Fisheries Research 209 (2019) 129–142

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Evaluating the performance of data-limited methods for setting catch targets through application to data-rich stocks: A case study using Northeast U.S. fish stocks

John Wiedenmann[a,*], Christopher M. Free[b], Olaf P. Jensen[b]

[a] Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, USA
[b] Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA

ARTICLE INFO

Handled by A.E. Punt

Keywords:
Data-poor
Data-limited
New England groundfish
Catch limits
Control rules
Stock assessment

ABSTRACT

Use of data-limited methods for setting target catches is increasing in the Northeast U.S., but there remains considerable uncertainty over which methods may be suitable for stocks in the region. We retrospectively evaluated the ability of data-limited methods to set target catches close to the overfishing limit for data-rich stocks in the Northeast U.S. Methods explored include options that would be used in truly data-poor cases (i.e., catch-only methods), but we also evaluated methods with different data requirements for stocks that have information beyond a catch time series. The majority of options we explored that used average catches over some portion of the time period, or adjusted the recent catches based on trends in an index were sensitive to the level of historical exploitation. Such methods produced target catches above the overfishing limit for stocks that had a history of overfishing, or target catches that were overly conservative for stocks with a history of light exploitation. Careful consideration of the level of historical exploitation rates, if possible, is therefore needed if using such approaches are to be applied. Catch curve methods, which require catch-at-age information, were the only approaches not sensitive to the level of historical exploitation, and were largely effective at setting target catches close to the overfishing limit, even for stocks with intense historical exploitation rates. However, there were cases where catch curve methods produced unsustainable target catches, particularly for stocks with episodic recruitments, such that care is needed when implementing catch curve methods.

1. Introduction

When possible, fisheries management actions are based on estimates of current stock status and management targets produced from complex, age-structured stock assessment models (Geromont and Butterworth, 2015). These models require large amounts of data, as well as analyst expertise and time to construct and run the model, and summarize model output. In the U.S., when such “data rich” assessments are not possible, catch limits must still be set for federally-managed fisheries, and a number of data-limited methods have been developed to set catch limits for cases with varying amounts of data.

The reasons preventing age-structured or less complex assessment models from being used vary. In truly data-poor cases, the necessary data are not available to run an assessment model, and the available catch time series may need to be used, often with assumptions about life history and relative stock status, to set target catches (MacCall, 2009; Berkson et al., 2011; Dick and MacCall, 2011). Stocks may have sufficient data to conduct an assessment, but the model results may be deemed too uncertain to be the basis for setting catch targets. One possible reason for this uncertainty is that some of the data may be uninformative, or different datasets may provide conflicting signals regarding population trend that cannot be reconciled given model assumptions. Such a case can be thought of as data-rich but information-poor, and more data-moderate approaches may be used that utilize available information beyond a catch time series, including indices of abundance (e.g., Geromont and Butterworth, 2014) and age-structured information (e.g., Thorson and Cope, 2015; for simplicity we herein refer to both data-poor and –moderate approaches for setting catch targets as data-limited methods).

Recent reviews conducted to determine the methods for setting target catches in U.S. fisheries revealed that data-limited methods were the most common basis for setting the acceptable biological catch (ABC) and annual catch limits (ACL; Berkson and Thorson, 2014; Newman et al., 2015). As of 2014, 30% of the ACLs were based on

* Corresponding author.
E-mail address: john.wiedenmann@gmail.com (J. Wiedenmann).

https://doi.org/10.1016/j.fishres.2018.09.018
Received 5 June 2018; Received in revised form 18 September 2018; Accepted 19 September 2018
0165-7836/

conventional, data-rich stock assessments, and 70% used data-limited methods (59% were data-poor and 11% were data-moderate; Newman et al., 2015). However, use of data-limited methods was not uniform across the Regional Management Councils, as regions such as the Caribbean and Western Pacific relied heavily on data-limited methods, while the Northeast U.S. (comprised of the Mid-Atlantic and New England regions) relied primarily on data-rich stock assessments (Berkson and Thorson, 2014; Newman et al., 2015).

While the Northeast U.S. may be thought of as data-rich region, use of data-limited methods is increasing. In the Mid-Atlantic, age-based assessments for Atlantic mackerel (*Scomber scombrus*) and black sea bass (*Centropristis striata*) did not pass review (Deroba et al., 2010; NEFSC, 2012), and explorations of a wide range of data-limited methods were used to help inform the determination of the ABC (Wiedenmann, 2015; McNamee et al., 2015**).** In New England, recent assessments did not pass review for the Georges Bank stocks of Atlantic cod (*Gadus morhua*) and yellowtail flounder (*Limanda ferruginea*), and for witch flounder (*Glyptocephalus cynoglossus*), and data-limited approaches were used to set the associated ABCs (Legault et al., 2014; NEFSC, 2015a,b). In all of these examples, the use of data-limited methods has been viewed as an interim measure until a new assessment model can be developed to address the issues identified in the failed assessments.

Although exploration of data-limited methods has increased in the Northeast U.S., there remains considerable debate about which methods may be suitable for stocks in the region. Developing support for or against particular data-limited methods requires both simulation testing (e.g., Wiedenmann et al., 2013; Carruthers et al., 2014; Geromont and Butterworth, 2015) and validation using information from stock assessments (Kokkalis et al., 2017; Sagarese et al., In press). Our aim in this paper was to identify effective data-limited methods for setting catch targets using information from data-rich stocks in the region. We retrospectively evaluated the performance of data-limited methods with varying data requirements encompassing methods that would be used for truly data-poor stocks, to more data-intensive methods that would be used for data-rich, information poor stocks. Using the most recent stock assessment as the source of information for historical stock dynamics, we compared the target catches from the data-limited methods to estimates of the overfishing limit (OFL; the catch that defines overfishing). Our focus was to identify which options, if any, were able to limit overfishing without being too conservative.

2. Methods

2.1. Data-limited methods

We applied 24 data-limited methods for setting target catches to 19 stocks managed by the New England and Mid-Atlantic Fishery Management Councils (NEFMC and MAFMC, respectively; see Table 1 for a list of the stocks). These stocks have a varied history of exploitation rates, although higher exploitation rates were generally observed in the 1990s than more recently (Fig. 1). The data-limited methods we used can be broadly classified into four categories: average catch methods, index-based methods, catch curve methods, and production models. These methods are detailed in Table 2, but we provide a brief summary of the general approaches here. Average catch methods set the target catch as some summary statistic (e.g., the mean or median catch) over part or all of the available catch data. Most of the average catch methods we explored only required a catch time series, although one method (DCAC; MacCall, 2009) also required some additional assumptions (Table 2). Index-based methods are an extension of the average catch methods, adjusting recent average catches based on trends in an index of abundance to set the target catch. These methods therefore require an index of abundance and total catch over time. Catch curve methods aim to estimate total mortality (Z) using numerical catch-at-length or catch-at-age data. Although length data may be more readily available in data-limited cases, we used only catch-at-age data because length data were often not reported in the assessments. Using catch-at-age data, Z is estimated by fitting a log-linear model to the fully-selected ages, and is then used with other assumptions depending on the method (Table 2) to adjust the recent average catch to generate a target catch. Finally, production models use an underlying surplus production model to estimate current biomass and reference points (more detail on the production models is provided below).

Our goal was not to test every possible data-limited method, but rather to understand the behaviors of a subset of methods in application to data-rich stocks. Therefore, the methods we used are not an exhaustive list of the possibilities. We omitted methods that required a complete time series of catch data (i.e., DB-SRA and its variants; Dick and MacCall, 2011) because complete catch histories were not available for any of the stocks in the region. In addition, we omitted the majority of methods that required assumptions about absolute current stock biomass (e.g., 10,000 mt), current relative status (e.g., the ratio of current to unfished biomass, or B/B_0), or relative change in abundance over the time series (e.g., a 30% decline). We included two methods that required such assumptions. The first is an average catch method that requires a user-specified assumption about stock depletion over the time period of available catches (DCAC) because this approach has been used across the U.S. (primarily in the Pacific; Newman et al., 2015; PFMC, 2016) and it has been suggested as a potential fallback method for some assessments in the Northeast U.S. (ASMFC, 2015a,b; Rago, 2017). DCAC adjusts the historical average catch to account for a one-time "windfall" catch that is the result of stock depletion, producing an estimate of yield that was likely to be sustainable over the same time period of available catch data. We explored fixed assumptions about depletion across stocks and across years in DCAC, assuming 60% and 80% declines in biomass relative to unfished biomass, B_0. We also explored a "data-rich" version of DCAC when biomass is known (MacCall, 2009), for comparison with the methods requiring multiple assumptions in the absence of biomass estimates (Table 2). The second method we used falls into the production model category (SPMSY), and require bounds for uniform distributions of relative status B/B_0 in the first and last years of available catch data. Martell and Froese (2013) provide guidance on the bounds based on the catch in those years relative to the maximum catch in the time series, and we used their recommended bounds here (Table 2).

2.2. Inputs and stock information

For each stock we used the most recent stock assessment that passed review as the primary source of information (Deroba, 2015; Legault et al., 2013; NEFSC, 2012, 2013, 2015a, 2017; Terceiro, 2016). We compared target catches from each data-limited method with the estimated OFL, so we needed all the necessary inputs for each method, as well as the estimated OFL over time for each stock. Time-varying estimates of the OFL were not provided in the assessments, but we calculated the OFL for the j^{th} stock in each year, t, with

$$OFL(j,t)=\sum_{a}^{a_{\max}}\frac{s(j,a,t)F_{MSY}(j)}{s(j,a,t)F_{MSY}(j)+M(j,a,t)}W(j,a,t)N(j,a,t)\left(1-e^{-(s(j,a,t)F_{MSY}(j)+M(j,a,t))}\right)$$

where a denotes age, N, s, and F_{MSY} are the model estimates of numerical abundance, fishery selectivity (proportion-at-age subject to fishing mortality), and limit fishing mortality rate, W is the observed weight in the catch, and M is the assumed rate of natural mortality. Note that this is an estimate of the OFL in hindsight from the most recent assessment for each stock, and is not the OFL that was specified for management purposes following earlier assessments.

Inputs to the data-limited methods obtained from the stock assessments were the annual observations of total catch (by weight) and numerical catch-at-age, and aggregate indices of abundance (kg per tow

Table 1
List of stocks explored in this analysis. Management refers to the regional fishery management council responsible for managing the stock (either New England, NEFMC, or Mid-Atlantic, MAFMC). The abbreviated name is how stocks are referenced in the text, and the code name is how they are referenced in Figs. 2, 6, and 8. Years refers to the years of catch and index data, used in our analysis. The first possible year of catch or index data for all stocks was 1978, and we excluded data from earlier years to omit the very large catches from the foreign fleets prior to the passing of the original Magnuson Act (Sosebee et al., 2006). For all stocks we also used assessment estimates from 1990 to the final year listed here to calculate the OFL (Eq. (1)).

