Fisheries and Oceans Canada / Pêches et Océans Canada

NOAA FISHERIES
NATIONAL MARINE FISHERIES SERVICE

CERT | ***TRAC***

Comité d'évaluation des ressources transfrontalières

Transboundary Resources Assessment Committee

Document de travail 2014/14

Working Paper 2014/14

Ne pas citer sans
autorisation des auteurs

Not to be cited without
permission of the authors

MINIMUM BOUNDS ON GEORGES BANK YELLOWTAIL FLOUNDER SPAWNING STOCK BIOMASS WITH A METAANALYSIS OF CATCHABILITY ACROSS NORTHEAST STOCK ASSESSMENTS

David Richardson[1], Rich Bell[1], John Manderson[2] and Jon Hare[1]

[1]Northeast Fisheries Science Center
Oceanography Branch
Narragansett Lab
Narragansett, RI 02882

[1]Northeast Fisheries Science Center
Behavioral Ecology Branch
Sandy Hook Lab
Highlands, New Jersey 07732

Ce document est disponible sur l'Internet à :

This document is available on the Internet at :

http://www.mar.dfo-mpo.gc.ca/science/TRAC/trac.html

ABSTRACT

The objective of this working paper is to evaluate the minimum spawning stock biomass of Georges Bank yellowtail flounder based on an analysis of different factors contributing the catchability of the NEFSC trawl survey. This catchability analysis is then applied to a suite of other stocks that are assessed using NEFSC trawl survey data. The catchability of Georges Bank yellowtail flounder was found to be highest during the night using the Bigelow survey gear. When the survey indices were standardized to nighttime Bigelow tows, the estimate of 2010-2012 spawning stock biomass ranged from 11,000-23,000 mt using wing swept areas and 4,200-9,000 mt using door swept areas as the effective trawl area. Estimates of the nighttime catchability of the Bigelow net for 7 other flatfish stocks from the NEFSC assessments ranged from 0.6-1.8 in the fall and 0.3-1.7 in the spring (a q of 1.0 corresponds to wing swept area and 2.55 to door swept area). The Georges Bank yellowtail stock exhibited a strong increase in implied catchability (survey SSB/assessment SSB) as was highlighted in the 2013 assessment. A similar pattern of increasing implied catchability in the mid-1990s was also evident in the other flatfish stocks on average, though the magnitude of change was much lower.

INTRODUCTION

For fisheries research surveys, the catchability coefficient (q) defines the relationship between a survey index and population size. Typically, catchability is estimated within a stock assessment model that incorporates survey estimates of population trends, catch data, and biological parameters. A recent large scale study, with real and simulated data, revealed that population trends are generally consistently estimated regardless of the stock assessment approach (Deroba et al. 2014). However, when the underlying data or biological processes did not conform to the standard assumptions of the model, the estimates of the scale of the population differed notably among different types of assessment models. The challenges in scaling single-species models transfer over to many multispecies models and analyses, as estimates of survey catchabilities from the stock assessments are often used scale the input time series in multispecies analyses (Brodziak et al. 2004). One potential way to improve and validate the scaling of both single-species and multi-species population models is through empirical studies of the catchability of survey gear.

Accurate calculations of the catchability coefficient of large-scale surveys, using field based studies, can be difficult due to the need for independent and unbiased information on the true density of animals within the study area. For some high-value species with limited mobility, such as sea scallops, camera based approaches have been used to accurately estimate survey catchability (Northeast Fisheries Science Center 2010). However, for mobile fish species the methods to develop accurate estimates of catchability are less well developed. An alternative approach is to establish maximum bound on the catchability of a survey using studies of the relative catchability of the reference survey gear to an optimized survey gear. These maximum bounds can be used to scale minimum biomass estimates. These minimum biomass estimates in turn can serve as a check on an assessment output, or in cases of species with limited catch histories, can serve as a guide to establishing conservative catch levels.

The process of establishing maximum bounds on catchability provides an important avenue for the knowledge of fishermen to be incorporated into the science that affects the stock assessment process. The fishing industry has invested substantial time and resources into optimizing fishing gear to target specific populations and into understanding the distribution of species. This focus on gear optimization stands in sharp contrast to the focus on standardization that is essential for developing a useful monitoring survey. How the catch rates of a standardized gear differ from those of an optimized gear can inform the maximum bounds on the catchability of the survey gear. Furthermore, practical considerations restrict monitoring surveys to fixed geographical bounds that define a survey area, even if portions of a population occur outside these boundaries. Fishing areas are generally bound by a different set of constraints, providing potentially useful information on whether a stock is fully available to a survey, and if not whether additional sampling would be appropriate.

Another important question in the stock assessment process is whether the assumption of constant survey catchability has been met for each survey used in the model. Changes in catchability can be driven by a suite of different processes, from technical factors influencing how the gear is towed, to biological factors that affect the availability of a species to the survey. Changes in catchability are usually most evident when two survey time series for the same stock exhibit different long-term trends, indicating at least one is not an accurate index of population trends. Population model results can provide another potential indicator of changing survey catchabilities. However, in many cases population model results may not provide enough information to distinguish changes in catchability from changes in biological processes, inaccurate catch data, or other factors. A much more difficult task than identifying when catchability changes may have occurred is identifying the underlying causes. Different approaches have been used to evaluate the causes of changes in the catchability of individual species (e.g. Manderson in prep). However as most trawl surveys are multispecies in nature, a change in catchability of one species will likely affect other species that share either similar morphologies, behavior, or habitats. Multispecies exploratory analyses can thus be an important first step in understanding whether survey gear catchability has changed, and if it has why.

Here we provide an analysis of the maximum bounds on catchability and the minimum bounds on biomass of a suite of stocks that are assessed by the Northeast Fisheries Science Center. These minimum biomass estimates are then compared to the biomass estimates that are derived from the established stock assessment process.   The intent of this work is to provide information that can guide catchability studies by the Northeast Cooperative Research Program, both in terms of the design of these studies and which species to target. We then evaluate whether the stock assessments are suggesting trends in catchability through time that are consistent among species with similar habitats or morphologies. While the work is designed to provide comparisons across the suite of assessed species in the region, we provide special focus on the Georges Bank yellowtail flounder.

SURVEY DESCRIPTIONS

NEFSC Bottom Trawl survey

The (NMFS) Northeast Fisheries Science Center (NEFSC) annual fall (September-November) and spring (March-May) bottom trawl surveys have occurred since 1963 and 1968 respectively. Details of the surveys and sampling gear can be found in Smith (2002) and in Table 1. Briefly, 300-400 stations are sampled in a random stratified design twice annually on the northeast U.S. continental shelf extending from Cape Hatteras, North Carolina to the Western Scotian Shelf, Canada. At each station, all fish species are identified and weighed, and all individuals are measured, with some subsampling of length measurements when species-specific catches are exceptionally large. In 2009, the vessel that historically performed these surveys, the *R/V Albatross IV*, was retired and replaced with the *R/V Henry Bigelow*. This change in survey vessel was also used as an opportunity to change the survey gear and towing protocols. Furthermore, the larger size of the *R/V Bigelow* required that certain historically surveyed inshore strata be dropped from the survey. The core area consistently sampled from 1973-2013 by the NEFSC survey encompasses 224,562 km^2. In anticipation of the vessel change, an extensive calibration experiment was performed to enable standardization of the catch rates of the two vessels and survey gear (Miller et al. 2010). The initial calibration analysis focused on aggregate calibration coefficients for weight and number per tow between the two vessels. For many commercially important species, length-specific calibration factors were subsequently calculated.

Massachusetts State Trawl Survey

The State of Massachusetts has conducted a spring (May) and fall (September) trawl survey since 1978. Details on survey gear and towing protocols are in Table 1. Survey tows are restricted to daylight hours. The survey covers an area of 6,285 square kilometers and is stratified using five geographic regions and six depth zones. Within Cape Cod Bay, this survey overlaps with some inshore strata that are also survey by the NEFSC trawl survey.

NEAMAP

The Northeast Area Monitoring and Assessment Program (NEAMAP) began in 2008 after pilot surveys in the previous years. This survey covers a 12,097 km^2 area of the inshore waters from Cape Hatteras North Carolina to Rhode Island. All tows are carried out on the *F/V Darana R.* a 90 foot commercial trawler. The survey was designed to cover areas inshore of the NEFSC *R/V Bigelow* survey including both areas formerly trawled by the *R/V Albatross* and areas inshore of these strata. Except for in the deeper sounds (i.e. Block Island Sound and Rhode Island Sound) the maximum depth of the NEAMAP strata is ≈20 m. The gear and towing protocol were designed to match the *R/V Bigelow* protocol, with the exception that roller gear is used instead of rockhopper gear (Table 1). Additionally all tows are performed during daylight hours.

Long Island Sound Trawl Survey

The Long Island Sound trawl survey run, by the State of Connecticut Department of Energy and Environment Protection, has been ongoing since 1984. This survey covers a 3,400 km^2 area using a stratified random sampling design and only daytime tows. Sampling occurs during both the spring (April-June) and fall (Sept-Oct.) aboard the 50 ft R/V John Dempsey (Table 1). Waters sampled by the survey range from 5-46 m in depth. There is no overlap of this survey with any other survey

ANALYSIS OF CATCHABILITY

The catchability equation

The relationship between a trawl survey index, catchability and population biomass is generally defined using the following equation (Northeast Fisheries Science Center 2009):

$$B_t = \frac{I_t A}{q * a} \qquad \text{[eq. 1]}$$

Where:

***I_t*:** Index value at year t (kg tow^{-1})

***B_t*:** Biomass of the population at year t (kg)

q: catchability

a: area covered by a single trawl (km^2 tow^{-1})

A: area covered by the survey (km^2)

Within this equation I_t, ***a*** and ***A*** are all values that are measured on a survey or are part of the survey design. Catchability can be further broken into two components. The first component, **availability,** is the proportion of the total population within the footprint covered by the survey. The second component, **detectability**, represents the proportion of fish within the footprint of an average individual trawl that are captured within the trawl. Fish in the water column, or that escape above, below or to the sides of a bottom trawl all contribute to detectability values that are less than 1. Catchability (q) is the product of availability and detectability.

$$q = \delta * \rho \qquad \text{[eq. 2]}$$

δ= detectability

ρ= availability

For most trawl surveys and species, catchability is unknown and difficult to quantify empirically. However, it is possible to use trawl survey data to set maximum bounds on both detectability and availability; these maximum bounds can then be used to establish minimum bounds on stock biomass.

