Title: Something strange in the neighborhood: Diverging signals in stock assessment data for Northeast U.S. fish stocks.

Abstract

In the Northeast U.S., many stock assessments have a history of problematic model diagnostics, with multiple age-based assessments recently being rejected in the peer review process, and are not suitable for management advice. The role in which diverging signals in the coastwide bottom trawl survey may be contributing to assessment problems was explored here for 18 stocks in the region. Specifically, trends in total mortality (*Z*) estimated from catch curve analysis and a relative measure of the harvest rate (total catch / survey index; called relative *F*) were evaluated. Across stocks, relative *F* has declined over time, on average, since the mid 1990s, yet *Z* has not for many stocks. Weak positive or even negative correlations between relative *F* and *Z* resulted for 13 stocks. This diverging signal appears to be contributing to assessment model performance, as larger retrospective patterns (a measure of assessment uncertainty) occurred for stocks with negative correlations between relative *F* and *Z*. While a variety of mechanisms could be involved in these diverging signals, the available evidence suggests that unreported catch and / or increasing natural mortality likely play a role to varying degrees for each stock.

Keywords: stock assessment, retrospective pattern, groundfish, climate change, predation, misreported catch

Introduction

Modern fisheries management typically relies on data-intensive stock assessment models to estimate current population size and management reference points. Age-based assessment models are commonly used when possible to account for interannual variability in cohort strength and growth, as well as size–selective removals by the fishery. Such models are fit to various data sources, including indices of abundance that provide trends in relative abundance over time, and catch-at-age data that provides information on the magnitude of different cohorts and on how quickly those cohorts disappear. A number of assumptions must be made within the assessment model to account for unobserved processes such as rate of natural mortality (M), but also to link the model to the data (e.g., assuming the index of abundance is proportional to overall abundance). Problems can arise when model assumptions are violated or when there are diverging signals in the data, which can lead to biased model estimates and the subsequent over- or under-exploitation of the resource (Kraak et al., 2009; Maunder and Piner, 2015; Van Beveren et al., 2017).

In the Northeast U.S., there exists a considerable amount of data for inclusion in stock assessments. The Northeast Fisheries Science Center (NEFSC) has conducted a biannual bottom trawl survey in the spring and fall each year since 1968 and 1963, respectively, covering the continental shelf from Cape Hatteras, NC to the Gulf of Maine (Reid et al., 1999). Additional surveys of inshore waters are operated annually by many of the states in the region. These surveys, combined with the coastwide NEFSC surveys provide fishery-independent information on trends in relative abundance. It is therefore not surprising that of the eight Regional Management Councils in the U.S., the New

England and Mid-Atlantic Fishery Management Councils (NEFMC and MAFMC, respectively) that manage Northeast U.S. fish stocks in federal waters have the highest proportion of catch limits based on data-rich methods (Newman et al., 2015).

Despite the wealth of information in the region, many age-based assessments have a history of problematic model diagnostics, including temporal patterns in the residuals of model fits to the data, as well as strong "retrospective patterns" (e.g., NEFSC, 2008). A retrospective pattern occurs when an assessment model consistently over- or underestimates historical population abundance (and other quantities) with every additional year of data added to the model and is indicative of some inconsistencies between the model assumptions and signals in the data, or of divergences among data sources within the model (Mohn, 1999). With a positive retrospective pattern in abundance (typically total or spawning biomass), historical estimates of abundance are overestimated (i.e., positively biased), and get revised downward as new information is added to the model. In negative retrospective patterns, historical estimates of abundance are underestimated, and get revised upward with new information in the model. Throughout this paper we refer to the sign of retrospective patterns in abundance only. In the Northeast U.S., many stocks have a history of positive retrospective patterns (NEFSC, 2002; 2005; 2008; Deroba et al., 2010; Brooks and Legault, 2016). In some cases, this retrospective bias has been considerable, as it was found that across fifteen stocks in the New England multispecies groundfish complex, nine had an average positive retrospective bias greater than 50% across assessments since 2002, with three of those having a bias greater than 100% across repeated assessments (Wiedenmann and Jensen, 2018).

It has been shown through simulation studies that retrospective patterns are not necessarily reflective of the direction of estimation error (Legault, 2009; Hurtado et al., 2015), but strong patterns, positive or negative, can lead to a lack of confidence in the model results, leading to rejection of the model as a basis for management advice (Punt et al., 2020). In the Northeast U.S., a few stocks with a history of age-based stock assessments have had recent assessments rejected in the peer review process. These stocks include Atlantic mackerel (*Scomber scombrus*), Georges Bank (GB) yellowtail flounder (*Limanda ferruginea*), GB cod (*Gadus morhua*), witch flounder (*Glyptocephalus cynoglossus*), and Gulf of Maine (GOM) winter flounder (*Pseudopleuronectus americanus*), whose assessments were rejected largely due to strong, positive retrospective patterns (Deroba et al., 2010; NEFSC, 2011; Legault et al., 2014; NEFSC, 2015b; NEFSC, 2017a; Punt et al., 2020). An updated assessment for mackerel has since passed peer-review (NEFSC, 2018), while the other stocks remain without approved age-based assessment estimates of abundance and reference points. The retrospective pattern is worsening for many other groundfish stocks (NEFSC, 2017b), leading to concerns that future assessments for these stocks might also be rejected.

Positive retrospective patterns can lead to other management problems beyond rejected assessments. When final year abundance in the assessment is overestimated in the model, catch advice based on these estimates is also overestimated (Brooks and Legault, 2016), which can lead to overfishing. For some New England groundfish stocks, positive retrospective pattern resulted in continued overfishing, and these stocks remain at low biomass (overfished) despite efforts to rebuild them (Wiedenmann and

Jensen, 2018). The current approach in the Northeast U.S. for dealing with retrospective patterns is to adjust the model estimate of abundance in the final year based on the average retrospective bias over the last five to seven years (called a rho adjustment; Legault, 2020). However, this adjustment is not always sufficient to prevent overfishing (Wiedenmann and Jensen, 2019), and does not help identify what is causing the retrospective pattern.

The underlying mechanism(s) causing retrospective bias remains unclear. Mechanisms shown to result in strong, positive retrospective patterns in biomass estimates include unreported catch, increased natural mortality (M), establishment of closed areas, changes in survey catchability (Legault, 2009), or fishery selectivity (Hurtado-Ferro et al., 2015). Diagnosing the underlying cause(s) of a retrospective pattern is of great importance, and generally requires an exploration into the signals in the data going into an assessment, as opposed to looking at assessment estimates themselves due to the potential biases in estimates from the retrospective pattern (ICES, 2020). Standard assumptions in most of the current assessments in the region are that 1) survey catchability is constant through time, 2) catch data are unbiased, and 3) M is constant through time. Several assessments have explored changes in these assumptions as a means of addressing the retrospective pattern (NEFSC, 2008; Legault et al., 2014; NEFSC, 2017a), but such changes have ultimately been discarded or were never used in the final assessments. The exception is GOM cod, which has two accepted model formulations; one with a fixed M =0.2 and one with an assumed increase in M through time (from 0.2 to 0.4 between 1988 and 2003; NEFSC, 2015b; 2017b).

If survey catchability is constant and catch data are unbiased, annual estimates of the total catch divided by the survey index of abundance are an unbiased measure of relative fishing intensity (herein called relative *F*; Sinclair, 1998). Higher values of relative *F* in certain periods are therefore indicative of higher exploitation rates, and vice-versa. If *M* is constant through time, then trends in relative *F* should be positively correlated with trends in total mortality (*Z*; where $Z = F + M$), which can be estimated using catch-at-age data from a survey or fishery via catch curve analysis (Sinclair, 2001). A lack of a correlation or a negative correlation between these two variables could indicate a violation of one or more of common assessment assumptions listed above, and could result in a retrospective pattern.

As an example, the stock assessment with arguably the worst retrospective pattern in the region is GB yellowtail flounder (Figure 1). In the mid 1990s, GB yellowtail exhibited a sharp decline in relative *F* but *Z* estimated from the survey has remained stable, rather than declining, throughout the time period (Figure 1 here; also see Legault and McCurdy, 2017). This disconnect in data signals given the current model assumptions (1-3 above) may be contributing to the strong retrospective patterns observed over time for this stock (Figure 1). Given that many stocks in the region have strong, positive retrospective patterns, it is possible that this discrepancy between relative *F* and *Z* exists for other stocks in the region. In this paper, the relationship between relative *F* and *Z* was evaluated across stocks in the Northeast U.S. to see if this discrepancy exists for other stocks, and how it relates to the magnitude of the retrospective pattern observed for each stock. Potential mechanisms driving the discrepancies, where found, are then discussed with the evidence supporting or refuting each mechanism.

Methods

Eighteen stocks in the region were included in this analysis (Table 1). The NEFMC is responsible for management of 15 of these stocks, with 14 part of the same multispecies Fishery Management Plan (FMP) of benthic / demersal groundfish, and the other a pelagic stock (Atlantic herring; *Clupea harengus*) part of its own FMP. The MAFMC is responsible for managing the remaining 3 stocks, with two benthic / demersal stocks (scup (*Stenotomus chrysops*) and summer flounder (*Paralichthys dentatus*)) in the same FMP, and one pelagic stock (Atlantic mackerel) in a separate FMP.

Data sources

For each stock, relative *F* was calculated using catch and survey data and total mortality (*Z*) was estimated using catch curve analysis of survey data. Total catch is comprised of landings, derived from dealer reports, and discards, which are estimated based on observer data. For the catch curve analysis (described in detail below), numerical abundance-at-age data were used from both the NMFS spring and fall bottom trawl surveys. Inshore surveys from states throughout the region collect many of the species included in this analysis, but they generally collect mostly younger fish. The lack of older fish in these surveys limits the ability to conduct catch curve analysis, so inshore surveys were excluded from the analysis. To calculate relative *F* for each stock, annual estimates of the total catch (landings + discards, in weight) were obtained from stock assessment reports and the average annual indices of abundance were obtained from the spring and fall survey datasets (both in mean numbers per tow and kg per tow). Although the survey extends back to the 1960s, only catch and survey data from 1978 onward were used. Prior to the Magnuson Act, foreign vessels caught large amounts of groundfish and other stocks in the region. The magnitude of these catches is uncertain and generally assumed to be a lower bound (Anderson, 2015), and such high total catches had the potential to skew the range of relative *F* values. For some stocks, however, total catch estimates were only available starting after 1978, particularly for stocks with large recreational components (where estimates started in 1982). The time period of this analysis was therefore restricted to either 1978 or the first year total catch data were available (Table 1).