Full Stock Name	Scientific Name	Abbreviated name (code)	Management	Years
Georges Bank Atlantic cod	*Gadus morhua*	GB cod (GBC)	NEFMC	1978–2010
Gulf of Maine Atlantic cod	*Gadus morhua*	GOM cod (GMC)	NEFMC	1982–2012
Georges Bank haddock	*Melanogrammus aegelfinus*	GB haddock (GBH)	NEFMC	1981–2012
Gulf of Maine haddock	*Melanogrammus aegelfinus*	GOM haddock (GMH)	NEFMC	1978–2012
Georges Bank yellowtail flounder	*Limanda ferruginea*	GB yellowtail flounder (GBYTF)	NEFMC	1979–2010
Cape Cod/Gulf of Maine yellowtail flounder	*Limanda ferruginea*	CC/GOM yellowtail flounder (GMYTF)	NEFMC	1978–2012
Southern New England/Mid-Atlantic yellowtail flounder	*Limanda ferruginea*	SNE/MA yellowtail flounder (SNYTF)	NEFMC	1981–2012
Georges Bank winter flounder	*Pseudopleuronectes americanus*	GB winter flounder (GBWIN)	NEFMC	1982–2012
Southern New England/Mid-Atlantic winter flounder	*Pseudopleuronectes americanus*	SNE/MA winter flounder (SNWIN)	NEFMC	1981–2012
witch flounder	*Glyptocephalus cynoglossus*	witch flounder (WCH)	NEFMC	1982–2010
American plaice	*Hippoglossoides platessoides*	Plaice (APL)	NEFMC	1980–2012
Acadian redfish	*Sebastes fasciatus*	Redfish (RED)	NEFMC	1978–2012
white hake	*Urophycis tenuis*	white hake (WHK)	NEFMC	1978–2012
pollock	*Pollachius virens*	pollock (PLK)	NEFMC	1978–2012
Atlantic herring	*Clupea harengus*	herring (HER)	NEFMC	1978–2012
Summer flounder	*Paralichthys dentatus*	Summer (SFL)	MAFMC	1982–2012
Scup	*Stenotomus chrysops*	Scup (SCP)	MAFMC	1978–2012
Bluefish	*Pomatomus saltatrix*	Bluefish (BLUE)	MAFMC	1982–2012
Black sea bass	*Centropristis striata*	BSB (BSB)	MAFMC	1980–2012

in the spring and fall coastwide bottom trawl survey) used in the assessment models. When long time periods of catch data were available, we omitted data prior to 1978 as very large catches occurred by foreign fleets prior to the passing of the Magnuson Act (Sosebee et al., 2006), and such large catches could influence methods that rely on an average catch over an appropriate time period. Catch-at-age data included a plus group, where catches across older ages are aggregated into a single age class. We explored the effect of including or excluding the plus group in the catch curve estimation of Z, and found that excluding the plus group generally resulted in smaller estimates of Z, with estimates close to or below 0 (indicating increased abundance-at age in the catch) produced more frequently than when the plus group was included (Fig. 2). We therefore included the plus group in the calculation of Z. For black sea bass (*Centropristis striatus)* only the numerical fall index was available, and for bluefish (*Pomatomus saltatrix*) we used the recreational CPUE index from the Marine Recreational Information Program (MRIP), as bluefish are likely poorly sampled in the bottom trawl survey.

The catch curve methods and DCAC required additional life history information (Table 2). DCAC requires estimates of M, F_{MSY}/M, and B_{MSY}/B_0. For B_{MSY}/B_0we used the spawning potential ratio (SPR) proxies used to define reference points for each stock, which was 0.4 for all but two stocks (Table 3), and this value is identical to the mean value across stocks estimated in the meta-analysis of Thorson et al. (2012). We used the assumed M from each assessment, as well as the ratio of the assessment-estimated F_{MSY} to the assumed M. Values for F_{MSY}/M were generally comparable to the mean family-level estimates from the meta-analysis of Zhou et al. (2012), although some of our estimates were considerably higher (Table 3). Using these values as inputs to DCAC should reduce uncertainty and potentially improve

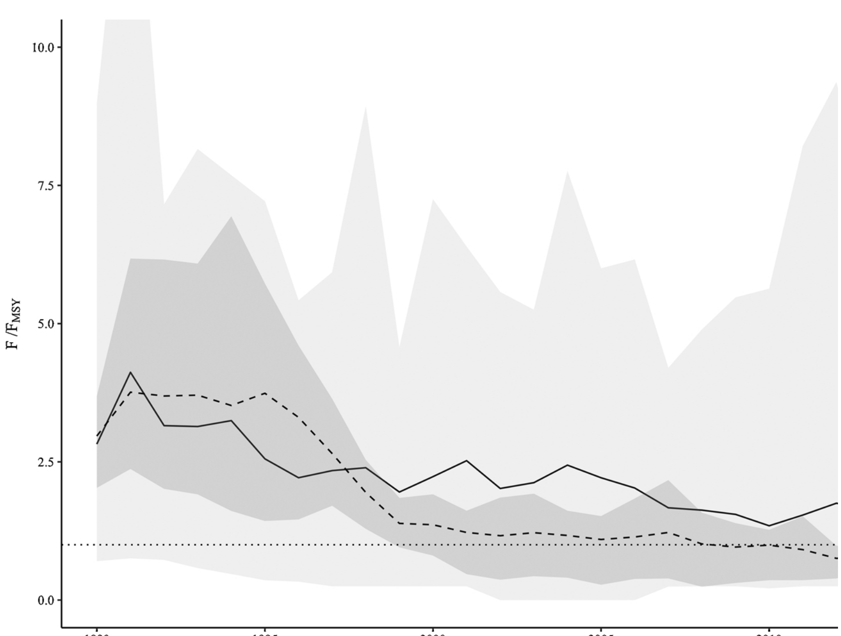

Fig. 1. The mean annual F relative to F_{MSY} across stocks used in this study from New England (solid line) and the Mid-Atlantic (dashed line). The light and dark shaded regions represent the range of observed F/F_{MSY} for New England and the Mid-Atlantic, respectively. The horizontal line at 1 represents F_{MSY}, above which overfishing is occurring.

Table 2
Brief description and equations for the data-limited control rules used, with the source for each control rule when available. Many of the approaches use multiyear averages of catch and index data, which is denoted $\bar{C_N}$, and $\bar{I_N}$, respectively, where N is the number of the most recent years used to calculate the average (typically 3, 5, 10, or all (Y) available years). For the index-based methods, when two indices of abundance were available for a stock (i.e., spring and fall survey), we calculated a single, unweighted average index across surveys for use in the methods. All of the catch curve methods used the last three years of available catch-at-age data, and catch data were summed across those years for each age to produce a single catch-at-age vector to estimate Z. For assumed inputs to the different methods, the assumed CV used to generate a distribution for each input is in parentheses (see Table 3 for input values and definitions).

Data-limited method abbreviation	Description	Inputs	Source
Average catch methods			
AvC, AC_3yr, AC5yr	$C_{targ} = \frac{1}{T}\sum_{t=Y-T+1}^{Y} C(t)$, where Y is total number of years available, and T is number of years to use (T = all years (Y), or the most recent 3, or 5 years)	Total catch (by weight)	
AC75_3yr, AC75_5yr	75% of the average catch over the last 3 or 5 years $C_{targ} = 0.75\frac{1}{T}\sum_{t=Y-T+1}^{Y} C(t)$, with T = 3 or 5.	Total catch (by weight)	
MC, MC_50	The median, and 50% of the median catch over the whole time period	Total catch (by weight)	
DCAC_20, DCAC_40	Depletion-corrected average catch. A method for adjusting average catches based on an assumed change in biomass over the time period. $C_{targ} = \sum_{t=1}^{t=Y} C(t)\left(Y + \frac{\Delta}{F_{MSY} \cdot B_{MSY}/B_0}\right)^{-1}$ Where F_{MSY} is calculated as the product of the assumed M and the assumed ratio of F_{MSY} to M. Δ is the assumed depletion over the time period relative to $B_0\,(B(1){-}B(Y)))/B_0$, and we assumed values of 0.8 and 0.6 for the DCAC_20 and DCAC_40 runs, respectively.	Total catch (by weight), assumed F_{MSY}/M (0.2), B_{MSY}/B_0(0.05),M(0.2), and Δ(0.2)	MacCall (2009)
DCAC_DR	The "data-rich" version of DCAC, calculated using estimates of the exploitable biomass (B_e) in the first (t=1) and last (t = Y) years of available catch data, $C_{targ} = \frac{(\sum_{t=1}^{t=Y} C(t) - (B_e(1) - B_e(Y)))}{Y}$	Total catch (by weight), estimates of exploitable biomass	MacCall (2009)
Catch Curve Methods			
BK_CC1 BK_CC3 BK_CC5	Variations of the Beddington and Kirkwood life history method combined with catch curve analysis. $C_{targ} = \frac{0.6k \cdot \bar{C_3}(1-e^{-F})^{-1}}{0.67 - L_{ratio}}$, where $\bar{C_3}$ is the average catch in the last 3 years, F is estimated using the assumed M and the catch curve estimate of Z (F = Z-M), k is the von-Bertalanffy growth rate, and L_{ratio} is the ratio of the length at first capture to L_∞. The differences across BK_CC1, BK_CC3, and BK_CC5 are the assumption about L_{ratio} (0.1, 0.3, and 0.5, respectively).	Total catch (by weight), numerical catch-at-age, assumed k(0.1), L_∞(0.1), t_0(0.1), b (0.1), c(0.1),M(0.2), Lat first capture (0.2).	Beddington and Kirkwood (2005)
YPR_CC	Nearly identical to Fdem_CC, $C_{targ} = F_{MSY}\bar{C_3}(1{-}e^{-F})^{-1}$, but with F_{MSY} based on the $F_{0.1}$ estimate from a yield-per-recruit model, assuming knife-edge selection at the length of full selection (Table 3).	Total catch (by weight), numerical catch-at-age, assumed a_{max}, k(0.1), L_∞(0.1), t_0(0.1), b(0.1), c(0.1),M(0.2), L_{FS}(0.2)	Carruthers and Hordyk (2017)
Fdem_CC	$C_{targ} = F_{MSY}\bar{C_3}(1{-}e^{-F})^{-1}$, where $\bar{C_3}$ and F are described in the BK_CC methods, and F_{MSY} is calculated as r/2, with r calculated using the demographic approach of McAllister et al. (2001).	Total catch (by weight), numerical catch-at-age, identical assumed inputs as YPR_CC, but also with h(0.2).	Carruthers and Hordyk (2017); McAllister et al. (2001)
M_CC	Nearly identical to Fdem_CC and YPR_CC, $C_{targ} = F_{MSY}\bar{C_3}(1{-}e^{-F})^{-1}$, but with F_{MSY} set equal to the assume value of M	Total catch (by weight), numerical catch-at-age, M(0.2)	
Index-based methods			
Islope1 Islope4	The average catch from the most recent 5 years ($\bar{C_5}$) is adjusted based on the slope (λ) of a log-transformed index of abundance over the same period. $C_{targ} = (1 + \varnothing\lambda)\eta\bar{C_5}$. For Islope1 $\varnothing$ = 0.4, and η = 0.8. For Islope4 $\varnothing$ = 0.2, and η = 0.6.	Total catch (by weight), survey indices of abundance.	Geromont and Butterworth (2014)
Itarget1 Itrarget4	Uses the recent 5 and 10 year average index ($\bar{I_5}$ and $\bar{I_{10}}$, respectively) and $\bar{C_5}$ to calculate C_{targ} with $C_{targ} = \begin{cases} 0.5\eta\bar{C_5}(1 + (\bar{I_5}{-}0.8\bar{I_{10}})/(\gamma\bar{I_{10}}{-}0.8\bar{I_{10}}))\bar{I_5} > 0.8\bar{I_{10}} \\ 0.5\eta\bar{C_5}(\bar{I_5}/0.8\bar{I_{10}})^2\bar{I_5} > 0.8\bar{I_{10}} \end{cases}$ For Itarget1 γ = 1.5, and η = 1. For Itarget4 γ = 2.5, and η = 0.7.	Total catch (by weight), survey indices of abundance.	Geromont and Butterworth (2014)
GB_slope	Similar to the Islope methods, with $C_{targ} = (1 + \lambda)\cdot\bar{C_5}$, with estimates of C_{targ} more extreme than ± 20% of the most recent catch capped at ± 20%.	Total catch (by weight), survey indices of abundance.	Carruthers and Hordyk (2017); Geromont and Butterworth (2014)
PlanB_3	Adjust the 3-year average catch ($\bar{C_3}$) based on the transformed slope (λ) of a log-linear fit to the last 3 years of a loess-smoothed index of abundance. $C_{targ} = \lambda\cdot\bar{C_3}$. The span for the loess fit was set to 9.9/Y.	Total catch (by weight), survey indices of abundance.	NEFSC (2015a)
Production models			
Schaefer production model (called Production)	A Schaefer surplus production model ($B(t)$ = B(t-1) + rB(t-1)(1-B(t-1)/K – C(t-1)) fit to the available indices of abundance and catch data through year Y, estimating r, K, and biomass in the first year with available data. The target catch in the final year Y is $C_{targ} = B(Y)r/2$, where r /2 is the estimated F_{MSY}.	Total catch (by weight), survey indices of abundance.	Schaefer (1954)
SPMSY	A "simple method for estimating MSY" that assumes an underlying production model, and randomly draws values of r and K and starting and ending estimates of relative depletion (B(1)/K and B(Y)/K) to find the combination of parameters that are sensible given the catch history (i.e., parameters that results in biomass ≤ catch in any given year are excluded). The target catch in the final year is KB(Y)/K• r /2.	Total catch (by weight), assumed B(1)/K and B(Y)/K, drawn from uniform distributions (bounds for the draws varied based on the catch in those years relative to the maximum catch, see Martell and Froese, 2013 for details).	Martell and Froese (2013)

