are believed to be an issue. The Ismooth, Islope, and ES-Frecent DLMs produced high Fs and limited stock recovery with unreported catches and when the stock was depleted. While @Hilbornetal2002 and @CoxKronlund2008 did not reach the same conclusion about the "hold-steady" DLM, those studies did not consider initial levels of depletion as low as in this study. These results highlight the importance of accurate catch reporting, as unreported catch can create a negative feedback loop with perpetually high Fs being produced by a management system that seemingly should result in sustainable catch advice.

Three methods were consistently risk-averse across scenarios, limiting the frequency and magnitude of overfishing and resulting in high stock biomass. These methods were the two catch curve options (CC-FM and CC-FSPR) and DynLin. The catch curve methods produced a wider range of average catches across scenarios, and also had greater interannual variability in catches compared to DynLin. While the lower exploitation rates from these approaches may be undesirable due to forgone yield, there may be circumstances where they are preferred. For example, for stocks that are believed to be heavily depleted, low exploitation rates would allow for a more rapid recovery.

A number of methods performed poorly, particularly when catches were unreported. These methods include three of the expanded survey biomass approaches (ES-Fstable, ES-FM, ES-FSPR), AIM, and Skate. The AIM model has been widely used across stocks in the region [@nefsc02a; @nefsc05; @nefsc08], although there is a decreasing trend in its use across model resistant stocks [@nefsc19]. The findings here suggest that alternative approaches should be considered in cases where AIM is still used and there is concern over unreported catches. The Skate method is used to manage the skate complex in the Northeast U.S. (a group of seven co-managed species). Interestingly, six of the seven species are considered in good condition with high survey biomass indices in recent years [@nefmc20]. That the Skate method performed poorly in our analysis but performs well for the skate complex illustrates how the performance of methods in this analysis may be sensitive to the scenarios and species life history considered. As may be the case for the Skate method, the performance of some methods may depend on the condition of the stock when the method is first applied, and less so on life-history. Therefore, care is needed when trying to generalize these results across stocks that may have different life histories, exploitation histories, and without unreported catches or increases in M.

In addition to the analytical differences among the thirteen DLMs, most of the DLMs and control rules had multiple options that could be adjusted to make them more or less risk averse. DynLin had a large number of user defined decision points. Given the large range of options already explored in the study, one suite of options was selected for each DLM-control rule and kept constant for all simulations. Further studies could explore the different options within an individual DLM to understand how they might affect performance.

Many other data-limited methods exist for setting catch advice that were not included in this evaluation, and they vary widely in complexity, data inputs, and assumptions required [e.g., @carruthers2018dlm]. Length based methods were not evaluated to keep the overall number of methods tractable, and due to the availability of age based information in the region. Methods that require only catch data or snap shots of survey data were not considered due to the availability of the relatively long and contiguous Northeast Fisheries Science Center's spring and fall, coastwide bottom trawl surveys, and the fact that "catch only" methods have been shown to perform poorly [e.g., @carruthers2014eval]. Complete catch histories are not available for stocks in the region (i.e., from the inception of fishing). Consequently, methods that required complete catch histories or required assumptions about relative depletion [e.g., DCAC in @maccall2009dca; DB-SRA in @dick2022dsra] were also omitted from consideration. The need for short run-times and the desire for methods that could be reviewed quickly prevented the use of modern state-space production models such as SPiCT [@pedersen2017spict] and JABBA [@winker2018jabba].

The SCAA was confronted with inconsistent data in this study, while the DLMs typically used only a single source of data and thus did not encounter inconsistencies. A recent examination of the data used in assessments in this region similarly found inconsistencies in data streams even before modeling. @wiedenmann2022strange found a negative relationship between relative F (catch/survey) and survey Z for stocks with strong retrospective patterns but the expected positive relationship for stocks without a retrospective pattern. It is exactly this sort of tension that creates retrospective patterns in integrated models, but is not found in DLMs that only use one type of data.

Despite conducting hundreds of thousands of simulations, there are still limitations to our study. We only examined one life history representative of groundfish in the region. We acknowledge that best practice is to select a DLM for a specific life history and fishery condition [e.g., @fischer2020dlm]. As is typically the case with large simulation studies, we were not able to tune any of the DLMs or the SCAA in any given realization, which would occur in practice for an actual stock assessment. We also examined only scenarios that started with Mohn's rho values near 0.5 for spawning stock biomass. This is a strong retrospective pattern, but some stocks in the region have even stronger retrospectives. Performance of the DLMs and SCAA would be expected to degrade with stronger retrospectives, but by how much is still an open area for research. Similarly, sources of retrospective patterns that create different relationships between the true values and estimated values should also be explored [see @deroba2014retro]. To make the results interpretable, we only examined a single source for the retrospective pattern at a time. In reality, there may be more than one factor leading to an observed retrospective pattern. How the multiple sources would interact to influence performance is another topic for future research. Development of harvest control rules specifically for situations where retrospective patterns are found in age-based assessments would also be beneficial. The large number of scenarios examined and the large number of realizations gives us confidence that our results are meaningful in general, but that the performance of any of the DLMs may differ in actual practice.