Analysis of detectability using day-night differences in catch levels

Detectability of many fishes in a trawl net varies substantially over a day-night cycle. For example, daytime catch rates of planktivores in bottom trawl surveys is often higher due to nighttime feeding in the water column, whereas flatfish often exhibit the opposite pattern of higher nighttime catch rates in surveys (Ryer et al. 2010). This day-night behavior is relevant to broader analyses of survey catchability for two reasons. First, the NEFSC trawl survey uses 24-hour operations, whereas the NEAMAP and most state surveys that sample inshore areas only trawl during daylight hours (Table 1). Second, the relative detectability of the NEFSC survey between the day and night can be used to scale the maximum detectability of this survey. We can assume that detectability during day and night is less than 1:

$$\delta_{day} \leq 1 \text{ and } \delta_{night} \leq 1 \qquad \text{[eqs. 3]}$$

From the survey data we can calculate the day and night catch rates to obtain the ratio of daytime to nighttime detectability:

$$\frac{\delta_{day}}{\delta_{night}} = \frac{Catch_{day}}{Catch_{night}} \qquad \text{[eq. 4]}$$

By setting daytime (nighttime) detectability to its assumed maximum value (1) we can calculate a maximum value for nighttime (daytime) detectability. In turn we can calculate a maximum value for the average detectability for the 24-hour survey:

$$\delta_{max} = \delta_{day,max} * Proportion\ day\ tows + \delta_{night,max} * Proportion\ night\ tows \qquad \text{[eq. 5]}$$

For all analyses we used the solar zenith angle to define day (<90.8) and night (>90.8) (Jacobson et al. 2011).

Analysis of catchability with two simultaneous non-overlapping surveys

When two surveys of a resource are available, the catch levels on one survey can be used to inform the catchability on the other survey, assuming that two criteria are met. First, the surveys must occur at approximately the same time to minimize the extent of "double-counting" of fish moving from one survey area to another, and 2) the surveys must not overlap in space. The NEFSC fall trawl survey and the NEAMAP, Massachusetts, and LIS fall trawl surveys fulfill these two criteria at a reasonable level of approximation. That is, these surveys can be assumed to measure different components of the same population at approximately the same time. This is not the case for the NEFSC spring surveys which offset in time from the inshore surveys.

With two paired surveys the catchability equations can be rewritten as follows:

$$B_t = \frac{A_B}{a_B(\rho_B \delta_B)}\ I_{B,t} = \frac{A_N}{a_N(\rho_N \delta_N)}\ I_{N,t} \qquad \text{[eq. 6]}$$

In this example the subscript *B* refers to the NEFSC fall trawl survey on the *R/V Henry Bigelow* and the subscript *N* refers to the NEAMAP survey on the *F/V Darana R*. This equation can be rearranged to put the two components of catchability on one side of the equations and the known/measured values on the other side:

$$\frac{A_B}{A_N}\frac{a_N}{a_B}\frac{I_B}{I_N} = \frac{(\rho_B}{(\rho_N}\frac{\delta_B)}{\delta_N)} \qquad \text{[eq. 7]}$$

Since these two surveys do not overlap in space and occur at approximately the same time we can assume that the combined availability of fish to both surveys is less than or equal to 1:

$$(\rho_B + \rho_N) \leq 1 \qquad \text{[eq. 8]}$$

When setting the maximum bounds of catchability of the NEFSC survey, Equation 8 can be rewritten as:

$$\rho_N = 1 - \rho_B \qquad \text{[eq. 9]}$$

Furthermore, in setting the maximum bounds on catchability of the NEFSC survey the most conservative assumption is that NEAMAP detectability is 1. These two assumptions allow the two components of NEFSC survey catchability to be written solely as a function of known or measured values of the two surveys.

$$\frac{A_B}{A_N}\frac{a_N}{a_B}\frac{I_B}{I_N} = \frac{\rho_B * \delta_B}{(1-\rho_B)*1} \qquad \text{[eq. 10]}$$

With equation 10 we can set a maximum bound on availability (ρ) for any given value of detectability (δ). The most conservative value of catchability (q) results from using the maximum value of detectability obtained from the day-night analyses described in equation 5.

Inclusion of Long Island Sound and Massachusetts survey data

The CT DEP Long Island Sound Survey and Massachusetts state fall trawl surveys occur concurrently with the NEAMAP and the NEFSC trawl survey but do not overlap in space (portions of the Massachusetts and NEFSC survey are an exception; strata of overlap are excluded in these cases). These two state surveys utilize substantially different nets from those used by the NEFSC and NEAMAP surveys. In order to further refine the maximum bounds on the NEFSC Bigelow survey catchability, we included these surveys in the analysis. The most conservative approach to including these surveys was to assume 1) that the three inshore surveys (NEAMAP, LIS, Mass) have a detectability of 1.0 and 2) that in aggregate the inshore surveys and the Bigelow survey are sampling the entire area occupied by the population. With these assumptions it is possible to rewrite equations:

$$B_t = \frac{A_B}{a_B \rho_B \delta_B c} I_{B,t} = \left(\frac{A_N * I_{N,t}}{a_N} + \frac{A_N * I_{M,t}}{a_M} + \frac{A_N * I_{LIS,t}}{a_{LIS}}\right) * \frac{1}{\rho_{inshore}\delta_{inshore}} \qquad \text{[eq. 11]}$$

Under the most conservative assumptions $\delta_{inshore} = 1$ and $(\rho_{inshore} + \rho_B) = 1$. As with the previous analysis we can calculate a maximum Bigelow availability (ρ_B) for every assumed value of Bigelow detectability (δ_B).

Albatross detectability from calibration study

The general convention within the NEFSC is to convert all trawl catches to the equivalent Albatross units. An extensive paired trawl study in 2008 was used to determine the calibration factors for the two vessels and nets (Miller et al. 2010). This calibration study of the Albatross to Bigelow sampling can be used to inform the maximum bounds on the detectability of the Albatross sampling:

$$\delta_A = \frac{\delta_B}{Cal_{B:A}} * \frac{SA_{Big}}{SA_{Alb}}$$ [eq. 12]

δ_A=catchability at length Albatross

$\delta_{B=}$catchability at length Bigelow

$Cal_{B:A}$= Calibration factor to convert Albatross N or kg tow^{-1} to Bigelow N or kg tow^{-1}

SA_{Alb}=Swept area of an Albatross tow

SA_{Big}= Swept area of a Bigelow tow

An important point with this equation is that there are two factors that inform the detectability estimates, the calibration factors on a tow-specific basis, and the differences in swept areas between the two nets. For most species, the Bigelow caught more individuals per tow than the Albatross even though the area covered by a tow was smaller. The maximum detectability of the Albatross net is obtained by substituting the maximum detectability of the Bigelow net into this equation.

Survey based estimates of Spawning Stock Biomass

For all of the analyses in this working paper, we modified the catchability equation to focus on Spawning Stock Biomass (SSB_t) rather than total stock biomass (B_t). Three factors motivated this change. First, SSB is generally a preferred measure of the status of a stock as it corresponds to the reproductive component of a population. Second, Spawning Stock Biomass is consistently reported in stock assessments allowing for comparisons across multiple stocks; total stock biomass is rarely reported. Third, the standard survey bottom trawls tend to have low selectivity for small immature fish; these small fish are not included in most SSB estimates. The modified catchability equation (i.e. Eq 1) focused on SSB can be written as:

$$SSB_t = \sum_{L=1}^{maxL} \frac{(N_{L,t} * W_L * P_L)}{(\delta_L * \rho_L)} * \frac{A}{a}$$

[eq. 13]

SSB_t = Spawning stock biomass in year t (kg)

N_L= Stratified Mean Number tow^{-1} at Length (L) in year t

W_L= Weight at Length (kg)

P_L= Proportion mature at Length

δ_L = detectability at Length

ρ_L = availability at Length

A = Total stock area covered by survey (km^2)

a= area covered by an tow (km^2 tow^{-1})

As with all of the previous equations, the focus of this is equation is to set minimum bounds on Spawning Stock Biomass. This can be done by rewriting Equations 3-11 using length-specific abundance rather than aggregate biomass.

Further information on basic biological parameters is required to implement this equation. O'Brien et al. (1993) fit logistic equations to the proportion of individuals mature at both length and age for a suite of 19 finfish species on the northeast shelf. This document separated out maturity by sex. In all cases, we used the parameters for the female maturity rather than male maturity:

$$P_L = \frac{1}{1+e^{-(\alpha+\beta L)}} \qquad \text{[eq. 14]}$$

The data used to estimate these maturity parameters were collected from 1985-1990.

Wigley et al (2003) fit length-weight equations for 74 species of fish caught on the spring and fall trawl surveys using the following equation:

$$W_L = aL^b \qquad \text{[eq. 15]}$$

Season and sex specific parameters were calculated. For this analysis we used the combined sex parameters for each season individually.

METAANALYSIS OF ASSESSMENT CATCHABILITIES

Calculation of Implied SSB

With some exceptions, most NEFSC assessments estimate catchability internally within the assessment model. That is, NEFSC trawl survey catchability is a model output that is calculated based on the assumptions and data input into the model. Furthermore, most assessments also output estimates of selectivity at age. The product of survey catchability and selectivity at age provides a measure of catchability at age. The reporting of catchability and selectivity values for NEFSC science center surveys differs among assessments based on whether index values or swept area biomasses were used. Furthermore, the presence of selectivity at age results in multiple catchability estimates per stock, again making interstock comparisons difficult.

For these reasons we defined a measure of survey catchability termed the implied catchability of SSB:

$$q_{SSB} = \frac{SSB_{Survey}}{SSB_{Assessment}} \qquad \text{[eq. 16]}$$

This measure has a number of advantages. Most prominent is that SSB is the most consistently reported value in any stock assessment. In practical terms the measure of q_{SSB} can be viewed as the implied catchability of the Bigelow survey during either the day or night (whichever is found to have a higher detectability) also taking into account the measured biomass of spawning size fish outside the survey area. Values less than 1 are expected due to less than 100% efficiency of either the Bigelow or inshore gear, and the presence of fish in unsurveyed areas. This measure assumes that selectivity is 1 across all mature length classes of fish for the reference survey (i.e. Bigelow day or night).