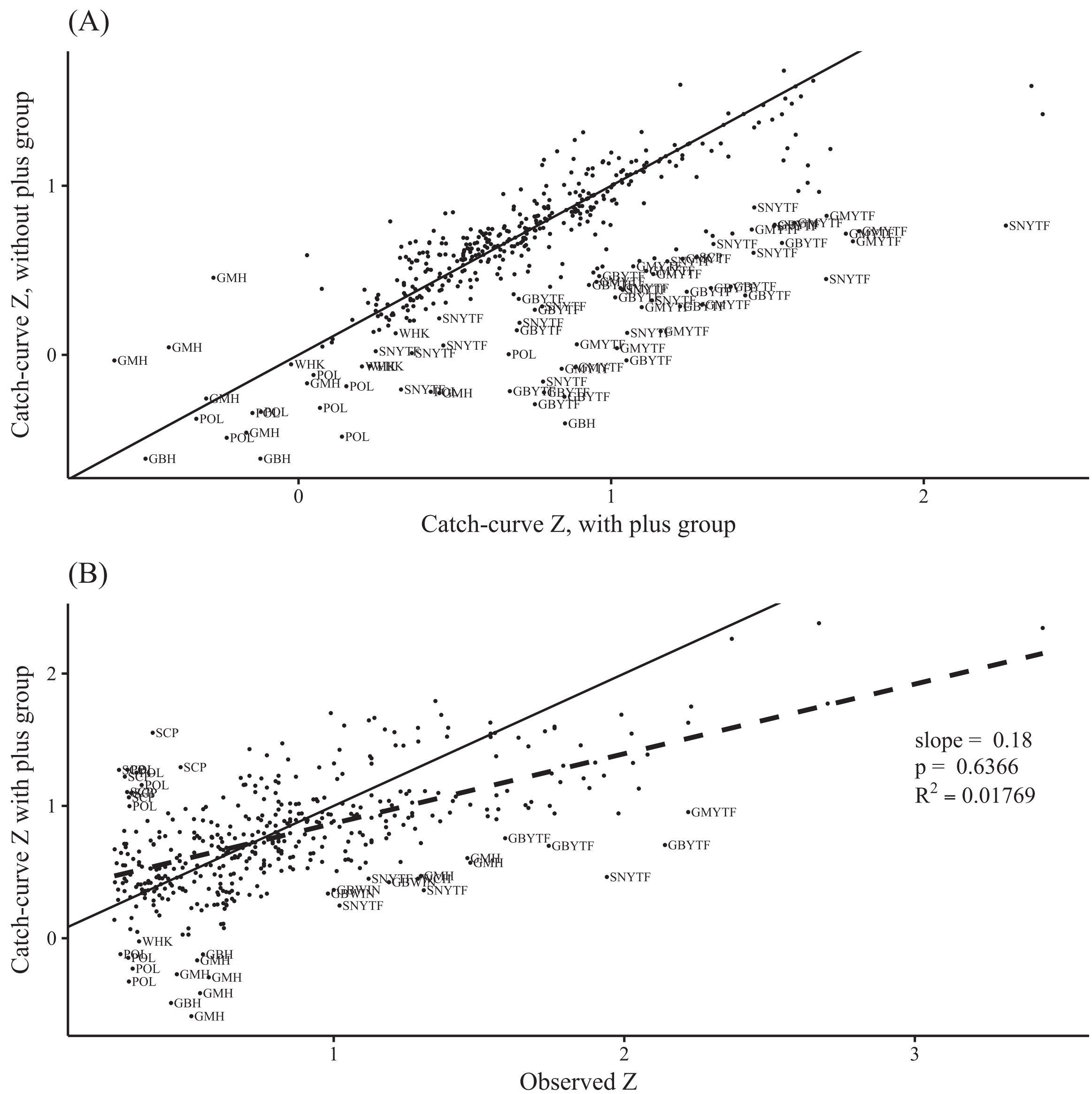

Fig. 2. Catch curve estimates of total mortality (Z) across years for each stock. Upper: Comparison of Z estimates when the plus group was included in the log-linear fit to when the plus group was omitted from the fitting. Lower: Comparison of the estimated Z including the plus group to the observed fully-selected Z obtained from the assessment. The solid line is the 1:1 line, and the dashed line (bottom plot only) is the linear fit, omitting all negative values of Z. Labels have been added to some of the points to identify specific stocks where 1) negative values of Z were estimated (with or without the plus group), 2) when there was a large discrepancy in between estimates with or without the plus group included (upper), and 3) when there was a large discrepancy between the estimated Z and the observed Z from the stock assessment (lower).

performance since these values were also used to calculate the OFL. MacCall (2009) suggests using DCAC only when $M \leq 0.2\text{yr}^{-1}$, and also using $F_{MSY}/M \leq 1$,otherwise the correction factor might be too small. Our estimates of M were mostly $\leq 0.2\text{yr}^{-1}$, but F_{MSY}/M values were sometimes > 1 (Table 3). To test the sensitivity of DCAC to our assumptions, we used the data-rich version that circumvents these assumptions using changes in biomass estimates to adjust the catch (Table 2).

Inputs for the catch curve methods beyond the catch-at-age data were used to estimate F_{MSY} using various approaches (Table 2). The inputs for the various methods included maximum age, steepness of the stock-recruit relationship, von Bertalanffy growth parameters, length-weight conversion parameters, and also the length-at-first-capture and –at-full selection in the fishery. Steepness values were obtained from Myers et al. (1999). Maximum age and the parameters for the von Bertalanffy model were taken from the current or past assessments when available, or from Fishbase (www.fishbase.org). Parameters for converting length to weight were obtained from Wigley et al. (2003). We defined length-at-full selection as the mean length calculated from the von Bertalanffy growth model corresponding to the age at 95% selection in the fishery. Defining length-at-first capture was challenging for each stock. For the lone method that required this input, we explored three versions where length-at-first capture was assumed to be 10, 30, and 50% of the asymptotic length (Table 2). Parameters values for each stock are listed in Table 3.

2.3. Application

The data extracted for each stock were then used in the data-limited methods to calculate target catches. We used the data-limited toolkit (DLMtool; Carruthers and Hordyk, 2017) for our analyses, which is an R (R Core Team, 2017) package developed to test and apply data-limited methods for real-world applications. DLM tool has two distinct components, a management strategy evaluation (MSE) simulation modules

Table 3
Stock-specific life history parameters use in DCAC and the catch curve methods. Parameters are as follows: a_{max} is the maximum age; h is the steepness of the stock-recruit relationship; M is the natural mortality rate, F_{MSY}/M is the ratio of F_{MSY} to M; B_{MSY}/B_0 is the fraction of unfished biomass where maximum production occurs; L_∞, k, and t_0 are the von Bertalanffy growth model parameters ($L(a) = L_\infty(1 - e^{-k(a-t_0)})$, b and c are the parameters relating length to weight ($W(a) = bL(a)^c$), and L_{50} and L_{FS} are the lengths at 50 and full selectivity, respectively. Values for F_{MSY}/M were based on the estimated F_{MSY} and the assumed M from the assessment, and the value in parentheses is the family-level mean from Zhou et al. (2012). The assumed M was age- and time-invariant for all stocks but summer flounder and Atlantic herring we used the mean value across fully-selected ages as our assumed M. Estimates of B_{MSY}/B_0 are based on the management SPR targets. In DLMtool all of these specified inputs were set as the mean of lognormal distribution for the methods that used them, and we used the DLMtool default CVs for each of these inputs to create the distributions (CVs listed in Table 2).

Management	Stock	a_{ma}	h[a]	M	F_{MSY}/M	B_{MSY}/B_0	L_∞	k	t_0	b ($\times 10^{-6}$)	c	L_{FS}
NEFMC	GOM Cod	16	0.84	0.20	0.925 (1.01)	0.40	150.9	0.11	0.13	5.13	3.16	60.0
	GB Cod	16	0.84	0.20	0.85 (1.01)	0.40	114.0	0.22	0.17	7.29	3.08	58.0
	GOM Haddock	22	0.74	0.20	1.50 (1.01)	0.40	64.2	0.40	−0.30	9.30	3.02	51.0
	GB Haddock	25	0.74	0.20	1.50 (1.01)	0.40	73.8	0.38	0.17	8.13	3.07	51.0
	GB yellowtail flounder	12	0.75	0.20	1.25 (1.16)	0.40	50.0	0.33	0.00	5.76	3.13	35.0
	SNE/MA yellowtail flounder	12	0.75	0.30	1.17 (1.16)	0.40	35.4	0.91	0.25	5.76	3.13	34.0
	CC/GOM yellowtail flounder	12	0.75	0.20	1.40 (1.16)	0.40	48.0	0.35	−0.10	5.76	3.13	36.5
	GB winter flounder	19	0.80	0.30	1.40 (1.16)	0.40	58.0	0.28	0.00	8.85	3.11	36.0
	SNE/MA winter flounder	16	0.80	0.30	1.08 (1.16)	0.40	46.5	0.32	0.00	10.40	3.04	33.4
	Plaice	30	0.80	0.20	1.00 (1.16)	0.40	62.2	0.17	0.00	2.86	3.31	40.0
	Witch	25	0.80	0.15	1.20 (1.16)	0.40	60.0	0.15	0.02	2.39	3.26	41.5
	Acadian redfish	50	0.47	0.05	0.76 (0.69)	0.50	35.9	0.16	−0.24	8.29	3.20	29.7
	White hake	20	0.79	0.20	1.00 (1.01)	0.40	135.3	0.09	−0.89	3.13	3.23	47.0
	Pollock	24	0.81	0.20	1.00 (1.01)	0.40	108.3	0.16	−0.44	7.43	3.09	68.0
	Atlantic herring	15	0.44	0.45	0.55 (0.88)	0.40	28	0.518	0.40	7.53	3.0314	25
MAFMC	Black sea bass	15	0.80	0.20	0.80 (0.92)	0.40	46.5	0.15	−0.51	15.60	3.1365	22
	Bluefish	14	0.80	0.20	0.85 (0.92)	0.40	113	0.126	−0.60	10.90	3.0548	41
	Summer flounder	14	0.80	0.25	1.24 (1.16)	0.35	85.5	0.14	−1.20	3.89	3.25	36.0
	Scup	15	0.95	0.20	0.80 (0.92)	0.40	46.5	0.15	−0.51	15.60	3.14	22.0

[a] Steepness values were obtained from Myers et al. (1999). When not provided at the species level, we used the value at the Family level. When the Family level was not provided (bluefish and black sea bass), we assumed a value of 0.8.

to test methods, and an application side where the available data for a stock are input to estimate the target catch for each method. We used the application portion of DLMtool (and not the MSE), which has a wide range of built-in methods of varying complexity, but it also allows users to specify their own unique options, or to modify the existing methods as needed. All but three of the methods we used in DLMtool were either existing or slight modifications of existing options. We added the data rich version of DCAC (DCAC_DR), thePlanB_3 index-based method, currently used as a fallback approach in New England (NEFSC, 2015a; code obtained here https://github.com/cmlegault/PlanBsmooth/wiki/Basics), and the M_CC catch curve method that sets F_{MSY} equal to the assumed M (Table 2). We also modified all of the catch curve methods to account for low estimates of Z. All of the catch curve methods estimate the mean F in the last three years using the estimated Z and assumed M ($F = Z - M$), and adjust the average catch over this period up or down is F is below or above the estimated F_{MSY}, respectively (Table 2). When $M > Z$, DLMtool uses a default F of 0.005yr^{-1}, but we used a minimum F of 0.05yr^{-1} for all catch curve methods, but also compared the impact of this minimum to the lower default value.

