NRC Research Press

ARTICLE

Could recent overfishing of New England groundfish have been prevented? A retrospective evaluation of alternative management strategies[1]

John Wiedenmann and Olaf P. Jensen

Abstract: We conducted a retrospective evaluation of alternative management strategies for stocks in the New England groundfish complex that have had recent history of target catches being set above the level that defines overfishing. In many cases the original target catches were unsustainable and would have resulted in stock collapses if the target catch had been removed. We evaluated (*i*) alternative harvest control rules, (*ii*) whether or not to do projections, (*iii*) whether the inputs to the projections (starting abundance and future recruitments) should be modified, and (*iv*) whether the target catches should be smoothed to prevent large changes from year to year. The greatest reductions in target catches resulted when no projections were done and the target catch was fixed over the period between assessments. Large reductions in target catches also occurred when a downward adjustment was made to the starting abundance in the projections based on the retrospective pattern. Neither approach alone was sufficient to prevent overfishing for most stocks, but when used in conjunction with one another or with an alternative control rule that reduced the target harvest rate as biomass fell below the target, the magnitude and frequency of overfishing was greatly reduced for most stocks. Attempts to adjust recruitment based on perceived changes over time were also effective for a few stocks, while attempts to smooth the target catches over often resulted in increases in the target catches.

Résumé : Nous avons réalisé une évaluation rétrospective de différentes stratégies de gestion pour des stocks dans le complexe de poissons de fond de la Nouvelle-Angleterre qui, dans le passé récent, ont été visés par des objectifs de prises supérieurs au niveau définissant une surpêche. Dans de nombreux cas, les objectifs de prises initiaux n'étaient pas durables et se seraient traduits par l'effondrement de stocks s'ils avaient été atteints. Nous avons évalué (*i*) différentes règles de contrôle des prises, (*ii*) la pertinence de faire des projections, (*iii*) si les intrants des projections (abondance initiale et recrutements futurs) devraient être modifiés et (*iv*) si les objectifs de prises devraient être lissés pour prévenir des variations importantes d'une année à l'autre. Les plus grandes réductions des objectifs de prises se produisent quand aucune projection n'est faite et que l'objectif de prises est fixé pour toute la période entre les évaluations. D'importantes réductions des objectifs de prises se produisent également quand un ajustement à la baisse est fait à l'abondance initiale dans les projections à la lumière du motif rétrospectif. Aucune de ces deux approches ne suffit à elle seule à prévenir la surpêche pour la plupart des stocks, mais quand elles sont utilisées ensemble ou avec une autre règle de contrôle qui réduit le taux de prises visé quand la biomasse passe sous l'objectif, la magnitude et la fréquence des cas de surpêche sont considérablement réduites pour la plupart des stocks. Les tentatives d'ajustement du recrutement selon les changements perçus au fil du temps sont aussi efficaces pour quelques stocks, alors que des tentatives de lisser les objectifs de prises se traduisent souvent par des augmentations des objectifs de prises. [Traduit par la Rédaction]

Introduction

The aim of modern single-species fisheries management is to maintain the population size of the target stock close to the level that produces maximum sustainable yield (MSY; Hilborn 2010). Managers attempt to attain this target population biomass (B_{MSY}) by setting catch limits that will achieve exploitation rates close to the level that achieves MSY, or F_{MSY}. In practice, however, achieving a target harvest rate can be challenging due to uncertainty in our understanding of the biology and of the population size of the target species (scientific uncertainty) and also uncertainty in how the fishery responds to regulations (management uncertainty). The relative importance and sources of these uncertainties may vary by system or by species within a system, and identifying appropriate management strategies for dealing with this ever present uncertainty is crucial for the sustainable management of global fisheries.

Different approaches have been used to identify effective management strategies. One approach is to develop a closed-loop simulation model called a management strategy evaluation (MSE) to test a suite of management strategies to the different uncertainties inherent in biology of the stock and in the assessment and management processes (Butterworth and Punt 1999; Punt et al. 2016). For each management strategy, MSE models quantify the long-term average performance across a range of measures of management success (and trade-offs among objectives) over thousands of runs for different model configurations representing alternative states of nature. MSEs have become an indispensable tool in fisheries management and have been used in a wide range

Received 30 March 2018. Accepted 20 July 2018.

J. Wiedenmann. Department of Ecology, Evolution and Natural Resources, Rutgers University.
O.P. Jensen. Department of Marine & Coastal Sciences, Rutgers University.
Corresponding author: John Wiedenmann (email: john.wiedenmann@rutgers.edu).

[1]This article is being published as part of the special issue "Under pressure: addressing fisheries challenges with Management Strategy Evaluation" arising from two related theme sessions sponsored by the American Institute of Fishery Research Biologists at the 147th Annual Meeting of the American Fisheries Society, Tampa, Florida, USA, August 2017.

 Published at www.nrcresearchpress.com/cjfas on 30 July 2018.

Table 1. New England groundfish stocks explored in this work.

Full stock name	Scientific name	Abbreviated name
Georges Bank Atlantic cod	*Gadus morhua*	GB cod
Gulf of Maine Atlantic cod	*Gadus morhua*	GOM cod
Georges Bank haddock	*Melanogrammus aegelfinus*	GB haddock
Georges Bank yellowtail flounder	*Limanda ferruginea*	GB yellowtail flounder
Cape Cod–Gulf of Maine yellowtail flounder	*Limanda ferruginea*	CC–GOM yellowtail flounder
Southern New England–Mid-Atlantic yellowtail flounder	*Limanda ferruginea*	SNE–MA yellowtail flounder
Southern New England–Mid-Atlantic winter flounder	*Pseudopleuronectes americanus*	SNE–MA winter flounder
Witch flounder	*Glyptocephalus cynoglossus*	Witch flounder
American plaice	*Hippoglossoides platessoides*	Plaice

of applications all over the world (see Punt et al. 2016 and Punt 2017 for many examples).

In contrast with MSEs, which are prospective in nature, retrospective investigations can also be used to identify suitable management strategies. Evaluation of historical assessment and management performance can provide insight into whether objectives have been met in the past and may provide guidance on future management practices. For example, Ralston et al. (2011) and Punt et al. (2018) evaluated variability in biomass estimates across repeated assessments for stocks off the western US and southeastern Australia, respectively, with the aim of identifying appropriate buffer sizes for setting target catches to account for scientific uncertainty. Wiedenmann and Jensen (2018) reviewed the assessment and management performance for stocks in the New England groundfish complex since 2004 and found that harvest rates for many stocks continued to exceed F_{MSY}, despite target catches for these stocks set to achieve harvest rates generally at or below 75% of F_{MSY} (the full control rule is to use 75% of F_{MSY}, or for stocks that are overfished use the lesser of 75% of F_{MSY} or the F that allows for rebuilding in a specified time frame; Federal Register 2009). Actual catches were below the target catches in most cases (29% below on average, across stocks and years), and the primary causes of high F values were overly optimistic projections used to set the catch targets, resulting from overestimated abundance in the assessment, but also declining recruitment (Brooks and Legault 2016; Wiedenmann and Jensen 2018).

A big question remains for New England groundfish: what could have been done differently to prevent overfishing in recent years? Answering this question requires (1) identifying the sources of this scientific uncertainty in the assessment estimate (i.e., what caused the assessments to overestimate abundance), but also (2) understanding whether alternative methods for setting catch target would have reduced the frequency and magnitude of overfishing. Addressing point (1) is imperative, but is also enormous in scope. Assessment scientists in the region have devoted substantial effort to deal with this issue (see Legault 2009 and Brooks and Legault 2016 for details of example efforts); yet, considerable uncertainty remains in many of the current assessments (NEFSC 2017*b*). Here we focus on addressing point (2), so that effective methods can be identified and used in the setting of future catch targets. In their analysis on sources of bias in projections for New England groundfish, Brooks and Legault (2016) also explored potential remedies to the bias. They explored adjustments to the starting abundance in the projection based on the retrospective pattern in the assessment model (called a ρ-adjustment) and found that such adjustments improved but did not eliminate the bias in the projected abundance. Based on their findings, Brooks and Legault (2016) recommended some additional potential solutions, including shorter projection periods (≤3 years), basing future recruitments on more recent time periods of recruitment (and not the whole time series), and a more conservative control rule.

Here we conducted a retrospective evaluation of alternative management strategies to determine how alternative methods for setting catch targets would have performed for a subset of New England groundfish stocks (Table 1). Our analysis is an extension of the work of Brooks and Legault (2016), but there are some important differences between approaches. Brooks and Legault (2016) relied on projections from retrospective peels of the 2008 assessment for each stock (removing 1–7 years of data and refitting the model), and the estimated abundance from these peels (which were the starting points for the projections) was not always comparable to the estimates in the same years from earlier assessments (NEFSC 2002, 2005). Also, they used the final abundance estimates from NEFSC (2008) as their reference for quantifying projection bias, but many of these estimates have changed dramatically with subsequent assessments (see figure 3 in Wiedenmann and Jensen (2018) for some examples). Here we explored a variety of alternative methods for setting catch targets from 2004 to 2012, using the original assessment output as the starting point in the projections (NEFSC 2002, 2005, 2008) and using the most recent assessment that passed review for each stock as the basis of updated understanding of stock status over time. Alternative methods we explored can be categorized as (*i*) approaches for adjusting the terminal abundance estimate from the assessment, (*ii*) using different expectations of future recruitment, (*iii*) fixing the target catch over the interval between assessment models, (*iv*) alternative harvest control rules, and (*v*) the effect of gradual changes in the catch targets that prevent dramatic increases or decreases from year to year. Methods were explored in combination with one another to identify combinations that performed well, with performance evaluated with respect to the ability to set target catches each year close to the level that would have achieved F_{MSY} in each year.