An interesting finding of this study is the linear versus diffuse patterns between SSB and catch across methods. These patterns have implications for the trade-offs among methods, with linear relationships resulting in more consistent exploitation rates across stock sizes. Therefore, these methods have higher certainty of a given catch at a given stock size. However, they also tended to result in lower stock sizes, on average, across methods. The more diffuse relationships resulted in more variable exploitation rates across stock sizes, with some situations where the population biomass was quite high but the catch was low (relative to MSY), resulting in a very low F. The reasons behind these different patterns remain unclear, and future work to explore these patterns is warranted.

One of the reasons for the difference in performance between the catch and natural mortality retrospective sources was how the reference points were calculated. In all cases, the initial conditions, including the natural mortality rate, were used to compute the reference points. This decision was made based on the fact that the increase in natural mortality was assumed to be unknown in the simulations. If the increase in natural mortality was known, the age-structured assessments would have accounted for it, different reference points might have been computed [@legault2016increaseM] and there may not have been a retrospective pattern at all [@legault2020rose], and no need to consider alternative DLMs. The reference points for the increased M scenarios would have been different if they were computed using the values from the final year of the base period, but the overall conclusions regarding the different DLMs would not change as this just results in a rescaling of the axis. These results are not shown to reduce confusion regarding the simulations.

Closed-loop simulation is a common tool for examining performance of catch advice from various stock assessment approaches in a feedback setting. It is often used as part of a full management strategy evaluation when working with stakeholders to develop management regulations that make trade offs between near term and long term catches, risk to the fish population, and mixed-fleet allocations [@carruthers2016simpleMPs; @goethel2019mse; @harlyan2019hcr]. We did not conduct a full management strategy evaluation with stakeholder input [@goethel2019stakeholder], but see that as a fruitful next step that could build on the conclusions from our closed-loop work. Using a generic groundfish life-history and monitoring standard performance metrics related to stock status and catch stability, we were able to cull the herd of potential DLMs and we would not carry the consistent poor performers forward for further study. The wide range of expertise reflected in the authorship was by design so that the simulation specifications and performance metrics were broadly useful. Before undertaking a full management strategy evaluation and engaging regional stakeholders, we would want to select a specific stock and jointly identify specific management regulations to be tested [@deroba2019dream]. Results of this work have been presented to both local fishery management councils, with generally positive feedback about the utility of the conclusions for identifying appropriate model approaches when an SCAA is rejected. Our work was similar to all other closed-loop simulations in that it was designed to address a specific situation, including much recent work comparing the performance of data-limited and data rich assessment approaches [e.g., @fulton2016datarich; @sagarese2019dlm; @bouch2020datapoor; @li2022dlm].

This study is a first attempt to identify suitable methods for setting catch advice when stock assessment models are rejected due to large, positive retrospective patterns. Although no single method performed best across scenarios, a number of generally suitable and unsuitable methods were identified under specific conditions. The results of this work can help scientists and managers select a subset of possible options for consideration to set catch advice when assessment models are rejected. The approach developed here can, and should be expanded to consider other cases not explored here, as performance of individual methods are very likely case-dependent.

Acknowledgements

We thank the Index-Based Methods and Control Rules Research Track review panel of Paul Rago (chair), Yong Chen, Robin Cook, and Paul Medley for feedback on preliminary results, three anonymous reviewers and the associate editor for reviewing an earlier version of this work. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA or the Department of Commerce.

Data and Code Availability

All data and code used in this work are available at https://github.com/cmlegault/IBMWG.

References

Tables

Table 1. Parameters used in all population models.

	Description	Unit	Yellowtail	Cod	Haddock	Source
ρ	Ford's growth coefficient	wk^{-1}	4.48	4.43	4.49	
M	Natural Mortality	wk^{-1}	0.2064	0.2728	0.3340	
W_R	Weight of fully recruited fish	kg	0.39	2.95	1.12	
W_{R-1}	Weight of pre-recruit fish	kg	0.13	0.39	0.19	
σ^2	Variance in recruited fish	kg^2	0.55	0.55	0.55	

```
## Warning: ‘funs()‘ was deprecated in dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with ‘tibble::lst()‘:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was generated.
```

Table 1. Maturity-, weight-, and selectivity-at-age of the simulated fish population.

Table 5: Parameters used in population models for each scenario.