DLMtool includes methods that use underlying production models, including DB-SRA (Dick and MacCall, 2011), which we did not use due to the full catch time series requirement, and SPMSY (the simple method for estimating MSY; Martell and Froese, 2013), which we did use. SPMSY is similar to DB-SRA, in that it estimates MSY-based reference points and the OFL in the last year, but it does not require a complete catch time series (Table 2). In addition to SPMSY, we included a Schaefer surplus production model in our analysis (Schaefer, 1954), implemented outside of the DLMtool framework. Parameters for the surplus production model (r, K, and starting biomass relative to K) were estimated by fitting the model to the available indices of abundance (and estimating catchability for each survey)using a maximum likelihood approach (assuming lognormal observation errors in the indices, with even weighting to each index when multiple were available) and assuming catch data are known for each stock (Fig. 3). The target catch was set to the estimate of the OFL in the last year (Y) of each model fit (OFL $= r/2 \cdot B(Y)$; Table 2). We considered other variations of production models where B_{MSY} is not necessarily $K/2$(Pella and Tomlinson, 1969; Fox, 1970), but ultimately decided on using the Schaefer model, as it allows for more direct comparisons with SPMSY (which assumes Schaefer dynamics). A production model fit to catch and survey data is a simpler form of an assessment, and we are making comparisons to estimates of the OFL from age-based assessments (Arnold and Heppel, 2015; Cope et al., 2015). The debate over which model may be "correct" has a long history in fisheries; we are not attempting to address the debate here. Rather, here we asked that if the true dynamics of a stock were those estimated in the age-based model, what would the impact have been if a production model were used to set target catches (Punt and Szuwalski, 2012)?

For each data-limited method, DLMtool produces a distribution of target catches (C_{targ}) based on the user-specified number of iterations. The stochastic calculation of the target catch varies by method, with some methods relying on user-specified levels of uncertainty (an assumed CV for many of the parameters). Other methods rely on the uncertainty in estimated values, such as the standard deviation of the average catch over some time period, or in the standard error of estimates of the slope and intercept parameters from a linear fit to the index of abundance over time, or in the log-transformed numerical catch at age in a catch curve analysis. For all inputs that required a specified CV, we used the default CVs specified in DLMtool across stocks. The highest default CV we used was 0.2, which was for inputs likely to be more uncertain than others (e.g., M or relative depletion; Table 2), and resulted in distributions generally ranging from 0.5 to 2 times the specified mean for such inputs (Table 2).

2.4. Performance

We calculated the distribution of target catches using 1000 iterations for each of the methods in DLMtool from 1990 to 2012 for each stock. We used the median of the distribution of the target catch for each stock/year/method as our value for comparison with the

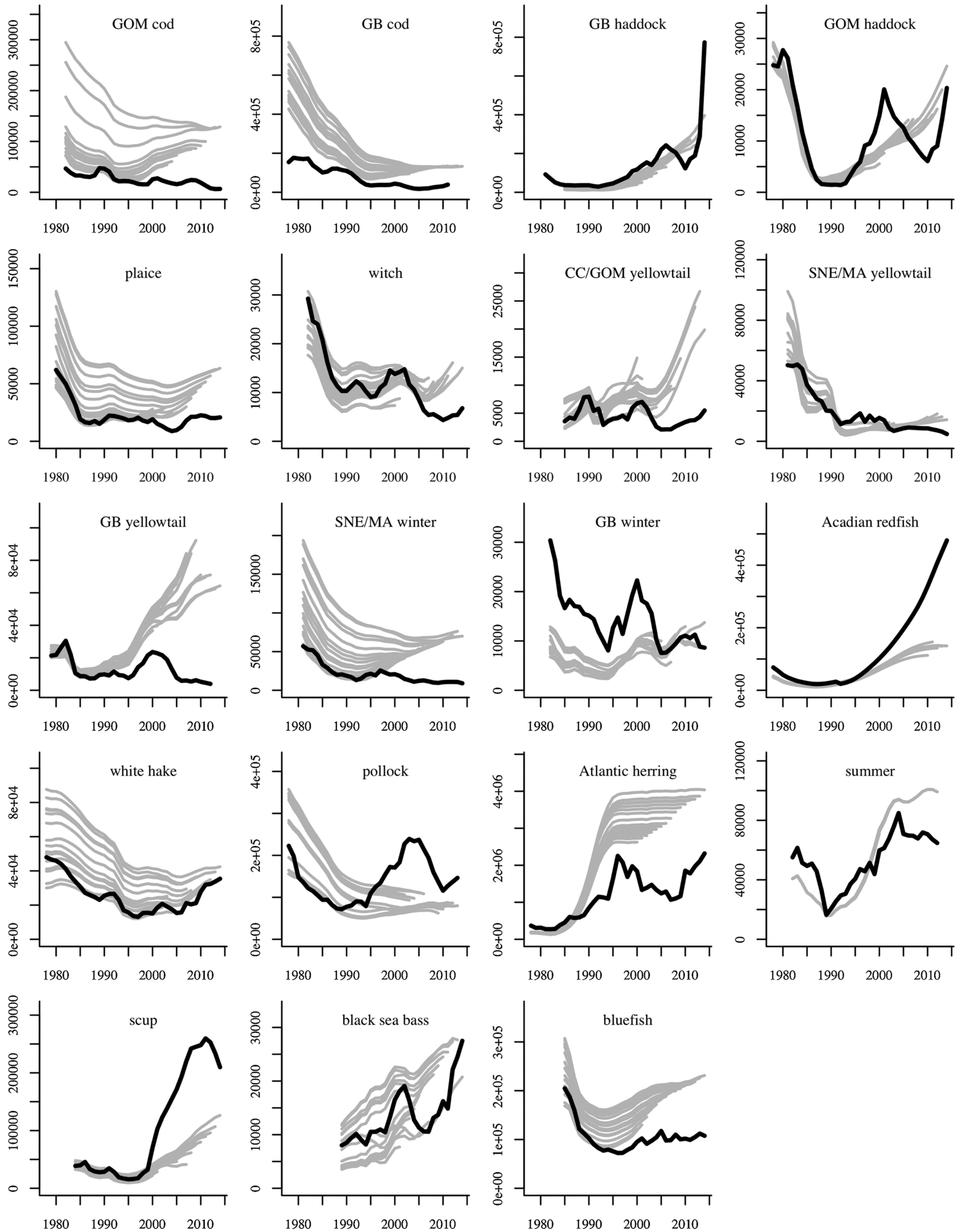

Fig. 3. Schaefer surplus production model fits (gray lines) of total biomass each year, along with the current estimates of total biomass for each stock. Multiple fits were done for each stock using different length time series (i.e., fit through 2000, 2001, 2002, and so on).

estimated OFL (Eq. (1)), with a one year lag. Inclusion of a lag was intended to mimic the process of setting target catches, where under the best of circumstances the target catch would be calculated using data from the previous year. We selected 2012 as the cutoff to reduce the impact of uncertainty in more recent assessment estimates based on retrospective patterns. Recent assessments for Georges Bank cod, Georges Bank yellowtail flounder, and witch flounder did not pass review due to increasingly strong retrospective patterns. We still included these stocks in our analyses, using the most recent assessment that passed review, and only using data through 2010, assuming that model estimates become more stable moving back in time. However, changes to future assessments for these or other stocks that dramatically change historical estimates would alter our estimates of the OFL, and potentially our conclusions.

We also compared target catches for each stock to the target catches set by management. We obtained management target catches from 2000- and 2004-onward for Mid-Atlantic and New England stocks, respectively, for comparison with the target catches estimated by the different data-limited approaches. From 2010-onward the target catches were considered the ABC, but prior to 2010 they were often referred to as the total allowable catch (TAC).For simplicity we refer to them as the original target catches (OTC), noting that they were not always set to achieve the OFL (or close to it), either in cases without an assessment or in cases of rebuilding.

Because we used static estimates from real stocks it is not possible to remove the target catch (i.e., there is no feedback between the catch, stock, and data like in MSE simulation models). Our annual estimates of the target catch must therefore be viewed as independent from one another, and we cannot calculate common MSE performance metrics such as the probability of overfishing or the change in biomass over time in relation to each method. Nevertheless, our approach is a useful exploration of what the target catch would have been under a data-limited method in any particular year from 1990 to 2012.

3. Results

Fig. 4 shows the range of median catch/OFL estimates for each method across stocks and years, separated by historical fishing intensity. For each method, a wide range of target catches (relative to the OFL) occurs for stocks with and without a history of overfishing. For stocks without a history of overfishing, most methods tended to produce target catches below the OFL (Fig. 4A). Exceptions to this were the Schaefer surplus production model, and the catch curve methods BK_CC3 and BK_CC5 (see Table 2 for more details on each method), which had a median catch/ OFL above 1. In contrast, most methods resulted in target catches above the OFL for stocks with a history of overfishing, with only the index-based approach Itarget4 and catch curve method BK_CC1 having a median catch/OFL below 1 (although other approaches had medians close to 1; Fig. 4B).

It is evident from Fig. 4 that the performance of the methods is sensitive to the exploitation history for each stock. This result is expected given that many of these approaches use an average catch over some time period as the foundation for setting the target catch. The time period of catches (and other inputs) used by each method varies, but was typically 3, 5, or all available years of data. For each stock in each year we calculated the mean F/F_{MSY} over the relevant period for a method (i.e., the last 3 years if the method uses an average catch over the last 3 years) and compared these estimates to the target catch/OFL from each method (Fig. 5). The average catch and index-based methods resulted in target catches/OFL that were positively correlated with the mean F/F_{MSY} over the same period (Fig. 5A–J). Weaker correlations ($R^2 < 0.5$) occurred for approaches that used the all available years of catch data compared to those that used only the most recent three or five years of data ($R^2 \geq 0.8$). The slopes of the fit differed greatly across methods, although most had positive slopes, indicating sensitivity to recent or historical fishing intensity. Many of the average catch approaches and both production model approaches had slopes > 1, resulting in a greater magnitude of overfishing for stocks that had experienced higher rates of historical overfishing, particularly those that used all available catch data (but excluding years prior to 1978). One approach that uses the average catch over the available time period is DCAC (MacCall, 2009), and we found that the assumed depletion level (DCAC_20 and DCAC_40) did not have a large impact on the target catch/OFL from this method (Fig. 5H–I), and performance using the