Methods

Our focus is determining whether alternative methods for setting catch targets could have prevented or reduced overfishing for groundfish stocks, but before describing our approach we must first define what we mean by overfishing. In the US, managers annually specify the catch that is estimated to achieve F_{MSY}, also called the overfishing limit, or OFL (hereinafter called the management OFL). If the catch exceeds the management OFL in a given year, overfishing technically occurs. However, a later assessment could show that the management OFL was underestimated, such that the estimated F in that year may be below F_{MSY}. Alternatively, the estimated F may be above F_{MSY} even if the catch did not exceed the management OFL that was later determined to be overestimated. Our focus here is on the historical definition of overfishing, and our definition of the OFL is the estimated catch that would have achieved F_{MSY} in hindsight based on the current understanding of past stock dynamics from the most recent stock assessment (hereinafter called the hindsight OFL; see section on Calculating performance of the alternative methods below for more details).

The work detailed herein was applied to nine of the 20 stocks in the New England groundfish complex (Table 1). We focused on

Table 2. Description of the types of modifications made to the calculation of the target catches, by the five main categories.

Category	Details
Abundance modification	(1) No adjustment, use original values (2) Do a retrospective adjustment using the estimated r (3) Adjust abundance using the estimated relative error from the previous assessment
Recruitment	(1) Use the default method for specifying recruitment in the projections (2) Use a shortened portion of the recruitment times series in projections if a regime shift (decrease only) is detected
Control rule	(1) Use the default target F (2) Use 65% F_{MSY} (3) Use 75% of F_{MSY} (4) Use the ramped P^* control rule (CV = 0.5) that sets the ABC = 0 at $S/S_{MSY} < 0.1$ (5) Use the ramped P^* control rule (CV = 1.0) that sets the ABC = 0 at $S/S_{MSY} < 0.1$ (6) Use the ramped P^* control rule (CV = 0.5) that sets a fixed ABC at $S/S_{MSY} < 0.2$ (7) Use the ramped P^* control rule (CV = 1.0) that sets a fixed ABC at $S/S_{MSY} < 0.2$
Projections	(1) Catches are based on projections over the entire management period (2) Project to first year of management period, fixed catches (3) No projections, catches fixed based on the terminal estimated biomass from the assessment
Catch smoothing	(1) Use the original target catch (2) Only allow catches to change by 20% annually (3) Target catches are a weighted average of the old estimates (from previous assessment) and the current assessment

Note: For each category, methods numbered 1 are the original method used to set the catch targets. Methods from different categories were explored in a factorial manner (for a total of 378 combinations of approaches) to determine the impact that combined approaches would have. For example, a retrospective adjustment of estimated abundance was explored using the different control rules, with or without projections, and so on. See text for explanation of terms.

those stocks that had age-based assessments and projections for each of the 2002, 2005, and 2008 stock assessments (NEFSC 2002, 2005, 2008), and projections following these assessments were the basis of setting catch targets for 2004–2005, 2006–2009, and 2010–2012, respectively. We refer to these periods where catches were set following each assessment as management periods 1, 2, and 3. Our overall approach involved first implementing alternative methods for setting catch targets for each stock and for each management period and then determining the impact that such catch targets would have had on each stock. We first detail the projection model and how alternative methods were implemented to create alternative catch targets and then detail how we quantified the impact that the different target catches would have had on each stock.

Projection model

In previous work (Wiedenmann and Jensen 2018), we developed a projection model in R (R Core Team 2017) that mimics the AGEPRO software used to conduct projections for New England groundfish (Brodziak et al. 1998). Details of our projection model can be found in Wiedenmann and Jensen (2018), but we provide a brief summary here. The model used the original input files for the projections used to set catch targets (obtained from New England Fishery Management Council staff), including the estimated starting abundance at age (based on bootstraps or Markov chain Monte Carlo draws from the assessment for each stock), assumed mass-, selectivity-, and maturity-at-age, as well as the target fishing mortality rate over the projection period. Mass-, maturity-, and selectivity-at-age were fixed over the projection interval and were based on averages over the most recent 3 to 5 years, depending on the stock. Target catch in a given year in the projection was set at the median over all iterations for that year, with the number of iterations specified in the input file.

We ran the projection model for each stock and management period using the original inputs as a baseline for comparison with the alternative approaches. The alternative methods we explored can be classified into five categories: (*i*) modifying the starting abundance in the projections, (*ii*) modifying how future recruitments are estimated, (*iii*) using alternative harvest control rules, (*iv*) whether or not to base future target catches on projected changes in abundance, and (*v*) methods for smoothing the target catches. Multiple approaches were explored within most categories (Table 2), and we applied the methods in a factorial manner (for a total of 378 combinations) to determine the effect that different combinations of approaches would have on setting the target catches. The methods within these five categories can be broadly classified as either adjustments to the population inputs into the projection model (1 and 2) or alternative ways of determining the target catch given the inputs (3, 4, and 5).

Modifying starting abundance-at-age

We evaluated three methods (i.e., runs 1–3) for setting the initial abundance-at-age ($N_{init}(a)$) in the projection model (Table 3). For run 1, we used the original (unmodified) distribution of abundance-at-age ($N_{orig}(a)$) for each projection. For runs 2 and 3, we used an adjustment factor, f, and calculated a modified abundance-at-age with

$$(1) \qquad N_{init}(a) = N_{orig}(a)/(1 + f)$$

For run 2, f was based on a measure of the average retrospective error in the terminal assessment estimates. The specific measure used was called Mohn's ρ (Mohn 1999), and we obtained estimates of ρ for each stock from NEFSC (2002, 2005, 2008). Estimates of ρ were generally positive (indicating a tendency for overestimation of terminal biomass), but there were also some negative estimates. Because our focus was on identifying options that would have produced more conservative target catches, we only made abundance adjustments when ρ was positive (i.e., $f = \max(0, \rho)$). For run 3, f was based on the relative error in terminal biomass from the previous assessment (REB). For example, the 2002 assessment estimated biomass through 2001 for each stock. If the 2005 assessment for a given stock estimated the biomass in 2001 to be 50% lower than what was estimated in the 2002 assessment, then the adjustment of terminal abundance for the 2005 assessment would be a 50% decrease (i.e, f = REB = 0.5). As in run 2, we only made adjustments when REB was positive ($f = \max(0, \text{REB})$). We did not have assessments for all stocks prior to the 2002 assessment, so we only conducted the REB adjustment following the 2005 and 2008 assessments. Note that within a given assessment, the estimate of ρ is an average of the REB in the last 5–7 years, but here our REB estimate compares across assessments and thus takes into account the impact that changes to the assessment model or data had on the terminal estimates.

Table 3. Ratio of the target catch under alternative methods to the hindsight overfishing limit (OFL) for each stock, averaged over the management periods 2004–2005, 2006–2009, and 2010–2012.

Alternative method	Management period	GOM cod	GB cod	Witch flounder	Winter flounder	Plaice	GB yellowtail flounder	CC–GOM yellowtail flounder	SNE–MA yellowtail flounder	GB haddock	Mean across stocks
Original target catch	2004–2005	3.05	1.85	10.17	2.60	4.74	12.41	2.78	8.85	1.26	5.30
	2006–2009	5.05	3.64	6.62	1.43	4.78	14.25	6.51	0.62	9.82	5.86
	2010–2012	14.02	1.57	2.91	0.13	2.00	14.91	2.90	2.79	10.46	5.74
Recruitment modification (2)	2004–2005	—	—	—	2.17	—	—	1.56	2.45	—	2.06
	2006–2009	—	—	—	1.21	—	—	5.74	0.63	—	2.53
	2010–2012	—	—	—	0.11	—	2.62	—	1.40	—	1.38
No projections (3)	2004–2005	2.74	2.58	3.20	2.06	4.29	10.62	3.20	2.04	0.76	3.50
	2006–2009	2.11	1.49	5.42	0.60	2.95	9.94	2.24	0.50	1.79	3.00
	2010–2012	3.39	1.04	1.18	0.07	4.75	1.23	1.08	1.00	23.02	4.08, 1.72
Ramped P^* control rule (7)	2004–2005	1.92	1.21	9.00	1.63	2.00	8.32	1.13	3.05	0.79	3.23
	2006–2009	3.05	2.22	10.04	14.01	1.16	1.71	1.69	0.61	3.09	4.17
	2010–2012	5.52	0.96	3.26	1.09	1.54	15.51	2.42	8.64	5.09	4.89
Abundance modification (2)	2004–2005	2.95	1.07	4.07	1.28	4.73	7.63	1.96	6.11	1.26	3.45
	2006–2009	4.77	3.49	1.77	0.41	4.27	11.18	4.75	0.48	9.45	4.51
	2010–2012	7.74	1.36	1.75	0.09	1.28	14.67	2.47	2.06	7.64	4.34
Abundance modification (3)	2004–2005	—	—	—	—	—	—	—	—	—	—
	2006–2009	4.57	3.23	5.46	0.86	4.08	0.63	4.13	0.52	9.77	3.69
	2010–2012	1.36	0.90	2.07	5.97	0.13	1.45	1.14	0.94	3.13	1.90
Abundance modification (2), no projections (3)	2004–2005	2.67	1.71	1.81	1.09	4.29	4.95	2.04	0.89	0.76	2.25
	2006–2009	2.02	1.44	0.18	1.62	7.77	2.69	1.75	0.37	1.74	2.18
	2010–2012	2.60	0.91	0.79	4.75	0.05	0.73	1.06	0.89	20.43	3.58, 0.75
No projections (3), ramped P^* control rule (7)	2004–2005	0.58	0.41	2.39	3.88	0.36	1.27	0.42	0.23	0.22	1.09
	2006–2009	0.61	0.27	8.26	1.96	0.14	0.74	0.24	0.13	0.97	1.48
	2010–2012	1.47	0.18	0.16	0.55	1.70	0.84	0.55	0.54	18.78	2.75, 0.75
Abundance modification (2), no projections (3), P^* control rule (7)	2004–2005	0.56	0.93	0.29	0.19	1.26	1.27	0.22	0.11	0.22	0.56
	2006–2009	0.55	0.26	1.12	1.01	0.06	0.62	0.20	0.11	0.95	0.54
	2010–2012	1.03	0.29	0.16	0.11	1.70	0.43	0.44	0.48	16.86	2.39, 0.55
Abundance modification (2), no projections (3), P^* control rule (5)	2004–2005	0.56	0.93	0.00	0.12	1.26	1.27	0.22	0.00	0.22	0.51
	2006–2009	0.55	0.16	1.12	1.01	0.00	0.62	0.12	0.05	0.95	0.51
	2010–2012	1.03	0.29	0.14	0.00	1.70	0.43	0.44	0.40	16.86	2.36, 1.84
Abundance modification (2), no projections (3), 65% F_{MSY} (2)	2004–2005	1.67	1.68	1.10	0.72	2.81	1.83	0.85	0.42	0.50	1.67
	2006–2009	1.47	1.05	3.05	7.92	0.21	1.15	0.69	0.40	1.04	1.47
	2010–2012	1.75	0.58	0.93	7.01	0.42	0.63	1.00	2.37	19.34	3.78, 1.84

