Estimation of survey efficiency and biomass for commercially important species from industry-based paired gear experiments

Timothy J. Miller[1] David E. Richardson[2] Philip J. Politis[1]

Christopher D. Roebuck[3] John P. Manderson[4,5]

Michael H. Martin[1,6] Andrew W. Jones[2]

[1]corresponding author: timothy.j.miller@noaa.gov, Northeast Fisheries Science Center, Woods Hole Laboratory, 166 Water Street, Woods Hole, MA 02543 USA

[2]Northeast Fisheries Science Center, Narragansett Laboratory, 28 Tarzwell Drive Narragansett, RI 02882 USA

[3]Fishing Vessel Karen Elizabeth, Narragansett, RI 02882, USA

[4]Northeast Fisheries Science Center, James J. Howard Marine Sciences Laboratory, 74 Magruder Road Highlands, NJ 07732 USA

[5]Current address: OpenOcean Research, Suite 101, 40 West Evergreen Avenue, Philadelphia, PA, 19118 USA

[6]Current address: Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Building 4 Seattle, WA 98115 USA

Abstract

Fishery-independent surveys provide valuable information about trends in population abundance for management of commercially important fish stocks. A critical component of the relationship of the catches of the survey to the size of a fish stock is the catch efficiency of the survey gear. Using a general hierarchical model we estimated the relative efficiency of a chain sweep to the rockhopper sweep used by the Northeast Fisheries Science Center bottom trawl survey from paired-gear experimental tows carried out between 2015 and 2017 using a twin-trawl vessel. For 10 commercially important species, we fitted and compared a set of models with alternative assumptions about variation of relative efficiency between paired gear tows, size and diel effects on the relative efficiency, and extra-binomial variation of observations within paired gear tows. These analyses provided evidence of changes in relative efficiency with size for all species and diel effects were important for all but one species. We then used the bottom trawl survey data from surveys between 2009 and 2019 with the relative catch efficiency estimates from the best performing models to estimate annual and seasonal chain sweep-based swept area biomass for 17 managed stocks. We estimated uncertainty in all results using bootstrap procedures for each data component. We also assessed the effect of calibration on uncertainty and correlation of the annual biomass estimates.

Keywords

gear efficiency, biomass estimation, hierarchical generalized additive models

1 Introduction

Ecosystem monitoring surveys such as fisheries-independent trawl surveys are used to obtain information on a range of species and are therefore not optimized with respect to sampling design or gear for any one species (Bijleveld et al., 2012; Wang et al., 2018). Gear and sampling protocols are designed to provide consistent and representative samples that allow indices of abundance at size and age to be developed for a suite of species (Azarovitz, 1981; Thiess et al., 2018). To provide indices of population abundance with minimal potential sources of bias, survey bottom trawl gear must be configured to be towed across as wide a variety of habitats as possible, including seafloor habitats with complex physical structures.

Indices of abundance at age and size derived from fishery-independent bottom trawl surveys are scaled to population size by the survey catchability (q) parameter (Arreguín-Sánchez, 1996). Catchability is typically estimated internally within stock assessment models that incorporate fisheries landings, indices of abundance, and life history parameters. However, the amount or quality of data and degree of contrast in the time series is often such that this parameter, and therefore the population size, is difficult to estimate (Maunder and Piner, 2015). In such cases, estimates of survey catchability from auxiliary data can inform the stock assessment. These external estimates can be used as a direct input into the assessment model (Somerton et al., 1999), can serve as a diagnostic measure of model accuracy (Miller et al., 2019), or contribute to an alternate means of providing catch advice when an assessment model is not considered acceptable (Legault and McCurdy, 2017).

Catchability can be decomposed into two components, the proportion of the population available to the survey sampling frame and the efficiency of the survey gear given an individual is available to the gear (Paloheimo and Dickie, 1964). Here efficiency is the fraction of available fish retained by the gear, equivalent to availability-selection in Millar and Fryer (1999). Estimates of these components allow relative abundance indices to be converted into absolute abundance indices without a population model. As such, investigations of gear mensuration (Kotwicki et al., 2011), species-specific gear efficiency (Thygesen et al., 2019), and availability of the stock to the survey design frame (Nichol et al., 2019) improve our understanding of catchability and therefore abundance of fish stocks.

Paired-gear studies where two gears are fished either concurrently or close together temporally and spatially have long been used to estimate the efficiency of one fishing gear relative to another (e.g., Gulland, 1964; Bourne, 1965). Of the two gears, one is often a reference gear that may be a gear currently used for annual surveys (e.g., Munro and Somerton, 2001). Typically neither of the gears is fully efficient and therefore the relative efficiency of gears is estimated (e.g., Miller, 2013; Kotwicki et al., 2017), but there are cases where one of the gears is assumed to be very nearly fully efficient (e.g., Somerton et al., 2013; Miller et al., 2019).