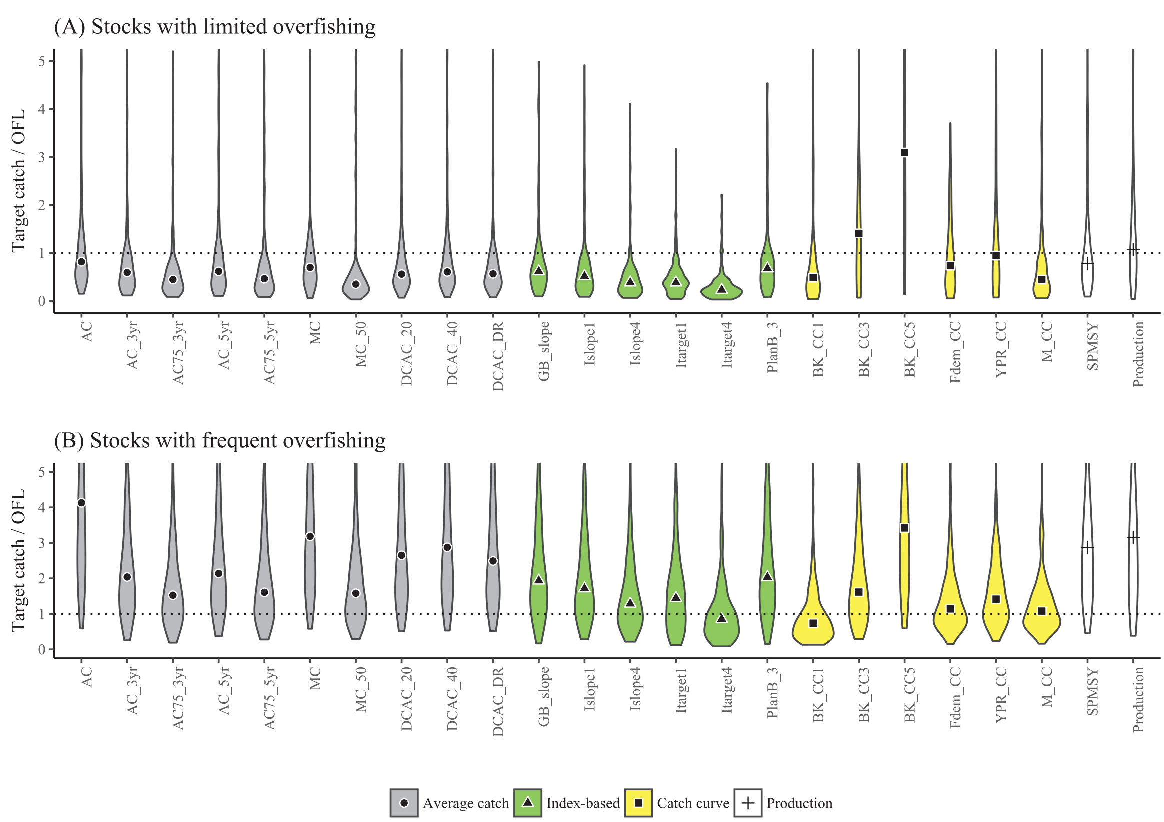

Fig. 4. The median target catch relative to the OFL from each control rule across stocks and years. The black shapes represent the median for each control rule. Top panel: stocks without a history of overfishing (defined as having less than half of the years from 1990 to 2012 with overfishing). Bottom panel: stocks with a history of overfishing (more than half of the years). The Production method refers to the Schaefer model.

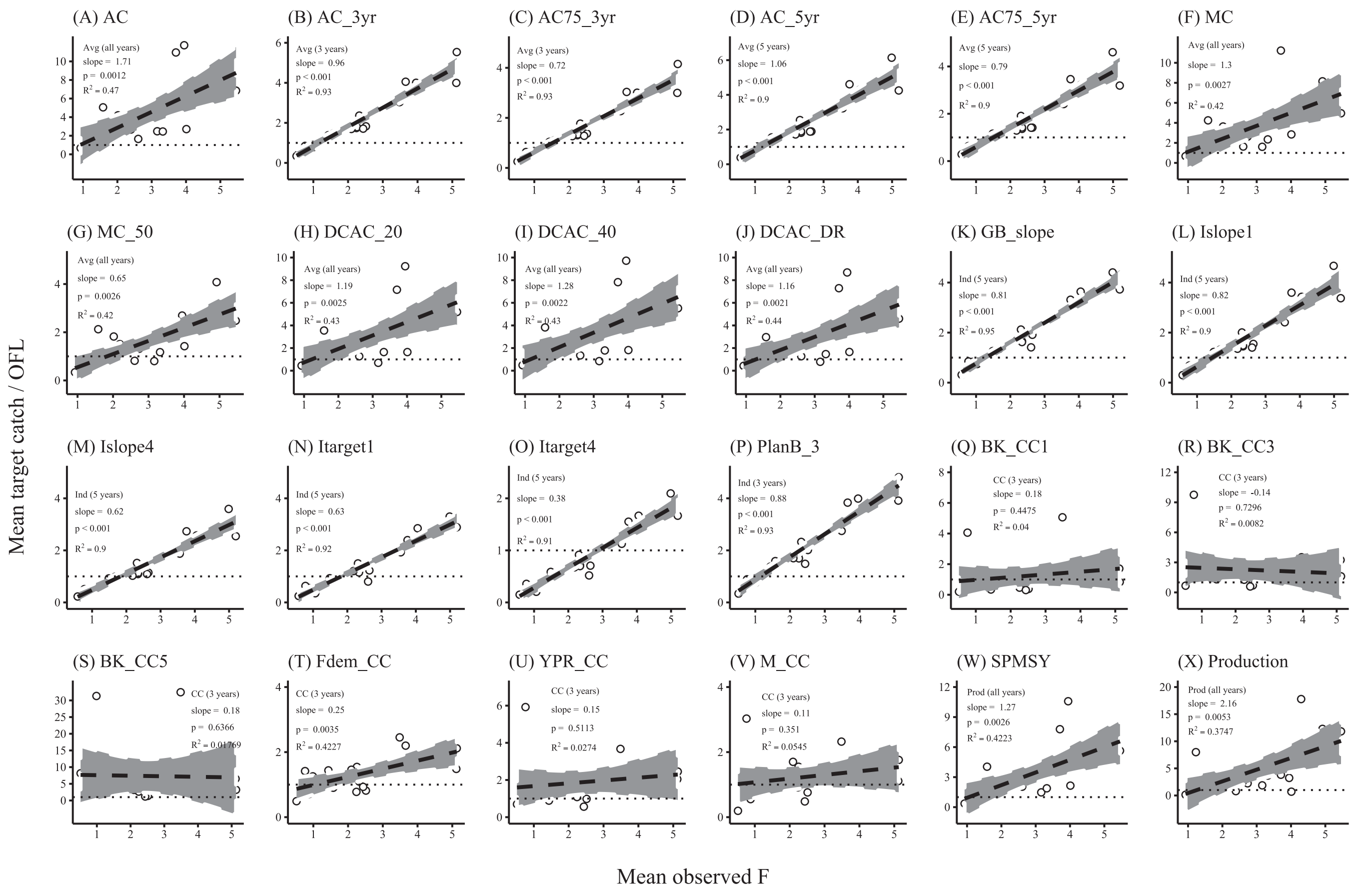

Fig. 5. For each method, the mean target catch relative to the OFL (averaged across years for each stock) as a function of the mean *F* during the relevant time period for each control rule. The relevant time period is defined as the years of data used in the particular control rule (typically the most recent 3 or 5 years, or all available years in some cases). The horizontal line at 1 indicates when the target catch is equal to the OFL. On each panel the approach category is listed (Avg = average catch (A–J); Ind = index-based (K–P); CC = catch curve (Q–V); Prod = surplus production model (W–X)), as well as the slope, p-value and R^2 for a linear fit. Most approaches had significant positive slope, indicating that the target catch/OFL increased with increasing mean *F*, although the magnitude of the increase varied greatly across methods (from 0.25 for Fdem_CC(T) to 2.19 for the Schaefer production model(X)). Most catch curve methods had slopes that were not significantly different from 0, indicating that the target catch/OFL was independent of the recent mean *F*.

data-rich version (DCAC_DR; where changes in assessment-estimated biomass are used to adjust the catch; Table 2) was similar to the other DCAC implementations (Fig. 5J).

Catch curve methods, on the other hand, were not correlated with the exploitation rate during the relevant period (non-significant slopes for all but Fdem_CC; Fig. 5Q–V). Target catches from these methods were often close to the OFL despite intense overfishing, but occasionally target catches were well above the OFL following low exploitation rates. Insensitivity to historical exploitation rates (which are often unknown) is a desirable behavior of a data-limited method, but it is problematic that the target catch from these methods was well above the OFL for some stocks. The stocks with very high target catch/OFL were Atlantic herring (*Clupea harengus*) and SNE/MA yellowtail flounder (Fig. 6A), but more stocks would have had very high target catches/OFL for certain methods if we had used the default minimum *F* in DLMtool (we used a minimum of 0.05yr^{-1} compared to the default of 0.005yr^{-1}; Fig. 6B). For Atlantic herring, pollock (*Pollachius virens*), GOM haddock (*Melanogrammus aegelfinus*), and white hake (*Urophycis tenuis)*, estimates of *Z* from the catch curve analysis were occasionally at or below the assumed *M*, resulting from high variability in recruitment. This problem was exacerbated by methods that resulted in high estimates of F_{MSY}, as assuming $F_{MSY} = M$ (the M_CC method) mitigated against very high catches for these stocks (Fig. 6B). SNE/MA yellowtail was not impacted by the assumed minimum *F* (Fig. 6B), and the other yellowtail flounder stocks also had relatively high target catch/OFL estimates, on average (Fig. 6A), and these were stocks where *Z* was consistently underestimated (albeit above the assumed *M*; Fig. 2B). Interestingly, these stocks have the fewest age classes used in the assessment (6), and the age-at-full selection in the catch was typically age 2 or 3, leaving only 3–4 points for the catch curve regression. This limited number of ages may be contributing to the consistent underestimation of *Z* for these stocks, which causes the target catches from the catch curve approaches to overestimate the catch relative to the OFL.

Our measure of performance thus far has been how close the target catches would have been to the OFL in a given year for a stock, and we found that many of the options would have resulted in continued under- or overexploitation, depending on the intensity of exploitation experienced (Fig. 5). Despite continued overfishing for a stock, the data-limited approaches could still be improvements over the existing management advice. Fig. 7 shows the proportion of times that the data-limited methods set catch targets closer to the OFL than the original target catches (we use OTC for simplicity, noting that the target catches were considered the ABC from 2010-onward, but were referred to as the TAC, in earlier years). The ratio of the OTC to the OFL is based on the current estimates of the OFL from the most recent assessment for a stock, and not what was estimated to be the OFL in earlier assessments at the time the target catch was set. In cases where the OTC was below the OFL (either due to using a buffer or due to earlier assessments/projections underestimating biomass, or both), data limited methods were more often than not more conservative than the OTC. When the OTC was above the OFL (largely due to assessments/projections