Note: For methods that did not use projections, averages across stocks were calculated with and without GB haddock for the 2010–2012 management period, as this stock had a very large ratio. The numbers in parentheses refer to the specific method detailed in Table 2.

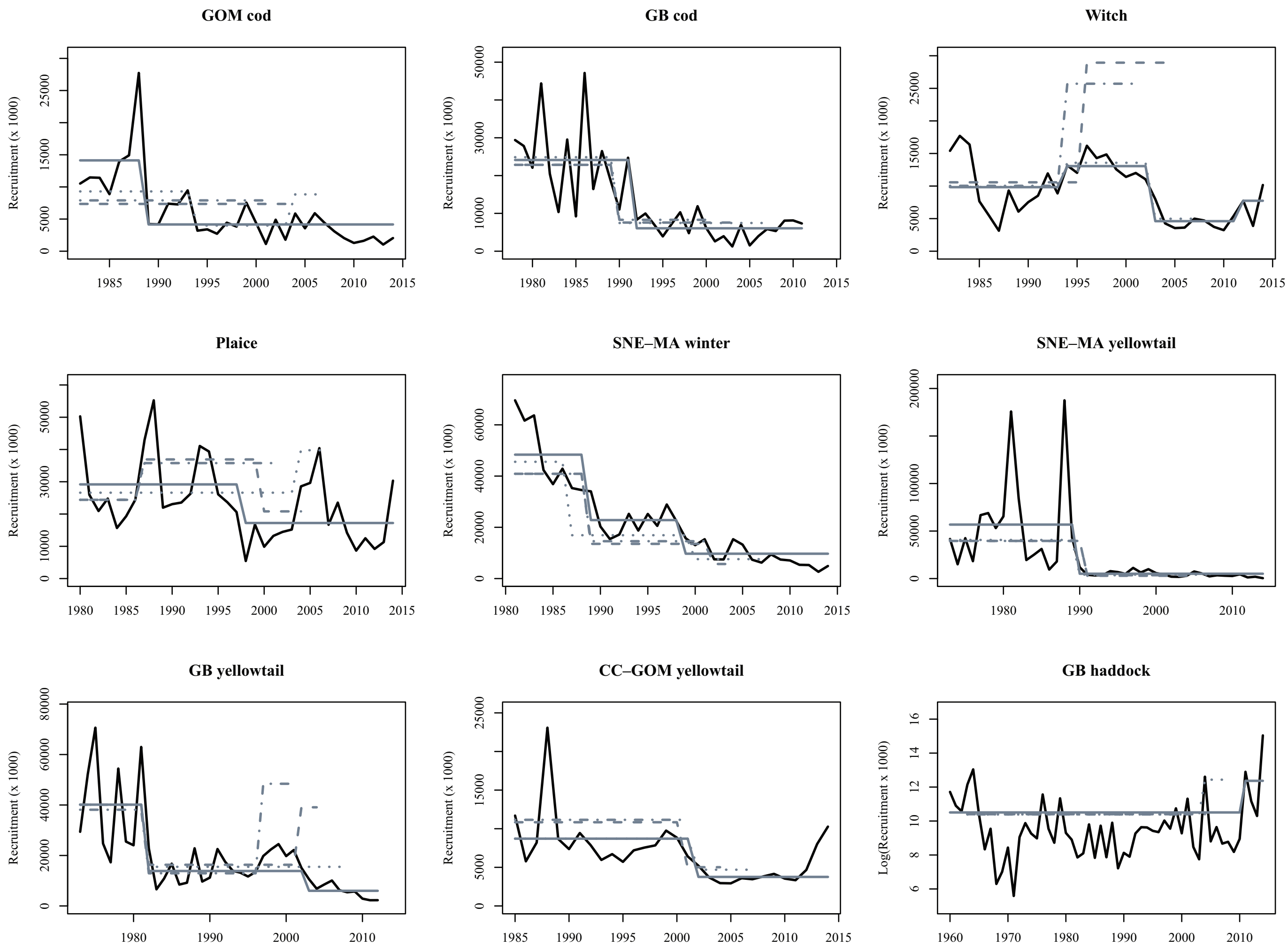
Modifying future recruitments

Forecasted recruitments in the original projections were done in one of two ways: they were either based on random draws around a Beverton–Holt stock–recruitment fit, or they were drawn from an empirical distribution based on some portion of the estimated time series of recruitment for a stock. For run 1 (Table 2), we used the original method for specifying future recruitments. For run 2, we use a truncated empirical distribution using recent recruitments if a decline in recruitment was identified. To determine whether a decline in recruitment occurred in recent years, we used the algorithm developed by Rodionov (2004) to detect recent climatic regime shifts. This method was used by Vert Pre et al. (2013) to determine whether temporal changes in stock productivity had occurred for a large number of global fish stocks. The algorithm works by calculating the mean recruitment over a specified initial time period, then calculates the mean for a subsequent period and assigns this period as a new regime if the mean is significantly different from the old mean according to the Student's t test. The algorithm continues sequentially until each time period is assigned to an existing or new regime. For each stock and each assessment, we used this regime-shift algorithm to determine whether the estimated recruitment had declined in recent years. We assumed a minimum initial interval of 5 years and omitted the terminal year estimate of recruitment due to the high uncertainty in the estimate. In some cases, no regime shift was detected for the entire time series, while in other cases, both increases and decreases were predicted in a single time series (Fig. 1). In cases where the mean recruitment from the terminal regime was lower than the mean from the previous regime, we used the empirical recruitment estimates from the terminal regime period only in the projection model. If no decline or an increase in recent recruitment was detected, no modification was made to the forecasted recruitment method. When a lower recruitment regime was detected, we did not adjust the reference points because this would have altered the performance of the threshold control rules that reduce the harvest rate as the biomass falls below the spawning biomass reference point S_{MSY} (see below) and also would have impacted the target F based on rebuilding projections for some stocks.

Alternative harvest control rules

The existing control rule for New England groundfish stocks has been to use the lesser value of 75% of F_{MSY} or $F_{rebuild}$ (if the

Fig. 1. Recruitment regimes estimated using the regime shift detection algorithm of Rodionov (2004) following NEFSC (2002, 2005, and 2008; dot-dashed, dashed, and dotted gray lines, respectively) compared with the estimated recruitment from the most recent assessment (solid black line). The solid gray line is the estimated regime using the most recent recruitment estimates. Recruitment for Georges Bank (GB) haddock is plotted on the log scale due to a few very large recruitments. Refer to Table 1 for description of other stocks.

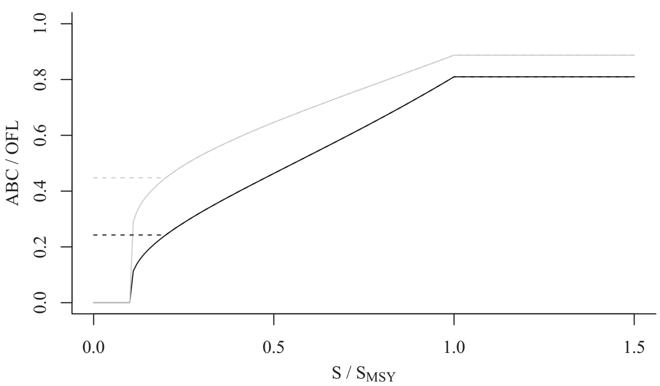


population is in need of rebuilding). We evaluated a total of seven control rules (runs 1–7), with the historical target *F* used in run 1. For runs 2 and 3, the target *F* was set to 65% and 75%, respectively, of the estimated F_{MSY} in each projection, regardless of whether or not it was lower than the estimated $F_{rebuild}$. Control rules for runs 4–7 were variations of the threshold-based P^* control rule (Shertzer et al. 2008). The general P^* approach uses the point estimate of the management OFL and assumes that the point estimate is the median of a lognormal distribution with a specified coefficient of variation (CV). The catch target is determined by selecting a percentile of the management OFL distribution below the median. Selecting the 40th percentile as the target implies a 40% chance of overfishing ($P^* = 0.4$) if the management OFL is unbiased. This approach results in the catch target being lower than the median (point estimate), and the size of this buffer increases with a lower percentile (target P^*) for a given CV or a higher assumed CV for a given percentile. We explored options where the target P^* declines (and buffer between the target catch and the OFL increases) as biomass falls below S_{MSY}. For runs 4 and 5, the control rules were based on the Mid-Atlantic Fishery Management Council's control rule, where the target P^* is 0.4 when biomass is above S_{MSY} and declines linearly as biomass drops below the target until reaching 10% of S_{MSY}, at which point the fishery is closed. The difference between the control rules in runs 4 and 5 was in the assumed CV of the management OFL distribution (0.5 and 1.0, respectively; Fig. 2). Use of a CV of 1.0 was based on the simulation testing of Wiedenmann et al. (2017) but also on the recommendation of Brooks and Legault (2016), while 0.5 was selected as a less conservative option. The control rules were nearly identical in runs 6 and 7, with the difference being that fishing was still allowed at low biomass levels. For the control rules in runs 6 and 7, the assumed CVs of 0.5 and 1.0 were used, respectively, and a fixed target P^* was applied when the stock was at or below 20% of S_{MSY} (Fig. 2).

