ICES Journal of
Marine Science

ICES Journal of Marine Science (2020), 77(7-8), 3016–3030. doi:10.1093/icesjms/fsaa184

Original Article

Rose vs. Rho: a comparison of two approaches to address retrospective patterns in stock assessments

Christopher M. Legault *

National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02556, USA

*Corresponding author: tel: +1 508 495 2025; e-mail: chris.legault@noaa.gov.

Legault, C. M. Rose vs. Rho: a comparison of two approaches to address retrospective patterns in stock assessments. – ICES Journal of Marine Science, 77: 3016–3030.

Received 10 March 2020; revised 9 September 2020; accepted 14 September 2020; advance access publication 17 November 2020.

Two approaches to address retrospective patterns in stock assessments are compared. The Rose approach is an ensemble of models that all remove the retrospective pattern through changes in data, parameter values, or model assumptions. It is time intensive and can result in a wide range of historical abundance trends. The Rho approach modifies the terminal year estimates of a single model that exhibits a retrospective pattern. It is fast and easy to apply but results in a discontinuous time series. Neither approach identifies the source of the retrospective pattern. The pros and cons of these two approaches are compared in terms of catch advice and stock status using four examples with varying strength and direction of retrospective patterns. The choice of which approach to use could be based on time and expertise available to conduct and maintain an assessment, with Rose preferred if a lot of both are available while Rho preferred otherwise. If the Rho approach is used, managers should consider adjusting their control rule or risk buffer to account for the difference between Rose and Rho results shown here.

Keywords: ensemble modelling, Mohn's rho, retrospective analysis, stock assessment

Introduction

Retrospective patterns are consistent deviations in estimated parameters of a stock assessment, such as spawning stock biomass (SSB) or fishing mortality rate (F), as recent years of data are removed (Mohn, 1999). Well-behaved stock assessments demonstrate random patterns of positive and negative deviations, not retrospective patterns. Strong retrospective patterns have been used to justify rejecting stock assessments (Punt *et al.*, 2020). Rules of thumb have been developed in different regions to define when a retrospective pattern is strong (Hurtado-Ferro, 2015; Brooks and Legault, 2016). Typically the source of the retrospective pattern cannot be clearly identified (ICES, 2008). This is a problem because incorrectly using the wrong "fix" for a retrospective pattern, meaning making a change to the data or assumed parameter values only to eliminate the retrospective pattern, can lead to poor catch advice (Szuwalski *et al.*, 2018). Other closed-loop simulations have found asymmetries in risk associated with retrospective patterns, meaning that the impact of positive biases is not the same as the impact of negative biases (Hordyk *et al.*, 2019). Ignoring a typical retrospective pattern of SSB increasing as years of data are removed in a stock assessment will often lead to overfishing because the catch will be set too high under the incorrect estimates of population size (Brooks and Legault, 2016).

Two approaches are considered here to address retrospective patterns, referred to as the Rose and Rho approaches. The Rose approach is an ensemble of models that have had the input data or assumed parameters modified to remove the retrospective pattern. The name Rose comes from the line in Shakespeare's Romeo and Juliet that a rose by any other name would smell as sweet. This is because the approach specifically avoids determining the cause of the retrospective pattern and instead intentionally uses multiple approaches to remove the retrospective pattern without concern about whether that source is in fact the correct one. So whether we call the additional source of mortality natural or fishing does not matter, only that some additional source of mortality is needed is important, for example. The Rose approach relies on the power of ensemble models to balance the risks and benefits across a number of plausible scenarios to produce more robust status determinations and catch advice. Recommendations to use model ensembles for catch advice from stock assessments have been made for decades (Patterson *et al.*, 2001) and are seeing a recent resurgence (Karp *et al.*, 2018). Although ensemble approaches are supposed to create more stable estimates as additional years of data are added, Stewart and Hicks (2018) did not always find this to be the case for stock assessment models due to high correlations among different model configurations. They suggest using a small number of models with low correlations in an ensemble. Ensemble approaches have been applied to estimate stock status of marine fisheries around the world (Anderson *et al.*, 2017; Rosenberg *et al.*, 2018). Ensemble approaches also provide a link to management strategy evaluation through the development of multiple operating models (Stewart and Martell, 2015). The Rose approach is a special case of ensemble modelling that is designed to explicitly address stock assessments that exhibit a strong retrospective pattern.

The Rho approach adjusts the terminal year estimates of fishing mortality and SSB for status determination and the terminal year plus one estimates of population abundance at age for catch projections according to the magnitude of the retrospective pattern in the original stock assessment (Brooks and Legault, 2016). It has been applied for a number of years in the Northeast United States [Northeast Fisheries Science Center (NEFSC), 2008]. The approach is used when the rho-adjusted value of either SSB or fishing mortality rate falls outside the uncertainty bound of the terminal year estimate. The Rho approach has been found to perform well in simulations when it adjusts the stock size towards the true value but leads to trade-offs between short- and long-term gains when it does not (Deroba, 2014). The rho adjustment has been found to adjust catch in the correct direction, but to be insufficient to fully eliminate overfishing in some New England groundfish stocks (Brooks and Legault, 2016; Wiedenmann and Jensen, 2018; Wiedenmann and Jensen, 2019).

There are two main hypotheses regarding the cause of retrospective patterns seen in many New England groundfish stock assessments: missing catch and increased natural mortality. Large-scale underreporting of landings has been documented in the New England groundfish fishery in recent years (United States of America, v. Carlos A. Rafael 2017; Bellanger *et al.*, 2019). Both missing catch and increased natural mortality have been explored as possible explanations for the retrospective patterns in two New England groundfish stocks (Rossi *et al.*, 2019). Large increases in natural mortality have been used in stock assessments in a neighbouring region to explain the inability of stocks to recover when fishing has been severely curtailed (Swain and Benoît, 2015). The lack of recovery of Northern cod has been attributed to both increased natural mortality and unreported catch, with different authors finding more evidence for the former (Cadigan, 2016) or latter (Rose and Walters, 2019) depending on the specific assessment approach used. Currently, there is no agreement among experts in the region regarding the relative strength of each hypothesis. Similar debates occur around the world (ICES, 2008; ICES 2020).

This article compares the catch advice and stock status resulting from the Rose and Rho approaches to address retrospective problems through four case studies. The Georges Bank yellowtail flounder (*Limanda ferruginea*) stock assessment has exhibited a strong retrospective pattern for many years, leading to the rejection of analytical stock assessments in 2014 in favour of an empirical approach relying on survey information (Legault *et al.*, 2014). The Gulf of Maine haddock (*Melanogrammus aeglefinus*) stock assessment has recently exhibited a strong retrospective pattern in the opposite direction from that typically seen and uses the Rho approach [Northeast Fisheries Science Center (NEFSC), 2020]. The white hake (*Urophycis tenuis*) stock assessment has only recently exhibited a retrospective pattern strong enough to warrant a rho adjustment [Northeast Fisheries Science Center (NEFSC), 2020]. The proposed age-based stock assessment for witch flounder (*Glyptocephalus cynoglossus*) was recently rejected due to a strong retrospective pattern and the stock is now assessed by an index-based approach [Northeast Fisheries Science Center (NEFSC), 2017a]. Through application of both approaches to these four stocks, general advice is provided for their use in other stock assessments.

Data and methods

The Georges Bank yellowtail flounder data come from the most recent assessment (Legault and Finley, 2019). Analytical approaches to assess this stock were abandoned in 2014 due to the strong retrospective pattern observed for many years in the virtual population analysis of this stock (Legault *et al.*, 2014). For this paper, a statistical catch-at-age model (ASAP; Legault and Restrepo, 1999) was formulated using the data from the recent assessment. This model formulation has not been reviewed or approved by the Transboundary Resources Assessment Committee, the scientific body charged with providing advice on this stock, and is used for demonstration purposes only. The main features of this assessment are three independent surveys and total catch (landings plus discards) available for years 1973–2018, all with age information available for ages 1–6+. Natural mortality was assumed to be 0.4, instead of 0.2 as used in previous virtual population analysis assessments, based on life history information compiled for the empirical benchmark in 2014 (see https://nefsc.noaa.gov/saw/trac/). After tuning the effective sample sizes and input survey coefficients of variation, a strong retrospective pattern remained. For this stock, there was additional information available that was not included in the assessments presented here. Specifically, a survey catchability study was conducted that allowed conversion of recent surveys to absolute biomass (Miller *et al.*, 2018). This information was used to examine the plausibility of estimates of scale from the different model runs.

The Gulf of Maine haddock data and assessment were from the most recent assessment [Northeast Fisheries Science Center (NEFSC), 2020]. The Gulf of Maine haddock assessment used the statistical catch-at-age model ASAP. The main features of this stock assessment are two independent surveys and total catch (landings plus discards available for years 1977–2018, with age information for ages 1–9+ available for all years). The natural mortality rate was assumed to be 0.2 for all years and ages. A strong retrospective pattern was present in this assessment, but the direction was opposite to that typically seen. The survey catchability study was not applicable to Gulf of Maine haddock due to use of a restrictor cable in front of the net opening that was expected to impact round fish catchability and because the location where the study was conducted was not prime habitat for Gulf of Maine haddock.

The white hake data and assessment were from the most recent assessment [Northeast Fisheries Science Center (NEFSC), 2020]. The white hake assessment also used the statistical catch-at-age model ASAP. The main features of this stock assessment are two independent surveys and total catch (landings plus discards) available for years 1963–2018, with age information available for ages 1–9+ beginning in 1989 for the catch, 1968 for one survey, and 1963 for the other survey. The natural mortality rate was assumed to be 0.2 for all years and ages. A strong retrospective pattern was present in this assessment, as well as the previous assessment [Northeast Fisheries Science Center (NEFSC), 2017b], but not the one prior to that [Northeast Fisheries Science Center (NEFSC), 2015]. The survey catchability study was not applicable to white hake for the same reasons given above for haddock.