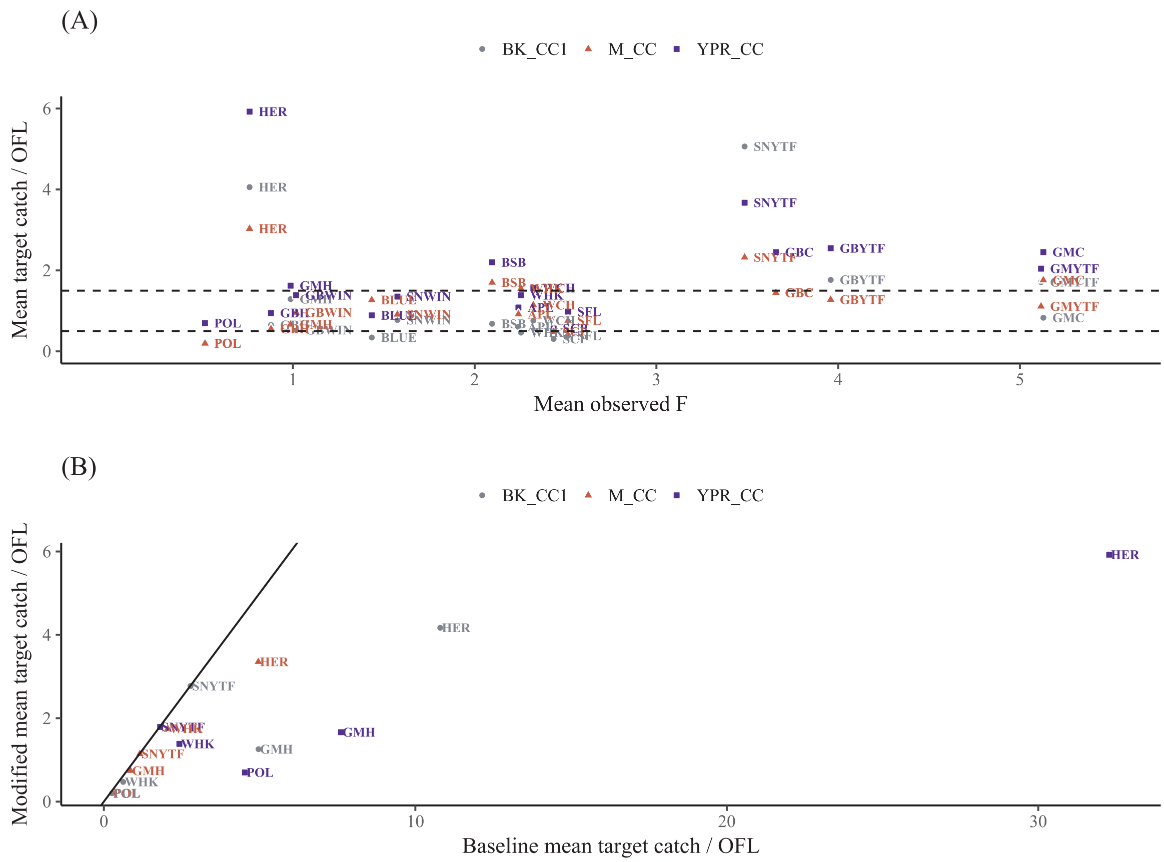

Fig. 6. A) Similar to Fig. 5, but for three catch curve methods, with individual stock name abbreviations showing (see Table 1). Each point represents the average across years (1990–2012) for each stock. The dashed horizontal lines shown when the control rule was able to get within ± 50% of the OFL, on average. B) The target catch for a subset of stocks, based on the assumed minimum *F* estimated from the catch curve analysis (estimated F = estimated Z – assumed M). The baseline method uses the DLMtool default minimum *F* of 0.005yr^{-1}, while the modified method uses a minimum *F* of 0.05yr^{-1}. The solid black line is the 1:1 line, such that points close to the line indicate insensitivity to the assumed minimum *F*. The target catch in A) was calculated using the modified, higher minimum *F*.

overestimating biomass; c.f. Wiedenmann and Jensen, 2018) many of the data-limited options were improvements over the OTC. The average catch approaches that used the recent average catch (3–5 years) were improvements over OTC 60–74% of the time. The index-based approaches also used the average catch in the last 3–5 years, and as a results were also an improvement over the OTC (66–73% of the time). All but one of the catch curve approaches (BK_CC5) were an improvement over the OTC more often than not, while the production model approaches were more frequently farther above the OFL than the OTC (Fig. 7).

The magnitude of the improvement (or worsening) of the data-limited target catch, on average, compared to the OTC is shown in Fig. 8 for a subset of methods. The data-limited methods were often closer to the OFL than the OTC when the OTC was well above the OFL. For the average catch and index-based methods, the largest improvements occurred for the most conservative options, but with the tradeoff of producing target catches well below the OFL when the OTC was at or below the OFL (Fig. 8A, B). The three catch curve methods shown (BK_CC1, M_CC and YPR_CC) produced catch targets that were much closer to the OFL when the OTCs were more than twice the OFL (Fig. 8C). The production models tended to produce target catches above the OFL, although interestingly the data-limited version SPMSY was generally more conservative than the Schaefer surplus production model that was fit to survey data (Fig. 8D).

4. Discussion

We evaluated the ability of several data-limited methods to set target catches close to the OFL for data-rich stocks in the Northeast U.S. Most options we explored were very sensitive to the level of historical exploitation, producing target catches above the OFL for stocks that had a history of overfishing, or target catches below the OFL for stocks with a history of light exploitation. The more conservative options reduced the magnitude of overfishing relative to the historical level for overexploited stocks, but at the cost of being too conservative for lightly exploited stocks. Catch curve methods were the only approaches we explored that were insensitive to the level of historical exploitation, and were largely effective at setting target catches close to the OFL for overexploited stocks.

Given our findings, which approaches are suitable or unsuitable to use when a data-poor /-moderate method is needed? The approaches we tested had different data requirements, from truly data-poor methods that required only a catch time series (the average catch methods), to more data-moderate approaches that required an index of abundance or catch-at-age data. Most stocks in our analysis experienced intense exploitation for at least part of their history, so approaches that used the average or median catch over the entire time period often resulted in very high target catches relative to the OFL. DCAC aims to adjust the average catch by an assumed depletion level, and we assumed relatively large levels of depletion over the catch time period across all stocks and all years (60% and 80%). For stocks that experienced light historical exploitation it is therefore not surprising that our application of DCAC was too conservative. However, for overexploited stocks, even the larger depletion assumption was insufficient in our analysis. Our data-rich application of DCAC performed similarly to our application using static levels of depletion, suggesting that this result is not due to the assumptions we used in the method. MacCall (2009) notes that DCAC estimates a catch that would be sustainable, on

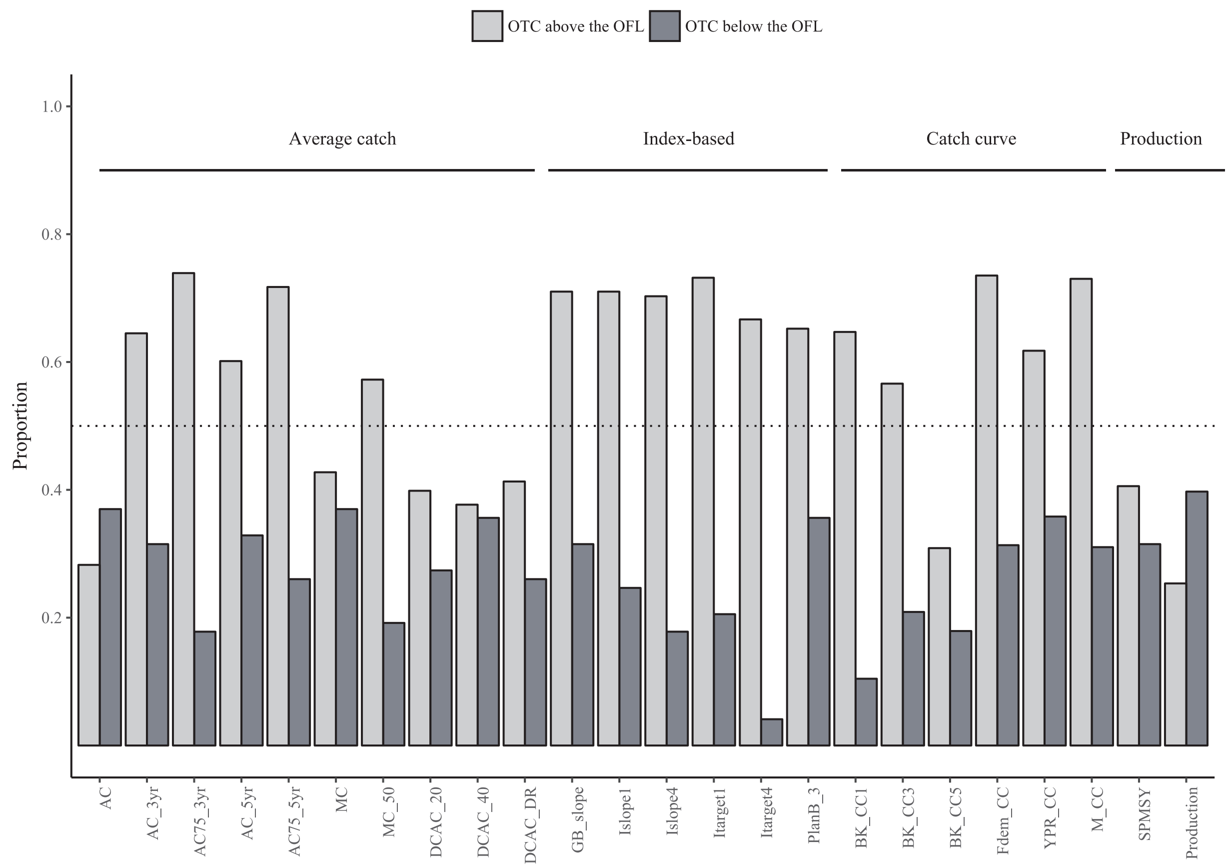

Fig. 7. Proportion of times (across years and stocks) when the target catch from the data-limited control rule was closer to the OFL than the original target catch (OTC) that was set for management, whether or not the OTC was above or below the OFL. The horizontal line at 0.5 separates when the method was more or less likely to be closer to the OFL than the original OTC.

average, over the period of available catch data, and cautions that the particular yield may no longer be sustainable for severely depleted stocks. Therefore, MacCall (2009) recommends against using DCAC for stocks undergoing rebuilding. Simulation studies have shown that DCAC tends to perform well when stocks are close to B_{MSY}, but that unsustainable catches can result when $B < < B_{MSY}$ (Wiedenmann et al., 2013; Carruthers et al., 2014). Our results are in agreement with these simulation studies, and support MacCall's caveat against using DCAC for stocks likely to be overfished, or at least for the need of an additional correction factor. Rago (2017) explored DCAC as a fallback for Atlantic halibut (*Hippoglossus*) in the Northeast U.S., a stock believed to be heavily overfished, and further adjusted the DCAC-estimated catch by multiplying by an assumed B/B_{MSY}, although DCAC was ultimately not recommended for management. Further exploration of the impacts of such adjustments is warranted to better understand the utility of DCAC for heavily depleted stocks. We note, however, that our results may be sensitive to the time periods of catch data input into DCAC, as they may not be representative of the "windfall" catch period used in the derivation of the method (MacCall, 2009). However, including catches from earlier time periods would have resulted in higher target catches for many stocks using DCAC (using the same assumed depletion levels) due to the very high catches from foreign fleets prior in earlier years (Sosebee et al., 2006).

The index-based approaches were sensitive to the intensity of recent exploitation, but all of the approaches would have resulted in comparable or more conservative target catches relative to recent levels (slopes < 1 in Fig. 5). Thus, the index-based methods would not have been worse than what was already occurring for a stock, and the more conservative options we explored would have reduced the magnitude of overfishing that was occurring in such cases. For example, both Islope4 and Itarget1 produced target catches for stocks close to the OFL when stocks had experience recent harvest rates between 1.5 to 2.5 times F_{MSY}, but these options were overly conservative when stocks were fully or under-exploited. The PlanB_3 approach was the least conservative index-based method we explored for stocks experiencing recent overfishing. This approach is currently used to set target catches for GB cod following problems with the age-based assessment (NEFSC, 2015a, 2017), and our findings suggest that perhaps a more conservative option may be better suited for this stock given that it is still believed to be overfished, although whether or not overfishing is occurring is unknown. Care is needed when selecting which index-based approach to use, with careful weighing of the evidence indicating whether or not overfishing is likely to be occurring, although determining recent exploitation rates may be incredibly difficult for a data-limited stock. Recent exploitation rates from other assessed species, either in the region or within the same fishery if possible, may be used as a proxy for the focal stock, as Free et al. (2017) showed that the best predictor of relative population size was the status of other stocks in the same fishery. A caveat to index-based approaches is that they do not aim to achieve MSY in the long run for a stock. For example, the more conservative options may allow for rebuilding of an overfished stock, but their long-term application would likely result in a considerable amount of forgone yield (Carruthers et al., 2015). Alternatively, the less conservative index-based options could preserve the status quo harvest rates, keeping the population relatively stable for an overfished population, but at a level below where maximum production occurs, resulting in a loss of long-term yield in such cases of "sustainable overfishing" (Hilborn et al., 2015).