Projected or fixed catch targets

In addition to using the standard projections (projection run 1), we explored two alternatives. One alternative (run 2) was to project only to the first year of the management period and fix the target catch for the remainder of the management period. For example, target catches were set for 2004 and 2005 following the 2002 assessment, so biomass would be projected from 2002 to 2004, and the catch at the target *F* (specified in the control rule) in 2004 would also be used in 2005. Alternatively, no projections were done (run 3) and the catch target for the management period was calculated using the target *F* and the initial abundance in the projection model. Using the same example, the target catch in 2004 and 2005 would be based on the catch at the target *F* using

Fig. 2. Two versions of the ramped P^* control rule, where the buffer size increases (smaller acceptable biological catch (ABC)/overfishing limit (OFL)) as biomass falls below spawning biomass target (S_{MSY}). In version one, the fishery is shut down when spawning biomass S falls below 10% S_{MSY} (solid lines). In version two, a fixed buffer is used when biomass falls below 20% S_{MSY} (dashed lines). Black and gray lines are for assumed coefficients of variation (CVs) of the OFL distribution of 1.0 and 0.5, respectively. The OFL used in the control rule is what was estimated as the OFL following each assessment (i.e., what we are calling the management OFL).



the estimated initial abundance-at-age in 2002. Runs 2 and 3 were explored as more conservative options because projections for these stocks generally predicted increases in biomass that were overly optimistic (Brooks and Legault 2016; Wiedenmann and Jensen 2018).

Alternative methods for setting catch targets: smoothing of catch targets

All of the methods described thus far only use output from the most recent assessment when estimating catch targets (although the REB abundance adjustment modified the output based on past levels of error). Using only the most recent assessment can result in a large change in the target catch between assessments if the current biomass estimate differs substantially from the projected biomass from the previous assessment. We evaluated three methods for smoothing the estimated catch targets: (1) use only the most recent information from the assessment when setting catch targets (i.e., no smoothing; the status quo approach); (2) constrain the catch targets based on the most recent assessment to only allow for annual changes of ±20%; and (3) use a weighted average of the catch target from the previous assessment and the current catch targets. For run 2, if $C^*_{\text{target}}(t+1)$ is the new target catch in year $t+1$ based on the projection, then the actual catch target in year $t+1$ will be

$$(2)\qquad C_{\text{target}}(t+1) = \begin{cases} \max.\left[0.8 \cdot C_{\text{target}}(t), C^*_{\text{target}}(t+1)\right] & C^*_{\text{target}}(t+1) > C_{\text{target}}(t) \\ \min.\left[1.2 \cdot C_{\text{target}}(t), C^*_{\text{target}}(t+1)\right] & C^*_{\text{target}}(t+1) < C_{\text{target}}(t) \end{cases}$$

For run (3), if C_{prev} is the target catch in the final year of the previous management period, then the target catch in year t is a weighted average of C_{prev} and $C^*_{\text{target}}(t)$:

$$(3)\qquad C_{\text{target}}(t) = (1-w) \cdot C_{\text{prev}} + w \cdot C^*_{\text{target}}(t)$$

where w represents a weighting factor. Values of w between 0 and 1 are possible, but we used an even weight ($w = 0.5$) for run 3. Note that we used the smoothed catches, regardless of whether or not they were above the estimated OFL at the time (covered in more detail in the Discussion).

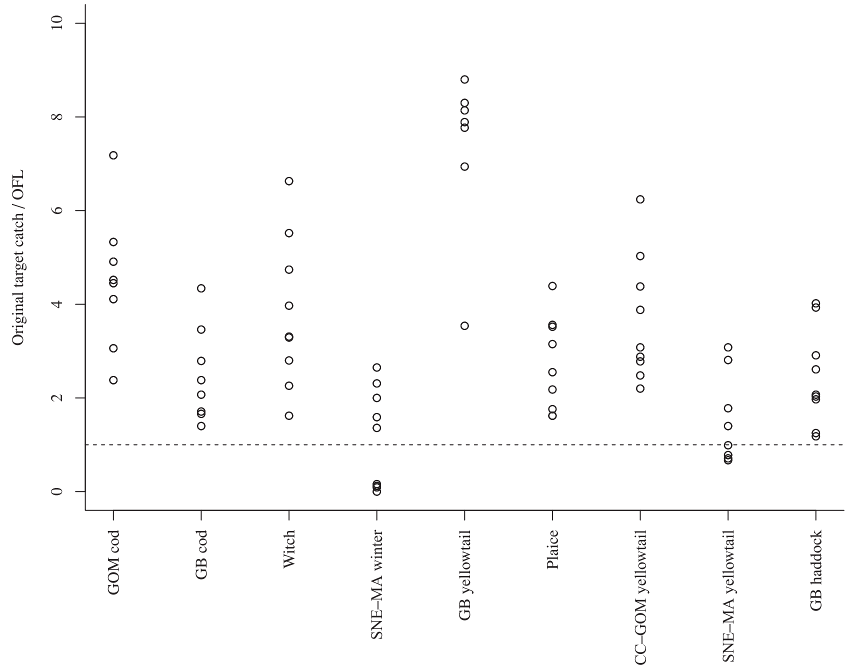
Calculating performance of the alternative methods

The alternative methods produced different target catches for each management period for each stock, and the next step was to try to understand what impact such changes in target catches would have had on the stock. Alternative target catches lower or higher than the original target catches could have resulted in growth or decline in the stock, respectively. In MSE simulations, the impacts of such catches would change the abundance of the stock over some period, which would potentially be identified in later assessments, and catches would be adjusted accordingly based on the management strategy being tested in the closed loop of the simulation. Here, we have one realization for each stock, with abundance in a given year driven by past catches and stock productivity.

For each stock, we obtained estimated abundance-, mass-, selectivity-, and maturity-at-age and management reference points from the most recent age-based stock assessment that has passed review for each stock. For six of the stocks, the most recent assessment was in 2017 (NEFSC 2017*b*), but for Georges Bank (GB) yellowtail flounder (*Limanda ferruginea*), GB cod (*Gadus morhua*), and witch flounder (*Glyptocephalus cynoglossus*), recent age-based assessments did not pass review. For these stocks, we used the most recent assessment that passed review (NEFSC 2012 for GB cod; Legault et al. 2013 for GB yellowtail flounder; and NEFSC 2015 for witch flounder), and we address the caveats for using estimates from these stocks in the Discussion. With estimates from the most recent assessment, we could calculate the hindsight OFL in each year and compare our new target catches with these estimates. Such an approach would tell us if catches were above or below the hindsight OFL in a given year, but it does not take into account the changes in stock size (and the OFL) that would have occurred under alternative target catches. An alternative approach could be that for each time series of catches (2004–2012) from the different methods, forecast the impact that such catches would have had on the stock. In other words, start the population for a given stock at the 2004 abundance (estimated from the most recent assessment), with future abundance based on the resulting mortality from the new target catches, and recruitments each year (which could be either fixed at their estimated values or allowed to change as stock size changes). While this approach would allow us to measure the cumulative impact of different catches on the stocks, our target catches are based on outputs from sequential assessments, and changes in stock status over time following one assessment would likely have changed the estimates from later assessments. For example, if an alternative method produced target catches much higher than the original target catches following the 2002 assessment, such catches could have driven the stock to very low levels, and subsequent assessments would likely have detected such a crash, and catch targets would have been adjusted accordingly.

Given the limitations of both of these approaches, we opted for a third approach. We applied the target catches from each management period to each stock and only evaluated the impact such catches would have had on the stock over that management period. The impact of each alternative method for setting catch targets was therefore independent across assessment periods. Doing so allowed for the population size (and the hindsight OFL) to change over time in response to new target catches, but such changes are quantified separately for each management period for each stock. For management periods 1, 2, and 3, we started the population at the estimated abundance from the most recent assessment in the first year of the period (2004, 2006, or 2010). Recruitment in subsequent years was based on the annual estimates from the most recent assessment. We explored allowing for dynamics changes in recruitment based on changes in spawning biomass (determined from stock–recruit fits to the model estimates), but found little difference in the results. For each stock, this approach produced dynamic estimates of the hindsight OFL

Fig. 3. Ratio of the original target catch set for management in a given year to the hindsight OFL in that year for each stock. Points for a stock represent the values across years (2004–2012). Note that the hindsight OFL used here is the static estimate based on the fixed estimates of abundance from the most recent assessment. We used the static and not the dynamic hindsight OFL here because in some cases the target catches were so high that it caused the population to crash, resulting in very low hindsight OFL estimates in some years and therefore very high ratios of the target catch/hindsight OFL.



for each management period independently, and we compared target catches under each alternative method with the resulting dynamic hindsight OFL to determine which methods, if any, would have consistently reduced the magnitude of overfishing for groundfish. We also calculated the proportion of times during each management period in which a given method would have caused overfishing (calculated across years and stocks for the management period) and would have caused a collapse of the stock. We defined a collapse as any year in which the target catch was ≥95% of the exploitable biomass in that year. In such cases, the maximum instantaneous fishing mortality rate resulting from the catch was capped at 3.0, preventing the stock biomass from dropping below 0.