The witch flounder data and assessment were from the most recent benchmark assessment [Northeast Fisheries Science Center (NEFSC), 2017a]. The ASAP statistical catch-at-age model proposed during this benchmark assessment was not accepted for use by the review panel due to a strong retrospective pattern but is used here for demonstration purposes. The main features of this stock assessment are five independent surveys and total catch (landings plus discards) available for years 1982–2015, with age information available for ages 1–11+ for all years for the catch and all years for each of the surveys (beginning in years 1982, 1982, 1982, 1984, and 2001). The natural mortality rate was assumed to be 0.15 for all years and ages. The strong retrospective pattern observed in this assessment had been observed in previous assessments as well [Northeast Fisheries Science Center (NEFSC), 2015]. The survey catchability study was available for this stock. However, only two of the five surveys used in the witch flounder assessment could be expanded based on the results of the catchability study, so these comparisons are not as direct as for the Georges Bank yellowtail flounder stock.

The commonly used version of Mohn's rho is

$$\rho = P^{-1} \sum\nolimits_{p=1}^{P} \frac{\hat{X}_{Y-p,Y-p} - \hat{X}_{Y-p,Y}}{\hat{X}_{Y-p,Y}},$$

where P is the number of years removed (peeled) from the terminal year (Y), $\hat{X}_{Y-p,Y-p}$ is the estimate of derived quantity X in year $Y-p$ using data through year $Y-p$, and $\hat{X}_{Y-p,Y}$ is the same estimated quantity using data through the full time series of year Y. In the Rho approach, the rho-adjusted estimate of a derived quantity, such as SSB, is computed as the quantity divided by the sum of one plus rho [e.g. rho-adjusted $SSB_{2019} = SSB_{2019}/(1+\rho)$]. Therefore, a derived quantity with a positive Mohn's rho is rho-adjusted to a smaller value while a quantity with a negative Mohn's rho is rho-adjusted to a larger value. For retrospective patterns in the typical direction, this means a rho-adjusted SSB is decreased while a rho-adjusted F is increased. The Rho approach as applied in the Northeast United States only applies the rho adjustment when the rho-adjusted value of either SSB or fishing mortality rate in the terminal year of the assessment falls outside the 90% confidence interval of the associated estimate in the stock assessment. For example, if the terminal year estimate of SSB is 100 mt with a 90% confidence interval of 80–125 mt, then a Mohn's rho of 0.3 would be sufficient to require a rho adjustment because $100/(1+0.3) = 76.9$, which is less than the lower bound of 80. Miller and Legault (2017) recommended always applying rho adjustment due to risk trade-offs because of the challenge of identifying the best confidence interval size to use when making the rho adjustment and the fact that small magnitude Mohn's rho values have small adjustments.

The standard practice in the region is a 7-year peel for retrospective analyses ($P=7$), but for this study, a 5-year peel was used and all Mohn's rho estimates reported here are for 5-year peels. Miller and Legault (2017) found that estimates of Mohn's rho stabilized after five peels. This saved considerable time compared to the standard 7-year peel used for these stock assessments in the Rose approach.

One way to derive the models used in the Rose approach is through the use of a search for multipliers that change the input catch data or assumed natural mortality rate in recent years. Legault (2009) demonstrated that this approach can remove the retrospective pattern from a stock assessment. There are an infinite number of ways to derive models that achieve a value of zero for Mohn's rho for SSB. For this paper, multipliers are applied to either catch or natural mortality either suddenly or with a linear ramp across a range of years. The multipliers comprise two components: weight and magnitude. The first component is the time series aspect, where the weight changes from zero to one either suddenly or through a linear ramp. In this paper, two linear ramps are considered, 4 years and 9 years. The 4-year ramp weight changes from zero to one following the sequence (0, 0.2, 0.4, 0.6, 0.8, 1.0), while the 9-year ramp weight follows the sequence (0, 0.1, 0.2, 0.3, …, 1.0). The first year the weight equals1.0 is defined as the change year in this paper. The second component of the multiplier is the magnitude that is found through a grid search. The application of the overall multiplier is $[1 + \text{weight} \times (\text{magnitude} - 1)]$ for each year.

The multiplier is either applied to the total catch in biomass, with no change to the catch proportions at age, or to the natural mortality rate. The natural mortality rate can have the multiplier applied evenly across all ages, be applied to only young ages (1–3 for Georges Bank yellowtail flounder, 1–4 for Gulf of Maine haddock and white hake, 1–5 for witch flounder), or be applied to only old ages (4–6+ for Georges Bank yellowtail flounder, 5–9+ for Gulf of Maine haddock, 6–9+ for white hake, 6–11+ for witch flounder). The multipliers were first applied for magnitude values of 2–10 in increments of 1 for retrospective patterns in the usual direction or from 0.1 to 0.9 in increments of 0.1 for Gulf of Maine haddock that had the opposite direction retrospective pattern. A finer search at increments of 0.1 (typical retrospective pattern) or 0.02 (opposite) was used to find the Rose model that reduced the Mohn's rho for SSB to as close to zero as possible at this precision of increments. This process could have been continued to result in Mohn's rho values arbitrarily close to zero with finer gridding of the multipliers, but the 0.1 (or 0.02 for the opposite direction retrospective pattern) was sufficient to get the selected Mohn's rho values within 0.03 of zero. This was deemed sufficient removal of the retrospective pattern for the Rose application. It should be noted that a Mohn's rho of zero does not mean that all peels fall exactly on the terminal year estimate; just that the positive and negative deviations cancel.

In this paper, the choice of catch or natural mortality multipliers is referred to as the fix, the time series aspect of the change is referred to as the ramp, the magnitude of the multiplier is referred to as the multiplier, and the first year when the full magnitude of the multiplier is applied is referred to as the change year. The change years selected for this study were 2000, 2005, 2010, and 2015 for Gulf of Maine haddock, white hake, and witch flounder to examine a systematic range. For Georges Bank yellowtail flounder, an additional 2 years were selected to allow for an earlier change (1995) and a change associated with the earliest peel in the Mohn' rho calculation (2013). A combination of fix, ramp, and change year is referred to as a scenario. To be eligible for consideration as a Rose model, a scenario must have a multiplier that results in Mohn's rho for SSB near zero, defined here as within 0.03 of zero. Selection of the Rose models may depend additionally on balance among different hypotheses regarding the source of the retrospective pattern. The Rose estimate is computed as the arithmetic mean of the Rose model values for a given quantity, such as terminal year SSB or a specific year's catch advice.

Stock status was determined using reference points defined by F and SSB per recruit and yield per recruit at 40% of unexploited conditions, common proxies for maximum sustainable yield reference points used in the region. When natural mortality changes within the assessment to eliminate the retrospective pattern, the original natural mortality rate is used in the calculation of the biological reference point because the species has not had enough time to adjust their life history characteristics to account for the new natural mortality rate and because the change in the natural mortality rate may be aliasing missing catch (Legault and Palmer, 2016). The selectivity pattern, weights at age, and maturity at age used in the per-recruit calculations were computed as the average values in the most recent 5 years of the assessment. The SSB reference point was computed simply as the SSB per recruit at F40% times the 75th percentile of recruitment over the entire time series as a pragmatic way to avoid the estimation of stock–recruitment curves when stock productivity may be changing due to changing natural mortality. Using the 75th percentile of recruitment keeps the recruitment for the SSB reference point within the estimated range, but closer to the upper end of the observations as would be expected fitting a stock–recruitment curve to the estimates with a prior on asymptotic recruitment remaining within the observed range. The stock status was determined using the ratio of the terminal year in the assessment and the reference point for both SSB and fully selected fishing mortality rate. The Rho approach used the rho-adjusted values of terminal year SSB and F for status determination while the Rose estimate was the mean of the Rose models' status determination calculations. Following the standard approach for the region, a stock is considered undergoing overfishing when F is greater than Ftarget and a stock is considered overfished when SSB is less than one half of SSBtarget.

For the Rho approach, two rho adjustments were considered for projections (Brooks and Legault, 2016). One was to use Mohn's rho for SSB applied to all ages in the population, and the other was to use the age-specific Mohn's rho values. Both rho adjustments were applied to the start of the terminal year plus one population numbers at age. The basis for using Mohn's rho for SSB applied to all ages is because it is how the determination is made that a rho adjustment is needed. The basis for using the age-specific Mohn's rho values is that these reflect changes by age. Either adjustment is used currently in the Northeast United States [Northeast Fisheries Science Center (NEFSC), 2020]. More research is needed to determine if there are situations in which one adjustment performs better than the other.

Catch advice from both the Rose and Rho approaches was derived for the next 3 years by applying the F40% in short-term projections assuming that the recruitment in each of the projection years was the median of the recent 10 years excluding the terminal year (because it often has higher estimation uncertainty than other years). In the Rose approach when natural mortality multipliers are used, they are continued in the projections. In the Rose approach when catch multipliers are used, the catch advice needs to be adjusted to account for changed input data. The catch advice is computed as the results from the projections divided by the expected catch multiplier in those years. In all the cases examined here, the most recent multiplier is used to determine the catch multiplier because all of the cases have stabilized in the recent years (there are at least 4 years at the end of the time series with the same multiplier). If a scenario was considered that included changes in the catch multiplier that continue to the terminal year, a decision would be required whether to continue the trend in the catch multiplier forward. This decision would be based on local expertise regarding possible sources of the error and how they might change in the near future. A default assumption of the catch multiplier in the terminal year used to adjust the catch advice will be reasonable in most situations, unless something dramatic changes in the near future such as a major change in catch reporting or monitoring.

Table 1. Mohn's rho based on 5-year peel for fully selected fishing mortality (F), SSB, and population numbers at age (*N* age) for Georges Bank yellowtail flounder (GBYT), Gulf of Maine haddock (GOMH), white hake, and witch flounder.