Measures of swept area

Two different measures of the average swept area of an individual tow of the bottom trawl are commonly used (Fig. 1). The first measure, the **wing swept area,** is a product of the average distance between the wings of the trawl gear and the distance towed. This is the standard measure of swept area used in most assessments, as it corresponds to the area of the bottom covered by the portion of the gear capable of catching fish. The second measure, the **door swept area**, is a product of the distance between the doors of the trawl gear and the distance towed. Certain species of fish have been shown to be herded into the trawl mouth due to interactions with the doors, sand clouds, or sweeps (Somerton and Munro 2001, Somerton et al. 2007). For herding to occur, fish must swim at a speed and in a direction to avoid being overtaken by the gear while in the path of the sweeps or doors, before eventually being overtaken by the gear when in the path of the trawl mouth (Winger et al. 2004). For the purposes of this analysis, we use wing swept areas in all of the calculation. Thus, it is possible for the implied q to be greater than 1 if notable herding occurs and this herding is not offset by lower gear efficiency. The door swept area for the Bigelow is 2.55x the wing swept area (Table 1). The actual maximum catchability is thus 2.55 corresponding to the absolute herding of fish between the doors and 100% capture efficiency of the gear. That is, absent empirical information on herding efficiency for a taxa, implied q values of 1-2.55 are reasonable and reflect a decision to define swept area based on the wings as is the standard in NEFSC assessments. Catchability values of 1-2.55 would be <1 if swept area was defined based on the doors.

Analytical Details and Confidence Intervals

All analyses were done on a stock specific basis (rather than species specific) using the same strata reported in the assessments for that stock (Table 2). For many species with only a single stock on the northeast shelf we used a core strata set that has been consistently samples from 1972-2013. Strata sets for the inshore surveys were also defined. The NEAMAP and LIS surveys were not subsetted into particular strata for any stock, as all stocks either included or did not include the entire NEAMAP or LIS survey area (Table 1). However, the Massachusetts survey occurs near the stock boundaries for many species, and thus strata subsets were required for this survey (Table 1). This need to subset the Massachusetts state survey was also driven by the overlap of NEFSC and Massachusetts strata in certain areas.

Calibration factors were applied according to standard protocols for the NEFSC trawl survey. Available calibrations factors for gear, doors and vessels were applied to all of the catch data. For the transition from the Albatross to Bigelow we used length specific calibration factors for cases in which these were used in the stock assessment and were readily available. In cases in which the length-specific calibration factors were not used we instead used the season-specific calibration factors for the aggregate numbers reported in Miller et al (2010). For one species, Winter Flounder, we set a maximum calibration factor of 3.09 for large fish as the reported calibration factors reach very high levels (>10) that suggest minimal catchability of large fish by the R/V Albatross. This decision resulted in lower estimated of the implied catchability of SSB.

The basic biological parameters for weight-at-length and proportion mature at length came from Wigley et al. (2003) and O'Brien et al. (1993). Weight-at-length was reported on a species and season specific basis but not a stock-specific basis. Maturity-at-length was reported on a sex-specific basis for each stock separately for the seasonal survey that most closely corresponded to the spawning season. In all cases we used the maturity equations for female fish. For most flatfish, females mature at a longer length than males whereas for most other species the size at maturity is comparable among sexes. Use of female maturity is conservative in these cases. However for Pollock, White Hake and both stocks of Atlantic Cod, males mature at a longer length. Some stock assessments account for varying maturity-at-age in the models and most include varying weight-at-age. For this analyses, age is not a factor as all analyses are done on 1cm length bins. However, variations in weight-at-length are not accounted for. These variations likely have a limited effect on the results. However variations in maturity-at-length could produce a more notable effect and trend in catchability. We currently have not accounted for trends in maturity.

Confidence intervals on the catchability estimates were obtained using the rescaling bootstrapping technique outlined in Smith (1997). This approach maintains the random stratified sampling design of the survey in estimating confidence intervals. The coherence in adjacent length classes is also maintained in the bootstrapping (e.g. a bootstrap sample with a high abundance of 25 cm individuals will likely also have a high abundance of 26 cm individuals). For our analyses we have up to six different survey estimates of biomass that contribute to the final estimate of the maximum bounds of catchability: 1) Daytime NEFSC, 2) Nightime NEFSC, 3) NEFSC 24 hour, 4) NEAMAP, 5) Long Island Sound, and 6) Massachusetts state trawl survey. For surveys 3-6 we used the 2009-2012 data when all of the surveys were operating concurrently and the Bigelow net and vessel were in use. We used the calibrated 1975-2012 data to obtain the nighttime and daytime catch levels. We calculated a total of 1000 bootstrap samples for each survey and proceeded through the calculations for each of these runs.

To be conservative we report the lower 95% confidence interval of the survey minimum SSB for each year; these lower values are in turn used to calculate the implied catchability. The general implications of this are that a correction for day:night detectability differences was not applied if the bootstrapping indicated that the daytime versus nightime catch rates were not significantly different at the 95% level. However, for the two components of catchability we also report the median and 95% confidence intervals. In reporting the stock-specific results we converted length-specific availability, detectablity and day night ratios to aggregate values using the three weighting factors of maturity at length, weight at length and average number at length across the time series.

In all analyses we did not account for the uncertainty in the calibration factors of the Bigelow to Albatross nets. This uncertainty will affect both the trend of the time-series and the scale of the biomass levels and implied catchability prior to 2009. For example, if the estimated calibration factor is higher than the actual calibration factor the scale of the implied catchability and biomass prior to 2009 will be higher than it should be, which will in turn lead to a recent trend in biomass that is more negative than it actually is.

For both the fall and spring NEFSC surveys we calculated an annual estimate of the implied catchability across the entire assessment time series. Median values of this implied catchability across the entire time series are reported. In many cases the seasonal timing of the survey is notably different from the season of spawning. In this analysis we made no attempt to correct for this difference in timing by applying mortality and growth functions to abundance at length data.

Evaluations of trends in implied catchability across assessments

Common trends in the implied catchability across the 21 stocks over two seasons (minus Atl. Mackerel in fall and Bluefish in Spring) are currently being evaluated using Dynamic Factor Analysis (Zuur et al. 2003). The run times on this analysis are long for a matrix of this size (40 surveys x 43 years) and the results are not final.

In the interim we focused on trends in the implied catchability (survey ssb/assessment ssb) of the seven other flatfish stocks (SNE and GOM yellowtail flounder, SNE and GB winter flounder, American plaice, witch flounder, summer flounder). The implied catchabilities were first log-transformed and then the mean over the time series for each stock was subtracted from each annual value. These values were then averaged across stocks for each year. The resulting averages provide a measure of the extent to which the implied catchabilities of flatfish were above or below average over the time series. Only the 1985-2010 period, common to all the assessments, was used in this analysis.

RESULTS

Georges Bank Yellowtail

Weight-at-length and maturity-at-length are displayed in Fig 2a,b and Fig 5a,b. Yellowtail flounder reach a weight of ≈0.5 kg at 40 cm in length. The length at 50% maturity is ≈26 cm. Over the entire time series, the peak abundance of GB yellowtail flounder occurred at 33-36 cm in both the NEFSC fall survey and spring surveys (Fig. 2c, Fig 5c). The stock boundaries for Georges Bank yellowtail do not include any of the state surveys (Fig 2d-f, Fig 5d-f; note the figures are generalized to all species/stocks to allow for intercomparisons even if the availability term is not used). Overall, the strata defining the GB yellowtail flounder stock encompass 37,773 km^2 (Fig 2f, 5f).

During the fall, nighttime catch rates were higher than daytime catch rates across nearly the entire length range. These differences were significant for all but the largest and least frequently sampled, length classes. A general pattern of decreasing contrast between the nighttime and daytime catch rates was evident with increasing size (Fig 3a). The diel patterns in catch rates corresponded to a median of 0.6-0.8 in the maximum detectability of yellowtail over the 24-hour survey operations during the fall. During the spring, survey nighttime catch rates were generally higher, but the differences were only significant at 26-32 cm (Fig 6a). In the spring, the maximum detectability of the Bigelow net showed a similar increasing trend with length as the fall (Fig 6b). Importantly the upper bounds (95% CI) of detectability were 1 for all but the 26-32 cm length class; these upper bounds are used in calculating the minimum spawning stock biomass. Due to the absence of inshore areas for the Georges Bank stock of yellowtail flounder, availability was 1 across all lengths (Fig 3C and 6C).

The Bigelow to Albatross calibration on a swept area (rather than per tow) basis was lower for the smaller size range of yellowtail flounder than the larger size range (i.e. the Bigelow catches proportionately more fish at smaller sizes). The length-specific calibrations indicate a maximum detectability by the Albatross net of ≈0.3 (Fig 3d, 6d; note section on *measures of swept area* when interpreting this). Combining the day-night detectability analysis, availability, and the calibration experiment yields a catchability of the Albatross net ranging from about 0.08 for 20 cm fish to 0.22 for 40 cm fish in the fall (Fig 3F). For the spring survey the catchability diverged little from the values suggested solely by the calibration study (Fig 6F).

The time series of swept-area spawning stock biomass from the spring and fall trawl surveys are shown in Fig 4a and 7a. These measures are based on a reference catchability of 1 during the nighttime portion of the Bigelow survey. The implied catchability of the survey shows an increasing trend (Fig 4B and 7B) consistent with the relative-q at each age class calculated in the recent assessment. The implied catchability over the time series was strongly skewed with a mean value of 2.3 in the fall and 1.82 in the spring and median values of 1.07 in the fall and 1.09 in the spring. Since about 2004 the implied catchability has been above the 2.55 value which corresponds to the area swept by the doors.

Metanalyses of stocks

Catch ratios between daytime and nighttime tows differed among stocks and species (Fig. 8). In general pelagic fish (butterfish, mackerel and bluefish) had higher catch rates during the day. Atlantic herring was an exception with more complex diel catch rates that varied notably by size and season (not shown) but that were not significant for the spawning portion of the population. Gadiformes also generally had higher daytime catch rates, though there were not significant in all cases, and were reversed for Georges Bank Atlantic cod in the fall which were caught significantly more during the night. Flatfish generally exhibited the opposite patterns of higher nighttime catch rates, particularly in the fall.

A majority of the stocks (16 of 29) were entirely (>98%) available to the NEFSC survey given documented catches in the inshore surveys (Fig. 9). The remaining species had availabilities that ranged as low as 60% for Scup. Notably, for summer flounder, winter flounder and windowpane flounder which are caught in higher proportions during the night in the NEFSC survey, the assumptions of 100% daytime detectability in the inshore surveys may have biased the availability estimates low.

The implied catchability values of the Bigelow net was lowest for pelagic species such as Atlantic herring and Atlantic mackerel (Fig. 10). These values account for day:night differences in catch rates and the abundances of fish in the inshore trawl surveys. Some level of herding by the Bigelow net was suggested for most of the flatfish stocks.

Analysis of trends in implied catchability

The individual trends in implied catchability are presented in Fig. 11. Flatfish specific trends are presented in Fig.12. On average the other 7 stocks of flatfish exhibited an increase in implied catchability starting in the mid 1990s. A similar pattern was evident in Georges Bank yellowtail flounder, except the magnitude of the change was much greater.