Beginning in 2009, the NMFS bottom trawl surveys were conducted with a new vessel, the NOAA ship *Henry B. Bigelow*, which uses a different net and protocols from the previous survey vessel (the *Albatross*). Prior to retiring the previous survey vessel, a large-scale study was conducted using both vessels to calculate size and species-specific conversion factors to account for these differences (Miller et al., 2010). For years 2009-onward the standardized survey estimates were used.

Estimating Z: Catch curve analysis

In catch curve analysis, the total mortality rate in a given cohort can be calculated as the inverse slope of log-transformed abundance against age for fully-selected age classes. When using fishery-independent data, catch curve analysis assumes mortality has not changed over time across cohorts, and that availability to the sampling gear is unchanged across years. When using fishery-dependent data, there are additional assumptions that fishing effort and selectivity are the same across years. Total mortality was calculated from the fisheries-independent survey data using a catch curve analysis developed by Sinclair (2001) that accounts for differences in year class strength by pooling multiple years of data together. The formula used is

$$\text{(1)} \qquad \log(N_{a,y})=b_0+b_1Y+b_2A+\varepsilon,$$

where $N_{a,y}$ is the stratified mean catch per age a in year y, Y is a class variable that indicates the year class, A is the covariate age, b_1 is a parameter vector of separate intercepts for each year class, and b_2 is the estimated total mortality for the time period (i.e., $Z_y = -b_2$). Four year moving windows of data were used by Sinclair (2001), such that the estimate of Z is an average for that time period, centered at the midpoint year (e.g., 2004.5 for years 2003-2006). Four years of data were used in this analysis, but three- and five-year windows were also explored with little differences found in the overall results. A range of ages were explored for each stock and for each survey, with age bounds selected that resulted in unbiased residuals in the youngest and oldest ages following the approach of Sinclair (2001; see Table 1 for age ranges). The approach to determine the age ranges was used to account for the fully-selected ages, and for the low catches of older ages. Annual Z values were calculated separately over the four-year intervals for both the spring and fall NMFS bottom trawl surveys, except for mackerel where only the spring survey was used. Annual estimates of Z are not independent from one another, as the four-year moving window approach uses some of the same years of data to calculate Z between successive years. For example, the Z estimates for a stock during the periods 2000-2003 and 2001-2004 both use data from 2001-2003. Therefore, it is not possible to conduct standard statistical tests that assume independence, and subsequent analysis of trends and patterns in *Z* estimates is largely descriptive. While it is possible to calculate *Z* for non-overlapping time periods, in some cases there was considerable variability in *Z* estimates between sequential periods with only one year of different data. Therefore, overall trends in *Z* for some stocks could be even more sensitive to the specific non-overlapping periods selected, so the moving window approach with overlapping periods was used to estimate *Z*.

Estimating relative F

For each stock, relative *F* in each year *t* was calculated by dividing the total catch (C_t; in mt) by the stratified mean weight (kg) per tow in the survey (I_t) in that year (relative $F_t = C_t / I_t$; Sinclair, 1998). Relative *F* was used as opposed to the assessment estimates of *F* because assessment estimates can become biased by violations of model assumptions that can result in retrospective patterns motivating this research. Relative *F* was calculated using weight instead of numbers due to the expected differences in selectivity of small, young fish between the fishery and survey. Using weight for both the fishery and survey in the calculation focuses on the adult portion of the population from each source. Weight-at-age from the fishery and survey often follow similar patterns over time and are not expected to contribute to problems with this metric. Relative *F* was calculated for both the spring and fall surveys (only the spring for mackerel) and standardized by dividing by the mean for the entire time series for each survey and for each stock. These standardized values were then used to calculate an average relative *F* over four-year intervals (e.g., 2003-2006 to calculate a value for the midpoint year of

2004.5) for comparison with *Z* estimates for the same period. Sinclair (1998) showed that estimates of relative *F* could become biased when the survey and fishery due not overlap temporally, and that surveys conducted in the middle of the year are ideal. The surveys used here generally occur in March-April and September-October, which are close to the midpoint of the year. While it is possible to average the spring and fall surveys each year, they were evaluated separately here because that is how they are used within the assessments for each stock. Also, the high correlation between seasonal surveys for most stocks (Table 2) suggests that averaging surveys would not alter the overall findings.

The relationship between relative F and Z, and retrospective patterns

For each stock, the Pearson correlation coefficient between relative *F* and *Z* was calculated, using estimates aggregated across surveys. Because estimates of *Z* are not independent, however, the significance of the correlation was not calculated, but these estimates were used to explore the general relationship (i.e., positive, negative) between relative *F* and *Z*. Stock-specific correlations between relative *F* and *Z* were then compared with estimates of Mohn's rho (Mohn, 1999) collected from recent stock assessments. Mohn's rho is a measure of the average retrospective uncertainty in terminal assessment estimates (typically over the most recent five to seven years), with more extreme rho values (positive or negative) indicating larger uncertainty in recent assessment estimates.

Results

Across stocks, relative *F* estimates were positively correlated between the spring and fall survey, with correlations ranging from 0.27 for herring to 0.95 for Georges Bank yellowtail flounder (Table 2). Annual relative *F* for each survey across stocks is shown in Figure 2, separated out by region (Gulf of Maine, Georges Bank, Southern New England / Mid-Atlantic, and pelagic stocks of mackerel and herring that span the whole region). Region-specific average relative *Fs* were calculated in each year across stocks to look for regional trends. For both the GB and SNE/MA regions, relative *F* declined sharply by the mid 1990s (Figure 2B and 2C). Some stocks in both regions had occasional annual spikes in relative *F* since the mid 1990s, but on average, relative *F* has remained low in these regions in more recent years. The average relative *F* since 1995 was 37% of the pre-1995 period on GB and 48% of the average of the earlier years in the SNE/MA region (Table 2). A more gradual decline in relative *F* started in the early 1990s in the GOM, and there is considerably more variability across stocks (Figure 2A). Nevertheless, the average relative *F* in the GOM has also remained low since the mid 1990s (46% of the earlier period; Table 2). Sharp declines in relative *F* occurred earlier for the pelagic stocks, between the late 1980s and early 1990s, with larger declines for Atlantic mackerel (Figure 2D). Across all stocks in the analysis, relative *F* is lower in more recent years. Since 1995, the average relative *F* for each stock has been 8 to 81% of the pre-1995 period (Table 2).

While relative *F* has declined across stocks in each region, *Z* has not for many stocks (Table 2 and Figure 3). There is considerable temporal variability in *Z* estimates, particularly in the GOM for species such as haddock and pollock (Figure 3A). When averaged across stocks in each region, total mortality varied without trend in both the

GOM and GB regions (Figures 3A and 3B). In the SNE/MA region, however, the average total mortality declined gradually, concurrent with the decline in relative *F*, though not as sharply, nor as much (Figure 3C). For the pelagic stocks, *Z* increased early on, but varied without trend since then (Figure 3D). Average *Z* for each stock was calculated for the periods pre-1995 and 1995 to present (across surveys), and *Z* in the recent period was higher for eight stocks, and lower for ten stocks (Table 2). For seven of the stocks with increases in *Z*, the increase ranged from 10% to 67%. For mackerel, a much larger increase occurred, with recent average *Z* 198% higher than the earlier period. For the stocks with decreases in *Z*, the recent average *Z* was between 46% to 93% of the pre- 1995 average (Table 2). For all stocks but herring, the correlations between estimates of *Z* in the spring and fall surveys were weaker than the correlations in relative *F* between surveys, with positive correlations for all but two stocks (pollock and GB haddock; Table 2).

The relationship between relative *F* and *Z* from each survey for each stock is shown in Figure 4, calculated over four-year intervals. Of the 18 stocks, only five had positive correlations > 0.2 between relative *F* and *Z*. These stocks are, in order of increasing correlation, GOM cod, plaice, SNE/MA winter flounder, scup, and summer flounder. Only scup and summer flounder had correlations ≥ 0.4. Eight stocks had weak to no correlation (defined here as within ± 0.2). The remaining stocks (Atlantic herring, witch flounder, mackerel, CC/GOM yellowtail flounder, and GB yellowtail flounder) had negative correlations (< -0.2) between relative *F* and Z calculated using the spring and fall survey data.

Notably, three of the four stocks with the largest negative correlations between relative *F* and *Z* have had recent stock assessments that did not pass peer review in large part due to strong retrospective patterns (GB yellowtail, mackerel, and witch flounder). GB cod also had an assessment that did not pass review, and it had a correlation between relative *F* and *Z* of -0.03. Stocks with positive correlations between relative *F* and *Z* had lower estimates of rho, while those with low to negative correlations had higher estimates of rho, in general, although there was variability in rho values for a number of stocks with similar correlation values (Figure 5).

For many stocks there were large declines in relative F, without declines in *Z* (Table 2). There are several possible causes for this disconnect, described in detail in the Discussion, including the evidence for or against each mechanism. One possibility is that relative *F* estimates are biased, and that the general declines observed across stocks (Table 2) did not actually occur. In other words, relative *F* may not have changed over the time period and that the trends observed are due to biased catch data. To result in no change in relative *F* over the entire period, catch data would need to have been overestimated earlier in the time period, or underestimated more recently. For the 13 stocks with weak to negative correlations (< 0.2; Table 2) the magnitude of bias in the catch data needed to make an apparent difference in relative *F* was estimated. For example, if biased catch data were the cause of an apparent 50% decline in relative *F* in the mid 1990s, that could result from earlier catches being inflated twofold, or more recent catches underestimated by half of the true catch. Using the relative *F* ratio ($relF_{ratio}$, calculated as the mean relative *F* from1995 onward / mean relative *F* before

1995 ; Table 2), the amount of catch needed each year in the over- and underreporting cases ($C_{over,t}$ and $C_{under,t}$, respectively) was calculated with:

(2) $C_{over,t} = C_{obs,t}(1 - relF_{ratio})$, and

(3) $C_{under,t} = C_{obs,t}\left(\frac{1}{relF_{ratio}} - 1\right)$,

where $C_{obs,t}$ is the observed catch in year t. The mean and ranges for $C_{over,t}$ and $C_{under,t}$ are presented in Table 3. Earlier in the time period, actual catches would have needed to be between 700 and 41,059 mt lower than the observed catch, on average, with the largest amounts occurring for the pelagic stocks of mackerel and herring, and also for GB cod. As a percentage, the amount of overreported catches needed ranged between 19 to 92% of observed catch, on average, during this period, (Table 3). The average annual magnitude of underreported catches needed was between 531 to 207,899 mt across stocks, with the largest values for mackerel and herring. Large amounts of underreporting would have also been needed for the GB stocks of haddock, yellowtail flounder, and cod (9,296 to 19,709 mt annually, on average; Table 3). As a percentage, the actual catch across stocks would have been between 24 to 1,147% higher than the observed catch, on average (Table 3).