Metric	GBYT	GOMH	White hake	Witch
F	−0.61	0.31	−0.19	−0.44
SSB	2.14	−0.25	0.24	0.55
N age 1	2.59	−0.46	0.77	0.27
N age 2	2.18	−0.35	0.21	0.64
N age 3	1.68	−0.27	0.01	0.60
N age 4	1.83	−0.17	0.04	0.40
N age 5	2.12	−0.14	0.16	0.34
N age 6	3.23	−0.12	0.25	0.42
N age 7	NA	−0.12	0.28	0.39
N age 8	NA	−0.12	0.30	0.49
N age 9	NA	−0.12	0.31	0.71
N age 10	NA	NA	NA	0.89
N age 11	NA	NA	NA	1.09

The *N* age 6 for GBYT, *N* age 9 for GOMH and white hake, and *N* age 11 for witch flounder denote plus groups. NA denotes not applicable.

All calculations were done using R (R Core Team, 2018). Code and data are available at https://github.com/cmlegault/rosevrho.

Results

The Mohn's rho values for SSB and F were much larger for Georges Bank yellowtail flounder than white hake or witch flounder, while the direction was opposite for Gulf of Maine haddock (Table 1, Figure 1, and Supplementary Figure S1). The rho adjustment reduced the terminal year SSB to 0.32, 0.80, and 0.64 of its original estimate for Georges Bank yellowtail flounder, white hake, and witch flounder, respectively, while increasing the terminal year SSB to 1.34 times its original estimate for Gulf of Maine haddock (Figure 1 and Supplementary Figures S2, S4, S6, and S8). The direction of the rho adjustment for terminal year F was opposite that of SSB for all four stocks (Supplementary Figures S1, S3, S5, S7, and S9).

The ability to remove the retrospective pattern using catch or natural mortality multipliers varied widely among the stocks with the magnitude of the multipliers associated with the direction and magnitude of the retrospective pattern in the original assessment (Figures 2–5). White hake and witch flounder could achieve

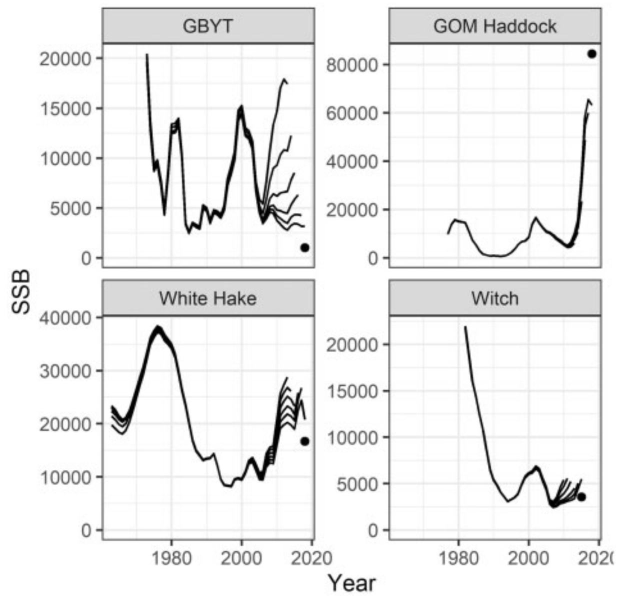

Figure 1. SSB retrospective patterns for Georges Bank yellowtail flounder (GBYT), Gulf of Maine (GOM) haddock, white hake, and witch flounder. The lines in each panel denote the SSB from the four original stock assessments and five retrospective peels (number of years removed from the full data set). The dot denotes the rho-adjusted value of the terminal year estimate.

Mohn's rho for SSB within 0.03 of zero for the majority of the 48 combinations of fix, ramp, and change year examined, resulting in 46 and 44 Rose models, respectively. Georges Bank yellowtail flounder could only sufficiently remove the retrospective pattern for one of the change years using catch multipliers ≤ 10 (Figure 2). To provide balance among the four fixes, the three natural mortality rate fixes were limited to a single change year selected based on the Mohn's rho for SSB closest to zero within each fix and ramp combination. This resulted in 12 Rose models for Georges Bank yellowtail flounder. Gulf of Maine haddock could only sufficiently remove the retrospective pattern for the natural mortality fix applied to all ages for multipliers greater than 0.1 (Figure 3). The 2015 change year for the sudden ramp used with this fix did not sufficiently remove the retrospective pattern, leaving only 11 Rose models for Gulf of Maine haddock.

The changes in Mohn's rho for SSB did not always follow a smooth pattern due to individual peels occasionally going to wildly different solutions. For example, in Figure 2, the Cmult fix and Ramp4 combination with change year 2005 scenario has a Mohn's rho value for a multiplier of 4 that does not follow the pattern of all the other multipliers. This prevented full automation from being used to search for the multiplier that resulted in a Mohn's rho of exactly zero. Generally, Mohn's rho values declined in a consistent pattern as the multiplier increased, allowing the gridding process to work smoothly.

The changes to the data or parameter values associated with the Rose models for the four stocks resulted in changed perceptions of the historical time series of SSB (Figures 6–9). The Georges Bank yellowtail flounder time series of SSB was generally quite similar in the early years, but the Rose models had higher estimates in early 2000s and lower estimates in the most recent years compared to the original assessment (Figure 6). The timing of the change in SSB from the original run depended on the change year with the change in SSB generally starting a few years prior to the change year and even more years prior the longer the ramp. The Gulf of Maine haddock time series of SSB followed the same pattern of a large increase at the end of the time series due to large recruitment events, but the original assessment shows a levelling off of the SSB in the most recent years while the 11 Rose models all show continued increase (Figure 7). The magnitude of the SSB increase in the early 2000s also changed by varying amounts across the Rose models for this stock.

The changes to SSB for white hake using the Rose models occurred both early and recently and were not consistent across the 46 Rose models (Figure 8). Some models, such as catch multipliers with a 9-year ramp and change year of 2000, resulted in higher SSB estimates for all years compared to the original run, although the terminal year was quite similar. Other models, such as all the natural mortality multipliers applied to all ages suddenly, estimated lower SSB in the earliest years, higher SSB in the 2000s, and lower SSB in the most recent years. This was the most common pattern across the Rose models. Some models, such as natural mortality multipliers applied to young age with a 4-year ramp with change year 2010, estimated SSB at or below the original run for all years. Witch flounder exhibited similar trends in SSB over time across the 44 Rose models, but there were noticeable shifts in scale for some models and the trend in the most recent years varied considerably (Figure 9). The changes in model estimates of SSB time series cannot be predicted a priori and must be examined for each model in the Rose approach.

For all four stocks, the directional changes in the status determination ratios F/Ftarget and SSB/SSBtarget were the same using both the Rose and Rho approaches (Figure 10 and Supplementary Figures S2–S9). For the three stocks with the typical retrospective pattern (Georges Bank yellowtail flounder, white hake, and witch flounder), the F/Ftarget increased and SSB/SSBtarget decreased compared to the original assessment, while F/Ftarget decreased and SSB/SSBtarget increased for Gulf of Maine haddock, the stock with the opposite retrospective pattern. Status determination for Georges Bank yellowtail flounder did not change from the original assessment with both approaches resulting in the classifications of not overfishing (F/Ftarget < 1.0) but overfished (SSB/SSBtarget < 0.5). The range of ratios varied considerably among the Rose models, but all were in the same direction relative to the original run. The changes to the terminal year and target values for Georges Bank yellowtail flounder for SSB and F are shown in Supplementary Figures S2 and S3, respectively. Only one of the Rose models has a different status of overfishing, but the remaining 11 Rose models, as well as the Rose estimate, all have F/Ftarget well below 1.0. The rho-adjusted ratios are within the range of Rose models estimates but do not increase F/Ftarget or decrease SSB/SSBtarget as much as the Rose estimate.

The Gulf of Maine haddock status plot shows the opposite directional change in the ratios from the other three stocks due to the opposite direction of the retrospective pattern in the original assessment. However, the Rose and Rho approaches change in the same direction from the original assessment. The Rho approach does not result in as large a change in the ratio as the Rose approach for SSB but has a slightly larger change for F. The few Rose models available for this stock are quite consistent with each other for the status determination, but the Rose models do vary

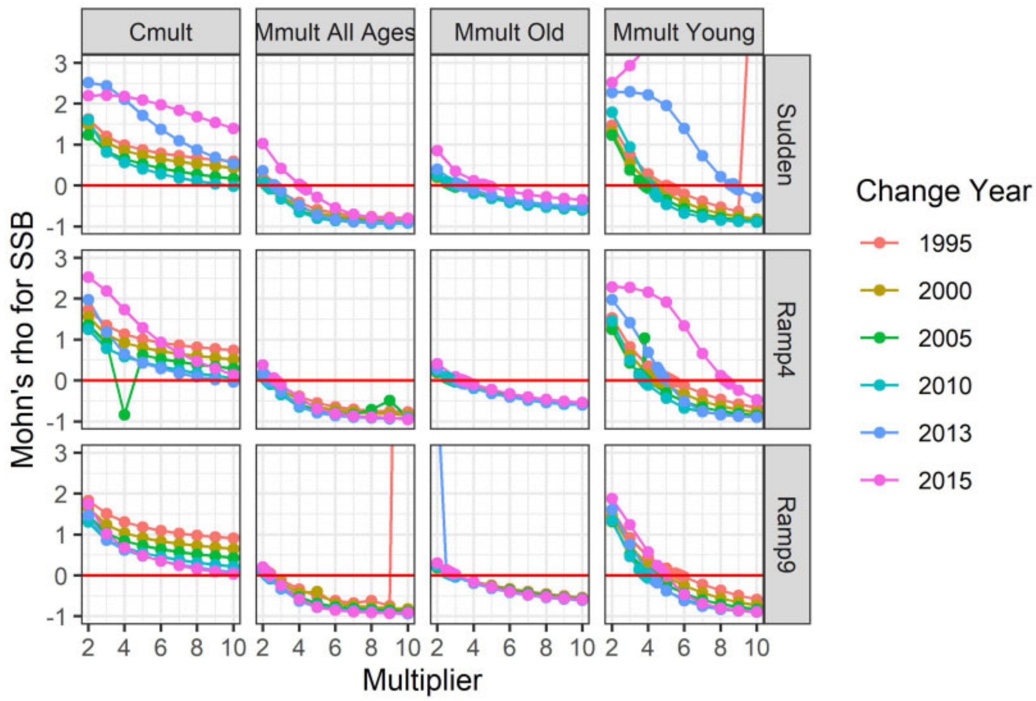

Figure 2. Mohn's rho for SSB from 5-year peels for Georges Bank yellowtail flounder for a range of multipliers for four fixes, three ramps, and six change years. The zero line (in red) is highlighted in each panel to denote where there is no retrospective pattern.