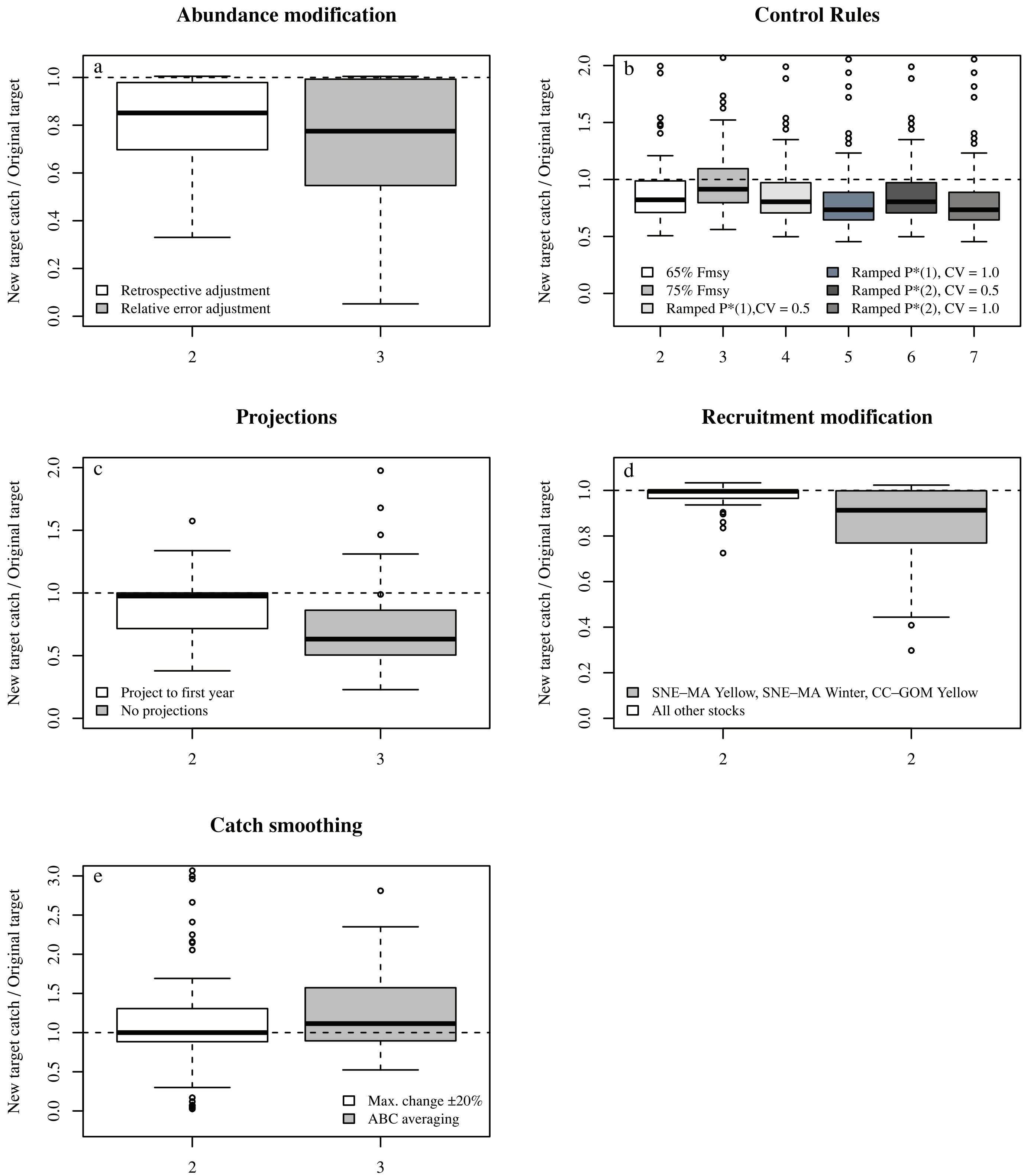
Results

The actual target catches used for management of New England groundfish from 2004 to 2012 frequently exceeded the hindsight OFLs for most stocks (Fig. 3). For seven of the nine stocks, the original target catch was above the hindsight OFL in all years evaluated. In total, the actual target catch was above the hindsight OFL 94% of the time, and it was more than twice the hindsight OFL 71% of the time across stock between 2004 and 2012. The actual target catches were only below the hindsight OFL for SNE–MA winter flounder (*Pseudopleuronectes americanus*) following the 2008 assessment (where the target F was near 0) and for SNE–MA yellowtail flounder following the 2005 assessment.

Because target catches were almost always above the hindsight OFL for these groundfish between 2004 and 2012, we first looked for the different approaches that consistently resulted in target catches that were consistently below the original target catches. In Fig. 4, the ratio of the alternative to the original target catches is shown for the different approaches. For a given alternative, all other approaches were unchanged to explore the sole impact of the alternative method. Both methods for adjusting the terminal abundance reduced the target catches in most instances. The retrospective adjustment of terminal biomass resulted in a median decrease of 15% (interquartile range (IQR) of 2%–31%), while the relative error adjustment resulted in a larger median reduction (23%), but a broader range too (IQR between a 2% increase to 45% decrease). For the alternative control rules, the only difference between the P^* control rules resulted from the assumed CV (0.5 or 1.0) and not the ways in which the buffer size differed at low biomass levels (see Fig. 2), although there were differences when biomass adjustments were used in conjunction with these control rules (see below). The ramped P^* control rules using a CV of 1.0 resulted in the largest overall reduction (26%; IQR 11%–35%), followed by the ramped P^* option with a CV of 0.5 (20% reduction; IQR 3%–30%). The largest single reduction occurred when no projections were used (fixing the target catch over the assessment interval based on the terminal biomass estimate), resulting in a 37% decrease (IQR 14%–50%) in target catches, on average across stocks. Accounting for declines in recruitment only had a substantial reduction in target catches for three stocks (SNE–MA winter flounder, SNE–MA yellowtail flounder, and CC–GOM yellowtail flounder), reducing the target catch by 29% on average for these stocks (IQR 5%–45%). Attempts to smooth the target catch tended to result in target catches that were comparable or higher than the original target catch that used the unsmoothed values, so we therefore excluded these catch smoothing methods from subsequent analyses (Fig. 4).

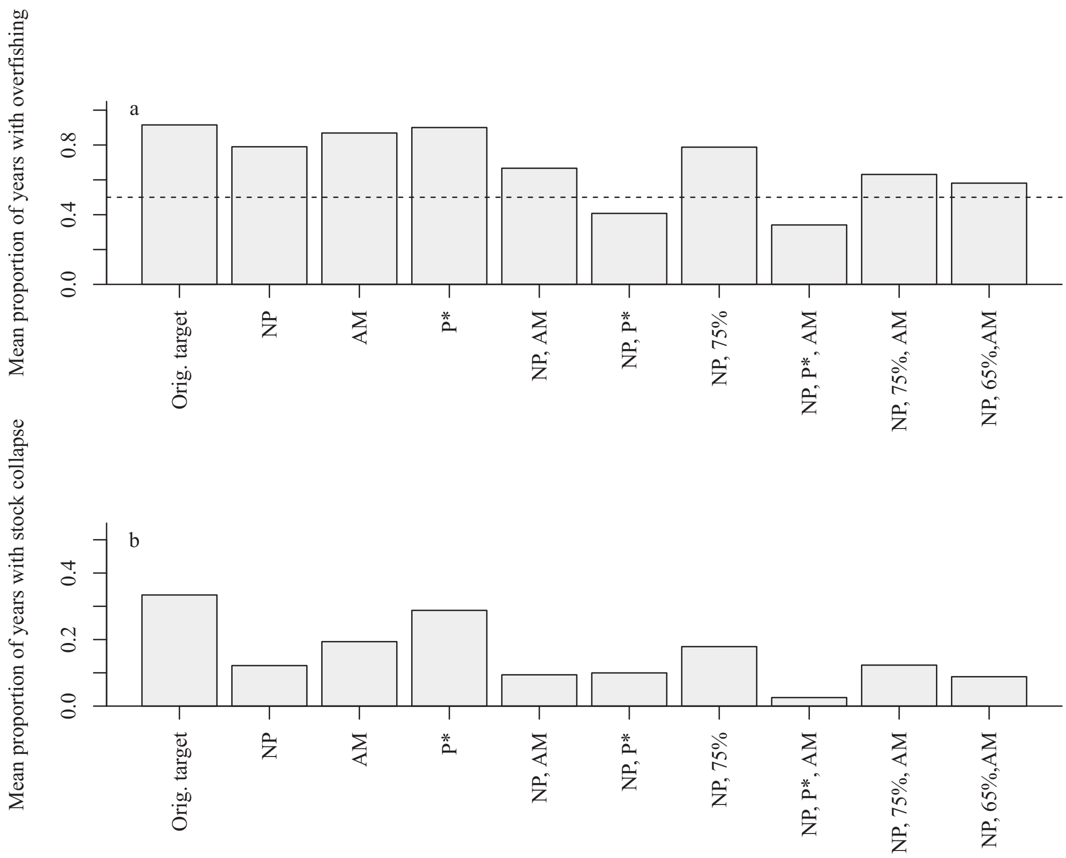
The results in Fig. 4 explore the effect that a single factor had on the target catches, but we also evaluated the impact of different methods in combination with one another. Although many methods resulted in declines in the target catches relative to the actual target catch, such declines may still have been insufficient to prevent overfishing. We therefore calculated the impact that combined methods had on limiting overfishing. For each stock we calculated the ratio of mean target catch to the hindsight OFL over each management period, with the hindsight OFL dynamic here based on the impact that the target catches would have had on stock abundance over the management period (Table 3). For each management period, we also calculated the proportion of times

Fig. 4. Ratio of the target catch under the alternative approach to the original target catch, across years and across stocks for a given method. The different approaches are separated by different shading, and the number (*x* axes) refers to the specific approach detailed in Table 2. For the recruitment modification plot, results are separated out by specific stocks.



that the target catch would have resulted in overfishing or resulted in a collapse of the stock (Fig. 5). In some cases, the target catch was so high that it was ≥95% of the exploitable biomass for a stock, and we called such cases a collapse of the stock. Under the original target catches, overfishing would have occurred in 100%, 86%, and 88% of the time in management periods 1 (2004–2005), 2 (2006–2009), and 3 (2010–2012), respectively, while a collapse would have occurred 27%, 42%, and 30% of the time for these

Fig. 5. The proportion of times that the target catch would have resulted in overfishing (*a*) and where the target catch would have been greater than 95% of the exploitable biomass (*b*) for alternative catch setting methods. We calculated the proportion individually for each management period, but the values were similar across periods for each method, so we averaged them together for each method. "Orig. target" refers to the original target catch set for management, NP refers to no projections, AM refers to modifying abundance, *P** refers to the ramped control rule (option 7) and 75% and 65% refer to control rule options 2 and 3, respectively (Table 2). When multiple approaches are listed, it means they were used in combination (e.g., NP, *P** means no projections were done and the *P** control rule was used). The horizontal dashed line at 0.5 in panel (*a*) delineates where overfishing is more likely to occur.