Discussion

Large declines in relative *F* occurred across stocks in the northeast U.S. since the mid 1990s, yet for a number of these stocks *Z* has changed little or even increased. As a result, there is little to no correlation, or even a negative correlation, between relative *F* and *Z* for these stocks. These diverging signals in the data appear to be a large contributor to the retrospective uncertainty in assessment estimates observed in the region. A key remaining question therefore is, what could be leading to the discrepancy between relative *F* and *Z* for these stocks? Possible mechanisms include 1) biased estimates of relative *F*, 2) biased estimates of *Z*, 3) natural mortality has increased to make up for the declines in fishing mortality, and 4) total mortality has been dominated by natural mortality over the entire period. The plausibility of these different mechanisms is discussed below, providing evidence for or against each. Although these mechanisms are detailed separately, they are not mutually exclusive, and multiple mechanisms may be involved to varying degrees across stocks.

Biased estimates of relative F

The declines in relative *F* observed here could have resulted from an increase in survey catchability, underreporting of catch data (landings and/or discards), or changes in fishery selectivity. If catchability increased in the NMFS bottom trawl survey in the mid 1990s, indices of abundance would be inflated, making our estimates of relative *F* (catch / index) biased low in more recent years. Residual patterns in the assessment fits to the NMFS spring and fall survey, coupled with strong retrospective patterns led to many of the stock assessments splitting the survey time series into separate time blocks, with different catchability estimates for each block (NEFSC, 2005; 2008). This approach reduced the magnitude of the retrospective error in a number of the assessments, but there was not a clear justification for changes in the survey catchability during that time period. Over time, as the retrospective pattern re-emerged in many assessments, this approach was ultimately abandoned (e.g., NEFSC, 2015b; 2017b).

One of the most well-documented responses of marine taxa to climate change is that species are shifting their distributions poleward and into deeper water, and this pattern has been documented for many species in the Northeast U.S. shelf ecosystem (Nye et al., 2009; Pinsky et al., 2013). For species that make seasonal inshore-offshore migrations, climate change could impact the timing of these migrations, potentially altering the availability of species to the spring and fall surveys (Langan et al., 2021). Changes in the distribution or phenology of stocks would need to increase the overall availability of species to the survey (thereby increasing catchability) to result in the declines in relative *F* that we observed. Additional research into these areas is warranted, though it is unlikely that all stocks in our analysis would have similar increases in availability to the survey in response to climate change.

An alternative mechanism impacting the catchability of all stocks would be through changes in the survey itself, although the evidence for such a mechanism is lacking. First, there have been no known changes in the survey design or gear that occurred in the mid 1990s when relative *F* estimates began to decline. Changes in the timing of the seasonal surveys towards earlier or later periods could impact the catchability of some stocks. The timing of the spring and fall surveys has varied over the last 50 years, although there does not appear to be an overall trend towards earlier or later survey periods that would lead to consistent changes in catchability (see Figure 6 in Legault and McCurdy, 2018).

Biased estimates of relative *F* could also result from biased catch data. The decline in relative *F* across many stocks in in the mid 1990s is coincident with the implementation of Amendment 5 to the New England multispecies (groundfish) FMP, which covers 14 of the stocks in our analysis. Amendment 5 was aimed at reducing fishing mortality by 50% across groundfish stocks through large reductions in allowable fishing days and increases in mesh size, but it also resulted in a large increase in data coverage by making vessel and dealer reporting mandatory (NEFMC, 1993). The large-scale effort reduction associated with Amendment 5 could have increased noncompliance in the groundfish fishery. Through surveys with fishermen, managers, scientists, and enforcement officers, King and Sutinen (2010) estimated an increase in noncompliance since the 1980s, with unreported catches between 12-24% of the total catch by the mid 2000s. They also noted strong economic incentives for noncompliance in the New England groundfish fishery based on the likelihood of detection and the penalties imposed versus potential profit. More recently, the owner of the most vessels in the entire New England groundfish fishery and holder of the largest quota for many stocks, was convicted of mislabeling landed stocks with limited quota (GB yellowtail, GB cod, plaice and witch flounder) as the more abundant stock in the region (GB haddock; Bellanger et al., 2019). Unfortunately, the magnitude and period of unreported catches from this single operation is unknown, nor is it known if this behavior was more widespread across the fishery.

Although there is evidence of unreporting of catches in the region, the magnitude needed to be the sole cause of the patterns we observed is likely much higher than is feasible for many stocks. The catches reported in Table 3 can be used as an approximation for the amount of unreported catch needed to make relative F unchanged over time assuming that the unreported catch has the same length and age characteristics of the reported catch. For some groundfish stocks, the amount of unreported catches needed each year may be feasible (hundreds to a few thousand metric tons each year; e.g., witch flounder). For other stocks, ten thousand plus metric tons of unreported catch each year for each stock seems unfeasible both from a fishery perspective and from an enforcement perspective. In other words, how could the fishery land or discard that much additional catch, and how could it go undetected? An alternative approach to estimating the amount of unreported landings has been explored in some of the assessments, where some multiplier increases catches until the retrospective pattern disappears (e.g. Legault et al., 2013; NEFSC, 2017a; Legault, 2020). Such approaches have similarly shown that large multipliers (three to five-fold increases) of observed catches are needed to remove the retrospective pattern.

Another way relative *F* could be biased is by large-scale changes in the fishery selectivity pattern over time changing the relationship between catch in the fishery and catch in the survey. For example, if the fishery switched gears from one that produced a flat-topped selectivity pattern to one that produced a strong dome in selectivity, while the survey maintained a constant selectivity pattern over time, the ratio of catches would no longer reflect a constant ratio of selectivity patterns. This change in fishery selectivity could create a bias in the relative *F*, but has not been explored extensively to date. There are no large-scale changes in fishery selectivity over time in any of the assessments examined here.

Biased estimates of total mortality

Estimates of *Z* were calculated via catch curve analysis using numbers-at-age from the spring fall surveys. As previously noted, many species in the region are moving into deeper waters in response to climate change (Nye et al., 2009; Pinsky et al., 2013). For demersal and benthic species, older, larger fish tend to occupy deeper waters compared to younger fish (Swain, 1993; Methratta and Link, 2007; Friedland et al., 2021). Climate-driven shifts of a population over time into deeper water could result in fewer older fish being available to the survey if they have shifted outside of the survey area. A decline in the availability of older fish to the survey would result in dome-shaped selectivity in the survey, violating the assumption of flat selectivity in the catch curve and resulting in higher estimates of *Z*. However, the approach used here to determine suitable age bounds for the catch curve analysis likely mitigates against this effect. This potential mechanism, and how the selection of age bounds impacts the *Z* estimate, has not been widely explored, making this is a potentially important area of future research.

Increases in natural mortality

If the trends in relative *F* are genuine, then increases in *M* could account for *Z* remaining high or even increasing in recent years. Due to the available evidence, the focus here is on climate change and increased predation as mechanisms for increasing *M*, but changes in prey abundance or availability, or inter- and intra-specific competition are also plausible mechanisms for increasing *M*.

The Northeast U.S. continental shelf is rapidly warming, with increases both in mean annual temperature, but also in the occurrence of marine heat waves (Mills et al., 2013; Pershing et al., 2015; Hare et al., 2016; see Figure 6A, B here). Warming is expected to have different impacts on stock productivity for different species (e.g., Blanchard et al., 2012; Free et al., 2019), but recent studies suggest that most groundfish stocks in the region will be negatively impacted by climate change (Hare et al., 2016; Klein et al., 2016). Increasing temperatures will increase metabolic demands of fish and reduce oxygen saturation in the water (Carazzo et al., 2019), and continued exposure to such conditions could lead to stress- or starvation-induced increases in M in adult fish. Periodic, rapid changes in temperatures (that may not be captured in an annual average) have also been linked with mass mortality events in marine systems (Genin et al., 2020).

In the Northeast U.S., there is a growing body of research linking M with changing environmental conditions. For example, age-specific values for M were estimated for GOM cod using assessment estimates of numbers-at-age and correlated these estimates with temperature (Pershing et al., 2015; though this finding has been debated; Palmer et al., 2016 and Swain et al., 2016). Environmental indices linked to M were incorporated into an assessment model for summer flounder, and it was found that the Gulf Stream Index (GSI; an annual measure of the northward position of the warm Gulf Stream) was the best predictor of M, and that including GSI as a driver improved the overall model fit (O'Leary et al., 2019). Time-varying M was estimated within an assessment model for weakfish (*Cynoscion regalis*), an inshore species of recreational importance in the region that had also exhibited discrepancies in relative F and Z (Jiao et al., 2012). Increases in the estimated M over time for weakfish were then linked to the Atlantic Multidecadal Oscillation (AMO), a basin-wide index of sea surface temperature. A growing body of literature suggests wide-ranging impacts of the AMO on ecosystems throughout the Atlantic (see Nye et al., 2014 for a review), and the AMO switched from a cool phase to a warm phase in the in the mid 1990s, coincident with the decline in relative F but not Z observed here.

Increases in *M* could also be driven by increased abundance of predators. In nearby Canadian waters, there is compelling evidence that large increases in marine mammal abundance, most notably grey seals (*Halichoerus grypus atlantica*), resulted in large increases in natural mortality for a number of demersal fish stocks (Chouinard et al., 2005; Benoit et al., 2011; O'Boyle and Sinclair, 2011; Swain and Benoit, 2015, Neuenhoff et al., 2019; Swain et al., 2019b). Many of these stocks were severely depleted by overfishing, and increased consumption by grey seals or other predators could be resulting in a "predator pit", preventing the recovery of these stocks despite management efforts (e.g., Swain and Benoit, 2015; Neuenhoff et al., 2019; Swain et al. 2019b; see Swain et al., 2011 for a detailed exploration into multiple hypotheses regarding the sustained high *Z* for cod in the southern Gulf of St Lawrence). In the mid 1980s, grey seals from Canada began colonizing a few isolated locations in our study region, (Moxley et al., 2017; Hayes et al., 2018). The population in U.S. waters is growing (Figure 6C), but the estimated size is well below the recent estimates in Canadian waters (Hayes et al., 2018). Smith et al. (2015) estimated the consumption of groundfish and other species by marine mammals in the Northeast U.S., and found overall consumption comparable to or higher than commercial fisheries landings. To reproduce the patterns observed here, increases in *M* would have started in the mid 1990s for most stocks when relative *F* started declining. This timing is coincident with the predator-induced increases in *M* in neighboring Canadian ecosystems (e.g., Swain and Benoit ,2015; Swain et al., 2019b), and with the increases observed in the study region (Figure 6C), adding support to the notion of increased *M* via predation. It is also worth noting that the two stocks with the highest correlations between relative *F* and *Z* are species primarily found in the Mid-Atlantic region where grey seals and harbor seals (*Phoca vitulina vitulina*) occur only seasonally, and at lower abundances than in New England waters (Hayes et al., 2018).