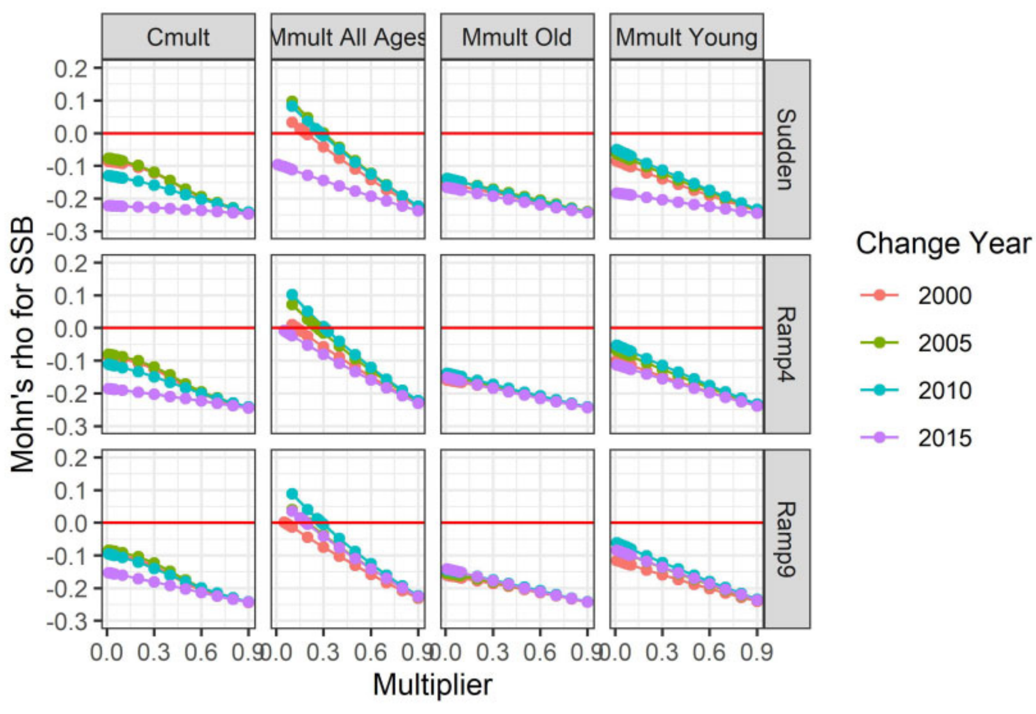

Figure 3. Mohn's rho for SSB from 5-year peels for Gulf of Maine haddock for a range of multipliers for four fixes, three ramps, and four change years. The zero line (in red) is highlighted in each panel to denote where there is no retrospective pattern.

considerably in the terminal year and reference points (Supplementary Figures S4 and S5).

The white hake status plot followed a similar pattern to the Georges Bank yellowtail flounder status plot with the rho adjustment not changing the ratios as much as the Rose estimate. This difference resulted in different status determinations with the rho-adjusted result being not overfishing and not overfished, while the Rose estimate was overfishing and overfished. All of the

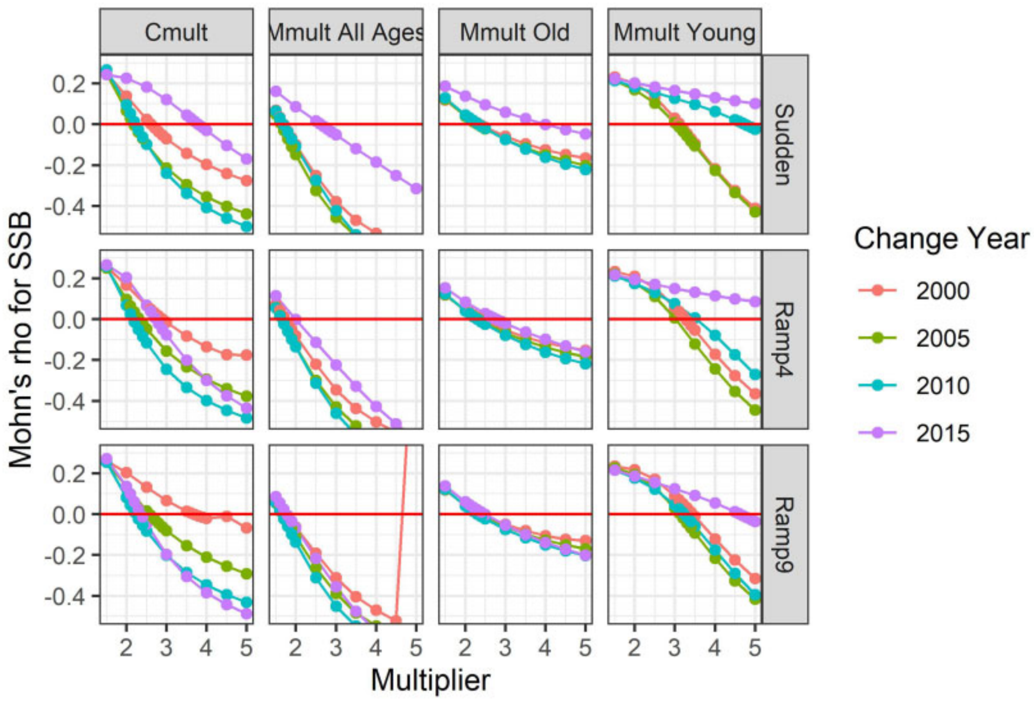

Figure 4. Mohn's rho for SSB from 5-year peels for white hake for a range of multipliers for four fixes, three ramps, and four change years. The zero line (in red) is highlighted in each panel to denote where there is no retrospective pattern.

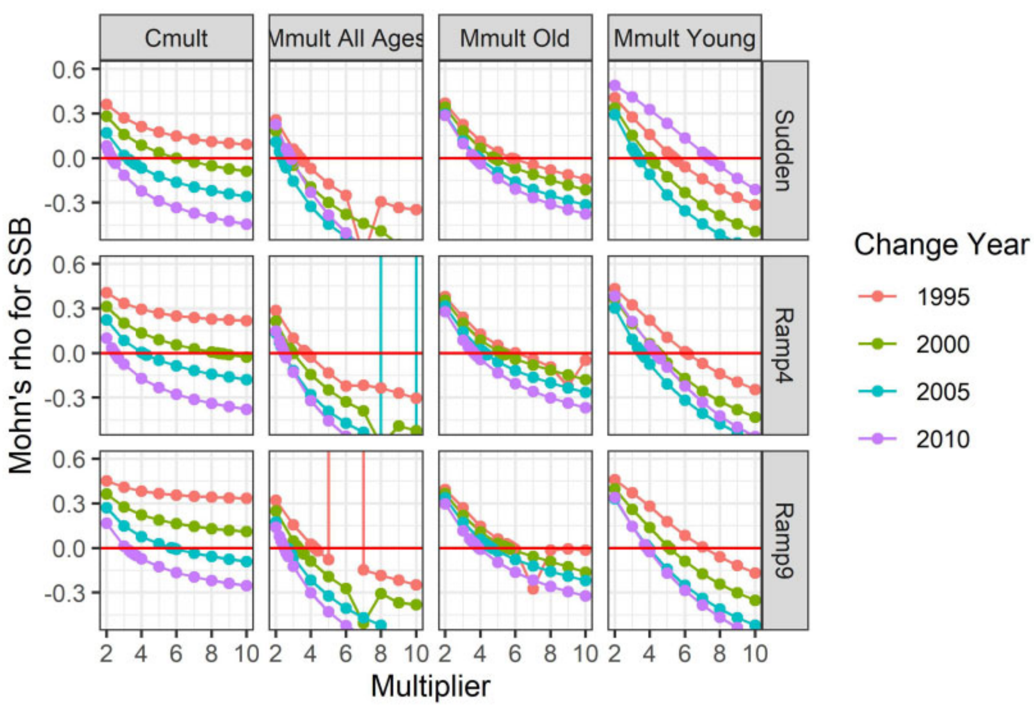

Figure 5. Mohn's rho for SSB from 5-year peels for witch flounder for a range of multipliers for four fixes, three ramps, and four change years. The zero line (in red) is highlighted in each panel to denote where there is no retrospective pattern.

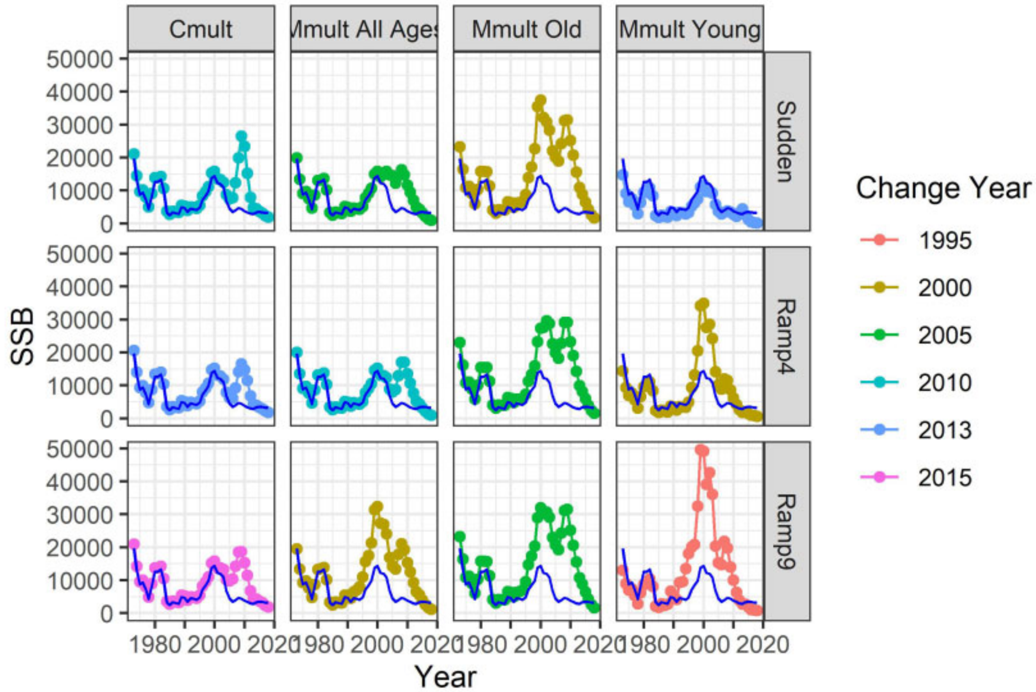

Figure 6. Georges Bank yellowtail flounder SSB (mt) for 12 Rose models that remove the retrospective pattern (coloured lines with dots) and the original run (blue solid line).