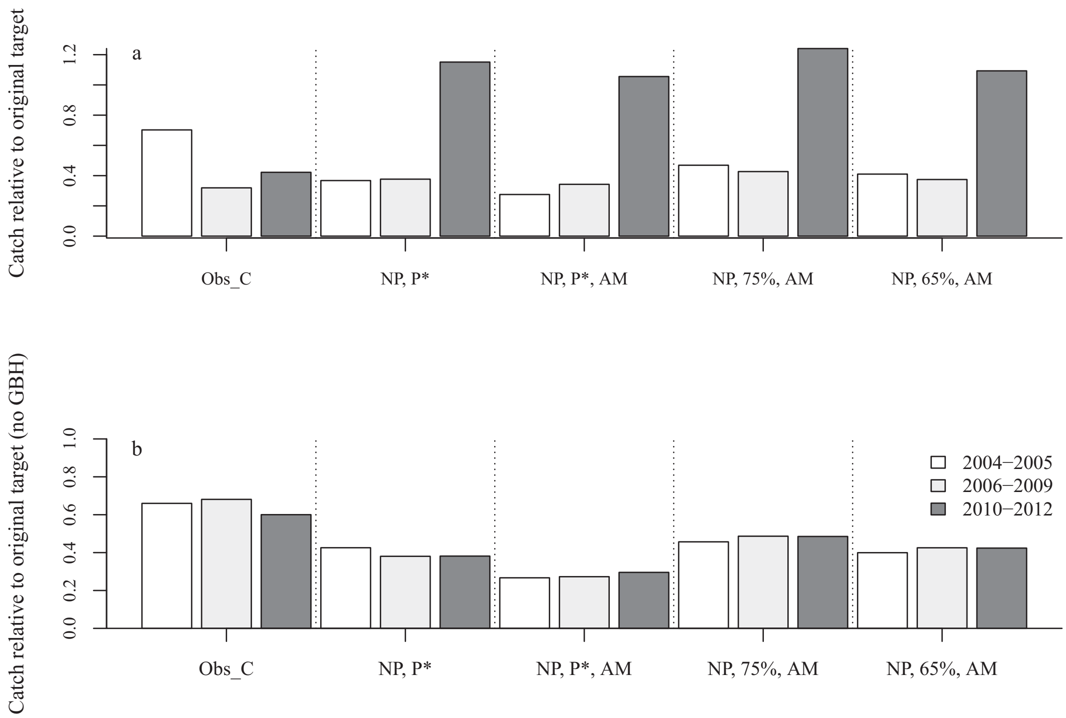


periods (means across periods are shown in Fig. 5). Not doing projections would have resulted in declines in the catch/hindsight OFL ratio in all instances except for GB haddock (*Melanogrammus aegelfinus*) following the 2008 assessment (the 2010–2012 management period; reasons for this are detailed in the Discussion). Across stocks, not doing projections resulted in mean decreases in the catch/hindsight OFL of 21%, 48%, and 42% for management periods 1, 2, and 3, respectively (the 42% mean for period 3 omits GB haddock in the calculation due to the very large ratio). Using retrospective adjustments of abundance (a Mohn's ρ-adjustment) resulted in respective 28%, 27%, and 26% mean decrease across management periods in the catch/hindsight OFL, while using the relative error adjustment resulted in a 27% decrease in period 2 and a 50% reduction in period 3 (Table 3). The ramped *P** control rule (option 7 in Table 2) alone would have reduced the mean catch/hindsight OFL in management periods 1 and 2 by 42% and 28%, respectively, but would have resulted in an increase in the target catch/hindsight OFL of 87% for management period 3. This large mean increase was driven primarily by SNE–MA winter flounder, which had a very low original target *F* for this management period; when this value was removed, the ramped *P** control rule only resulted in a 4% increase. Modifying the future recruitments only had an impact on SNE–MA yellowtail flounder, CC–GOM yellowtail flounder, and SNE–MA winter flounder, reducing the catch/hindsight OFL by 44%, 9%, and 24% for each management period, with the largest reductions occurring for SNE–MA yellowtail in management periods 1 and 3 and for CC–GOM yellowtail in period 1 (Table 3).

Although these methods alone tended to reduce the magnitude of overfishing, they did not end overfishing in most cases (Table 3), but they did reduce the frequency of stock collapses (Fig. 5). Only when certain methods were combined was the frequency of overfishing reduced below 50%, meaning overfishing was less likely to occur across these groundfish stocks. Using the ramped *P** control (option 6 in Table 2) rule without projections, overfishing would have occurred 41% of the time, on average, with stock collapses occurring 10% of the time (the collapses were witch flounder in the 2006–2009 management period and GB haddock in the 2010–2012 management period). When the ramped *P** control rule was used without projections and with a ρ-adjustment to abundance, overfishing would have occurred even less frequently (25% of the time), and the only collapse would have been for GB haddock in the 2010–2012 management period. Using the 75% or 65% F_{MSY} control rule resulted in overfishing more often than not, even when used with no projections and an abundance modification (63% and 58%, respectively), although these methods did reduce the frequency of stock collapses (Fig. 5).

These combined alternative methods were able to reduce the frequency and magnitude of overfishing across most stocks, but limiting overfishing is not the only objective of fisheries management, otherwise the solution would be to just shut the fishery down. In Fig. 6, we show the aggregate yield across management

Fig. 6. For each management period (2004–2005, 2006–2009, and 2010–2012), the aggregate catch by method across stocks including (*a*) and excluding (*b*) GB haddock is shown. Catch is relative to the original target catch set for management. Obs_C refers to the observed catch, NP refers to no projections, AM refers to modifying abundance, *P** refers to the ramped control rule (option 7), and 75% and 65% refer to control rule options 2 and 3, respectively (Table 2). When multiple approaches are listed, it means they were used in combination (e.g., NP, *P** means no projections were done and the *P** control rule was used).



periods for a subset of methods (yield is shown with and without haddock due to the scale of haddock catches relative to all other groundfish; Figs. 6*a* and 6*b*, respectively). Although reductions in catches were very large relative to the actual target catch (being 51%–82% lower depending on the method and management period, excluding GB haddock), the actual catches that occurred for these stocks were also lower than the target (34%–40% lower, excluding haddock). Therefore, some of the combined methods did not result in as drastic reductions relative to the actual catch that occurred. For example, fishing at 75% of the estimated F_{MSY}, without projections and with a ρ modification resulted in target catches 31%, 29%, and 19% lower than the aggregate catch (without haddock) in management periods 1, 2, and 3, respectively. In contrast, the ramped *P** control rule (option 6 in Table 2) without projections resulted in reductions of 35%, 44%, and 36% for the same management periods and reductions of 60%, 60%, and 51% when combined with the ρ-adjustment. These reductions are relative to the actual catches, which still resulted in overfishing for many stocks. When compared with the OFL, some of the combinations of methods resulted in considerable foregone yield to the fishery. For example, for the more conservative combination of methods explored (the ramped *P** control rule with a CV of 1.0, no projections, and a ρ-adjustment to abundance), target catches were below 30% of the hindsight OFL in most management periods (and below 50% in all periods) for GB cod, SNE–MA winter flounder, SNE–MA yellowtail flounder, and CC–GOM yellowtail flounder (Table 3). This version of the ramped *P** control rule allowed for fishing even at very low levels of biomass, but when the version that shut down the fishery when biomass falls below 10% of S_{MSY} was used (in conjunction with abundance modifications and no projections), target catches would have been set to 0 for GB cod and SNE–MA yellowtail in management period 1 and for SNE–MA winter flounder in periods 2 and 3 (Table 3).

Discussion

We conducted a retrospective evaluation of alternative management strategies for stocks in the New England groundfish complex that have had recent history of target catches resulting in overfishing. In many cases the original target catches were unsustainable and would have resulted in stock collapses had the target catch been achieved. We evaluated (*i*) alternative harvest control rules, (*ii*) whether or not to do projections, (*iii*) whether the inputs to the projections (starting abundance and future recruitments) should be modified, and whether the target catches should be smoothed to prevent large changes from year to year. The greatest reductions in target catches resulted when no projections were done and the target catch was fixed over the management period. Large reductions in target catches also occurred when a downward adjustment was made to the starting abundance in the projections. Neither approach alone was sufficient to prevent overfishing for most stocks, but when used in conjunction with one another or with an alternative control rule that reduced the target harvest rate as biomass fell below the target, the magnitude and frequency of overfishing was greatly reduced for most stocks. Attempts to adjust recruitment based on perceived changes over time were also effective for a few stocks, while attempts to smooth changes in the target catches were counterproductive and often resulted in increases in the target catches.