We found that catch curve methods were very effective overall, producing target catches close to the OFL, on average, independent of

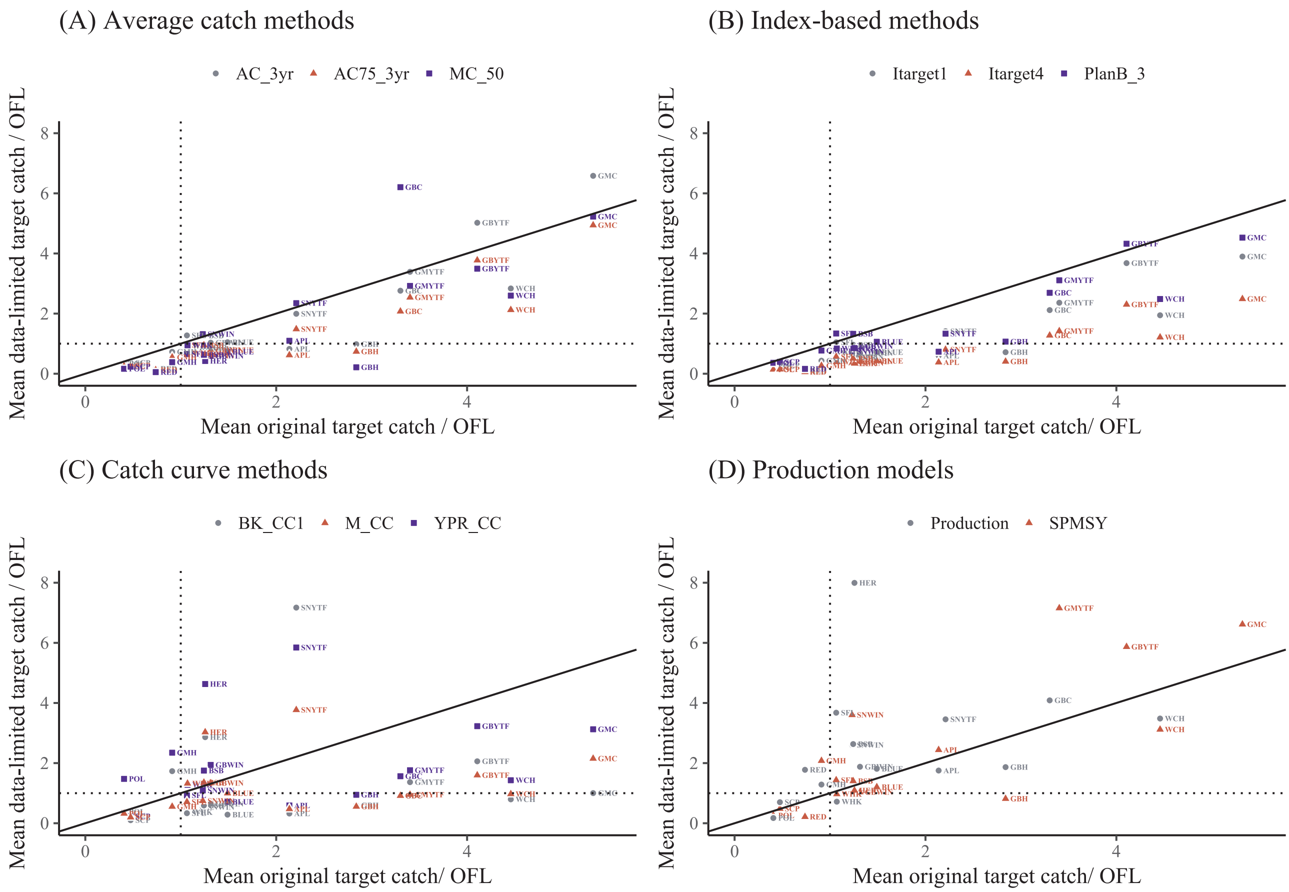

Fig. 8. Ratio of the mean original target catch (OTC) to the OFL and the median data-limited estimated catch to OFL for a subset of methods in each category. The mean values for each stock are calculated across all years where target catches are available for each stock (2000–2012 for Mid-Atlantic stocks, and 2004–2012 for New England stocks). The solid black line represents the 1:1 line, while the dashed horizontal and vertical lines indicate when the target catch and TAC are above or below the OFL, respectively. Limits of the y-axis are the same for each plot for ease of comparison, but some points are not shown in D as a result.

the exploitation history for a stock. While catch-at-age data may not be available in many data-limited cases, when it is, our findings support the use of catch curve methods (which are currently used for several species in Southeast Australia; Wayte, 2009). In particular, the M_CC method performed very well across stocks, and by simply assuming $F_{MSY} = M$ (or potentially lower values based on Zhou et al., 2012), this method avoids requiring many of the inputs used to estimate F_{MSY} in the other approaches (Tables 2 and 3). In some cases, however, catch curve methods also produced very large target catches, so our findings are not a blanket endorsement for these methods. The poor performance of catch curve methods in some instances does not rule out their use, however, as there are commonalities in the reasons for the high target catches in most cases. Large catches resulted when the catch curve greatly underestimated the total mortality for the stock, which tended to occur for stocks 1) with pulsed recruitment events, and 2) with a limited number of age classes with which to estimate Z. Expanding the catch-at-age matrix to include more ages, if possible, could address 2). For 1), we found that using a modest minimum F threshold in the catch curve estimation greatly improved the performance of the catch curve methods for many stocks. Another possible solution to 1) is to omit the large age class from the estimation of Z in a given year, or to estimate Z by following cohorts through the catch across multiple years. Further exploration into alternative ways to apply catch curve methods is warranted given our findings.

Interestingly, simulation studies of catch curve methods using the MSE portion of DLMTool have generally found them to perform poorly, resulting in a high risk of overfishing and low long-term yield (Miller, 2016; Sagarese et al., In press), and as a result they were not explored in greater detail in these studies. It is possible that the behavior that we observed, where these methods occasionally produced very large target catches (> 5 x OFL) using the default minimum F (0.005yr^{-1}) may be behind the overall poor average performance in the simulation studies. Infrequent, anomalously high catch levels applied over a multiple years in a simulation would result in frequent overfishing and cause the population to crash, resulting in low long-term yield (metrics often used to determine suitability of the methods). For Atlantic mackerel, Wiedenmann (2015) explored the MSE portion of DLMTool and similarly found poor performance of the catch curve methods, although the MSE was not used as a justification to include or exclude methods in the target catch determination, and the catch curve methods were explored in further detail. Target catches from the catch curve methods for mackerel were often conservative compared to the other methods explored. An age-based assessment for mackerel recently passed review (NEFSC, 2018), and estimated the OFL in 2017 to be 22,000 mt, compared to the catch curve-estimated catches between 13,000–26,000 mt (Wiedenmann, 2015), indicating that the catch curve methods were relatively close to the OFL. Thus, we recommend that catch curve methods are explored as an option when catch-at-age data are available, but to proceed with caution when very low estimates of Z result, or when an anomalously large target catch is produced.

Approaches that used a production model in the control rule (SPMSY, and our fit of the Schaefer model to the available survey indices) were also sensitive to the exploitation history, producing higher target catches (relative to the OFL) for more depleted stocks. This result is likely due to the "one way trip" declines for many stocks (Fig. 3) that do not provide sufficient information about the strength of densitydependence. The lack of recovery despite low catches for some stocks also suggests a change in stock productivity, violating the underlying assumptions of the production model, potentially resulting in inflated estimates of the OFL.

In reviewing the recent management performance for New England groundfish, Rothschild et al. (2014) noted the poor performance of the projection estimates relative to the updated age-based assessment estimates, and suggested surplus production models may be an alternative to age-based assessments for groundfish. We fit the Schaefer surplus production model to the available spring and fall indices and catch data, and compared estimates to the results from age-based assessments. It is interesting that SPMSY, which was not fit to index data, was generally more conservative than the Schaefer production model, although both production models in our analysis tended to produce higher estimates of total biomass and the OFL compared to the age-based models. This result is in agreement with other explorations of surplus production model applications to New England groundfish (Rothschild, 2013; Deroba et al., 2015), but does not resolve the question of which modeling approach is more accurate. The underlying population dynamics in production and age-structured models are abstractions of the natural world, and the ability of each model to accurately estimate total biomass and reference points will depend on the relative information in aggregate indices and in age structured data, and also on which, and to what extent model assumptions are violated. Here we used estimates from the most recent age-based assessments as our measure of the underlying population dynamics, as these estimates represent the current best available science for each stock. If production models were to become the standard assessment method, then our estimates of the OFL would be revised upward for many stocks, changing our interpretation of the ability of many of these data-limited methods to estimate the OFL.

An interesting finding of our work is that many of the data-limited approaches produced target catches that were improvements (i.e., closer to the OFL) over the OTCs from projections based on age-based assessments, particularly when the OTC was higher than the OFL. Wiedenmann and Jensen (2018) found that for New England groundfish (all NEFMC stocks listed in Table 1 except Atlantic herring), the target catches set were aimed at achieving harvest rates generally at or below F_{MSY}, but overly optimistic projections, primarily from overestimated terminal abundance in earlier assessments, resulted in the OTC being well above the OFL for many stocks (Brooks and Legault, 2016; Wiedenmann and Jensen, 2018). Across groundfish stocks, actual catches were 29% below the OTC, on average, yet the achieved F was 151% above the original target F(see Fig. 1 and Table 3 in Wiedenmann and Jensen, 2018). Many of the approaches we evaluated here use recent catches (not the target), such that using the average catch over the last 3 or 5 years was an improvement over the OTC, but more substantial improvements occurred for some of the catch curve methods and the more conservative index-based approaches. Geromont and Butterworth (2015) explored what they called empirical approaches (analogous to the Islope1 and Itarget1 methods) for four stocks (including two stocks used here) and found that the catches were generally comparable and less variable than those from the more complex age-based assessments. They did not argue for the abandonment of age-based assessments, but rather that simple, empirical methods could be used in the interim between assessments, freeing up resources by allowing for a greater interval between age-based assessments (5–10 years). Our findings support their recommendation, and having a longer interval between assessments could allow for more resources devoted to addressing many of the uncertainties in the assessments for these stocks.

An important caveat to our approach is that the target catch from each method is not removed from the population over time. In a MSE simulation model, the catch estimated each year from a data-limited method is removed from the population, such that there is feedback between unsustainable options that would drive the population to low levels, and vice-versa. Large changes in population status would likely be reflected in the survey index, catch-at-age data, and other metrics that inform the methods. Those methods that are updated with new information might therefore correct themselves in the long run in response to large changes in the population that occurred earlier in the time period. While MSEs are an indispensible tool for evaluating benefits and tradeoffs among management alternatives (Punt et al., 2016; Punt, 2017), retrospective evaluations like we performed here are a useful compliment to MSEs to identify effective management strategies. Many of our findings about average catch and index-based approaches are consistent with previous MSE work (Wiedenmann et al., 2013; Carruthers et al., 2014, 2015), but our findings on catch curve methods suggest better performance than in some recent MSE analyses using DLMtool (Miller, 2016; Sagarese et al., In press). Thus, both MSE and retrospective approaches may provide useful insights into performance of data-limited methods, and both approaches should be used to test new methods, or existing methods on stocks or fisheries that have not been explored.

Acknowledgements

The authors would like to thank Holly Kindsvater, Kiva Oken, and Abigail Golden for comments on an earlier version of this manuscript. We would also like to thank André Punt and an anonymous reviewer for reviewing the manuscript. This works was supported in part by funds under a NJ Sea Grant Omnibus Grant to JW and OPJ. Thus, this publication is the result of research sponsored by the New Jersey Sea Grant Consortium (NJSGC) with funds from the National Oceanic and Atmospheric Administration (NOAA) Office of Sea Grant, U.S. Department of Commerce, under NOAA grant numbers NA14OAR4170085 and the NJSGC. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the NJSGC or the U.S. Department of Commerce. This work was also supported by funds from the New England Fishery Management Council, although the statements, findings, and conclusions do not necessarily reflect the Council's views. NJSG-18-938.

References

Arnold, L.M., Heppel, S.S., 2015. Testing the robustness of data-poor assessment methods to uncertainty in catch and biology: a retrospective approach. ICES J. Mar. Sci. 72 (1), 243–250. https://doi.org/10.1093/icesjms/fsu077.