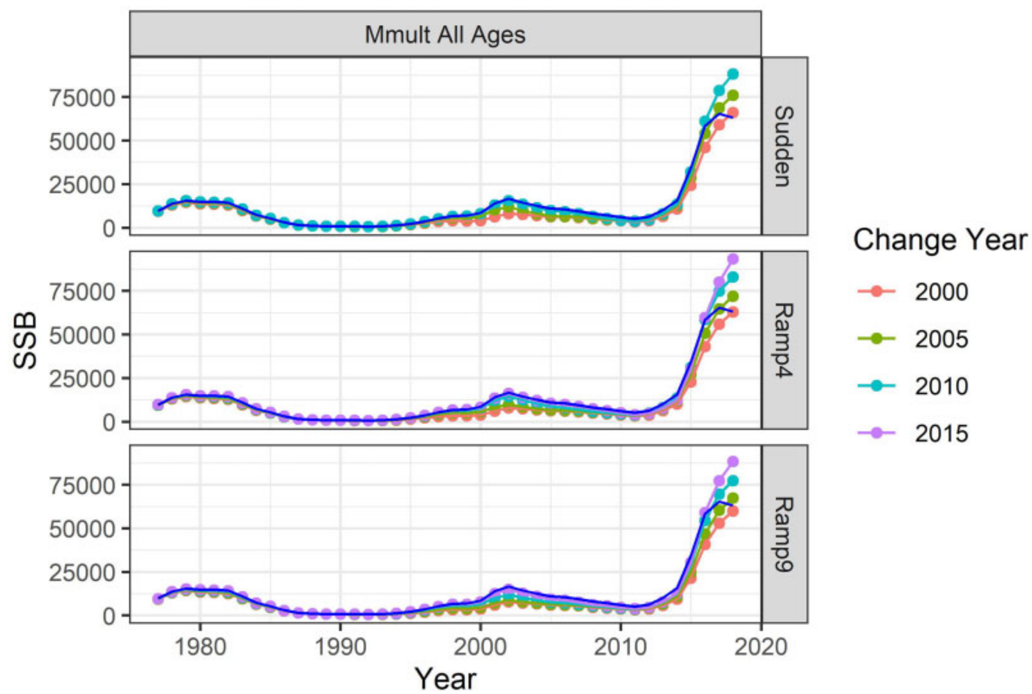

Figure 7. Gulf of Maine haddock SSB (mt) for 11 Rose models that remove the retrospective pattern (coloured lines with dots), and the original run (blue solid line).

Rose models had larger changes in both ratios than the rho adjustment for this stock. Most of the Rose models had the same status as the Rose estimate, but not all did, with some resulting in not overfishing but overfished and one model resulting in overfishing but not overfished. However, the vast majority of Rose models resulted in overfishing and overfished. The different fixes generally group together in the plot.

The witch flounder status plot shows the original status of not overfishing changing to undergoing overfishing for both the Rose and Rho approaches, while the stock is overfished in all three cases.

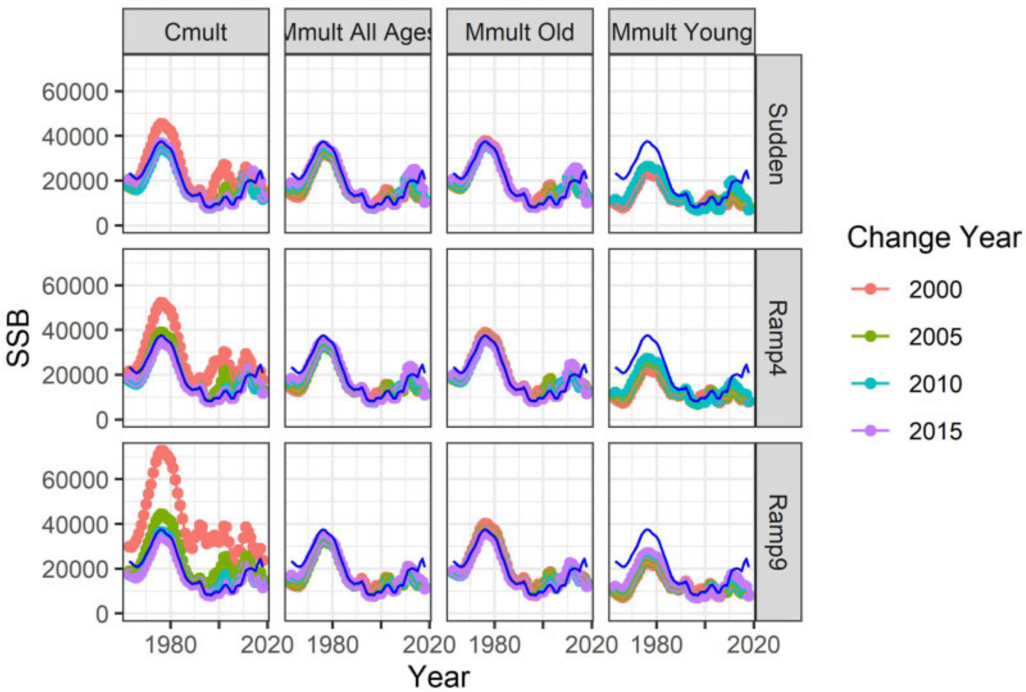

Figure 8. White hake SSB (mt) for 46 Rose models that remove the retrospective pattern (coloured lines with dots) and the original run (blue solid line).

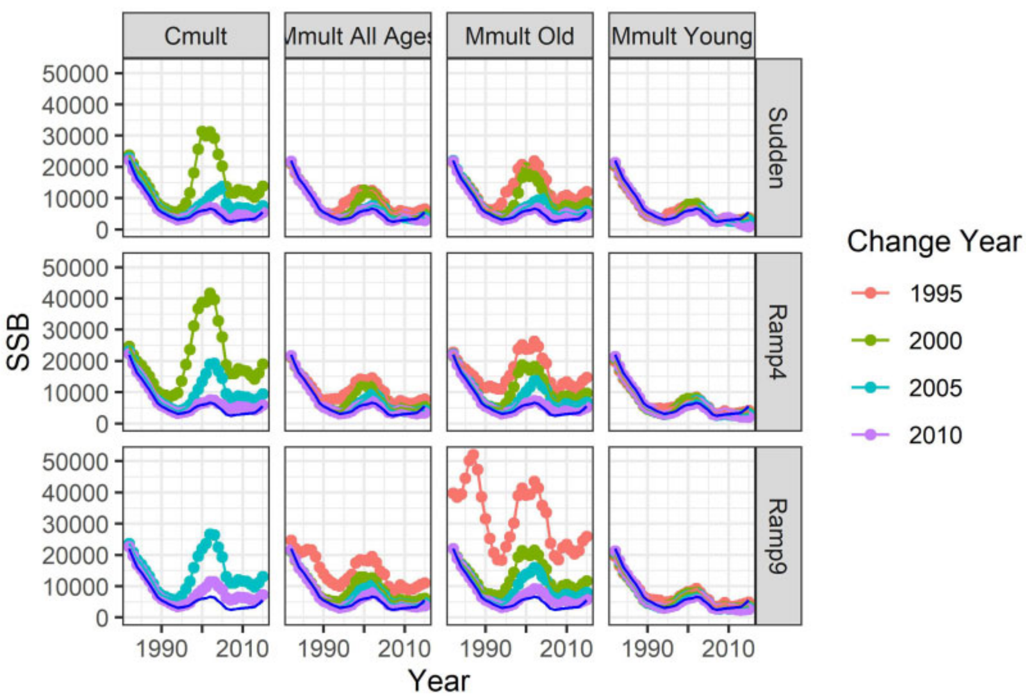

Figure 9. Witch flounder SSB (mt) for 44 Rose models that remove the retrospective pattern (colored lines with dots) and the original run (blue solid line).

Similar to white hake, the Rose models group by fix, with the catch multipliers and natural mortality multipliers applied to young ages fixes resulting in overfishing while the natural mortality multipliers applied to all ages or old ages generally resulting in not overfishing. Thus, the status determination for witch flounder could be changed between overfishing and not overfishing depending on the fixes used in the Rose approach. As in the other three stocks, the Rho approach did not move the ratios as much as the Rose approach.