Whether or not full efficiency of one of the gears is assumed, paired-gear studies are essential for generating abundance time series from fishery-independent surveys when there are changes in the vessel and(or) gears over time due to gear failures or improved technology (Pelletier, 1998). These studies are also helpful for combining surveys conducted close together in space or time using alternative gears (Kotwicki et al., 2013).

Within the northeast US there has been a heightened focus on bottom trawl survey operations and gear efficiency. This focus has in part resulted from low quotas for several groundfish stocks limiting fishing opportunities. To help provide clarity on the trawl operations and build trust in survey indices the New England and Mid-Atlantic Fisheries Management Councils developed a Northeast Trawl Advisory Panel. This panel is composed of members from industry, regional academics, as well as state and federal scientists. Together the group designed a set of experiments to better understand the efficiency of the bottom trawl survey gear for northeast US groundfish stocks.

In conducting paired-gear studies it is ideal to have the two gears deployed as close together spatially and temporally as possible to reduce variation between the gears in densities of the species being encountered. The twin-trawl rigging (Krag et al., 2015) where two trawls can be fished simultaneously approaches this ideal (ICES, 1996), and is the data-collection platform chosen by the Trawl Advisory Panel. The Panel decided to rig one of the twin trawls as the the gear used by the bottom trawl survey which uses a rockhopper sweep. The other trawl was rigged similarly except with a chain sweep in an attempt to eliminate any escapement of fish under the gear. Assuming the chain sweep-based gear is fully efficient allows the efficiency of the rockhopper sweep-based gear used by the bottom trawl survey to be estimated from these experiments.

The analytical methods to estimate the efficiency of the bottom trawl gear are based on those used by Miller (2013) to estimate size effects on relative catch efficiency of the NOAA Ship *Henry B. Bigelow* (*Bigelow*) to the NOAA Ship *Albatross IV* for a variety of commercially important species. We extend the model to consider different size effects for tows conducted during the day or night since both the spring and fall bottom trawl surveys conducted in the Northeast US are 24-hour operations. We apply these methods to paired gear observations and estimate relative efficiency of the chain sweep and rockhopper sweep gears. We also apply the estimated efficiency of the rockhopper gear to survey data to estimate spring and fall biomass indices from 2009-2019 for 17 commercially important fish stocks in the Northeast US (Table 1).

Often overlooked aspects of the application of relative catch efficiency estimates include the impact on the precision of abundance indices and the correlation among annual indices that the application induces. These indices are typically used as measures of relative abundance in stock assessment with the precision of the indices used to weight the observations within the assessment model. Furthermore, the observation error for each of the annual and seasonal indices is typically assumed to be independent of the others. Here we compare the precision of the biomass indices calibrated to the chain sweep gear to that of the and uncalibrated indices using the rockhopper sweep gear and measure the correlation of calibrated indices for each stock.

2 Methods

2.1 Data collection

Data were collected during three field experiments carried out in 2015, 2016, and 2017, respectively, aboard the *F/V Karen Elizabeth*, a 23.8m (78ft) stern trawler capable of towing two trawls simultaneously side by side (Figure 1). One side of the twin-trawl rig towed a NEFSC standard 400 x 12 cm survey bottom trawl rigged with the NEFSC standard rockhopper sweep (Politis et al., 2014) (Figure 2). The other side of the twin-trawl rig towed a version of the NEFSC 400 x 12cm survey bottom trawl modified to maximize the capture of flatfish. The modifications included reducing the headline flotation from 66 to 32, 20cm, spherical floats, reducing the port and starboard top wing-end extensions by 50cm each, and utilizing a chain sweep. The chain sweep was constructed of 1.6cm ($\frac{5}{8}$in) trawl chain covered by 12.7cm diameter x 1cm thick rubber discs on every other chain link (Figure 2). Two rows of 1.3cm ($\frac{1}{2}$in) tickler chains were attached to the 1.6cm trawl chain by 1.3cm shackles. To ensure equivalent net geometry of each gear, 32m restrictor ropes, made of 1.4cm ($\frac{9}{16}$in) buoyant, Polytron rope, were attached between each of the trawl doors and the center clump. 3.4m^2 Thyboron Type 4 trawl doors were used to provide enough spreading force to ensure the restrictor ropes remained taut throughout each tow. Each trawl used the NEFSC standard 36.6m bridles. All tows followed the NEFSC standard survey towing protocols of 20 minutes at 3.0 knots. Port and starboard net spreads were measured separately with two sets of Simrad ITI acoustic net mensuration sensors measuring from the port wing-end to the center clump and the starboard wing-end to the center clump. In 2015, 108 (45 day, 63 night) paired tows were conducted in eastern Georges Bank and off of southern New England (Figure 3). In 2016, 117 (74 day, 43 night) paired tows were conducted in western Gulf of Maine and the northern edge of Georges Bank. In 2017, 103 (61 day, 42 night) paired tows were conducted in the western Gulf of Maine and off of southern New England. Paired tows were denoted as "day" and "night" by whether the sun was above or below the horizon at the time of the tow.