Projections are an integral part of natural resource management (Clark et al. 2001), and in fisheries projections are used in the setting of target catches and in rebuilding analyses to determine rebuilding schedules under different harvest rates (e.g., Brodziak et al. 1998; Punt 2010). Projections require many assumptions, and catch targets estimated in later years of the projection period are conditioned on achieving the target *F* in earlier years. If assumptions are not met early in the projected time period, this error becomes compounded in later years because the achieved *F* continues to deviate from the target, and this snowballing effect can be particularly problematic for stocks that are projected to increase in size under the target *F*. The groundfish stocks we evaluated were all below the spawning biomass target (S_{MSY}), at least initially, and projections all used target harvest rates that resulted in predicted increases in biomass and catch over time in the model. Updated assessments, however, indicated that terminal biomass was consistently overestimated in earlier assessments for these stocks (Wiedenmann and Jensen 2018). These projections started with inflated estimates of abundance and expected this abundance to increase over time (Brooks and Legault 2016). Target catches from the projections therefore increased over time, often to levels that would have caused the stock to collapse had the target catch been realized. The exception to this pattern was GB haddock, which after a few very large recruitments during our study period was estimated to be above the biomass target in 2007 (NEFSC 2008). For this stock, projections indicated a decline in catch over time as the biomass approached the target, such that not doing projections resulted in higher target catches than if they had been done. Our finding is therefore that projections should not be the basis for setting catch targets for groundfish that are depleted and expected to increase. That is, we should not base future catch targets on predicted increases in abundance that may or may not be realized. This policy was recently adopted by the Scientific and Statistical Committee of the New England Fishery Management Council for the setting of the acceptable biological catch (ABC) for groundfish. Although our finding is specific to the New England groundfish stocks we evaluated, exploration into the use and efficacy of projections in other regions is warranted.

Prior to this work, the primary approach for dealing with uncertainty in the stock assessment estimates in New England groundfish was to do a retrospective (ρ) adjustment to the terminal biomass for assessments with a strong retrospective pattern (e.g., NEFSC 2008, 2012, 2015, 2017*b*). Operationally, a strong pattern has been defined as when the adjusted terminal biomass falls outside of the confidence bounds of the terminal biomass estimate (Brooks and Legault 2016). We evaluated doing adjustments regardless of the magnitude of the adjusted biomass, in part because estimates of assessment uncertainty were not always available in the assessments, but also because past work showed that ρ is a poor predictor of changes in biomass estimates in future assessments (Brooks and Legault 2016; Wiedenmann and Jensen 2018). In some cases, terminal biomass estimates from assessments with small positive retrospective patterns were revised substantially downward in later assessments. Although our work here and earlier work of Brooks and Legault (2016) showed that such adjustments alone were insufficient to end overfishing in most cases, their continued use is warranted, as they did reduce the magnitude of overfishing when applied.

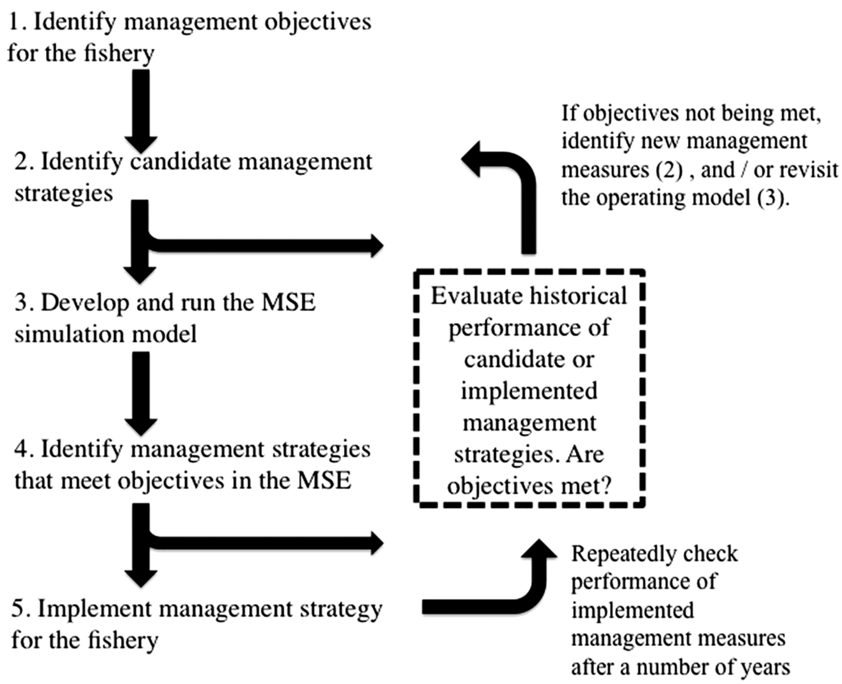
Research has shown that recruitment and overall stock productivity is largely driven by environmental forcing (Gilbert 1997; Vert Pre et al. 2013), such that attempts to accurately forecast large changes in recruitment could result in more sustainable catch limits (Szuwalski and Punt 2013; Punt et al. 2014). For New England groundfish, however, we found that such attempts were only effective for a few stocks. This finding was driven by the fact that declines in recruitment were often not detected by the regime shift algorithm we used, despite overall declines in recruitment for many stocks (Fig. 1). We evaluated different parameterizations of the algorithm and found little difference in our results. There are a number of possible approaches we could have explored to detect trends in recruitment, but their impact would likely have been small for the stocks where no decline was detected. Many of the assessments overestimated total abundance and recruitment in the final years of the model, such that no decline in recruitment was apparent in the time series used in the regime detection algorithm. Thus, trying to account for changes in recruitment is particularly challenging for stocks with strong retrospective patterns. Also, even if changes in recruitment had been accurately captured in the projections, the impact on the target catches was smaller than if the starting abundance was correct (Brooks and Legault 2016). This reduced effect is due to the fact that it takes 3 or more years for most of the recruits to enter the fishery, such that the impact of accurate recruitment forecasts only becomes apparent towards the end of longer projection intervals (Brooks and Legault 2016).

Many MSE-based studies have evaluated the performance of different harvest control rules and in general found that control rules that are more conservative as biomass declines perform well overall and are often better suited for limiting overfishing and reducing the risk of becoming overfished and also allowing for more rapid recovery of depleted stocks than options that apply a fixed harvest rate (Punt 2003; Punt et al. 2008; Benson et al. 2016; Wiedenmann et al. 2017). Our findings for New England groundfish are in agreement with these simulation-based studies, as using the ramped P^* control rule had a lower risk of overfishing or of stock collapse compared with using 75% or even 65% of F_{MSY}. However, use of the ramped control rule in conjunction with a biomass adjustment often resulted in target catches being very conservative for some groundfish stocks (catch < 20% of the hindsight OFL). In particular, using a ramped control rule that closes the fishery at low biomass would have set target catches to 0 in some years in instances in our analysis. While very low catches or even closure of a fishery may be warranted in certain cases, they can present challenges in a mixed stock fishery such as groundfish where species are caught together. A "choke" stock with tight catch restrictions could limit the ability to catch the total allocation for stocks in good condition. This is already occurring for some groundfish stocks (e.g., catches of GB haddock are limited in part by restrictive catch limits for GB yellowtail flounder and GB cod). Thus, although our work showed that ramped control rules performed better than the current control rule at limiting overfishing, adoption of such a control rule would require careful evaluation of the impacts on the fishery as a whole. MSE studies tend to focus on control rule performance using single-species models, but development of models that capture dynamics of mixed stock fisheries such as New England groundfish is needed.

Our finding that approaches to smooth the target catch performed poorly is not surprising given the history of overfishing for the groundfish stocks in our analysis. That is, when prior catches resulted in overfishing for a stock, using a weighted average of past and current catch targets will likely result in continued overfishing. In our analysis, we used the smoothed target catch, regardless of whether or not it was above the estimated OFL at the time, which is currently not allowed under US law (gradual reductions in the ABC, called phasing-in, are allowed if it is below the estimated OFL; Federal Register 2015). During the period of our analysis, gradual changes in the target fishing mortality rate were used in rebuilding plans for some New England groundfish stocks that allowed for $F > F_{MSY}$ in early years of the rebuilding period (see Wiedenmann and Jensen 2018 for more details), so using smoothed target catches above the OFL would not necessarily have been rejected outright. In our work, the smoothed target catches were often below the original target catch (Fig. 4*e*), but still resulted in overfishing in many cases due to overestimated abundance in assessment. Thus, care is needed when using a phased-in or smoothing approach, particularly for stocks with strong retrospective patterns where abundance (and the OFL) may be overestimated.

A caveat of our work is that we are using the updated model estimates to evaluate the performance of historical catch targets. In MSE simulations, the true abundance of the stock is known in the operating model, so the impact that different management strategies have on a stock is known. Here, we relied on the most recent assessment estimates as our measure of "truth", but it is

Fig. 7. Conceptual model of how a retrospective evaluation of management strategies fits in the broader context of using management strategy evaluations (MSEs) to identify robust management strategies. There are many points throughout the process where a retrospective evaluation could be done, either in the identification of which strategies should be included in the MSE, or once the MSE is completed, robust options could be evaluated for historical performance on stocks within a region. In cases where a management strategy has been implemented for many years, a retrospective evaluation could be conducted to see if objectives are being met. If not, then revisiting the MSE is warranted, either to modify the operating model or to include alternative management options.



likely that these estimates may change in future assessments. Many of the assessments for New England groundfish have strong retrospective patterns (some of which are worsening), with estimates in final years of an assessment revised downward in later assessments (NEFSC 2017*b*). Our analysis used estimated abundance from the most recent assessment through 2012, and some of these estimates may change in later assessments, either due to a retrospective pattern or due to large-scale changes in the assessment model. Recent work has shown estimated abundance in recent years from age-based assessments may be biased low based on comparisons with swept-area biomass estimates calculated from the NEFSC spring and fall bottom trawl surveys. Swept-area biomass estimates for flatfish stocks in the groundfish complex have either been comparable or higher than age-based estimates in recent years, although the magnitude of the difference varies by stock (NEFSC 2017*b*). Large changes in historical biomass estimates (up or down) could affect our results, and we recommend that an evaluation of this sort be conducted regularly as new information arises so that managers and scientific advisors can weigh such information in the process of setting future catch limits.