Similarly, the catch advice from the two rho adjustments was in the same direction as the Rose estimate, compared to the

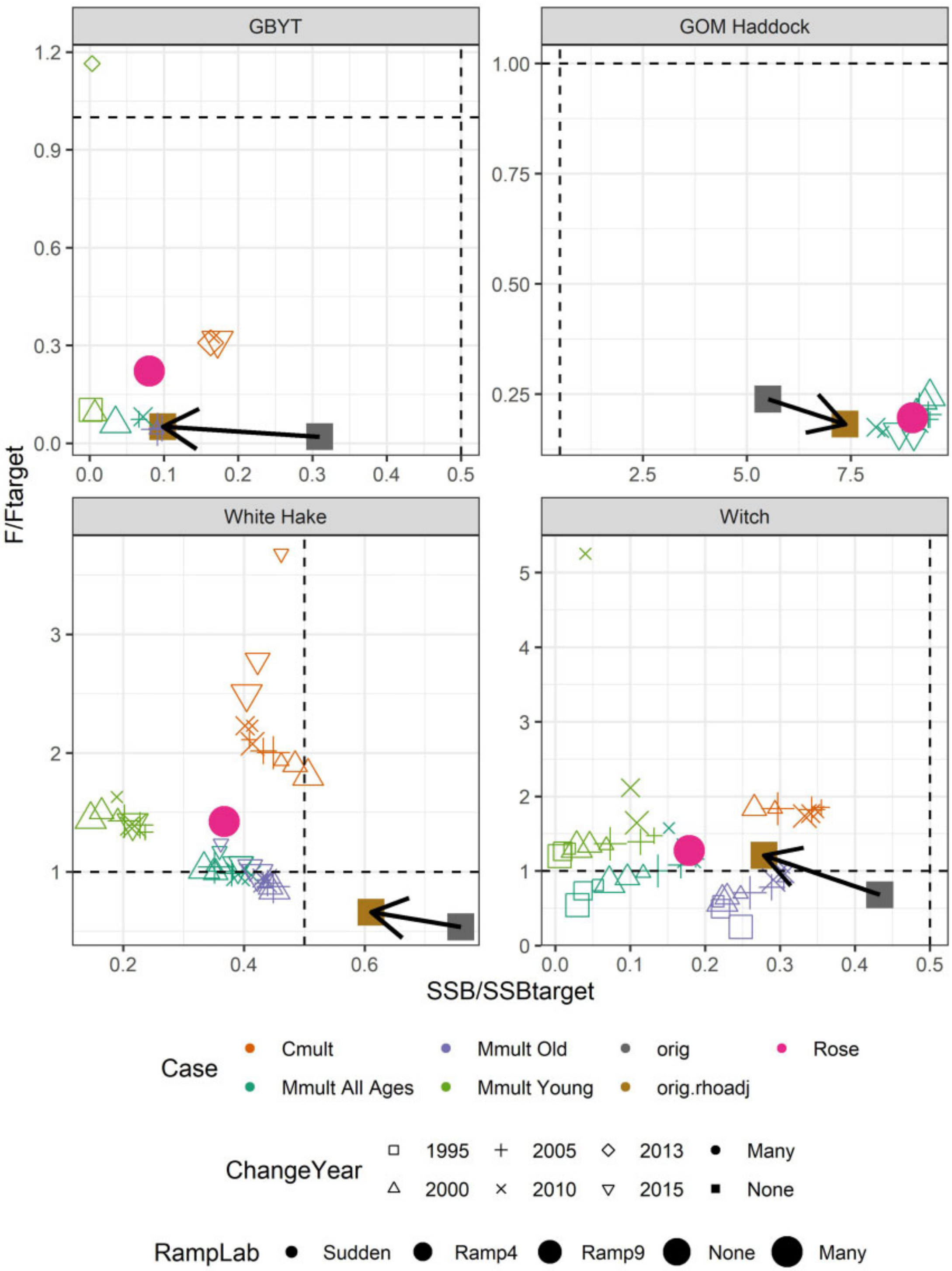

Figure 10. Status plot for Georges Bank yellowtail flounder (GBYT), Gulf of Maine (GOM) haddock, white hake, and witch flounder showing ratios of fully selected F and SSB to target values (F40% and associated SSB40%) for the original model, rho-adjusted (arrow shows change from original to rho-adjusted), the Rose models (open symbols), and the Rose estimate (filled circle). The horizontal dashed lines in each panel denote the overfishing threshold (F > Ftarget) while the vertical dashed lines in each panel denote the overfished threshold (SSB < 0.5 SSBtarget).

original assessment, but smaller in magnitude (Figure 11). For Georges Bank yellowtail flounder, the catch advice for the original run, based on F40%, was higher than both rho-adjusted values and all the Rose models in the first year, 2019, but there were a few Rose models that resulted in higher catch advice by the third year, 2021. The two rho adjustments resulted in nearly identical catch advice in all 3 years, which is due to the Mohn's rho not varying considerably by age (Table 1). The Rose models produced

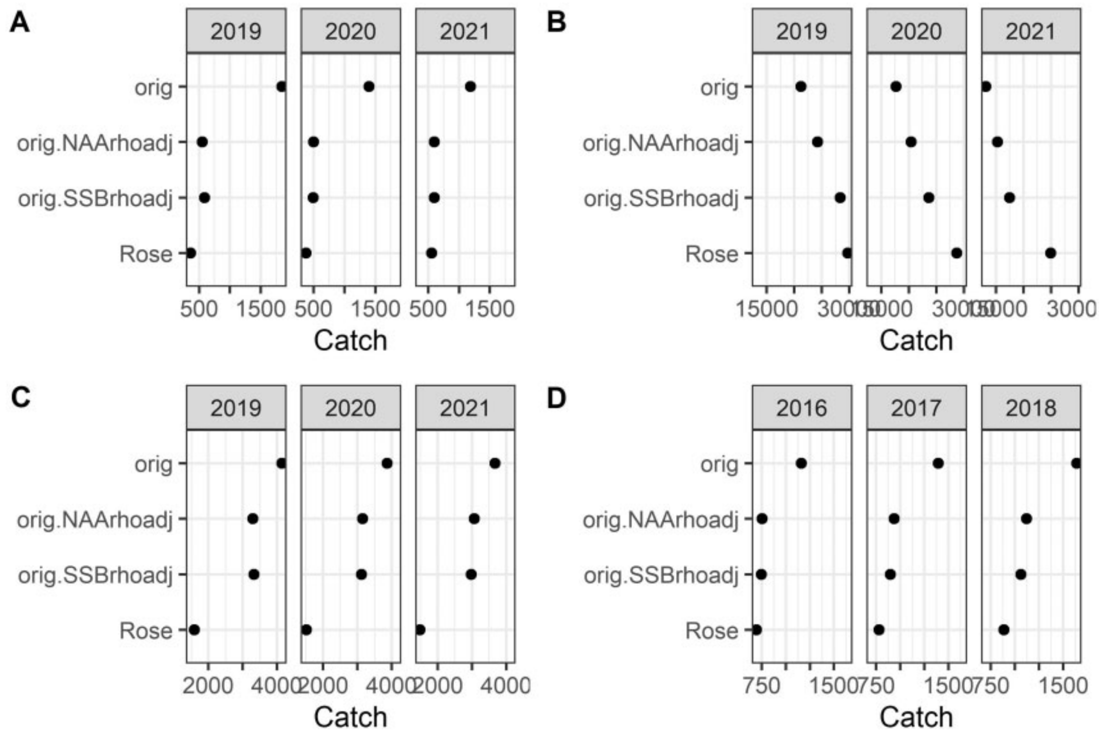

Figure 11. Catch advice (mt) for (a) Georges Bank yellowtail flounder, (b) Gulf of Maine haddock, (c) white hake, and (d) witch flounder for 3 years fishing at the target F40% from the original model, two rho-adjusted models, and Rose approach (see Supplementary Figure S10 for an expanded version of this figure including all the Rose models).

a wide range of catch advice across the 12 scenarios, but the Rose estimate was below rho-adjusted catch advice in all 3 years.

For Gulf of Maine haddock, the catch advice from both rho adjustments and the 11 Rose models was higher than the original run, due to the opposite direction of the retrospective pattern in the original assessment. The difference between the two rho adjustments was greatest for this stock, of the four examined in this paper, due to a combination of variability in Mohn's rho by age and strong cohorts entering the exploited ages. This increase in catch advice when confronted with a retrospective pattern is not universally accepted (see Discussion below) but does result from both the Rose and Rho approaches for this stock.

For white hake, the original run produced catch advice that was higher than both rho adjustments and all 46 Rose models. The two rho-adjusted approaches produced similar catch advice but varied a bit over time due to the large difference in Mohn's rho at age 1 (Table 1). These differences would become much larger if additional years were projected due to this difference in Mohn's rho at age 1, until this cohort passed through the projection years. Similar to the status determinations, all the Rose models resulted in lower catch advice than the two rho adjustments. Patterns in the catch advice from the Rose models can be seen for white hake (Supplementary Figure S10). In general, lower catch advice resulted from the catch multipliers than for the natural mortality fixes and for more recent change years than earlier change years, although there were exceptions. The Rose catch advice was approximately half the rho-adjusted catch advice for this stock for all 3 years.

The witch flounder catch advice followed similar patterns to the white hake catch advice, with the lowest catch advice provided by the Rose approach, the highest catch advice provided by the original assessment, and the rho-adjusted catch advice intermediate between these two. There was one Rose model that provided higher catch advice than the original assessment (Supplementary Figure S10), unlike the Georges Bank yellowtail flounder and white hake results. The pattern of decreasing catch advice for Rose models with more recent change years is more clearly seen for witch flounder than white hake, but again there are exceptions to this general result.

The Georges Bank yellowtail flounder and witch flounder stock assessments have additional data that were not included directly in the statistical catch-at-age models that can be used as a diagnostic. Specifically, the survey catchability study allows the survey biomass to be expanded to a population estimate, while the assessment model used stratified mean catch per tow as the survey value. This expansion of survey biomass is the basis for the current empirical approach used for management for both stocks [Legault and Finley, 2019; Northeast Fisheries Science Center (NEFSC), 2020]. The expanded survey biomass could have been used in the stock assessment model with a fixed survey catchability (or strong prior) but was not to demonstrate this type of external information comparison. For Georges Bank yellowtail flounder, the original run shows quite a different pattern in biomass in recent years compared to the expanded survey biomass, while most of the Rose models show a more similar decline (Supplementary Figure S11). The rho-adjusted value in the terminal year is nearly identical to the Rose estimate and the expanded survey biomass for this stock (Figure 12). For witch flounder, the original assessment shows a different trend from the expanded survey biomass, but the rho adjustment of the terminal year makes the trend more similar, while the Rose models follow a similar trend to the expanded survey biomass (Supplementary Figure S12). The original model and most of the Rose models are well below the expanded survey biomass. Due to the different surveys included in the assessment and the expanded survey biomass, whether these differences in scale are sufficient to reject specific Rose models requires local knowledge and expertise. For witch flounder, the Rose estimate in the terminal year is in the opposite direction of the rho adjustment compared to the original model, and all three values are well below the expanded the survey biomass (Figure 12).