While it is likely that *M* has increased over time for many stocks, the data are not available to independently estimate time-varying *M* across individual stocks. Estimation of time-varying *M* can be done within an assessment model (e.g., Lee et al., 2011; Jiao et al., 2012; Legault, 2020), and the increasing use of state-space assessment models will allow for ecosystem variables to be linked to *M* and other population process (recruitment, growth, maturity; e.g., Miller et al., 2017; O'Leary et al., 2019). However, estimation of *M* is conditioned on other assumptions within the model, and estimated patterns in *M* may not be genuine, and result from something else, such as unreported catch (Rossi et al., 2019). Cadigan (2016) developed a state-space assessment model for northern cod that allowed for time-varying *M* and survey catchability, as well as unreported catches, but this work required tagging data to reliably estimate changes in *M*. Estimation of *M* over time can also be done in cases where fisheries are closed and *F*~0 (Swain and Mohn, 2012; Sinclair et al., 2015; Swain and Benoît, 2015; Swain et al., 2019a), but this condition does not exist for most stocks in the region (Wiedenmann et al., 2019) and is confounded with the potential for unreported landings or discards. Independent estimates of *M* via tagging studies could be useful to determine the relative contribution of *M* versus unreported catch. Exploration of other indicators of change in *M* across stocks in the region is certainly warranted. Such an approach could explore a range of datasets across stocks looking for commonalities, including changes in size-at-age and / or body condition, changes in diet composition of the focal species, or changes in consumption estimates of the focal species by predators. Long-term information on the diet composition and foraging areas of marine mammals is lacking across the region, but studies that can improve the magnitude of current consumption estimates at the stock level would be very valuable to help understand the role marine mammals are playing across the ecosystem.

An approach analogous to the missing catch analysis but to estimate the amount of additional *M* needed to produce the patterns here is not possible given the information used in this study. However, Legault (2020) conducted an analysis where a range of *M* values were explored over different time periods in the assessment model to determine the magnitude of an increase in *M* needed to remove the retrospective pattern. This analysis was conducted for four stocks of New England groundfish, three of which have positive retrospective patterns (GB yellowtail flounder, witch flounder, and white hake). The amount of extra *M* needed to remove the retrospective pattern ranged between 1.5 to 5 times the assumed *M*, but the exact value depended on when and how quickly the increase occurred for a given stock. Larger increases in *M* were needed for the stocks with larger retrospective patterns (GB yellowtail and witch flounder; Legault, 2020).

Natural mortality has always dominated total mortality

Another mechanism for the lack of correlation between relative *F* and *Z* is that *Z* has always been dominated by natural mortality, such that changes in catch have little impact on *Z*. Recent estimates of swept area biomass for some stocks from the NMFS bottom trawl survey have been higher than the age-based assessment estimates, and the estimated harvest fractions (catch / swept-area biomass) are very low in recent years for some stocks (1-5% and 2-7% for GB yellowtail and witch flounder, respectively; Legault and Finley, 2019; NEFSC, 2017a). These estimates, however, ignore the potential for unreported catch. If *Z* were largely comprised of *M* over the entire time period, then this would mean that *M* is considerably higher than previously assumed for many stocks. For example, average *Z* estimates for all yellowtail flounder stocks and SNE/MA winter flounder ranged between 0.84 and 1.30, while the assumed *M* for these stocks is between 0.2 and 0.25 (Legault et al., 2013; NEFSC, 2017b). Also, if catches were small compared to total biomass, then it is likely that the fishery would frequently meet or exceed the catch limit in most years. However, observed catches (notwithstanding the potential for unreported catches) for New England groundfish have been well below the target in most years for most stocks (Wiedenmann and Jensen, 2018), casting additional doubt on the possibility that total biomass was much higher than previously estimated.

Conclusion

In summary, patterns in relative *F* across stock in the region suggest an overall decline in the fishing pressure during the 1990s. However, total mortality has not decreased for many stocks, leading to diverging signals in the data used in the stock assessments. As a result, strong, positive retrospective patterns exist across many assessments, particularly for stocks where this discrepancy is acute. Multiple mechanisms may be causing these diverging signals, but the most plausible ones are unreported catch and increased natural mortality (from predation or climate change). However, it is unlikely that a single cause can explain the divergence across all stocks, or perhaps even any given stock, and the importance of different mechanisms almost certainly varies among stocks and over time for individual stock. Continued explorations into the relative contribution of these or other mechanisms is certainly justified to help address the retrospective patterns common in many the assessments for Northeast U.S. fish stocks. Management strategy evaluation (MSE) simulation studies (e.g., Punt et al., 2016) that explore different mechanisms separately will be useful to identify robust management practices regardless of the presence or the cause of the divergence in the data.

References.

Anderson, E.D. (2015). Lessons from a career in fisheries science. *ICES Journal of Marine Science,* 72, 2169-2179.

Bellanger M., Holland, D.S., Anderson, C.M., & Guyader, O. (2019) Incentive effect of joint and several liability in fishery cooperatives on regulatory compliance. *Fish and Fisheries*, *20*, 715– 728.

Benoît, H.P., Swain, D.P., Bowen, W.D., Breed, G.A., Hammill, M.O., & Harvey, V. (2011) Evaluating the potential for grey seal predation to explain elevated natural mortality in three fish species in the southern Gulf of St. Lawrence. *Marine Ecology Progress Series*, *442*, 149-167.

Blanchard, J.L., Jennings, S., Holmes, R., Harle, J., Merino, G., Allen, J.I., et al. (2012) Potential consequences of climate change for primary production and fish production in large marine ecosystems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 2979–2989.

Brooks, E.N., & 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.

Cadigan N. (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.

Carozza, D.A., Bianchi, D., & Galbraith E.D. (2019) Metabolic impacts of climate change on marine ecosystems: Implications for fish communities and fisheries. *Global Ecology and Biogeography*, *28*, 158–169.

Chouinard, G.A., Swain, D.P., Hammill, M.O., & Poirier, G.A. (2005) Covariation between grey seal (*Halichoerus grypus*) abundance and natural mortality of cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences*, *62*, 1991-2000.

Deroba, J. (2015) Atlantic herring operational assessment report for 2015. US Department of Commerce, Northeast Fisheries Science Center Reference Document. 15-16: 30 p. Available: http://www.nefsc.noaa.gov/publications/

Deroba, J., Shepherd, G., Gregoire, F., Nieland, J., & Rago, P. (2010) Stock Assessment of Atlantic Mackerel in the Northwest Atlantic for 2009. Available: https://pdfs.semanticscholar.org/f7f5/58957942e5184d9bc03dfb911439840622ba.pdf?_ga=2.81166935.1782902224.1579530247-676141752.1579530247

Free, C.M., Thorson, J., Pinsky, M., Oken, K., Wiedenmann, J. & Jensen, O.P. (2019) Impact of historical warming on marine fisheries production. *Science*, *363*, 979-983.

Friedland, K.D., Smolinski, S., & Tanaka, K.R. (2021) Contrasting patterns in occurrence and biomass centers among fish and macroinvertebrates in a continental shelf ecosystem. *Ecology and Evolution*, *11*, 2050–2063.

Genin, A., Levy, L., Sharon, G., Raitsos, D.E., & Diamant, A. (2020) Rapid onsets of warming events trigger mass mortality of coral reef fish. *Proceedings of the National Academy of Sciences*, *117*, 25378-25385.

Hare, J.A., Morrison, W.E., Nelson, M.W., Stachura, M.M., Teeters, E.J., Griffis, R.B., et al. (2016) A vulnerability assessment of fish and invertebrates to climate change on the Northeast U.S. continental shelf. *PLoS ONE*, *11*, e0146756.

Hayes, S.A., Josephson, E., Maze-Foley, K., & Rosel, P.E. (2018) US Atlantic and Gulf of Mexico Marine Mammal Stock Assessments. NOAA Technical Memo. 2019; NMFS-NE 258; 291 p.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., et al. (2015) Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES Journal of Marine Science*, *72*, 99–110. doi:10.1093/icesjms/fsu198.

International Council for the Exploration of the Sea (ICES). 2020. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting). ICES Scientific Reports. 2:28. 38 pp. http://doi.org/10.17895/ices.pub.5997.

Jiao,Y., Smith, E., O'Reilly, R., & Orth, D. (2012) Modeling nonstationary natural mortality in catch-at-age models: an example using the Atlantic weakfish (*Cynoscion regalis*) fishery. *ICES Journal of Marine Science*, *69*, 105–118.

King, D.M., & Sutinen, J.G. (2010) Rational noncompliance and the liquidation of Northeast groundfish resources. *Marine Policy*, *34*, 7-21.

Klein, E.S., Smith, S.L., & Kritzer, J. (2017) Effects of climate change on four groundfish species. *Reviews in Fish Biology and Fisheries*, *27*, 317-338.

Kraak, S.B.M., Daan, N., & Pastoors, M.A. (2009) Biased stock assessment when using multiple, hardly overlapping, tuning series if fishing trends vary spatially. *ICES Journal of Marine Science*, *66*, 2272–2277.

Methratta, E.T., & Link, J. (2007) Ontogenetic variation in habitat association for four groundfish species in the Gulf of Maine – Georges Bank region. *Marine Ecology Progress Series, 338*, 169-181.

Langan, J.A., Puggioni, G., Oviatt, C.A., Elisabeth Henderson, M., & Collie, J.A. (2021) Climate alters the migration phenology of marine species. *Marine Ecology Progress Series*. *660*, 1-18.

Lee, H.H., Maunder, M.M., Piner, K.R., & Methot, R.M. (2011) Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twelve stock assessments. *Fisheries Research*, *109*, 89–94.

Legault, C.M., Chair. (2009) Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. Northeast Fisheries Science Center Reference Document 2. 09-01. 30 p. Available: http://www.nefsc.noaa.gov/nefsc/publications/s-a/

Legault, C.M., Alade, L., Gross, W.E., & Stone, H.H. (2013) Stock Assessment of Georges Bank Yellowtail Flounder for 2013. TRAC Reference Document. 01. 132 p. Available: https://apps-nefsc.fisheries.noaa.gov/rcb/publications/trac.html

Legault, C.M., Alade, L., Gross, W.E., & Stone, H.H. (2014) Stock Assessment of Georges Bank Yellowtail Flounder for 2014. TRAC Reference Document. 01. 214 p. Available: https://apps-nefsc.fisheries.noaa.gov/rcb/publications/trac.html

Legault, C.M., & McCurdy, Q.M. (2017) Stock Assessment of Georges Bank Yellowtail Flounder for 2017. TRAC Working Paper 2017; 03. 58 p. Available: https://apps-nefsc.fisheries.noaa.gov/rcb/publications/trac.html

Legault, C.M., & McCurdy, Q.M. (2018) Stock Assessment of Georges Bank Yellowtail Flounder for 2018. TRAC Working Paper 2018; 03. 64 p. Available: https://apps-nefsc.fisheries.noaa.gov/rcb/publications/trac.html

Legault, C.M., & Finley, M. (2019) Stock Assessment of Georges Bank Yellowtail Flounder for 2019. TRAC Working Paper 2019; 03. 60 p. Available: https://apps-nefsc.fisheries.noaa.gov/rcb/publications/trac.html

Legault, CM. (2020) Rose vs Rho: A comparison of two approaches to address retrospective patterns in stock assessments. *ICES Journal of Marine Science*, *77*, 3016-3030.