DISCUSSION

With some notable exceptions (e.g. acoustic surveys, camera surveys of scallops), the primary goal of research surveys is to provide an index of abundance rather than an absolute estimate of abundance. The function of a stock assessment model is to integrate a diverse set of data to determine the scale of the population. However, a variety of factors can make for imprecise scaling of population levels, including low fishing mortality rates over most of the time series, highly variable natural mortality, and inaccurate catch data. In these cases, external information can serve a role in population scaling, even if only to establish minimum estimates on stock biomass.

The purpose of this analysis was two-fold. First, we sought to quantify the different components of catchability for a suite of commercial species, in order to evaluate which species further empirical studies, presumably through cooperative research, could be useful in scaling the population. Second, we sought to determine whether consistent time trends were evident in the implied catchability of the NEFSC surveys given the assessment results. Importantly, trends in catchability are not indicative of errors in the assessment, which are integrating many data sources to derive a population trend. Rather they do indicate tension among data sources or model assumptions. In multi-species and ecosystem analyses, attempts should be made to understand these single-species assessment trends in catchability prior to proceeding with the assumption of fixed survey catchability.

For this analysis, attempts were made to develop conservative estimates of the implied catchability of the assessments. Specifically, the lower 95% CI of the minimum survey spawning stock biomass time series was used in the calculations. Additionally, for some species the nighttime catch rates were higher than the daytime catch rates on the NEFSC survey. We did not assume that this day:night difference persists in the inshore surveys and thus did not correct the inshore daytime only survey catch rates. If this day-night catch rate difference does persist inshore, than the availability of the stock to the NEFSC survey would be lower than our estimate due to low detectability in the inshore surveys. Spawning stock biomass would be in turn be scaled up if detectability in the inshore surveys is lower.

Other factors retained in the analyses could result in both over estimates and underestimate of survey spawning stock biomass. Mismatches between the statistical areas used for fisheries-dependent data versus the strata used for fisheries-independent data is one example (e.g. Northeast Fisheries Science Center 2012). Another example is the mismatch between the season of the survey versus the season used to calculate spawning stock biomass. To account for this the survey biomass would have to be reduced by total mortality and increased for growth to match the reference season; these calculations were not made in this analysis. The use of a fixed maturity at length equation calculated using 1985-1990 data is another example. Many of the maturity at age equations have been updated in recent assessments and in some assessments maturity-at-age is time varying. Finally we assumed that detectability and availability at length is stable across the time series. For availability-at-length the benchmark of 2009-2012 was used in the calculations.

This analysis highlights some species that cooperative research studies of catchability could potentially provide useful information to scale spawning stock biomass estimates in the assessment. Maximum catchability estimates could be further reduced if it could be demonstrated that another gear or tow protocol yields higher catch rates than the *R/V Bigelow* on an area swept basis. If day-night differences in catch rates are already being used to refine catchability than the survey design must account for this. Using SNE windowpane flounder as an example, approximately 10 times as many windowpane flounder were caught during the night versus during the day in NEFSC tows. In theory a gear could be designed that catch 10 times as many windowpane as the NEFSC survey during the day but the same number at night. In this case the 24-hour catch rate of the comparison gear would be about 1.8 times the 24-hour catch rate of the reference NEFSC gear. This increased catch rate would not contribute to refining the maximum catchability used in this analysis; rather only differences observed during nighttime tows would be useful. For many flatfish species comparative daytime tows would represent wasted effort in a cooperative study, or at a minimum would require that the adjustment for detectability use either the day-night differences or the gear comparison study, but not both.

One clear conclusion that emerged in this analysis is the potential utility of quantifying herding in flatfish on the northeast shelf. Six of the eight flatfish stocks we evaluated had median implied catchability estimates over the past decade that exceeded 1 (corresponding to the wing swept area) but were less than 2.55 (the door swept area). The implied catchability of one additional stock, GB yellowtail flounder, also exceeded the door swept area in that time period. These results are not unique to this region. On the west coast of the United States catchability estimates from the stock assessments were 2.97 for petrale sole (*Eopsetta jordani*), 1.22 for English sole (*Parophrys vetulus*) and 1.79 for rex sole (*Glyptocephalus zachirus*) using wing swept areas; door swept areas were approximate 3 times larger on this survey (Bryan et al. 2014). Herding is a known phenomena for flatfish and many other species (Ramm and Xiao 1995, Somerton and Munro 2001, Somerton et al. 2007) but has not been quantified for the R/V Bigelow net. Additionally, many of the studies that have documented herding in flatfish have used exclusively daytime experimental tows. Reduced herding of flatfish has been suggested for nighttime tows (Ryer 2008, Ryer et al. 2010). For most flatfish we used the nighttime tows as the reference in calculating biomass. If herding actually is limited at night than the wings rather than the doors would provide a more appropriate measure of the area swept by the trawls.

This analysis indicated some consistency in trends in the implied catchability across similar species. For example both Atlantic mackerel and Atlantic herring had a mid-1980s increase in implied-catchability, as is discussed in their respective assessments. On average, flatfish stocks, other than Georges Bank yellowtail flounder, exhibited an increase in implied catchability in the mid 1990s that matched the timing of the increase in Georges Bank yellowtail implied catchability. However the magnitude of the change was lower for these other stocks. This coherence in trends implies that the mechanisms underlying this change in implied catchability, and thus the stock assessment retrospective issue, are shared across a number of flatfish stocks despite some notable differences in distribution.

References

Brodziak, J. K. T., Legault, C. M., Col, L. A., and Overholtz, W. J. 2004. Estimation of demersal and pelagic species biomasses in the northeast USA continental shelf ecosystem. ICES CM Document **CM 2004/FF:07**.

Bryan, D. R., Bosley, K. L., Hicks, A. C., Haltuch, M. A., and Wakefield, W. W. 2014. Quantitative video analysis of flatfish herding behavior and impact on effective area swept of a survey trawl. Fish. Res. **154**:120-126.

Deroba, J. J., Butterworth, D. S., Methot, 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., and Hulson, P.-J. F. 2014. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science: Journal du Conseil.

Jacobson, L. J., Seaver, A., and Tang, J. 2011. AstroCalc4R: Software to Calculate Solar Zenith Angle; Time at Sunrise, Local Noon, and Sunset; and Photosynthetically Available Radiation Based on Date, Time, and Location. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 11-14, Woods Hole, MA.

Miller, T., Das, C., Politis, P., Miller, A., Lucey, S., Legault, C., Brown, R., and Rago, P. 2010. Estimation of Albatross IV to Henry B. Bigelow Calibration Factors. Northeast Fisheries Science Center Reference Document, 10-5.

Northeast Fisheries Science Center. 2009. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03, Woods Hole, MA.

Northeast Fisheries Science Center. 2010. 50th Northeast Regional Stock Assessment Workshop (50th SAW): Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-17, Woods Hole.

Northeast Fisheries Science Center. 2012. 53rd Northeast Regional Stock Assessment Workshop (53rd SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 12-05, Woods Hole, MA.

O'Brien, L., Burnett, J., and Mayo, R. 1993. Maturation of nineteen species of finfish off the northeast coast of the United States, 1985-1990. NOAA Technical Report, NMFS 113.

Ramm, D. C. and Xiao, Y. 1995. Herding in groundfish and effective pathwidth of trawls. Fish. Res. **24**:243-259.

Ryer, C. H. 2008. A review of flatfish behavior relative to trawls. Fish. Res. **90**:138-146.

Ryer, C. H., Rose, C. S., and Iseri, P. J. 2010. Flatfish herding behavior in response to trawl sweeps: a comparison of diel responses to conventional sweeps and evelvated sweeps. Fish. Bull. (Seattle) **108**:145-154.

Smith, S. J. 1997. Boostrap confidence limits for groundfish trawl survey estimates of mean abundance. Can. J. Fish. Aquat. Sci. **54**:616-630.

Smith, T. D. 2002. The Woods Hole bottom-trawl resource survey: development of fisheries-independent multispecies monitoring. *in* 100 years of science under ICES. *Edited by*, E. D. Anderson. ICES, Helsinki. pp. 474-482.

Somerton, D. A. and Munro, P. 2001. Bridle efficiency of a survey trawl for flatfish. Fish. Bull. (Seattle) **99**:641-652.

Somerton, D. A., Munro, P. T., and Weinberg, K. L. 2007. Whole-gear efficiency of a benthic survey trawl for flatfish. Fish. Bull. (Seattle) **105**:278-291.

Wigley, S. E., McBride, H. M., and McHugh, N. J. 2003. Length-Weight Relationships for 74 fish species collected during NEFSC Research Vessel Bottom Trawl Surveys, 1992-99. NOAA Tech Memo NMFS NE 171.

Winger, P. D., Walsh, S. J., He, P., and Brown, J. A. 2004. Simulating trawl herding in flatfish: the role of fish length in behaviour and swimming characteristics. ICES Journal of Marine Science: Journal du Conseil **61**:1179-1185.

Zuur, A. F., Tuck, I. D., and Bailey, N. 2003. Dynamic factor analysis to estimate common trends in fisheries time series. Canadian Journal of Aquatic Sciences **60**:542-552.

Table 1.