Discussion

The retrospective patterns show consistent directions for all four stocks, with all SSB peels above and all F peels below the assessment with full data for Georges Bank yellowtail flounder, white hake, and witch flounder, while Gulf of Maine haddock has the opposite direction retrospective pattern. This consistency in directional change of peels is a common pattern seen in many stock assessments (ICES, 2020) and highlights uncertainty in the estimated scale of biomass in the models, particularly in the terminal year of the assessment. To avoid providing catch advice that is unsustainable, it is important to examine consecutive assessment estimates for consistent patterns of over or under estimation relative to the most recent result, as this trend would be expected to continue into the future and impact subsequent assessments and catch advice. Mohn's rho has been demonstrated to be an important diagnostic tool for identifying a retrospective pattern, with large positive or negative values corresponding well with a consistent retrospective pattern, while values of Mohn's rho close to zero usually indicate a well-behaved model in this respect. However, it is important to realize that a non-zero Mohn's rho can result from a situation where not all the peels are in the same direction. In fact, a single peel can dominate the estimation of Mohn's rho in the opposite direction of all the other peels. This situation is not considered to exhibit a true retrospective and should be treated differently than the cases examined here (ICES, 2020).

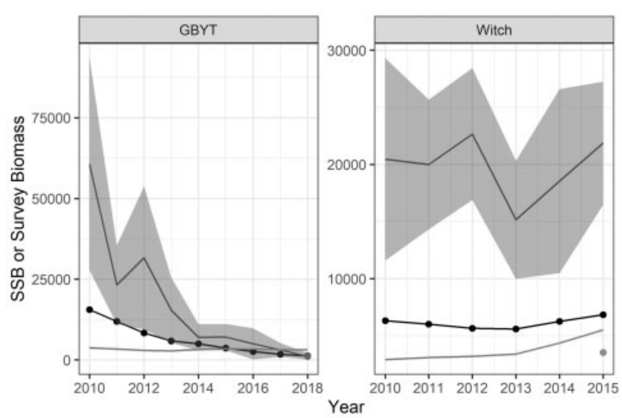

Figure 12. SSB from the original model (red line), rho-adjusted terminal year (red dot), and the Rose estimate (black line with dots) compared to estimated biomass from survey expansion using catchability study results (blue line with 95% confidence interval) in recent years for Georges Bank yellowtail flounder (GBYT) and witch flounder.

The four fixes used to remove the retrospective pattern in the Rose approach are labelled catch and natural mortality, but these are just labels that refer to how the additional mortality is applied—proportional to the observed removals or at a constant rate over time. The catch fix could actually be removals from the population by predators other than humans that happens to follow the same time trend as the fishery removals. The natural mortality fixes could be removals due to fishing that occur at a constant rate over time, instead of reflecting changes in the fishing intensity associated with the observed catch. The age distribution of the additional removals is also important, as demonstrated by comparing results from the three natural mortality fixes. The natural mortality at young ages fix could be a discarding of small fish associated with a fleet that is independent of the directed fleet creating the observed catch. Similarly, the natural mortality at old ages fix could be illegal, unreported, or unregulated catch of large fish by a fleet that is independent of the directed fleet creating the observed catch. The labels catch and natural mortality for the fixes are useful shorthand, but the actual source of the retrospective pattern may be more complex than these simple labels indicate, with the possibility of multiple factors occurring simultaneously.

There are other possible fixes to eliminate the retrospective pattern for use in the Rose approach. Maunder and Piner (2017) noted a number of model misspecifications that can cause bias in assessments. Legault (2009) demonstrated that splitting survey time series in virtual population analysis assessments reduced the retrospective pattern. However, this fix does not work as well for reducing retrospective patterns in statistical catch-at-age assessments. An alternative for statistical catch at age is to use multipliers for recent indices similar to the catch and natural mortality multipliers demonstrated here. This approach was attempted for all four stocks but was not successful, with the Mohn's rho for SSB not decreasing to zero for most change year and ramp combinations with multipliers up to 10 or down to 0.1. Further research is needed on how to adjust survey indices, for example by changing survey catchability directly for recent years, for use in the Rose approach. Other possible fixes that could be explored are changes in selectivity (Stewart and Martell, 2014), accounting for spatial impacts (Goethel *et al.*, 2015; Cao *et al.*, 2017), or using state-space models (Rossi *et al.*, 2019; Perretti *et al.*, 2020). Future research should examine a wider range of possible fixes to retrospective patterns and look for rules of thumb for how many are sufficient to form a robust Rose model.

The Rose approach used the arithmetic mean to derive the Rose estimate of quantities from the multiple Rose models in this paper, but other measures of central tendency could be used. Using the arithmetic mean is equivalent to giving equal weight to all of the Rose models. How to weight the component models in an ensemble is an active area of research (Karp *et al.*, 2018; Spence *et al.*, 2018). The white hake and witch flounder examples demonstrate the importance of selecting appropriate Rose models. The F and SSB status determination ratios separated by scenario groupings (Figure 10), meaning that the Rose estimate could differ substantially depending on the composition of the set of Rose models considered. In the white hake case, the status determination would not be expected to change due to the majority of the Rose models having the same status determination. However, as noted above, the selection of Rose models could change the overfishing status for witch flounder. The process for selecting the Rose models can be influential and requires local knowledge in addition to the modelling expertise.

Despite some patterning emerging from different fixes (see Figures 6–11 and Supplementary Figures S2–S9), the changes in model estimates of SSB and F over the entire time series cannot be predicted a priori and must be examined for each model in the Rose approach (Figures 6–9). Trying to apply a general rule of thumb, such as adding catch will cause F to increase, does not work because the flexibility of modern stock assessment models allows for compensation in many places during the estimation of parameters. However, the Rose models should not be selected based on outcomes, especially catch advice, but rather based on plausible hypotheses and an attempt to balance different possibilities. One should avoid "stacking the deck" to create a desired outcome in terms of either stock status determination or catch advice. This can be accomplished by selecting the Rose models before the status determinations or catch advice is calculated and by applying local knowledge to limit the range of possible multipliers for different fixes.

Closed-loop simulation testing of both the Rose and Rho approaches over a wide range of situations would improve the understanding of when one method is preferred to the other. For example, Van Beveren *et al.* (2020) demonstrated the problems of missing catch when trying to rebuild a stock through closed-loop simulation. Consistently, misreported catch does not appear to cause management problems (Rudd and Branch, 2017) but also does not create a retrospective pattern. A change in the missing catch over time can lead to overfishing (Rudd and Branch, 2017) and retrospective patterns (Legault, 2009). Ignoring a retrospective pattern will almost always result in poor advice (Brooks and Legault, 2016; Punt *et al.*, 2020). Due to the long run times and difficulty in fully automating because of occasional odd patterns in the Mohn's rho trend as multipliers change, incorporating the Rose approach in closed-loop simulation testing will be challenging. The odd patterns in the Mohn's rho trend as multipliers change might be resolved by setting stricter criteria for convergence and model acceptance (e.g. maximum gradient less than a specified value). Longer projection horizons could also be explored through closed-loop simulation to examine the impact of fixes on future catch advice when the underlying cause of the original retrospective pattern remains the same or changes.

The catch advice provided in these examples is for demonstration purposes. The white hake Rho approach catch advice is similar to, but the Rose approach catch advice is below, the catch advice from the recent assessment. The catch advice for Gulf of Maine haddock from both the Rose and Rho approaches is similar to the catch advice from the current assessment because the current assessment used the Rho approach, just with a different number of peels. The witch flounder catch advice from the Rose and Rho approaches is similar to the catch advice from the index-based assessment used for this stock once the age-based assessment was rejected due to a strong retrospective pattern. The Georges Bank yellowtail flounder catch advice presented here from both approaches is much higher than from the recent assessment. This is due to the difference in how the catch advice is derived and the use of a much higher base natural mortality rate (0.4) compared to the previously used value (0.2) when virtual population assessment was used. The higher natural mortality rate is based on an evaluation of biological characteristics conducted during the empirical benchmark. For comparison, the reference fishing mortality rate from the last accepted VPA was 0.25, which is much lower than the 1.00 value in the base assessment here. The F40% values present in Supplementary Figure S3 reflect the higher M pushing selectivity to the right compared to a lower value of M. Furthermore, if the ASAP model was used for catch advice, there would need to be some accounting for the requirement to rebuild the stock, which would likely result in lower catch advice than presented here. Rebuilding of highly depleted stocks, such as Georges Bank yellowtail flounder, may be more impacted by Allee effects (aka depensation) in the stock–recruitment relationship than F reductions (Winter *et al.*, 2020).

The Rose and Rho approaches have pros and cons. For the Rose approach, the main benefits are that it by definition produces an ensemble of models all of which do not exhibit a retrospective pattern, with all the benefits of ensemble modelling and none of the issues of models with retrospective patterns. The main benefit of such a model ensemble is that it directly addresses model structure uncertainty, although not completely because there are always other model formulations that could be developed to explain the data. As the name implies, it does not specifically identify the source of the retrospective pattern in the original run, but instead relies on a number of different model fixes to create the ensemble. This leads to the downsides of this approach, the long run times (about a day per stock on a standard laptop computer) needed to develop the ensemble of models, and the potential for "stacking the deck" by selecting only a limited number of fixes that do not sufficiently explore all the possible fixes for the retrospective pattern in the original run. For example, the Georges Bank yellowtail flounder, Gulf of Maine haddock, white hake, and witch flounder examples required 896, 693, 555, and 607 ASAP model runs, respectively, with six runs each (terminal year and five peels). When the model is next updated, a full examination across fixes would need to be done again to ensure that all the Rose models have no retrospective patterns. As noted above, the occasional odd Mohn's rho value as a multiplier changed prevented full automation of the approach, increasing the amount of time required to conduct the analyses. This is a common issue in statistical catch-at-age models and can be addressed through approaches such as jittering, but these typically require multiple addition runs, thus further increasing run times. How to determine when a sufficiently large and diverse range of models are available for the Rose approach is a common issue found in all ensemble modelling approaches (Millar *et al.*, 2015; Karp *et al.*, 2018). Removing models from the Rose due to other diagnostic issues, such as time series of all positive then all negative residuals in fits to indices, can be done. However, this should be done before the catch advice is known to avoid "cherry-picking" the models to be used in the ensemble based on the result. The Rose model also requires local knowledge to limit the range of possible multipliers examined for different fixes.