Maunder, M.N., & Piner, K.R. (2015) Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science*, *72*, 7–18.

Miller, T.J., Das, C., Politis, P.J., Miller, A.S., Lucey, S.M., Legault, C.M., et al. (2010) Estimation of Albatross IV to Henry B. Bigelow calibration factors. Northeast Fisheries Science Center Reference Document. 10-05; 233 p. Available: https://nefsc.noaa.gov/publications/

Miller, T.J., Hare, J.A., & Alade, L.A. (2016) A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to southern New England yellowtail flounder. *Canadian Journal of Fisheries* *and Aquatic Sciences*. *73*, 1261–1270. dx.doi.org/10.1139/cjfas-2015-0339.

Miller, T.J., O'Brien, L., Fratatoni, P.S. 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. *Canadian Journal of Fisheries and Aquatic Sciences*. *75*, 2159–2171. dx.doi.org/10.1139/cjfas-2017-0124

Mills, K.E., Pershing, A.J., Brown, C.J., Chen, Y., Chiang, F.-S., Holland, D.S., et al. (2013). Fisheries management in a changing climate: Lessons from the 2012 ocean heat wave in the Northwest Atlantic. *Oceanography*, *26*, 191-195.

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

Moxley, J.H., Bogomolni, A., Hammill, M.O., Moore, K.M.T., Polito, M.J., Sette, L. et al. (2017) Google Haul Out: Earth Observation Imagery and Digital Aerial Surveys in Coastal Wildlife Management and Abundance Estimation. *BioScience*, *67*, 760–768.

Neuenhoff, R.D., Swain, D.P., Cox, S.P., McAllister, M.K., Trites, A.W., Walters, C.J., et al. (2019) Continued decline of a collapsed population of Atlantic cod (*Gadus morua*) due to predation-driven Allee effects. *Canadian Journal of Fisheries and Aquatic Sciences*. *76*, 168-184.

Newman, D., Berkson, J., & Suatoni, L. (2015) Current methods for setting catch limits for data-limited fish stocks in the United States. *Fisheries Research*, *164*, 86-93.

Northeast Fisheries Science Center (NEFSC). (2002) Assessment of 20 Northeast groundfish stocks through 2001: a report of the Groundfish Assessment Review Meeting (GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, October 8-11. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 02-16. Available: http://www.nefsc.noaa.gov/saw/.

Northeast Fisheries Science Center (NEFSC). (2005) Assessment of 19 Northeast groundfish stocks through 2004. Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, 15-19 August 2005. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 05-13; 499 pp. Available: http://www.nefsc.noaa.gov/saw/.

Northeast Fisheries Science Center (NEFSC). (2008) Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 08-15; 884 pp. Available: http://www.nefsc.noaa.gov/nefsc/publications/.

Northeast Fisheries Science Center (NEFSC). (2011) 52nd Northeast Regional Stock Assessment Workshop (52nd SAW) Assessment Report. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 11-17; 962 p. Available: https://nefsc.noaa.gov/publications/

Northeast Fisheries Science Center (NEFSC). (2013) 55th Northeast Regional Stock Assessment Workshop (55th SAW) Assessment Report. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 13-11; 845 p. Available: http://www.nefsc.noaa.gov/nefsc/publications/

Northeast Fisheries Science Center (NEFSC). (2015a) 60th Northeast Regional Stock Assessment Workshop (60th SAW) Assessment Report. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 15-08; 870 p. Available: http://www.nefsc.noaa.gov/publications/

Northeast Fisheries Science Center (NEFSC). (2015b) Operational Assessment of 20 Northeast Groundfish Stocks, Updated Through 2014. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 15-24; 251 p. Available: http://www.nefsc.noaa.gov/publications/.

Northeast Fisheries Science Center (NEFSC). (2017a). 62nd Northeast Regional Stock Assessment Workshop (62nd SAW) Assessment Report. U.S. Department of Commerce.

Northeast Fisheries Science Center Reference Document, 17-03; 822 p. Available: http://www.nefsc.noaa.gov/publications/.

Northeast Fisheries Science Center (NEFSC). (2017b)Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document, 17-17; 259 p. Available: http://www.nefsc.noaa.gov/publications/.

Northeast Fisheries Science Center (NEFSC). (2018) 64th Northeast Regional Stock Assessment Workshop (64th SAW) Assessment Report. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 18-06; 529 p. Available: http://www.nefsc.noaa.gov/publications/.

Nye, J.A., Link, J.S., Hare, J.A., & Overholtz, W.J. (2009) Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Marine Ecology Progress Series*, *393*, 111–129. https://doi.org/10.3354/meps08220.

Nye, J.A., Baker, M., Bell, R., Kenny, A., Kilbourne, K.H., Friedland, K.D., et al. (2014) Ecosystem effects of the Atlantic Multidecadal Oscillation. *Journal of Marine Systems*, *133*, 103-116.

O'Boyle, R., & Sinclair, M. (2012) Seal−cod interactions on the Eastern Scotian Shelf:

reconsideration of modelling assumptions. *Fisheries Research*, *115-116*, 1−13.

O'Leary, C.A., Miller, T.J., Thorson, T.J., & Nye, J.A. (2019) Understanding historical summer flounder (*Paralichthys dentatus*) through the incorporation of oceanography-dependent vital rates in Bayesian hierarchical models. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*, 1275-1294.

Palmer, M.C. (2017) Vessel trip reports catch-area reporting errors: Potential impacts on the monitoring and management of the Northeast United States groundfish resource. US Department of Commerce, Northeast Fisheries Science Center Reference Doc. 17-02; 47 p. Available: http://www.nefsc.noaa.gov/publications/.

Palmer, J., Deroba, J., Legault, C.M., & Brooks, E.N. (2016) Comment on "Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery." *Science*, *352*, 423.

Pershing, A.J., Alexander, M.A., Hernandez, C.M., Kerr, L.A., Le Bris, A., Mills, K.E., et al. (2015) Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *350*, 809-812.

Pinsky, M.L., Worm, B., Fogarty, M.J., Sarmiento, J.L., & Levin, S.A. (2013) Marine taxa track local climate velocities. *Science*, 341,1239–1242.

Punt, A.E., Butterworth, D., de Moor, C., De Oliveira, J., & Haddon, M. (2016)

Management Strategy Evaluation: Best Practices. *Fish and Fisheries*, *17*, 303-34.

Punt, A.E., Tuck, G.N., Day, J., Canales, C.M., Cope, J.M., de Moor, C.L., et al. (2020) When are model-based stock assessments rejected for use in management and what happens then? *Fisheries Research*, *224*, 105465.

Reid, R.N., Almeida, F.P., & Zetlin, C.A. (1999) Fishery Independent Surveys, Data Sources and methods. NOAA Technical Memorandum NMFS-NE-122. Available: https://www.nefsc.noaa.gov/publications/.

Rossi, S.P., Cox, S.P., Benoît, H.P., & Swain, D.P. (2019) Inferring fisheries stock status from competing hypotheses. *Fisheries Research*, *216*, 155-166.

Sinclair, A. F. (1998) Estimating trends in fishing mortality at age and length directly from research survey and commercial catch data. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*, 1248–1263.

Sinclair, A.F. (2001) Natural mortality of cod (*Gadus morhua*) in the Southern Gulf of St. Lawrence. *ICES Journal of Marine Science*, *58*, 1-10.

Sinclair, M., Power, M., Head, E., Li, W.K.W, McMahon, M., Mohn, R. (2015) Eastern Scotian Shelf trophic dynamics: A review of the evidence for diverse hypotheses. *Progress in Oceanography*, *138*, 305-321.

Smith, L.A., Link, J.S., Cadrin, S.X., & Palka, D.L. (2015) Consumption by marine mammals on the Northeast U.S. continental shelf. *Ecological Applications*, *25*, 373-389.

Swain, D.P. (1993) Age- and density-dependent bathymetric pattern of Atlantic cod (*Gadus morhua*) in the Southern Gulf of St. Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*, 1255-1264.

Swain, D.P., & Mohn, R.K. (2012) Forage fish and the factors governing recovery of Atlantic cod (*Gadus morhua*) on the eastern Scotian Shelf. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 997-1001.

Swain, D.P., & Benoît, H.P. (2015) Extreme increases in natural mortality prevent recovery of collapsed fish populations in a Northwest Atlantic ecosystem. *Marine Ecology Progress Series*, *519*, 165-182.

Swain, D.P., Benoît, H.P., Hammill, M.O., McClelland, & G., Aubry, É. (2011) Alternative hypotheses for causes of the elevated natural mortality of cod (*Gadus morhua*) in the southern Gulf of St. Lawrence: the weight of evidence. Department of Fisheries and Oceans, Canadian Science Advisory Secretariat Research Document, 2011/036. iv + 33 p.

Swain, D.P., Benoît, H.P., Cox, S.P., & Cadigan, N.G. (2016) Comment on "Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery." *Science*, *352*, 423.

Swain, D.P., Ricard, D., Rolland, N., & Aubry, É. (2019a). Assessment of the southern Gulf of St. Lawrence Atlantic Cod (*Gadus morhua*) stock of NAFO Div. 4T and 4Vn (November to April), March 2019. Department of Fisheries and Oceans, Canadian Science Advisory Secretariat Research Document, 2019/038. iv + 105 p

Swain, D.P., Benoît, H.P., Hammill, M.O., & Sulikowski, J.A. (2019b). Risk of extinction of a unique skate population due to predation by a recovering marine mammal. *Ecological Applications*. *29,* 1282- 1299. e01921.

Terceiro, M. (2016). Stock Assessment of Summer Flounder for 2016. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document. 16-15; 117 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at http://www.nefsc.noaa.gov/publications/

Van Beveren, E., Duplisea, D. E., Castonguay, M., Doniol-Valcroze, T., Plourde, S., Cadigan, N. (2017). How catch underreporting can bias stock assessment of and advice for northwest Atlantic mackerel and a possible resolution using censored catch. *Fisheries Research*, *194*, 146-154.