	NEFSC Bigelow	**NEFSC Albatross**	**NEAMAP**	**LIS**	**MASS**
Net	400 x 12 4-Seam Trawl	Yankee 36 or 41	400 x 12 4-Seam Trawl	Wilcox 14 m high-rise otter trawl	3/4 North Atlantic-type two seam (Whiting) trawl
Sweep	Rockhopper	Roller	Cookie sweep		7.6 cm rubber disc sweep
Sampling	24 -Hour	24-hour	Day	Day	Day
Speed knot	3.0	3.5	3.0	3.5	2.5
Speed (km h^{-1})	5.55	6.50	5.55	6.48	4.63
Tow Duration (min)	20	30	20	30	20
Wing Spread (m)	12.80	12.50	12.80	8.00	8.40
Door Spread (m)	32.66	26.50	32.66		
Swept Area (a_s) km^2	0.024	0.038	0.024	0.0259	0.013
Survey Area (A_s) km^2	224,562	224,562	12,097	3,400	6,285

Table 2. Strata set used in the analyses

Species	Stock	NEFSC Strata set	NEAMAP	MASS strata	LIS
Atlantic cod	GOM	26-30, 36-40	-	25-36	-
Atlantic cod	GB	13-25	all	11-21	-
Haddock	GOM	26-28, 36-40	-	25-36	-
Haddock	GB	13-25	-	11-21	-
Pollock	Unit	Core Offshore,Core Inshore	all	All	-
Butterfish	Unit	Core Offshore,Core Inshore	all	All	all
Bluefish	Unit	1-14, 16, 19-20,23,25,61-70, Core Inshore	all	All	Need to get
Summer Flounder	Unit	1,2,5,6,9,10,61,65,69,73, CoreInshore	all	All	Need to get
Witch Flounder	Unit	Core Offshore,Core Inshore	-	All	Need to get
American Plaice	Unit	Core Offshore,Core Inshore		All	Need to get
Winter Flounder	SNE	1,2,5,6,9,10,25,69-74 Inshore: 2,5,8,11,14,17,20,23,26,29,45,46,56	all		Need to get
Winter Flounder	GB	13-23	-	-	Need to get
Yellowtail Flounder	SNE	1,2,5,6,9,10	all	11-16	Need to get
Yellowtail Flounder	GB	13-21	-	-	-
Yellowtail Flounder	GOM	25-26,39-40 Inshore: 56,59,60,61,64,65,66	-	25,26	-
Scup	Unit	Core Offshore,Core Inshore	all	All	Need to get
Black Sea Bass	Unit	Core Offshore,Core Inshore	all	All	Need to get
Atlantic Herring	Unit	Core Offshore,Core Inshore	all	All	-
Acadian Redfish	Unit	24,26-30, 36-40	-	All	-
Atlantic Mackerel	Unit	Core Offshore,Core Inshore	-	All	-

Figure 1. Diagram of bottom trawl gear. The area in orange corresponds to the wing swept area typically used as a measure of the area sampled by the bottom trawl gear. The door swept area also includes the area in blue. The use of door swept areas assumes that the sampled fish are herded by the sweep and doors into the area in front of the mouth of the net before eventually falling back into the net cod end. Modified from

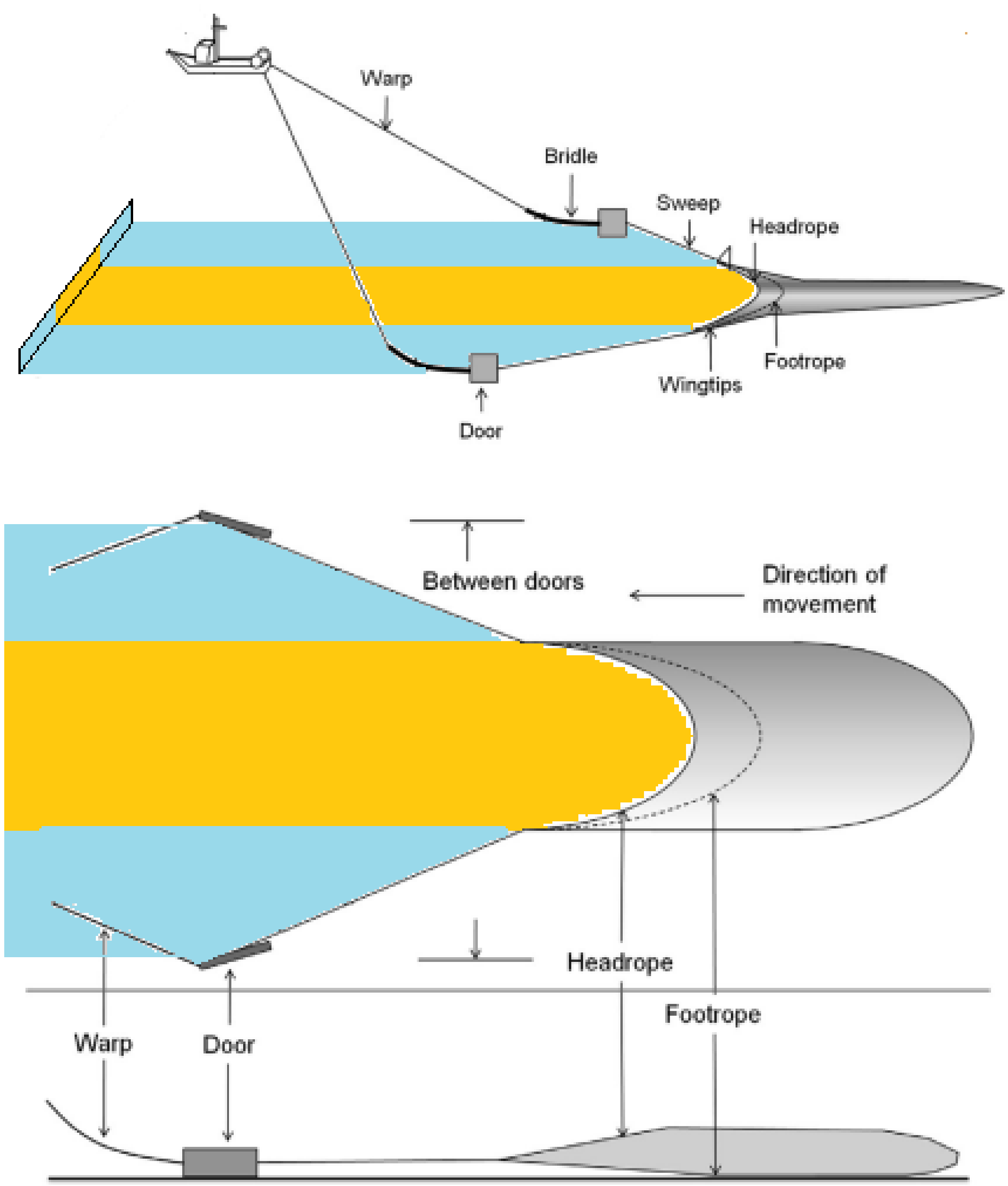


Fig. 2. Biological and sampling parameters for the Georges Bank Yellowtail stock during the fall. The NEAMAP and Massachusetts trawl surveys do not sample Georges Bank yellowtail flounder.

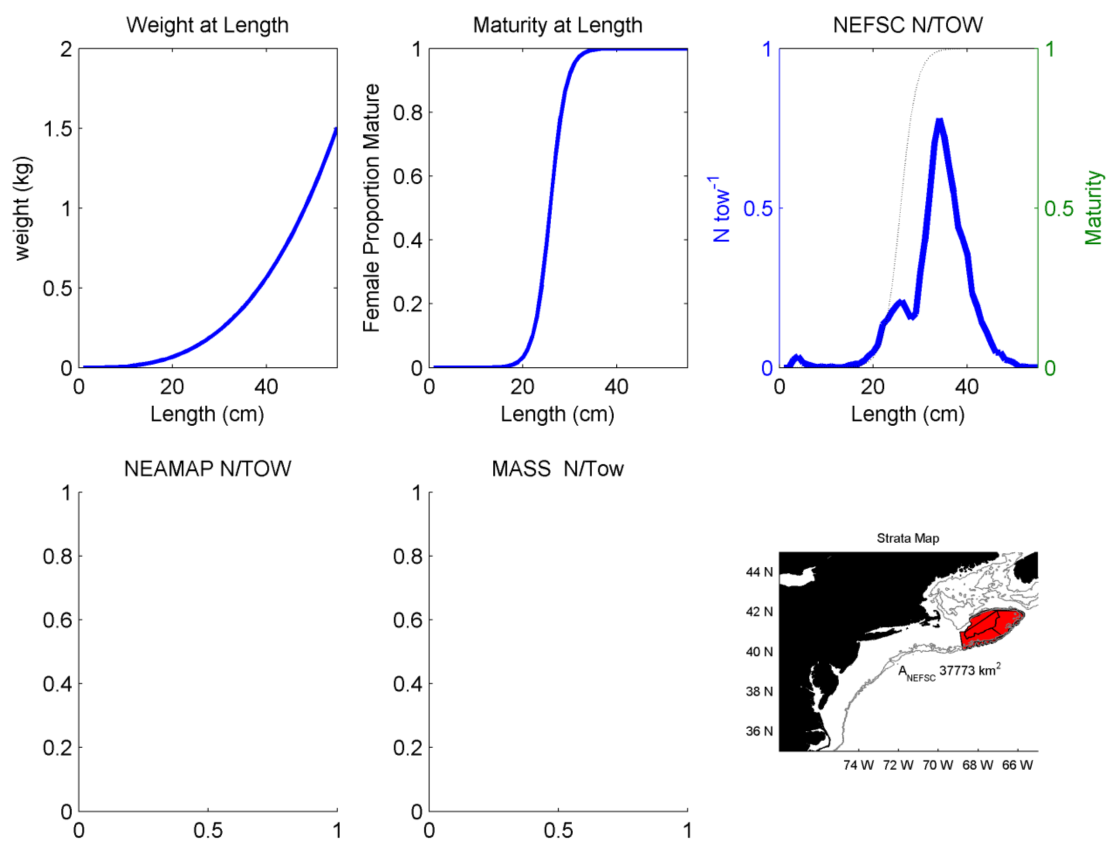


Fig 3. Components of length-specific catchability of Georges Bank yellowtail flounder in the fall. Black line corresponds to the median value from the bootstrapping and the grey bars the 95% confidence intervals. Availability is assumed to be 1 as no other survey covers the stock area. In calculating the survey spawning stock biomass estimated the upper 95% CI of the maximum Albatross Q was used.