ASMFC, 2015a. . Black Drum Stock Assessment for Peer Review. Atlantic States Marine Fisheries Commission, Stock Assessment Report. 361p. Available from. . http://www.asmfc.org/fisheries-science/stock-assessments.

ASMFC, 2015b. Tautog Benchmark Stock Assessment and Peer Review Reports. Atlantic States Marine Fisheries Commission, Stock Assessment Report. 283p. Available from. . http://www.asmfc.org/fisheries-science/stock-assessments.

Beddington, J.R., Kirkwood, G.P., 2005. The estimation of potential yield and stock status using life history parameters. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 163–170.

Berkson, J., Thorson, J.T., 2014. The determination of data-poor catch limits in the United States: is there a better way? ICES J. Mar. Sci. 72 (1), 237–242. https://doi.org/10.1093/icesjms/fsu08.

Berkson, J., Barbieri, L., Cadrin, S., Cass-Calay, S.L., Crone, P., Dorn, M., Friess, C., Kobayashi, D., Miller, T.J., Patrick, W.S., Pautzke, S., Ralston, S., Trianni, M., 2011. Calculating Acceptable Biological Catch for Stocks That Have Reliable Catch Data Only (Only Reliable Catch Stocks – ORCS). NOAA Technical Memorandum NMFS-SEFSC-616. 56 P. Available:. https://repository.library.noaa.gov/view/noaa/4004.

Brooks, E.N., 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. https://doi.org/10.1139/cjfas-2015-0163.

Carruthers, T.R., Hordyk, A., 2017. DLMtool: Data-Limited Methods Toolkit. http://cran.r-project.org/web/packages/DLMtool/index.html.

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., Cope, J., 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. 153, 48–68.

Carruthers, T., Kell, L., Butterworth, D., Maunder, M., Geromont, H., Walters, C., McAllister, M., Hillary, R., Levontin, P., Kitakado, T., Davies, C., 2015. Performance review of simple management procedures. ICES J. Mar. Sci. 73 (2), 464–482. https://doi.org/10.1093/icesjms/fsv212.

Cope, J.M., Thorson, J.T., Wetzel, C.R., DeVore, J., 2015. Evaluating a prior on relative stock status using simplified age-structured models. Fish. Res. 171, 101–109.

Deroba, J., 2015. Atlantic Herring Operational Assessment Report 2015. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-16; 30 p. Available at:. . http://www.nefsc.noaa.gov/publications/.
Deroba, J., Shepherd, G., Gregoire, F., Nieland, J., Rago, P., 2010. Stock Assessment of Atlantic Mackerel in the Northwest Atlantic for 2010. Transboundary Resources Assessment Committee, Reference Document 2010/01. 59 pp.. .
Deroba, J.J., Butterworth, D.S., Methot Jr., R.D., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y.-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.-J.F., 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72 (1), 19–30. https://doi.org/10.1093/icesjms/fst237.
Dick, E.J., MacCall, A.D., 2011. Depletion-Based Stock Reduction Analysis: a catch- based method for determining sustainable yields for data-poor fish stocks. Fish. Res. 110, 331–341.
Fox Jr., W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99 (1), 80–88. https://doi.org/10.1577/1548-8659(1970)99<80:AESMFO>2.0.CO;2.
Free, C.M., Jensen, O.P., Wiedenmann, J., Deroba, J.J., 2017. The refined ORCS approach: a catch-based method for estimating stock status and catch limits for data-poor fish stocks. Fish. Res. 193, 60–70.
Geromont, H.F., Butterworth, D.S., 2014. Generic management procedure for data-poor fisheries: forecasting with few data. ICES J. Mar. Sci. 72 (1), 251–261. https://doi.org/10.1093/icesjms/fst232.
Geromont, H.F., Butterworth, D.S., 2015. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. ICES J. Mar. Sci. 72, 262–274. https://doi.org/10.1093/icesjms/fsu017.
Hilborn, R., Fulton, E.A., Green, B.S., Haartman, K., Tracey, S.R., Watson, R.A., 2015. When is a fishery sustainable? Can. J. Fish. Aquat. Sci. 72, 1433–1441. https://doi.org/10.1139/cjfas-2015-0062. dx.
Kokkalis, A., Eikeset, A.M., Thygesen, U.H., Steingrund, P., Andersen, K.H., 2017. Estimating uncertainty of data limited stock assessments. ICES J. Mar. Sci. 74 (1), 69–77. https://doi.org/10.1093/icesjms/fsw145.
Legault, C.M., Alade, L., Gross, W.E., Stone, H.H., 2013. Stock Assessment of Georges Bank Yellowtail Flounder for 2013. TRAC Ref. Doc. 2013/01; 132 p. Available from. http://www.nefsc.noaa.gov/saw/trac/.
Legault, C.M., Alade, L., Gross, W.E., Stone, H.H., 2014. Stock Assessment of Georges Bank Yellowtail Flounder for 2014. TRAC Ref. Doc. 2014/01. 214 p. Available from. http://www.nefsc.noaa.gov/saw/trac/.
MacCall, A.D., 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES J. Mar. Sci. 66, 2267–2271.
Martell, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience. Fish Fish. 14 (4), 504–513. https://doi.org/10.1111/j.1467-2979.2012.00485.x.
McAllister, M.K., Pikitch, E.K., Babcock, E.A., 2001. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. Can. J. Fish. Aquat. Sci. 58, 1871–1890.
McNamee, J., Fay, G., Cadrin, S., 2015. Data Limited Techniques for Tier 4 Stocks: An alternative approach to setting harvest control rules using closed loop simulations for management strategy evaluation. Final Report to the Mid Atlantic Fishery Management Council. Available: https://static1.squarespace.com/static/511cdc7fe4b00307a2628 ac6/t/55a661a5e4b060ebc9d03cf0/1436967333432/DLanalysis_bsb_final.pdf. .
Miller, T., 2016. Blueline Tilefish Working Group Report. Report to the Mid-Atlantic Fishery Management Council. Available: http://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/56e046a37c65e4cd0ba8f8ce/1457538725993/BLT+Subcommittee+Report+20160309.pdf. .
Myers, R.A., Bowen, K.G., Barrowman, N.J., 1999. Maximum reproductive rates of fish at low population sizes. Can. J. Fish. Aquat. Sci. 56, 2404–2419.
Northeast Fisheries Science Center (NEFSC), 2012. 53rd Northeast Regional Stock Assessment Workshop (53rd SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-05. 559 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. . http://www.nefsc.noaa.gov/nefsc/publications/.
Northeast Fisheries Science Center (NEFSC), 2013. 55th Northeast Regional Stock Assessment Workshop (55th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 13-11. 845 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. . http://www.nefsc.noaa.gov/nefsc/publications/.
Northeast Fisheries Science Center (NEFSC), 2015a. Stock Assessment Update of 20 Northeast Groundfish Stocks Through 2014. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-XXXX. ; 238 p. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. http://www.nefsc.noaa.gov/nefsc/publications/.
Northeast Fisheries Science Center. (NEFSC), 2015b. 60th Northeast Regional Stock Assessment Workshop (60th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-08. 870 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. . http://www.nefsc.noaa.gov/publications/.
Northeast Fisheries Science Center (NEFSC), 2018. 64th Northeast Regional Stock Assessment Workshop (64th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 18-03. 27 p. Available:. . http://www.nefsc.noaa.gov/publications/.
Newman, D., Berkson, J., Suatoni, L., 2015. Current methods for setting catch limits for data-limited fish stocks in the United States. Fish. Res. 164, 86–93.
Northeast Fisheries Science Center, 2017. Operational Assessment of 19 Northeast Groundfish Stocks, Updated through 2016. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 17-17. 259 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. http://www.nefsc.noaa.gov/publications/.
Pella, J.J., Tomlinson, P.K., 1969. A generalized stock productionmodel. Int. Am. Trop. Tuna Comm. Bull. 13, 416–497.
Pacific Fishery Management Council (PFMC), 2016. Status of the Pacific Coast Groundfish Fishery. 309 pp. Available:. Stock assessment and fishery evaluation. http://www.pcouncil.org/wp-content/uploads/2017/02/SAFE_Dec2016_02_28_2017.pdf.
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. https://doi.org/10.1093/icesjms/fsv193.
Punt, A.E., Szuwalski, C., 2012. How well can F_{MSY} and B_{MSY} be estimated using empirical methods of surplus production? Fish. Res. 134–136, 113–124.
Punt, A.E., Butterworth, D., de Moor, C., De Oliveira, J., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish. 17, 303–334.
R Core Team, 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rago, P., 2017. Halibut Assessment Report for 2017 to the New England Fishery Management Council. Available:. . http://s3.amazonaws.com/nefmc.org/Halibut-Assessment-Report-draft-12-01-2017.pdf.
Rothschild, B.J., Jiao, Y., 2013. Comparison between maximum sustained yield proxies and maximum sustained yield. Open Fish Sci. J. 6, 1–9.
Rothschild, B.J., Keiley, E.F., Jiao, Y., 2014. Failure to eliminate overfishing and eliminate optimal yield in the New England groundfish fishery. ICES J. Mar. Sci. 71 (2), 226–233. https://doi.org/10.1093/icesjms/fst118.
Sagarese, S.R., Harford, W.J., Walter, J.F., Bryan, M.D., Isely, J.J., Smith, M.W., Goethel, D.R., Rios, A.B., Cass-Calay, S.L., Porch, C.E., Carruthers, T.R., Cummings N.J., In press. Lessons learned from data-limited evaluations of data-rich reef fish species in the Gulf of Mexico: Implications for providing fisheries management advice for data-poor stocks. *Can. J. Fish. Aquat. Sci.*
Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. I-ATCC 1, 27–56.
Sosebee, K., Traver, M., Mayo, R., 2006. Aggregate Resource and Landings Trends. Available from. www.nefsc.noaa.gov/sos/agtt/.
Terceiro, M., 2016. Stock Assessment Update of Summer Flounder for 2016. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-13. 18 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. http://www.nefsc.noaa.gov/publications/.
Thorson, J.T., Cope, J.M., 2015. Catch curve stock-reduction analysis: an alternative solution to the catch equations. Fish. Res. 171, 33–41. https://doi.org/10.1016/j.fishres.2014.03.024.
Thorson, J.T., Cope, J.M., Branch, T.A., Jensen, O.P., 2012. Spawning biomass reference points for exploited marine fished, incorporating taxonomic and body size information. Can. J. Fish. Aquat. Sci. 69, 1556–1568.
Wayte, S.E. (Ed.), 2009. Evaluation of New Harvest Strategies for SESSF Species. CSIRO Marine and Atmospheric Research, Hobart and Australian Fisheries Management Authority, Canberra 137 p.. http://www.afma.gov.au/wp-content/uploads/2010/06/HSE-AFMA-Report-June2009.pdf.
Wiedenmann, J., 2015. Application of data-poor harvest control rules to Atlantic mackerel. Final Report to the Mid-Atlantic Fishery Management Council.
Wiedenmann, J., Jensen, O.P., 2018. Uncertainty in stock assessment estimates for New England groundfish and its impact on achieving target harvest rates. Can. J. Fish. Aquat. Sci. 75 (3), 342–356. https://doi.org/10.1139/cjfas-2016-0484.
Wiedenmann, J., Wilberg, M., Miller, T., 2013. Evaluation of harvest control rules for data poor fisheries. N. Am. J. Fish. Manage. 33, 845–860.
Wigley, S.E., McBride, H.M., McHugh, N.J., 2003. Length-Weight Relationships for 74 Fish Species Collected During NEFSC Research Vessel Bottom Trawl Surveys, 1992-9. NOAA Tech Memo NMFS NE 171. 26 p.. .
Zhou, S., Yin, S., Thorson, J.T., Smith, A.D.M., Fuller, M., 2012. Linking fishing mortality reference points to life history traits: and empirical study. Can. J. Fish. Aquat. Sci. 69, 1292–1301.