Use of the rho adjustment is a much simpler approach than the Rose approach, requiring only the original model run and a retrospective analysis. It allows the full evaluation of the original run and provides a single model for use in management, as many management bodies appear to prefer currently. Like the Rose, it does not identify the source of the retrospective pattern. It simply assumes that whatever caused the retrospective pattern to occur will continue into the future and adjusts the terminal year estimates accordingly. The major downside of the rho adjustment is the discontinuity that it creates between the terminal year estimate and the rest of the time series of SSB and F (Figure 1 and Supplementary Figure S1). This discontinuity makes statements about recent trends quite challenging. For example, in the white hake case, the stock assessment shows the SSB generally increasing in recent years but the rho adjustment calls into question this perception and makes it difficult to state with any confidence whether the recent trend is up, but to a lower level, or in fact down. While the rho adjustment can be written out in an equation, it is still an *ad hoc* adjustment with not a lot of theoretical underpinning. The projections using the Rho approach only change the starting point but do not account for whatever caused the retrospective in the stock assessment. However, the directional changes in the SSB and F ratios do match the Rose approach but tend to not be as strong a change. This property could be considered by managers, for example to increase a risk buffer to account for this possible undercorrection compared to the Rose approach. However, since the amount of difference between the Rose and Rho approaches differs by stock and this difference does not appear directly related to the magnitude of the retrospective pattern, specific guidance for how much to change the catch advice from the Rho approach cannot yet be provided. Guidance for stocks exhibiting the opposite retrospective pattern (e.g. Gulf of Maine haddock) is not clearly defined currently, with some arguing to increase catch to address the retrospective pattern and others arguing it would be too risky given the costs of being wrong (ICES, 2020). Recent work using a number of groundfish stocks in the region has demonstrated that while the Rho approach has improved management advice in terms of reducing overfishing compared to ignoring the retrospective pattern, it has not been sufficient to eliminate overfishing (Brooks and Legault, 2016; Wiedenmann and Jensen, 2019).

Conclusions

The Rose and Rho approaches both change SSB and F status ratios in the same direction relative to the original run. The Rose estimate moved more than the rho-adjusted point in all four stocks examined. Catch advice from both approaches was less than from the original run for stocks that exhibit the typical direction of the retrospective pattern (Georges Bank yellowtail flounder, white hake, and witch flounder). The Rose catch advice was lower than rho-adjusted catch advice for these three stocks. The stock that had the opposite direction retrospective pattern (Gulf of Maine haddock) had the opposite directional change in status determination and catch advice, with again the Rose approach results being further away from the original assessment than the Rho approach results. Both approaches can be used to address stocks with retrospective patterns, and either is preferred to ignoring a retrospective pattern. The choice of which approach to use could be based on time and expertise available to conduct and maintain an assessment, with Rose preferred if lots of time and local knowledge of the stock and modelling are available while Rho preferred if either is limited. If the Rho approach is used, managers should consider adjusting their control rule or risk buffer to account for the difference between Rose and Rho results shown here.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Data availability statement

The data underlying this article are available at https://github.com/cmlegault/rosevrho.

Acknowledgements

I thank Irene Andrushchenko (DFO) for discussions that led to naming the Rose approach and the members of the ICES WKFORBIAS workshop for discussions on both the Rose and Rho approaches, as well as Liz Brooks, Editor Jan Jaap Poos, José De Oliveira, and an anonymous reviewer for constructive comments on an earlier draft of this article.

References

Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., Afflerbach, J., *et al.* 2017. Improving estimates of population status and trend with superensemble models. Fish and Fisheries, 18: 732–741.

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

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

Cao, J., Thorson, J. T., Richards, R. A., and Chen, Y. 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1781–1793.

Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences, 73: 296–308.

Deroba, J. J. 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's rho. North American Journal of Fisheries Management, 34: 380–390.

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

Hordyk, A. R., Huynh, Q. C., and Carruthers, T. R. 2019. Misspecification in stock assessments: common uncertainties and asymmetric risks. Fish and Fisheries, 20: 888–902.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., *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.

ICES. 2008. Report of the working group on methods of fish stock assessment (WGMG), 7–16 October 2008. Woods Hole, USA. ICES Document CM 2008/RMC: 03. 147 pp.

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.

Karp, M. A., Blackhart, K., Lynch, P. D., Deroba, J., Hanselman, D., Gertseva, V., Teo, S. *et al.* (Eds). 2018. Proceedings of the 13th National Stock Assessment Workshop: Model Complexity, Model Stability, and Ensemble Modeling. U.S. Dept. of Commerce, NOAA. NOAA Technical Memoranda NMFS-F/SPO-189. 49 pp.

Legault, C. M. 2009. Report of the Retrospective Working Group, 14–16 January 2008, Woods Hole, Mass. US Department of Commerce Northeast Fisheries Science Center Reference Document 09–01.

Legault, C. M., Alade, L., Gross, W. E., and Stone, H. H. 2014. Stock assessment of Georges Bank yellowtail flounder for 2014. TRAC Reference Document 2014-01. 219 pp. http://www.nefsc.noaa.gov/saw/trac/.

Legault, C. M., and Finley, M. 2019. Stock assessment of Georges Bank yellowtail flounder for 2019. TRAC Working Paper 2019-03. 60 pp. http://www.nefsc.noaa.gov/saw/trac/.

Legault, C. M., and Palmer, M. 2016. In what direction should the fishing mortality target change when natural mortality increases within an assessment? Canadian Journal of Fisheries and Aquatic Sciences, 73: 349–357.

Legault, C. M., and Restrepo, V. R. 1999. A flexible forward age-structured assessment program. ICCAT Collected Volume of Scientific Papers, 49: 246–253.

Maunder, M. N., and Piner, K. R. 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research, 192: 16–27.

Millar, C. P., Jardim, E., Scott, F., Osio, G. C., Mosqueira, I., and Alzorriz, N. 2015. Model averaging to streamline the stock assessment process. ICES Journal of Marine Science, 72: 93–98.

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

Miller, T. J., Politis, P., Blaylock, J., Richardson, D., Manderson, J., and Roebuck, C. 2018. Relative efficiency of a chain sweep and the rockhopper sweep used for the NEFSC bottom trawl survey and chainsweep-based swept area biomass estimates for 11 flatfish stocks. *In* SAW 66 Summer Flounder Data/Model/Biological Reference Point (BRP) Meeting. National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, MA, 17–21 September 2018.

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.

Northeast Fisheries Science Center (NEFSC). 2008. Assessment of 19 northeast groundfish stocks through 2007 U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document, 884: 08–15.

Northeast Fisheries Science Center (NEFSC). 2015. Operational assessment of 20 northeast groundfish stocks, updated through 2014. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document 251: 15–24.

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 pp.

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 pp.

Northeast Fisheries Science Center (NEFSC). 2020. Operational Assessment of 14 Northeast Groundfish Stocks, Updated Through 2018. U.S. Department of Commerce. Northeast Fisheries Science Center Reference Document 2020-XX. 208 pp.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Perretti, C. T., Deroba, J. J., and Legault, C. M. 2020. Simulation testing methods for estimating misreported catch in a state-space stock assessment model. ICES Journal of Marine Science, 77: 911–920.

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

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rose, G. A., and Walters, C. J. 2019. The state of Canada's iconic Northern cod: a second opinion. Fisheries Research, 219: 105314.

Rosenberg, A. A., Kleisner, K. M., Afflerbach, J., Anderson, S. C., Dickey-Collas, M., Cooper, A. B., Fogarty, M. J., *et al.* 2018. Applying a new ensemble approach to estimating stock status of marine fisheries around the world. Conservation Letters, 11: e12363.

Rossi, S. P., Cox, S. P., Benoît, H. P., and Swain, D. P. 2019. Inferring fisheries stock status from competing hypotheses. Fisheries Research, 216: 155–166.

Rudd, M. B., and Branch, T. A. 2017. Does unreported catch lead to overfishing? Fish and Fisheries, 18: 313–323.

Spence, M. A., Blanchard, J. L., Rossberg, A. G., Heath, M. R., Heymans, J. J., Mackinson, S., Serpetti, N., *et al.* 2018. A general framework for combining ecosystem models. Fish and Fisheries, 19: 1031–1042.

Stewart, I. J., and Hicks, A. C. 2018. Interannual stability from ensemble modelling. Canadian Journal of Fisheries and Aquatic Sciences, 75: 2109–2113.

Stewart, I. J., and Martell, S. J. D. 2014. A historical review of selectivity approaches and retrospective patterns in the Pacific halibut stock assessment. Fisheries Research, 158: 40–49.

Stewart, I. J., and Martell, S. J. D. 2015. Reconciling stock assessment paradigms to better inform fisheries management. ICES Journal of Marine Science, 72: 2187–2196.

Swain, D. P., and 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.

Szuwalski, C. S., Ianelli, J. N., and Punt, A. E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science, 75: 596–609.

United States of America, v. Carlos A. Rafael. 2017. 282 F. Supp. 3d. United States District Court for the District of Massachusetts.

Van Beveren, E., Duplisea, D. E., Marentette, J. R., Smith, A., and Castonguay, M. 2020. An example of how catch uncertainty hinders effective stock management and rebuilding. Fisheries Research, 224: 105473.

Wiedenmann, J., and 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., and 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.

Winter, A.-M., Richter, A., and Eikeset, A. M. 2020. Implications of Allee effects for fisheries management in a changing climate: evidence from Atlantic cod. Ecological Applications, 30: e01994.

Handling editor: Jan Jaap Poos