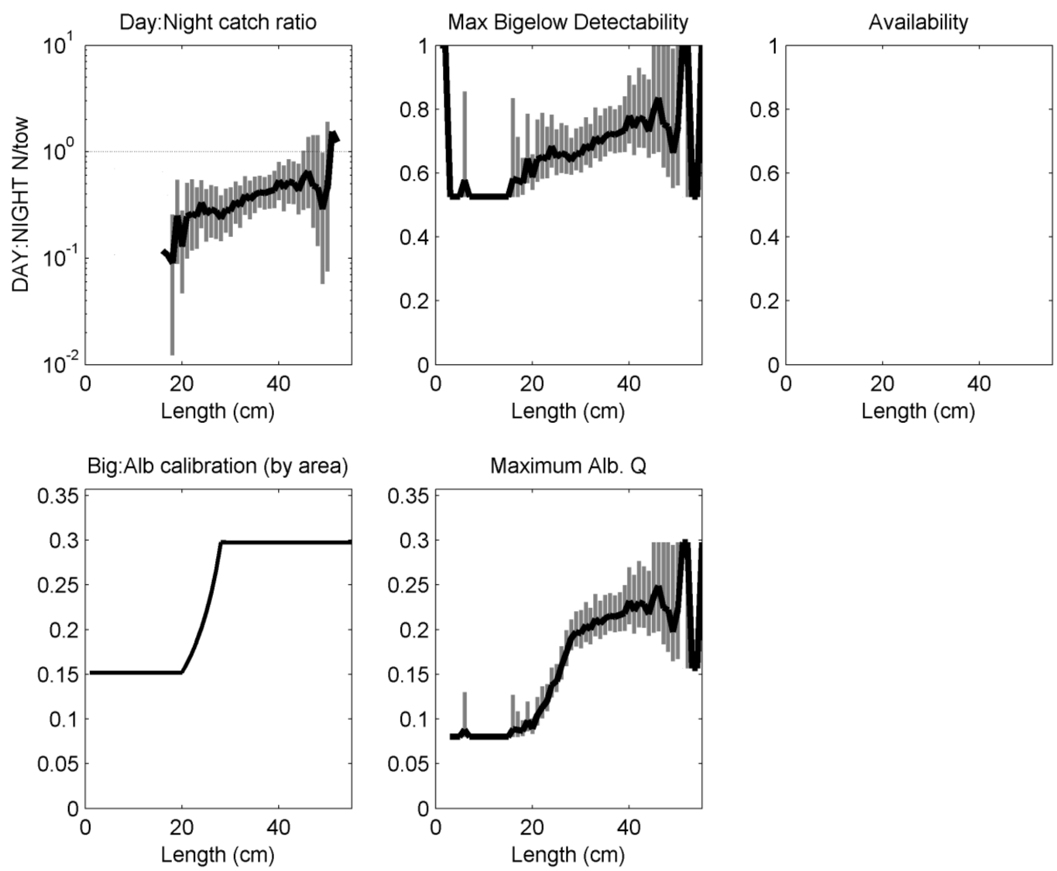


Fig. 4. Spawning stock biomass estimates from the survey and assessment. The implied Q represents the ratio of the Wing-Swept area SSB and the Assessment SSB. The dashed red line in panel B corresponds to the door swept area. The histograms are of the implied-Q across the time series.

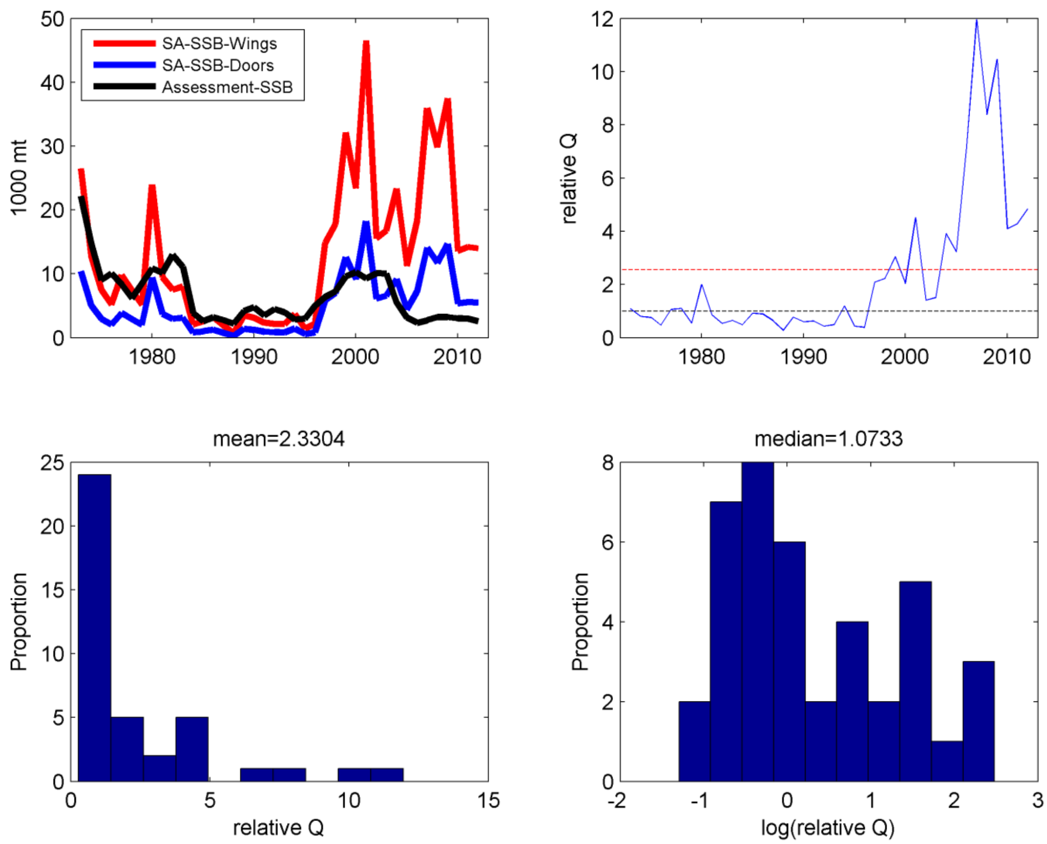


Fig 5. Same as figure 2 except for the Spring Georges Bank yellowtail survey.

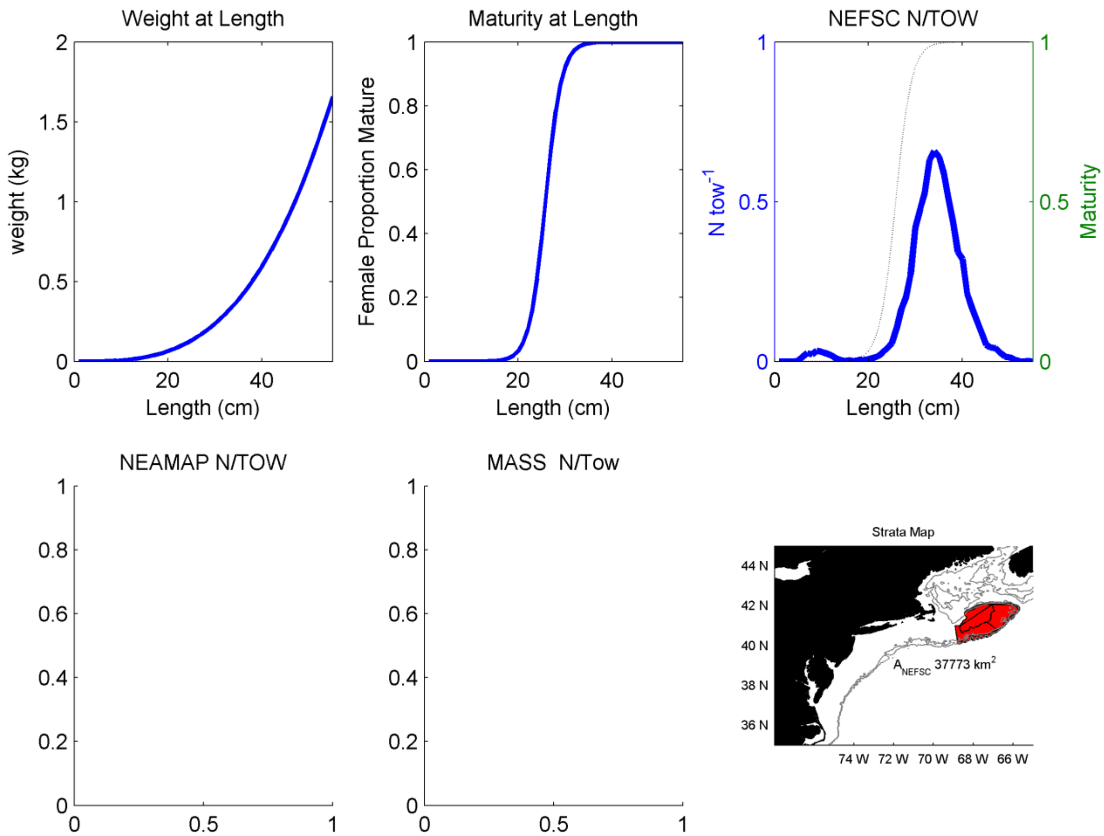


Fig 6. Same as figure 3 except for the spring Georged Bank yellowtail survey.

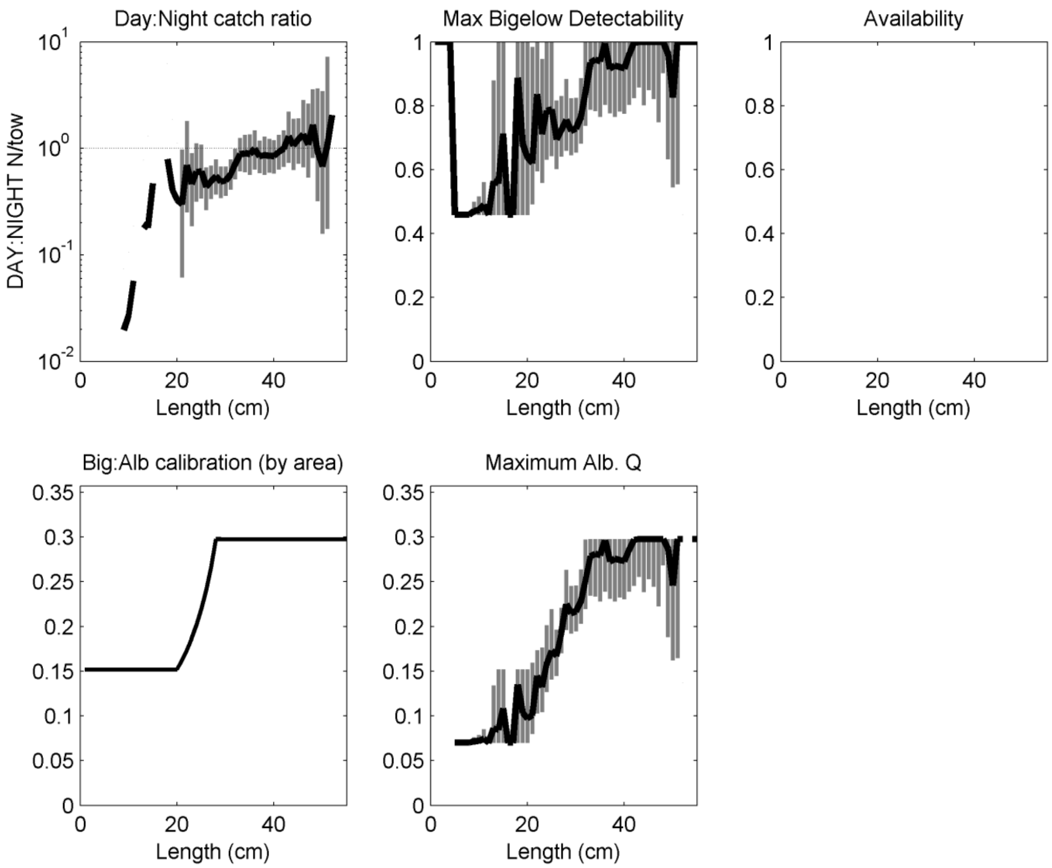


Fig 7. Same as figure 4 except for the spring Survey.