Parameter	Description	Unit	Yellowtail	Cod	Haddock
Constant Population					
M+F	Adjusted Mortality (Natural + Fishing)	1/wk	0.764	0.83	0.309
P0	Initial Biomass	kg	3190	21500	180000
a	Max recruitment rate	kg	30400	27900	73600
ß	Recruitment half saturation value	kg	4300	10500	40500
Decreasing Population					
M+F	Adjusted Mortality (Natural + Fishing)	1/wk	0.764	0.623	0.334
P0	Initial Biomass	kg	50000	21500	180000
a	Max recruitment rate	kg	1.07e+12	3.89e+08	4.97e+08
ß	Recruitment half saturation value	kg	2.3e+12	9.8e+08	2.08e+09
Increasing Population					
M+F	Adjusted Mortality (Natural + Fishing)	1/wk	0.564	0.372	0.134
P0	Initial Biomass	kg	3190	21500	180000
a	Max recruitment rate	kg	40000	45000	1e+05
ß	Recruitment half saturation value	kg	43000	62800	405000

Age	Maturity	Weight (kg)	Fishery Selectivity (before change if applicable)	Fishery Selectivity (after change if applicable)
1	0.04	0.15	0.07	0.02
2	0.25	0.5	0.17	0.05
3	0.60	0.9	0.36	0.12
4	0.77	1.4	0.61	0.27
5	0.85	2.0	0.81	0.50
6	0.92	2.6	0.92	0.74
7	1.00	3.2	0.97	0.89
8	1.00	4.1	0.99	0.96
9	1.00	5.9	1.00	0.99
10+	1.00	9.0	1.00	1.00

Table 2. Naming convention and details of the data-limited methods evaluated.

Method	Details
Ismooth	$C_{targ,y+1:y+2} = \overline{C}_{3,y}(e^{\lambda})$ where $\overline{C}_{3,y}$ is the most recent three year average; $\overline{C}_{3,y} = \frac{1}{3}\sum_{t=1}^{t=3} C_{y-t}$ and λ is the slope of a log linear regression of a LOESS-smoothed average index of abundance (spring and fall) with span = 0.3: $\hat{I}_y = loess(\hat{I}_y)$ and $LN(\widehat{I_y}) = b + \lambda y$
Islope	$C_{targ,y+1:y+2} = 0.8\overline{C}_{5,y}(1 + 0.4e^{\lambda})$ where $\overline{C}_{5,y}$ is the most recent five-year average catch through year $y-1$: $\overline{C}_{5,y} = \frac{1}{5}\sum_{t=1}^{t=5} C_{y-t}$ and λ is the slope of a log-linear regression of the most recent five years of the averaged index.
Itarget	$C_{targ,y+1:y+2} = \left[0.5C_{ref}\left(\frac{\overline{I}_{5,y} - I_{thresh}}{I_{target} - I_{thresh}}\right)\right]$ $\overline{I}_{5,y} \geq I_{thresh}$ $C_{targ,y+1:y+2} = \left[0.5C_{ref}\left(\frac{\overline{I}_{5,y}}{I_{thresh}}\right)^2\right]$ $\overline{I}_{5,y} < I_{thresh}$; C_{ref} is the average catch over the reference period (years 26 through 50): $C_{ref} = \frac{1}{25}\sum_{y=26}^{y=50} C_y$; I_{target} is 1.5 times the average index over the reference period: $I_{target} = \frac{1}{25}\sum_{y=26}^{y=50} \overline{I}_y$; $I_{thresh} = 0.8\ I_{target}$, and is the most recent five year average of the combined spring and fall index: $\overline{I}_{5,y} = \frac{1}{5}\sum_{t=1}^{t=5} \overline{I}_{y-t+1}$
Skate	$C_{targ,y+1:y+2} = F_{rel}\overline{I}_{3,y}$ where $F_{rel} = median\left(\frac{\overline{C}_{3,\mathbf{Y}}}{\overline{I}_{3,\mathbf{Y}}}\right)$ is the median relative fishing mortality rate calculated using a 3 year moving average of the catch and average survey index across all available years ($\mathbf{Y}$): $\overline{C}_{3,y} = \frac{1}{3}\sum_{t=1}^{t=3} C_{y-t}$ and $\overline{I}_{3,y} = \frac{1}{3}\sum_{t=1}^{t=3} I_{y-t+1}$