For six of the nine stocks in our analysis, the most recent age-based assessment (NEFSC 2017*b*) is considered the best available science for each stock, despite concerns raised about the swept-area estimates. For three of the stocks (GB yellowtail flounder, witch flounder, and GB cod), the recent age-based assessments have been rejected as the basis for management advice by peer review panels. For GB cod and yellowtail flounder, the age-based assessments were rejected because of very strong retrospective patterns. For witch flounder, a strong retrospective pattern combined with the higher swept-area estimates led to rejection of the most recent age-based assessment (NEFSC 2017*a*). For these stocks, we relied on the most current assessment that passed review as our source of updated estimates (NEFSC 2012, 2015; Legault et al. 2013). The performance of different methods for setting catch targets was similar for these stocks as it was for the other stocks, such that if we removed these three stocks from our analysis, the qualitative conclusions about the effectiveness of the different options would remain the same. Removing these stocks would lower our estimated frequencies of overfishing and stock collapse, however, as both GB yellowtail flounder and witch flounder experienced continued overfishing even for some of the more effective combination of methods (Table 3). The use of older assessment estimates for these stocks would only be problematic for our analyses if the remedies to the current problems drastically changed historical estimates.

Our identification of effective methods for setting catch targets for New England groundfish was based on a retrospective evaluation of historical catch targets. This approach is not an MSE, but we argue for using such an approach as part of the overall process of development, testing, and evaluation of management strategies (of which MSE is an integral component) to make sure that those methods identified in the MSE, or those that have been in place for some time, would have met the management objectives for most stocks and to identify possible alternatives when objectives would not have been met (Fig. 7). The frequency, magnitude, and direction of the uncertainty in catch targets may change with future assessments, such that the successful approaches we identified in this work may no longer be effective for setting future catch targets. We therefore recommend that this sort of retrospective analysis be done on a regular basis to determine the performance of recent catch targets.

Acknowledgements

We thank Tom Nies and Jamie Cournane from the New England Fishery Management Council for providing all the necessary management and assessment documents and projection files, for helpful discussions throughout the scope of this project, and for comments on earlier versions of this work. We thank Elizabeth Brooks and an anonymous reviewer for comments on this manuscript. We are also grateful to members of the Scientific and Statistical Committee of the New England Fishery Management Council for helpful discussions on this work. This work was funded by the New England Fishery Management Council, and the views expressed herein are not necessarily those of the Council.

References

Benson, A.J., Cooper, A.B., and Carruthers, T.R. 2016. An evaluation of rebuilding policies for U.S. fisheries. PLoS ONE, **11**(1): e0146278. doi:10.1371/journal.pone.0146278.

Brodziak, J., Rago, P., and Conser. R. 1998. A general approach for making short-term stochastic projections from an age-structured fisheries assessment model. *In* Proceedings of the International Symposium on Fishery Stock Assessment Models for the 21st Century. *Edited by* F. Funk, T. Quinn, II, J. Heifetz, J. Ianelli, J. Powers, J. Schweigert, P. Sullivan, and C.-I. Zhang. Alaska Sea Grant College Program, Univ. of Alaska, Fairbanks.

Brooks, E.N., and Legault, C.M. 2016. Retrospective forecasting — evaluating performance of stock projections for New England groundfish stocks. Can. J. Fish. Aquat. Sci. **73**(6): 935–950. doi:10.1139/cjfas-2015-0163.

Butterworth, D.S., and Punt, A.E. 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. **56**: 985–998. doi:10.1006/jmsc.1999.0532.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Jr., Pizer, W., Pringl, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., and Wear, D. 2001. Ecological forecasts: an emerging imperative. Science, **293**(5530): 657–660. doi:10.1126/science.293.5530.657.

Federal Register. 2009. Magnuson-Stevens Act Provisions; Annual Catch Limits; National Standard Guidelines; final rule. 74:11, January 16. GPO, Washington, D.C. pp. 3178–3213.

Federal Register. 2015. Magnuson-Stevens Act Provisions; National Standard Guidelines; final rule. 80:12, January 20. GPO, Washington, D.C. pp. 2786–2811.

Gilbert, D.J. 1997. Towards a new recruitment paradigm for fish stocks. Can. J. Fish. Aquat. Sci. **54**(4): 969–977. doi:10.1139/f96-272.

Hilborn, R. 2010. Pretty Good Yield and exploited fishes. Mar. Pol. **34**: 193–196. doi:10.1016/j.marpol.2009.04.013.

Legault, C.M., Chair. 2009. Report of the Retrospective Working Group, January 14–16, 2008, Woods Hole, Massachusetts [online]. US Dept. Commer., Northeast Fish Sci Cent Ref Doc. 09-01. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/nefsc/publications/.

Legault, C.M., Alade, L., Gross, W.E., and Stone, H.H. 2013. Stock Assessment of Georges Bank Yellowtail Flounder for 2013 [online]. TRAC Ref. Doc. 2013/01. Available from http://www.nefsc.noaa.gov/saw/trac.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES J. Mar. Sci. **56**: 473–488. doi:10.1006/jmsc.1999.0481.

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, 2002 [online]. Northeast Fish. Sci. Cent. Ref. Doc. 02-16. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/saw/.

NEFSC. 2005. Assessment of 19 Northeast groundfish stocks through 2004 [online]. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, 15–19 August 2005. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 05-13. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/saw/.

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, 2008 [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 08-15. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online from http://www.nefsc.noaa.gov/nefsc/publications/.

NEFSC. 2012. Assessment or Data Updates of 13 Northeast Groundfish Stocks through 2010 [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 12-06. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/nefsc/publications/.

NEFSC. 2015. Stock Assessment Update of 20 Northeast Groundfish Stocks Through 2014 [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 15-24. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/nefsc/publications/.

NEFSC. 2017*a*. 62nd Northeast Regional Stock Assessment Workshop (62nd SAW) Assessment Report [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 17-03. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/publications/.

NEFSC. 2017*b*. Operational Assessment of 19 Northeast Groundfish Stocks, Updated Through 2016 [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 17-17. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online from http://www.nefsc.noaa.gov/publications/.

Punt, A.E. 2003. Evaluating the efficacy of managing West Coast groundfish resources through simulations. Fish. Bull. **101**: 860–873.

Punt, A.E. 2010. SSC default rebuilding analysis: Technical specifications and user manual (Version 3.12b). University of Washington, Seattle.

Punt, A.E. 2017. Strategic management decision-making in a complex world: quantifying, understanding, and using trade-offs. ICES J. Mar. Sci. **74**(2), 499–510. doi:10.1093/icesjms/fsv193.

Punt, A.E., Dorn, M.W., and Haltuch, M.A. 2008. Evaluation of threshold management strategies for groundfish off the U.S. West Coast. Fish. Res. **94**: 251–266.

Punt, A.E., A'mar, T., Bond, N.A., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haltuch, M.A., Hollowed, A.B., and Szuwalski, C. 2014. Fisheries management under climate and environmental uncertainty: control rules and performance simulation. ICES J. Mar. Sci. **71**(8): 2208–2220. doi:10.1093/icesjms/fst057.

Punt, A.E., Butterworth, D., de Moor, C., De Oliveira, J., and Haddon, M. 2016. Management strategy evaluation: best practices. Fish Fish. **17**: 303–334. doi:10.1111/faf.12104.

Punt, A.E., Day, J., Fay, G., Haddon, M., Klaer, N., Little, L.R., Privitera-Johnson, K., Smith, A.D.M., Smith, D.C., Sporcic, M., Thomson, R., Tuck, G.N., Upston, J., and Wayte, S. 2018. Retrospective investigation of assessment uncertainty for fish stocks off southeast Australia. Fish. Res. **198**: 117–128. doi:10.1016/j.fishres.2017.10.007.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J.D., and Conser, R.J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fish. Bull. **109**: 217–231.

R Core Team. 2017. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.

Rodionov, S.N. 2004. A sequential algorithm for testing climate regime shifts. Geophys. Res. Lett. **31**(9). doi:10.1029/2004GL019448.

Shertzer, K., Prager, M., and Williams, E. 2008. A probability-based approach to setting annual catch levels. Fish. Bull. **106**: 225–232.

Szuwalski, C.S., and Punt, A.E. 2013. Fisheries management for regime-based ecosystems: a management strategy evaluation for the snow crab fishery in the eastern Bering Sea. ICES J. Mar. Sci. **70**(5): 955–967. doi:10.1093/icesjms/fss182.

Vert Pre, K.A., Amoroso, R.O., Jensen, O.P., and Hilborn, R. 2013. The frequency and intensity of productivity regime shifts in marine fish stocks. Proc. Natl. Acad. Sci. **110**: 1779–1784. doi:10.1073/pnas.1214879110.

Wiedenmann, J., and Jensen, O.P. 2018. Uncertainty in stock assessment estimates for New England groundfish and its impact on achieving target harvest catches. Can. J. Fish. Aquat. Sci. **75**(3): 342–356. doi:10.1139/cjfas-2016-0484.

Wiedenmann, J., Wilberg, M., Sylvia, A., and Miller, T. 2017. An evaluation of acceptable biological catch (ABC) harvest control rules designed to limit overfishing. Can. J. Fish. Aquat. Sci. **74**(7): 1028–1040. doi:10.1139/cjfas-2016-0381.