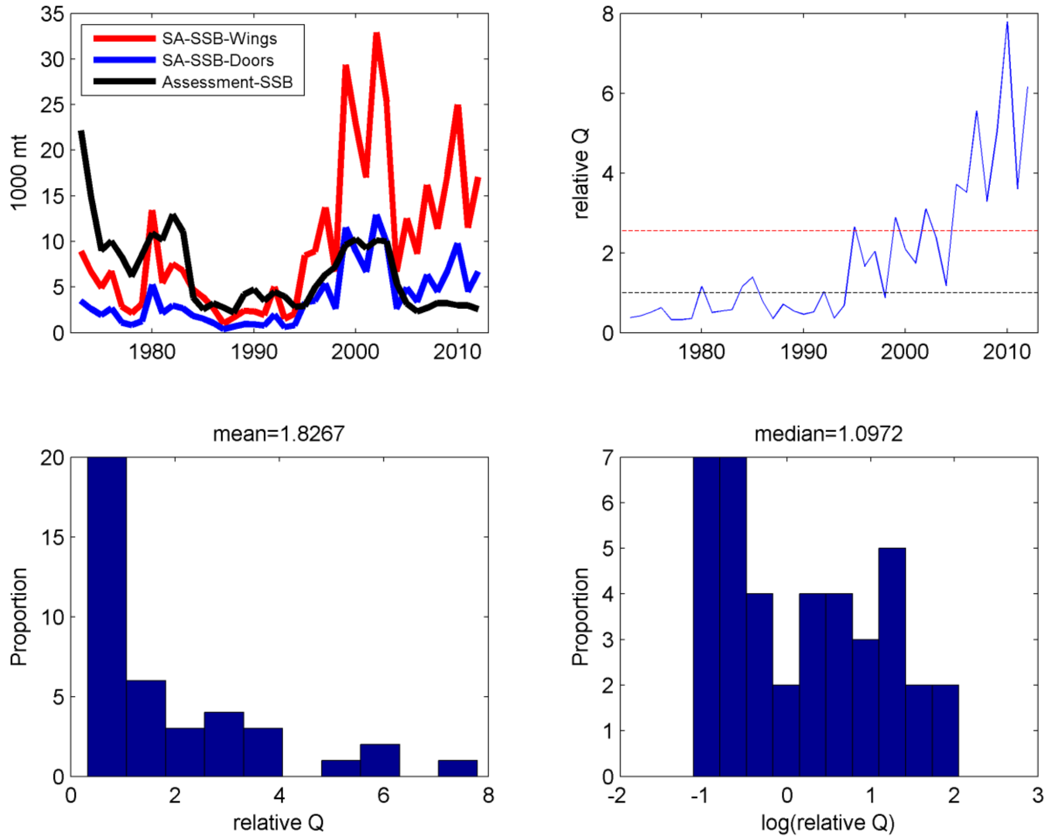


Fig. 8. Day:Night catch ratios of the NEFSC survey for each stock with 95% confidence intervals. The length-specific Day:Night ratios are weighted by the maturity at length, weight at length and Number at length to provide a single aggregate number that corresponds to the day:night ratio of spawning stock biomass.

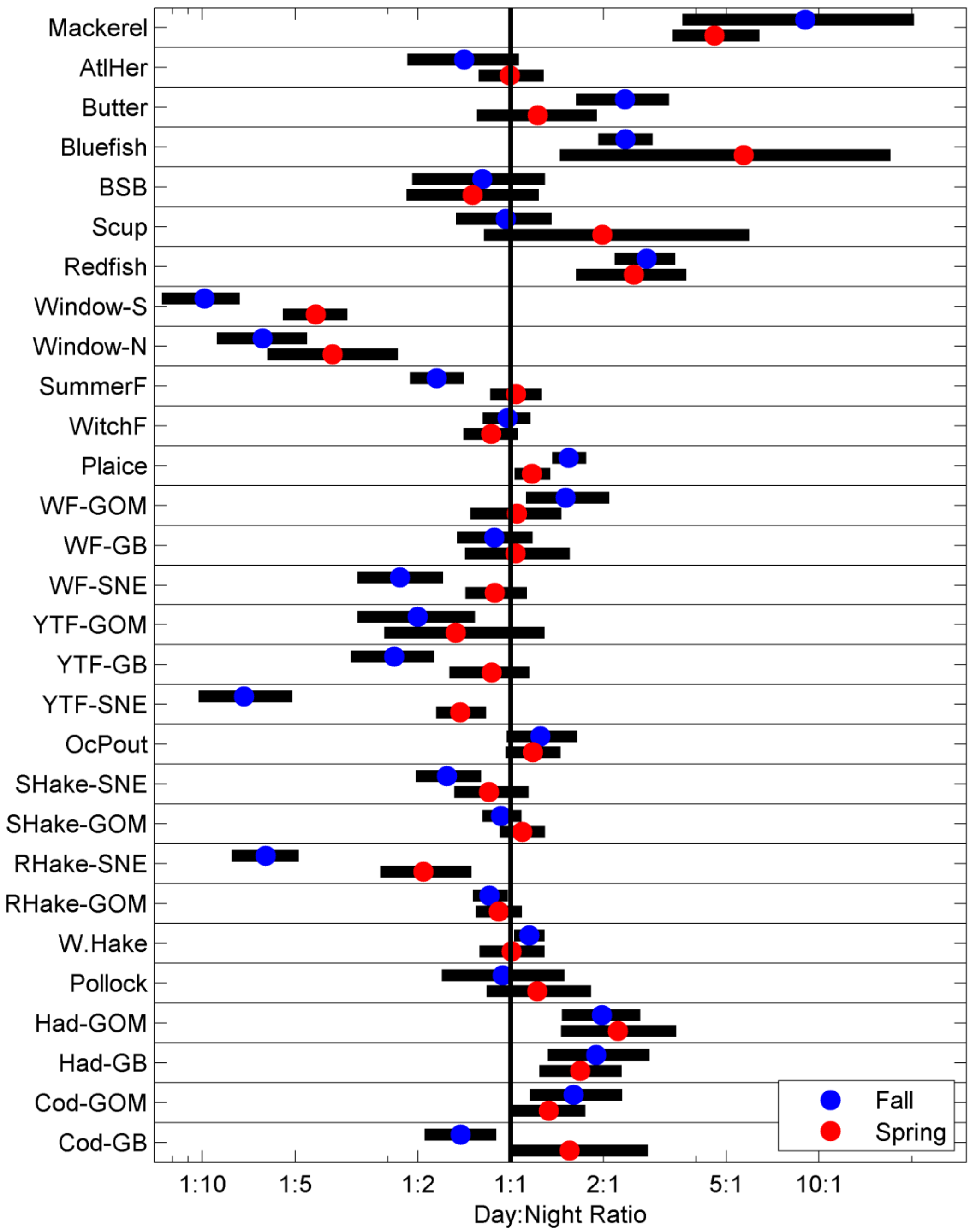


Fig. 9. Availability estimates of the NEFSC fall trawl survey based on a comparison of NEFSC, NEAMAP, and Massachusetts trawl survey data. Long Island Sound data is only included for butterfish. The length-specific Availability is weighted by the maturity at length, weight at length and Number at length to provide a single aggregate number that corresponds to the availability of spawning stock biomass. Detectability of the daytime inshore surveys is assumed to be 1 even if the NEFSC survey shows much higher nighttime catch levels.

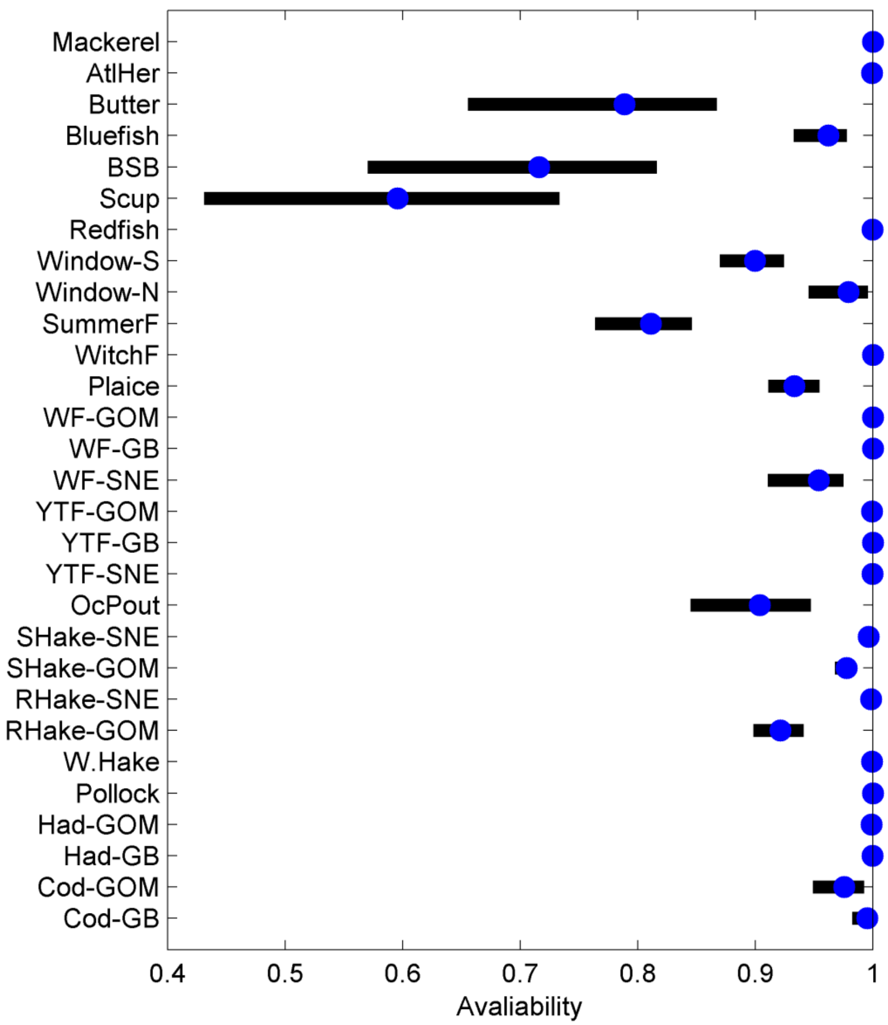


Fig. 10. Median implied catchability estimates of 21 NEFSC stock assessments across either the duration of the assessment or since 2000. The standard of comparison is either the nighttime or daytime Bigelow survey accounting for the measured abundance of fish outside the survey area (fall only). To be conservative, the upper 95% CI of catchability at length was used. The wing swept area corresponds to a catchability of 1 and the door swept area a catchability of 2.55. Values from 1-2.55 are possible due to herding.