Method	Details
An Index Method (AIM)	AIM first calculates the annual relative F: $F_{rel,y} = \frac{C_y}{\frac{1}{3}\sum_{t=1}^{t=3}\bar{I}_{y-t+1}}$ and the annual replacement ratio: $\Psi_y = \frac{\bar{I}_y}{\frac{1}{5}\sum_{t=1}^{t=5}\bar{I}_{y-t}}$. These values are used in a regression: $LN(\Psi_y) = b + \lambda LN(F_{rel,y})$ to determine $F_{rel,*}$, which is the value of $F_{rel,y}$ where the predicted $\Psi = 1$ or $LN(\Psi) = 0$. $F_{rel,*}$ is called either the "stable" or "replacement" F, and is used to calculate the target catch: $C_{targ,y+1:y+2} = \bar{I}_y F_{rel,*}$.
Dynamic Linear Model (DynLin)	@Langan2021DLM.
Expanded survey biomass method 1 $F_{40\%}$ (ES-FSPR)	$C_{targ,y+1:y+2} = B_{\bar{I},y}\mu_{targ}$ where $B_{\bar{I}}$ is the average of estimated fully-selected biomass from each survey: $B_{\bar{I},y} = \frac{1}{2}\left(\frac{I_{spr,y}}{q_{spr}} + \frac{I_{fall,y-1}}{q_{fall}}\right)$ and target exploitation fraction, μ_{targ} is calculated as: $\mu_{targ} = \frac{F_{targ}}{Z_{targ}}\left(1 - e^{-Z_{targ}}\right)$; $F_{targ} = F_{40\%}$ and $Z_{targ} = F_{targ} + M$
Expanded survey biomass method 2 $F =$ AIM replacement (ES-Fstable)	Same as the above expanded survey method, but with μ_{targ} equal to the stable exploitation fraction $F_{rel,*}$ calculated using the AIM approach (see above).
Expanded survey biomass method 3 $F = M$ (ES-FM)	Same as the above expanded survey methods, but with the target exploitation rate set to the assumed M: $F_{targ} = M$.
Expanded survey biomass method 4 $F =$ recent average (ES-Frecent)	Same as the above expanded survey methods, but with the target exploitation fraction set to the most recent three year average exploitation fraction: $\mu_{targ} = \frac{\sum_{y-2}^{y}\mu_y}{3}$ $\mu_y = \frac{C_{y-1}}{B_{\bar{I},y}}$

Method	Details
Catch curve Method 1 $F_{40\%}$ (CC-FSPR)	$C_{targ,y+1:y+2} = \frac{F_{targ}}{Z_{avg,y}} B_{cc,y} \left(1 - e^{-Z_{avg,y}}\right)$ where B_{cc} is the estimated biomass: $B_{cc,y} = \frac{C_{y-1}}{\frac{F_{avg,y}}{Z_{avg,y}}\left(1-e^{-Z_{avg,y}}\right)}$ with $Z_{avg,y} = \frac{Z_{spring,y}+Z_{fall,y-1}}{2}$; $F_{avg,y-1} = Z_{avg,y-1} - M$ and, $F_{targ} = F_{40\%}$. Survey catch at age used in catch curve to estimate Z.
Catch curve Method 2 M (CC-FM)	Same as catch curve method 1 above, but with $F_{targ} = M$.
Ensemble	Median of catch advice provided by AIM, CC-FSPR, ES-Frecent, ES-FSPR, Islope, Itarget, Ismooth, and Skate methods.

Table 3. Summary of the scenarios evaluated within the study design.

Factors	Variants
retrospective source	catch or natural mortality
fishing history	F_{MSY} in second half of base period or overfishing throughout base period (2.5xF_{MSY})
fishery selectivity blocks	constant selectivity or selectivity changes in second half of base period
catch advice multiplier	applied as is from DLM (1) or reduced from DLM (0.75)

List of Figures

Figure 1. Inner quartiles and medians for all performance measures across all scenarios and runs for each method. Vertical lines are shown at a value of 1 for the performance measures that are relative to the MSY reference points (A,B,C).

Figure 2. Relationship between long-term average spawning biomass and average catch (relative to MSY levels) for each method. Each point represents the median for a given scenario, separated by the source of the retrospective pattern (catch or M).

Figure 3. Median performance measures for each method, separated by the source of the retrospective error (catch = black, M = gray) and the exploitation history in the base period (always overfishing at $2.5xF_{MSY}$ (circle), or F reduced to F_{MSY} during base period (triangle)). Vertical lines are shown at a value of 1 for the performance measures that are relative to the MSY reference points (A,B,C).

Figure 4. Median F/F_{MSY} for each method, with results separated by the exploitation history in the base period (always overfishing at $2.5xF_{MSY}$ (circle), or F reduced to F_{MSY} during base period (triangle)) showing A) short- (gray) versus long-term (black) values, and B) with (black) or without (gray) a buffer applied when setting the catch (catch multiplier = 0.75 or 1).

Figure 5. Relationship between long-term average catch and spawning stock biomass relative to their reference points by method. Each point represents the average for years 21-40 in the feedback period for a single iteration of a scenario. The scenario shown is where catch was the source of the retrospective pattern with F reduced to F_{MSY} in the second half of the base period, there was a single selectivity block, and where no buffer was applied to the catch advice (catch multiplier = 1).