Wiedenmann, J. & Jensen, O.P. (2018) Scientific uncertainty in assessment estimates for

New England groundfish stocks and its impact on achieving target catches. *Canadian Journal of Fisheries and Aquatic Sciences*, *75*, 342-356.

Wiedenmann, J. & Jensen, O.P. (2019) Could recent overfishing of New England groundfish have been prevented? A retrospective evaluation of alternative management strategies. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*, 1006-1018.

Wiedenmann, J., Free, C.M., & Jensen, O.P. (2019) Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks. *Fisheries Research*, *209*, 129–142.

Table 1. List of the 18 stocks in our analysis, with the federal management council listed. All stocks management by the New England Council (NEFMC) except herring are part of the Northeast Multispecies FMP, and are collectively referred to as groundfish. Scup and summer flounder are in a separate FMP in the Mid-Atlantic (MAFMC). The three letter abbreviations (e.g., GBC) are used in Figure 5. Ages refers to the age range used to calculate *Z* in Eqn. 1 for the spring and fall surveys.

Full stock name	Abbreviated name	Scientific name	Council	Ages (spring)	Ages (fall)	Years used*
Georges Bank Atlantic cod	GB cod (GBC)	*Gadus morhua*	NEFMC	3-9	2-10	1978-2014
Gulf of Maine Atlantic cod	GOM cod (GMC)	*Gadus morhua*	NEFMC	4-11	3-12	1982-2014
Georges Bank haddock	GB haddock (GBH)	*Melanogrammus aegelfinus*	NEFMC	3-11	2-12	1978-2014
Gulf of Maine haddock	GOM haddock (GMH)	*Melanogrammus aegelfinus*	NEFMC	2-12	2-11	1978-2014
Georges Bank yellowtail flounder	GB yellowtail (GBY)	*Limanda ferruginea*	NEFMC	4-11	3-11	1981-2014
Cape Cod / Gulf of Maine yellowtail flounder	CC / GOM Yellowtail (GMY)	*Limanda ferruginea*	NEFMC	3-9	3-9	1978-2014
Southern New England / Mid-Atlantic yellowtail flounder	SNE / MA Yellowtail (SNY)	*Limanda ferruginea*	NEFMC	3-10	2-8	1985-2014
Georges Bank winter flounder	GB winter (GBW)	*Pseudopleuronectes americanus*	NEFMC	4-9	3-10	1978-2014
Southern New England / Mid-Atlantic winter flounder	SNE / MA winter (SNW)	*Pseudopleuronectes americanus*	NEFMC	3-9	3-9	1978-2014
Witch flounder	Witch (WCH)	*Glyptocephalus cynoglossus*	NEFMC	5-13	4-12	1978-2014
American plaice	Plaice (PLA)	*Hippoglossoides platessoides*	NEFMC	5-12	4-11	1978-2014
Acadian redfish	Redfish (RED)	*Sebastes fasciatus*	NEFMC	7-27	6-27	1978-2014
Pollock	Pollock (POL)	*Pollachius virens*	NEFMC	3-10	2-12	1988-2014
White hake	White hake (WHK)	*Urophycis tenuis*	NEFMC	4-9	3-9	1978-2014

Atlantic herring	Herring (HER)	*Clupea harengus*	NEFMC	4-8	4-7	1978-2014
Atlantic mackerel	Mackerel (MAC)	*Scomber scombrus*	MAFMC	-	2-8	1978-2014
Scup	Scup (SCP)	*Stenotomus chrysops*	MAFMC	1-8	1-8	1984-2014
Summer flounder	Summer (SFL)	*Paralichtys dentatus*	MAFMC	1-10	1-10	1982-2014

*The range of years where age-based survey data, total survey, and total catch data were all available for a stock

Table 2. Estimates of relative *F* and *Z* averaged across time periods prior to 1995, and from 1995 onward. The ratio reported for relative F and Z are the value from the more recent period divided by the earlier period. For each time period and stock, the averages were calculated across the spring and fall NMFS bottom trawl survey. At the bottom, region-specific averages of the ratios are calculated, with herring and mackerel lumped in their own pelagics group due to their wide distribution across the regions. For mackerel (MAC), the 2010 assessment (Deroba et al. 2010) split the survey time series to deal with the retrospective pattern, which resulted in a much lower reported Mohn's rho (-0.38). We used the rho estimate from the assessment without a survey split. For GOM cod (GMC), two assessment formulations were accepted, one with a fixed M=0.2 across all years, and one with a higher M=0.4 in more recent years (NEFSC 2015b). We used the rho estimate from the M = 0.2 model.

Table 2.

	Correlation between				
Stock	Relative *F* and *Z* (combined)	Relative *F* and *Z* (spring only)	Relative *F* and *Z* (fall only)	Relative *F* between spring and fall	*Z* estimates between spring and fall
GB cod	-0.02	-0.13	0.07	0.69	0.34
GOM cod	0.20	0.02	0.42	0.65	0.47
GB haddock	0.01	-0.04	0.05	0.77	-0.04
GOM haddock	0.05	0.05	0.14	0.56	0.43
GB yellowtail	-0.35	-0.40	-0.30	0.95	0.29
CC / GOM yellowtail	-0.36	-0.54	-0.22	0.79	0.05
SNE / MA yellowtail	0.15	0.05	0.21	0.65	0.36
GB winter	0.13	0.26	0.11	0.36	0.23
SNE / MA winter	0.40	0.53	0.41	0.94	0.32
Witch	-0.31	-0.48	-0.17	0.56	0.31
Plaice	0.35	0.37	0.35	0.83	0.48
Pollock	0.02	0.13	-0.07	0.70	-0.23
White hake	0.15	0.16	0.14	0.34	0.23
Redfish	0.13	-0.69	0.24	-	-
Summer	0.85	0.83	0.91	0.78	0.72
Scup	0.48	0.32	0.73	0.35	0.20
Mackerel	-0.44	-0.44	NA	-	-
Herring	-0.20	-0.47	-0.17	0.27	0.34

Table 2. cont.

stock	Relative *F* (avg. <1995)	Relative *F* (avg. ≥1995)	Relative *F* ratio	Survey *Z* (avg. <1995)	Survey *Z* (avg. ≥1995)	Survey *Z* ratio	Mohn's rho	Source
GB cod	1.43	0.62	0.43	0.66	0.73	1.10	0.68	NEFSC 2013
GOM cod	1.51	0.66	0.44	0.65	0.77	1.20	0.54	NEFSC 2015b
GB haddock	1.47	0.58	0.40	0.56	0.52	0.93	0.50	NEFSC 2015b
GOM haddock	1.84	0.26	0.14	0.35	0.18	0.52	-0.04	NEFSC 2015b
GB yellowtail	1.77	0.31	0.18	1.03	1.25	1.22	1.99	Legault et al. 2013
CC / GOM yellowtail	1.62	0.69	0.42	0.87	1.45	1.67	0.98	NEFSC 2015b
SNE / MA yellowtail	1.51	0.45	0.30	1.06	0.84	0.79	1.06	NEFSC 2015b
GB winter	1.38	0.66	0.48	0.72	0.56	0.78	0.83	NEFSC 2015b
SNE / MA winter	1.77	0.44	0.25	1.31	1.03	0.79	0.21	NEFSC 2015b
witch	1.10	0.89	0.81	0.40	0.50	1.25	0.91	NEFSC 2015b
plaice	1.42	0.62	0.44	0.80	0.66	0.82	0.32	NEFSC 2015b
pollock	1.21	0.82	0.68	0.30	0.36	1.23	0.28	NEFSC 2015b
white hake	1.20	0.83	0.69	1.04	0.82	0.79	0.18	NEFSC 2015b
redfish	1.95	0.16	0.08	0.13	0.17	1.30	0.26	NEFSC 2015b
summer	1.90	0.40	0.21	1.29	0.59	0.46	0.11	Terceiro 2016
scup	1.65	0.64	0.39	1.61	1.14	0.70	-0.08	NEFSC 2015a
mackerel	1.74	0.34	0.20	0.50	1.48	2.98	1.68	Deroba et al. 2009
herring	1.35	0.81	0.61	1.11	0.55	0.49	0.67	Deroba 2015

Region	Avg.	Region	Avg.
GB	0.37	GB	1.01
GOM	0.46	GOM	1.10
SNE/MA	0.48	SNE/MA	0.90
Pelagics	0.40	Pelagics	1.73
All stocks	0.40	All stocks	1.06

Table 3. For the 14 stocks with weak to negative correlations between relative *F* and survey *Z* (≤ 0.2; see Table 2), we calculated the amount of catch overestimation prior to 1995, or the amount underestimation since 1995 needed to produce no change in the mean relative *F* between time blocks.

Stock	Avg. catch up to 1994	Over-reported catch (mt) Average	Min	Max	Avg. catch 1995-on	Under-reported catch (mt) Average	Min	Max
GB cod	40,129	22,679 (57%)	9,066	35,327	7,153	9,296 (130%)	2,371	17,816
GOM cod	13,884	7,805 (56%)	5,065	11,775	5,604	7,196 (128%)	1,889	10,590
GB haddock	11,561	6,956 (60%)	2,809	16,629	13,048	19,709 (151%)	3,688	39,125
GOM haddock	2,892	2,487 (86%)	161	6,581	997	6,108 (613%)	2,089	9,660
GB yellowtail	5,556	4,579 (82%)	1,469	10,318	3,118	14,616 (469%)	745	34,773
CC / GOM yellowtail	1,871	1,078 (58%)	539	2,620	1,276	1,735 (136%)	645	3,559
SNE / MA yellowtail	7,857	5,508 (70%)	515	15,571	756	1,773 (235%)	682	3,398
GB winter	2,953	1,528 (52%)	595	2,219	1,830	1,961 (107%)	902	3,567
witch	3,677	700 (19%)	297	1,287	2,259	531 (24%)	226	813
pollock	16,202	5,256 (32%)	1,407	8,161	6,743	3,238 (48%)	1,781	5,866
white hake	6,980	2,155 (31%)	1,558	2,844	2,830	1,264 (45%)	672	2,126
redfish	4,715	4,337 (92%)	502	13,572	1,416	16,243 (1147%)	4,395	58,360
mackerel	51,009	41,059 (80%)	20,434	69,566	50,378	207,899 (413%)	52,002	463,955
herring	87,288	35,907 (41%)	17,800	49,579	112,716	75,311 (67%)	53,764	96,377