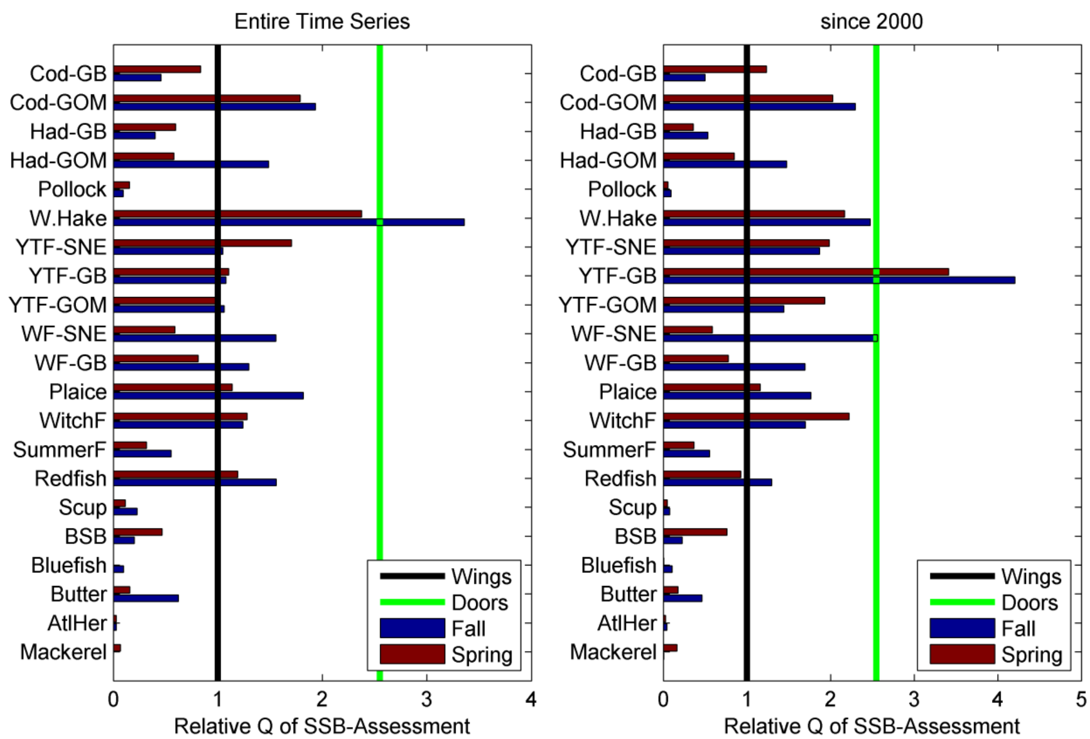


Fig 11. Implied catchability through time of 21 assessed NEFSC stocks across two surveys. (Note: the mackerel assessment was not accepted but are included here for comparison). Black lines are present to indicate a catchability of 1 corresponding to wing swept area and green lines a catchability of 2.55 for door swept area.

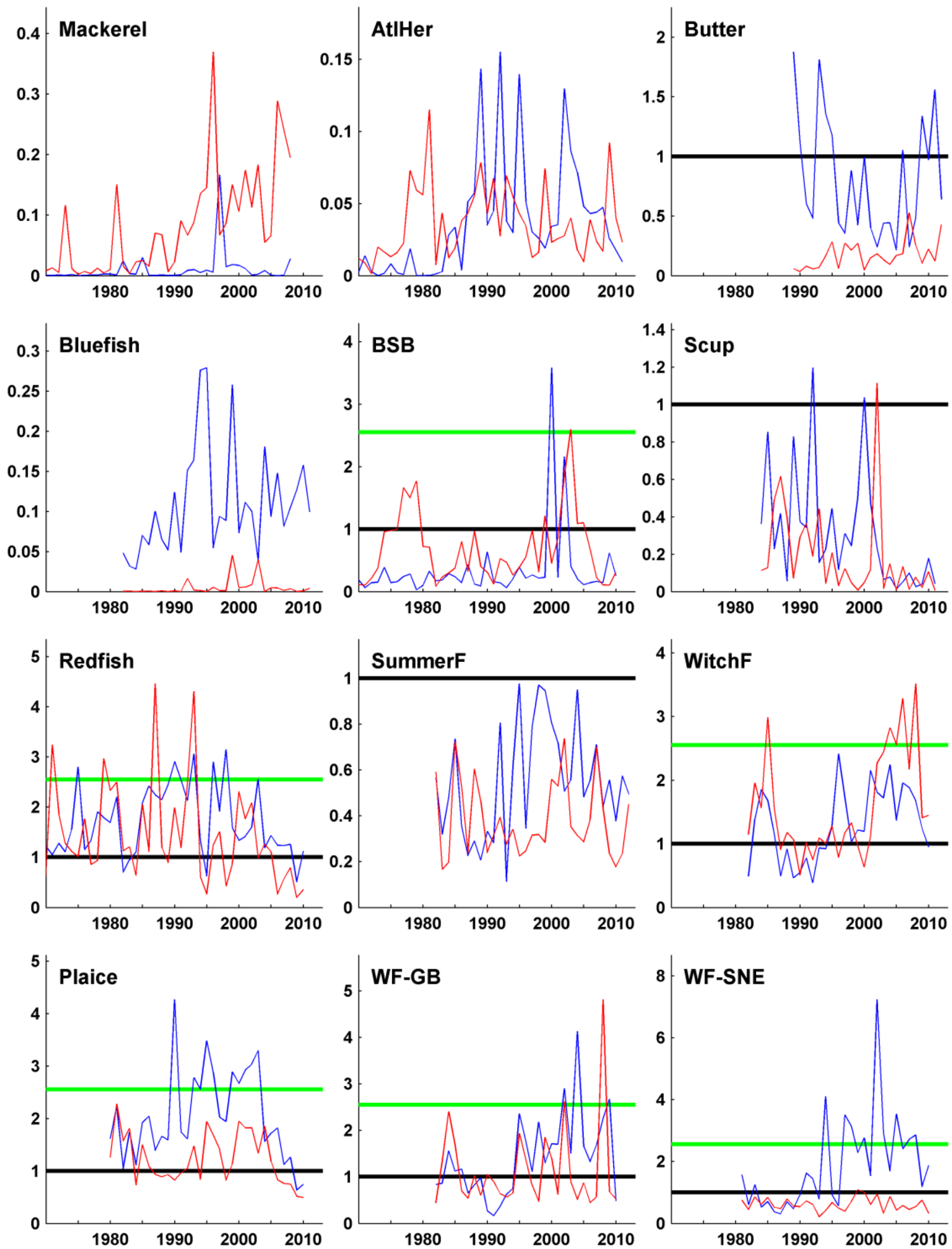
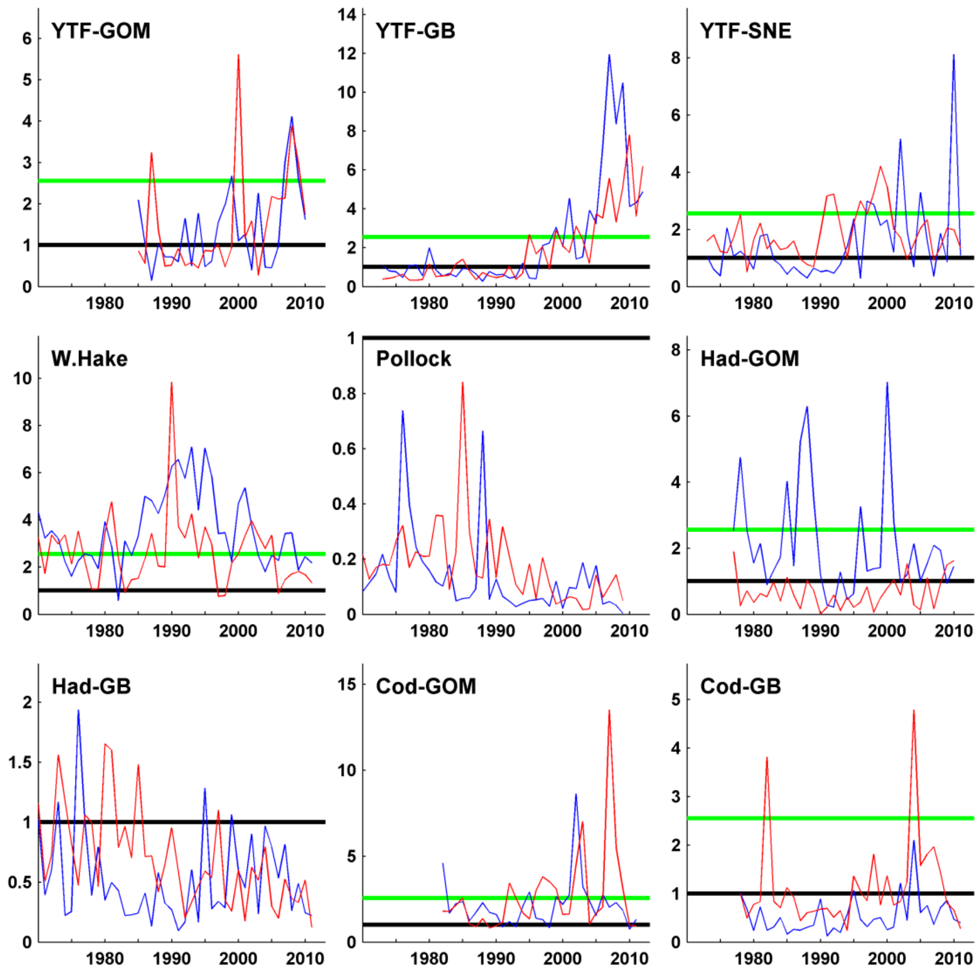


Fig 11 cont



YTF-GOM
YTF-GB
YTF-SNE
W.Hake
Pollock
Had-GOM
Had-GB
Cod-GOM
Cod-GB
1980
1990
2000
2010

Fig. 12. Plot of the de-meaned log of the implied-Q values for 7 flatfish stocks across the spring and fall trawl survey (grey) and for Georges Bank yellowtail (red). The demeaned log of the IQ values provide an indication of how much above or below each survey value is from what is expected based on the assessment results. For example, a value of 0.69 (log 2) indicates a survey value double the expected whereas a value of -0.69 (log 0.5) is a survey value half of the expected. The solid lines represent the average values for each year.