Figure Captions

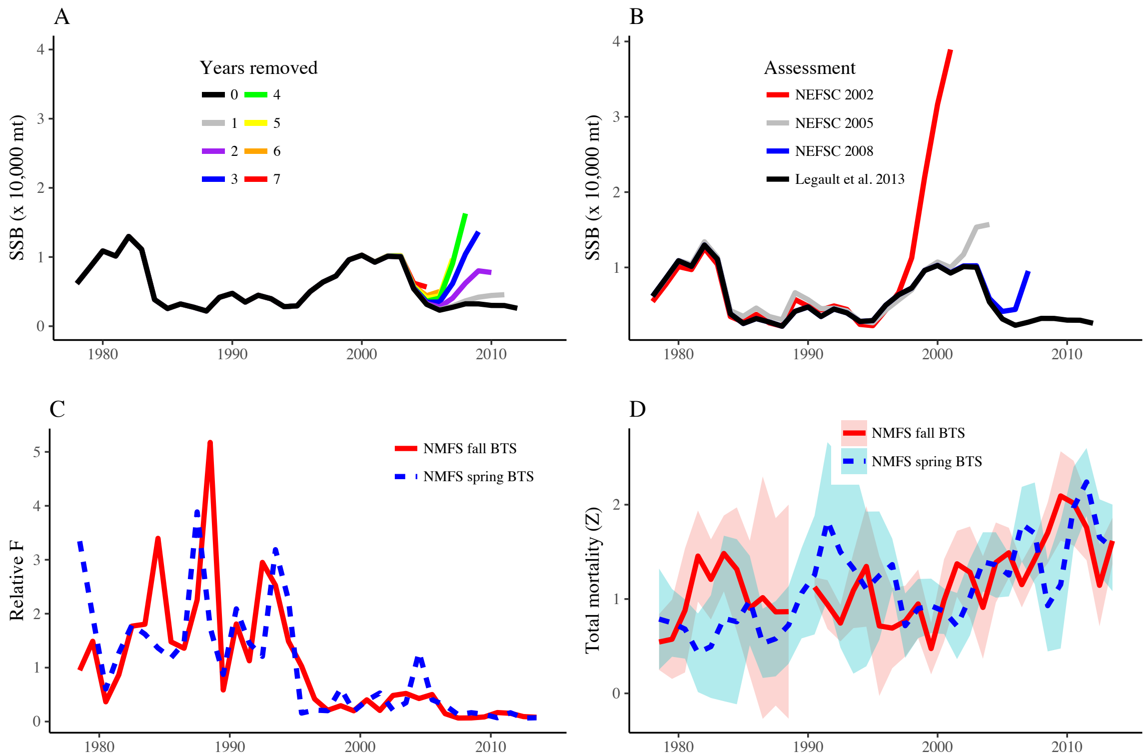

Figure 1. For GB yellowtail flounder: A) Positive retrospective pattern in spawning biomass (SSB) within the 2013 assessment (Legault et al. 2013). B) Historical retrospective pattern across assessments from 2002 onward. C) Relative *F* (total catch in mt divided by the average survey index of abundance in kg tow^{-1}) in the spring and fall bottom trawl survey (BTS), standardized by dividing by the mean value for the entire time series. D) Total mortality (*Z*) from the spring and fall surveys, calculated using the catch curve method of Sinclair (2001) using four years of consecutive data and plot at the midpoint of the year range (e.g., Z calculated using data from 1980-1983 is plotted at 1981.5). The shaded regions represent the 90% confidence intervals for *Z*.

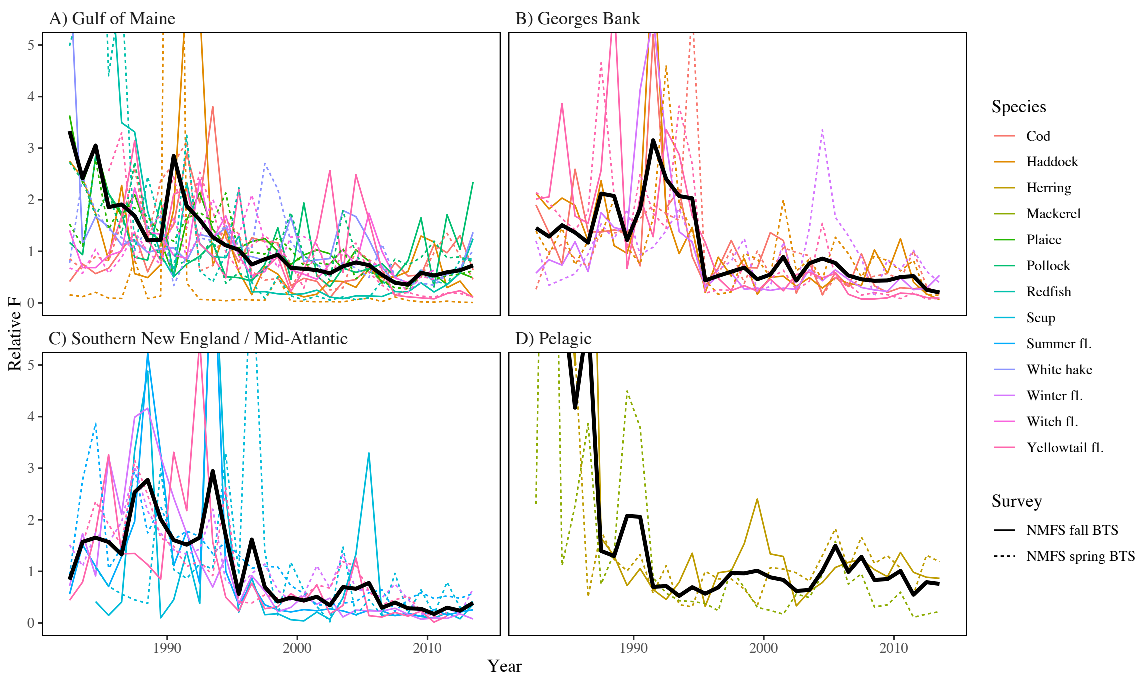

Figure 2. Relative *F* in the spring (dashed lines) and fall (solid lines) bottom trawl survey (BTS) for stocks in the Gulf of Maine (A), Georges Bank (B) and Southern New England / Mid-Atlantic (C) regions, and pelagic (D) stocks of herring and mackerel that are found throughout all of these regions. Pollock, white hake, witch flounder, plaice, and redfish are largely found in the Gulf of Maine region (where they were classified here), although they are also found on Georges Bank. Colored lines represent the individual species, with the same color used for stocks of a species across regions, and the thick black line is the average value across stocks in the region. The y-axis is truncated at a maximum value of 5 to better illustrate the overall trends, as there were some very large values (> 10) in some years for some stocks.

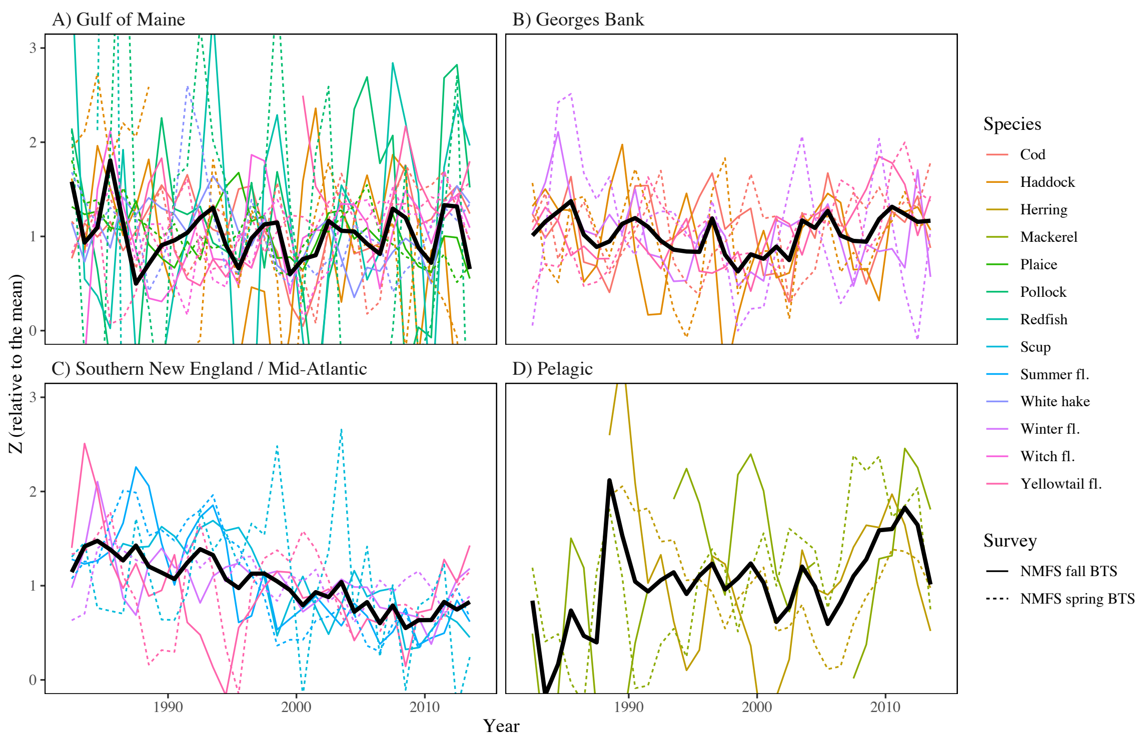

Figure 3. Similar to Figure 2, but showing total mortality (Z; scaled to time series mean for each stock) in the spring (dashed lines) and fall (solid lines) bottom trawl survey (BTS) for stocks in the Gulf of Maine (A), Georges Bank (B) and Southern New England / Mid-Atlantic (C). Colored lines represent the individual species, with the same color used for stocks of a species across regions, and the thick black line is the average value across stocks in the region. The y-axis is truncated between 0 and 3 to better illustrate the overall trends in *Z*.

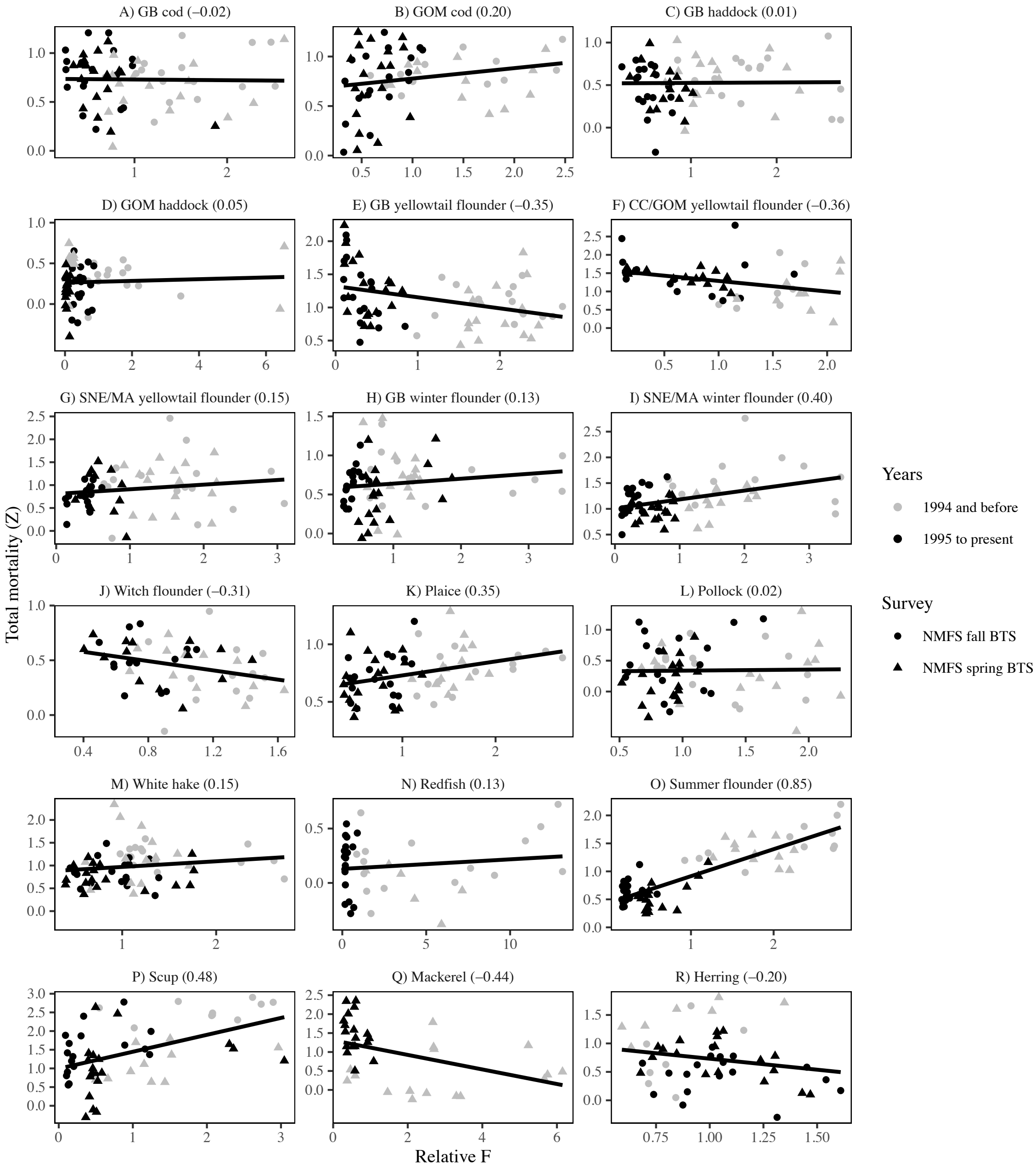

Figure 4. Relationship between the relative *F* (C/ I) and the estimate of total mortality (*Z*), with both metrics calculated using four years of data. The relative *F* is standardized to mean value in the time series, and for all stocks except herring we used the weight-based survey index. Gray points represent years from 1995 onward while black points are from prior to 1995. The circles and crosses denote the spring and fall bottom trawl surveys, respectively. The solid black line is a linear fit to the data aggregated across surveys. The axes are the same for each panel, so some points are not shown for certain stocks. The line shown is the linear fit to the data to show the overall trend in the relationship. The numerical value in parentheses next to the stock name is the correlation coefficient between relative *F* and *Z*.

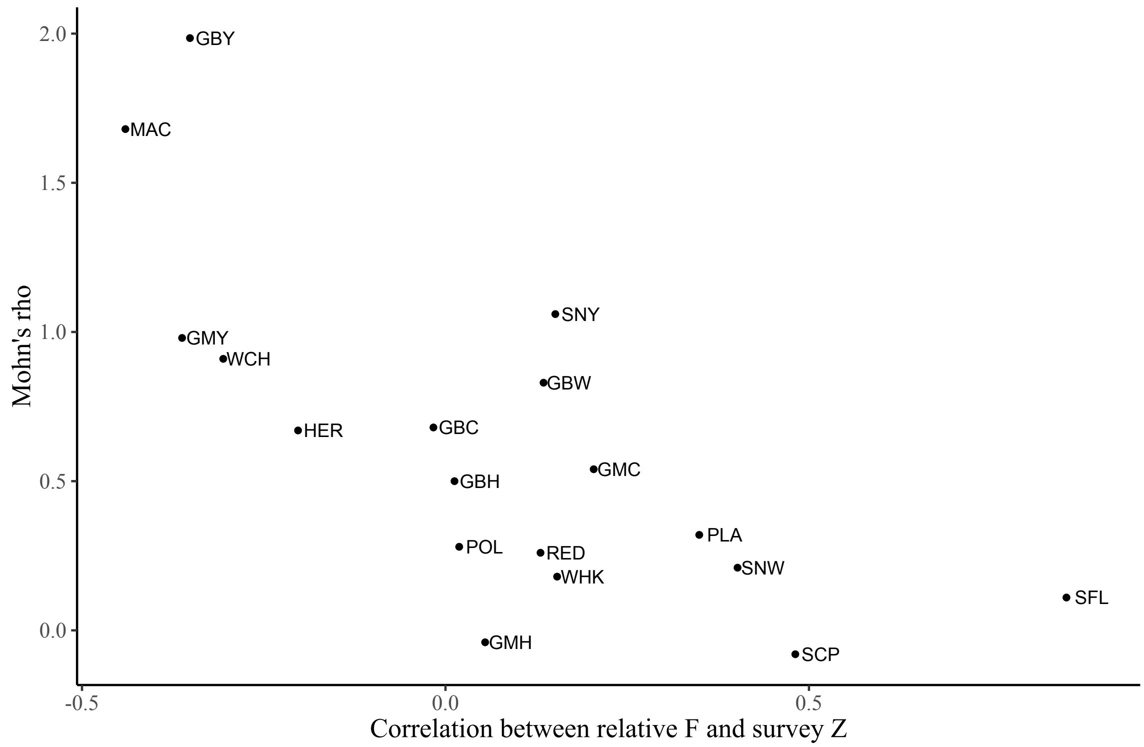

Figure 5. The estimated Mohn's rho for spawning biomass (a measure of the retrospective pattern) from a recent assessment for each stock (denoted by the 3 letter abbreviation; see Table 1) as a function of the correlation between the relative *F* and estimated *Z* from the surveys. Sources for the estimates of Mohn's rho are listed in Table 2.

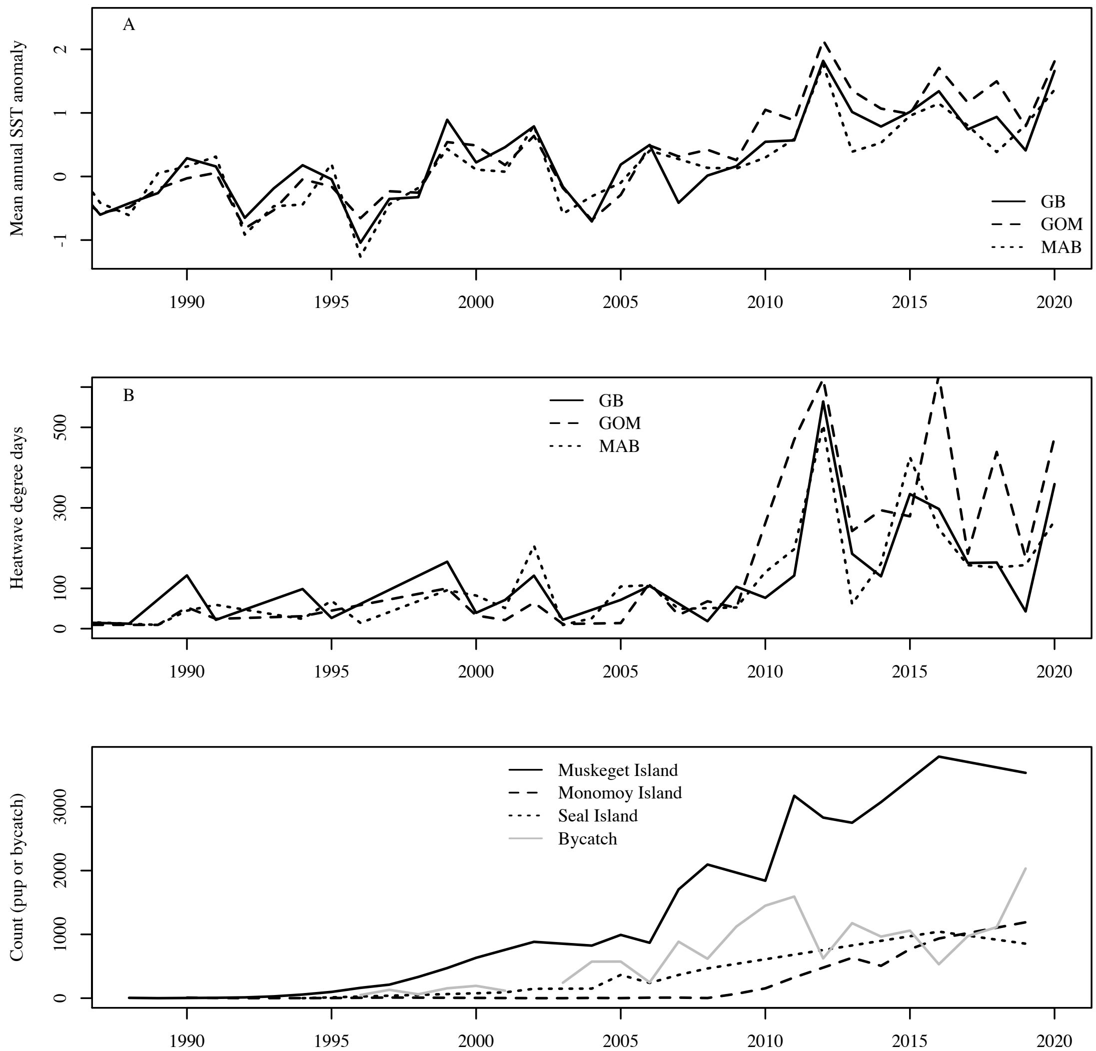

Figure 6. (A) Satellite-derived mean annual sea surface temperature (SST) anomalies (˚C; relative to the 1982-2010 average) by region. (B) Annual degree days when the daily temperature exceeded +1 standard deviation from the long-term average temperature for that day. (C) Five-year rolling averages of grey seal pup counts on three islands in the region (black lines) and of the estimated adult seals bycatch in fisheries in the region (red line). Pups counts and bycatch have been suggested for use as proxies for trends in adult population size in the region (Hayes et al., 2018). Muskeget and Monomoy islands are part of the state of Massachusetts (USA) and Seal Island is part of the state of Maine (USA). All the data used in the figure were obtained from https://github.com/NOAA-EDAB/ecodata.