In order to reduce shipboard processing time and maximize the number of tows, only select taxa were enumerated and measured, rather than the full processing of all species as occurs on the trawl survey (Politis et al., 2014). All flatfish species (order Pleuronectiformes), thorny skate (*Amblyraja radiata*), barndoor skate (*Dipturus laevis*) and goosefish (*Lophias americanus*) collected in each net of each tow were independently sorted, weighed and measured in all years. If the catch of a species was greater than ≈150 individuals, a subsample of ≈150 individuals was measured. Red hake (*Urophycis chuss*) were not quantified during the 2015 and 2016 sampling because other species were prioritized, but were fully processed in 2017. Winter skate (*Leucoraja ocellata*) and little skate (*L. erinacea*) were weighed in all years and but were not separated to species nor measured. Sea scallops were weighed in 2015 and 2016, but not 2017.

2.2 Paired-tow analysis

We employed the hierarchical modeling approach from Miller (2013) to estimate the efficiency (ρ) of the rockhopper sweep used by the NEFSC bottom trawl survey relative to the chain sweep-based gear for ten species (Summer flounder, *Paralichthys dentatus*; American plaice, *Hippoglossoides platessoides*; windowpane flounder, *Scophthalmus aquosus*; winter flounder, *Pseudopleuronectes americanus*; yellowtail flounder, *Limanda ferruginea*; witch flounder, *Glyptocephalus cynoglossus*; red hake; goosefish; barndoor skate; thorny skate) from three separate research trips carried out aboard a twin trawl vessel. We first fit and compared the same set of 13 models as Miller (2013) with different assumptions about variation of relative efficiency between paired gear tows, size effects on the relative efficiency, and extra-binomial variation of observations within paired gear tows. The binomial (BI_0 to BI_4) and beta-binomial (BB_0 to BB_7) models that were fitted for all species are described in Table 2 including pseudo-formulas analogous to those used to specify and fit mixed or generalized additive models in R (R Core Team, 2019; Wood, 2006). We then also included diel effects on relative catch efficiency and interactions with size effects with the best performing model of the original 13 models for each species. To fit these diel effects, we generalized the modeling framework somewhat in that we allowed multiple (cubic regression spline) smooth effects, differing by day and night, on relative catch efficiency. We implemented the models using the Template Model Builder package (Kristensen et al., 2016) in R and we used the "nlminb" optimizer to fit the models by maximizing the Laplace approximation of the marginal likelihood (R Core Team, 2019).

If the best model included smooth length effects and the estimated smoothing parameter implied a linear functions of length (on the transformed mean), then simple linear functions (i.e., completely smooth) were assumed for further models that included diel effects on relative efficiency. As such, there was one less (smoothing) parameter estimated for these models.

We compared two alternative ways of estimating uncertainty in relative catch efficiency. The first estimation approach uses the inverted hessian of the marginal log-likelihood and the delta-method to estimate uncertainty in the predicted relative catch efficiency at size. The second method is a bootstrap method where we refit models to bootstrap resamples of the paired station data. Specifically, we resampled the paired tows with replacement so that the total number of paired tows was the same for a given species, but the total number of length measurements varied depending on which of the paired tows entered the sample for a particular bootstrap. We made 1000 bootstrap samples and estimated relative catch efficiency at size from each bootstrap data set if the fitted model converged and the hessian at the maximized log-likelihood was invertible.

For models BI_4, BB_6, and BB_7, there are two fixed effects parameters associated with the spline coefficients that are treated as random effects for station-specific smoothers and the correlation of these pairs of random effects is estimated. However, this parameter was not estimable for red hake for BB_6 and assumed equal to zero.

2.3 Length-weight analysis

We fit length-weight relationships to the length and weight observations for each survey each year. We assumed weight observation j from survey i, was log-normal distributed,

$$\log W_{ij} \sim \mathrm{N}\left(\log \alpha_i + \beta_i \log L_{ij} - \frac{\sigma_i^2}{2}, \sigma_i^2\right) \tag{1}$$

We used a bias correction to ensure the expected weight $E(W_{ij}) = \alpha_i L_{ij}^{\beta_i}$. We estimated parameters by maximizing the model likelihood programmed with the Template Model Builder package and R and generated predictions of weight at length

$$\widehat{W}(L) = \widehat{\alpha} L^{\widehat{\beta}}. \tag{2}$$

Like the relative catch efficiency, we made bootstrap predictions of weight at length by sampling with replacement the length-weight observations within each annual survey and refitting the length-weight relationship to each of the bootstrap data sets.

2.4 Biomass estimation

For the 17 managed stocks that are populations of the species in the Northeast US where we have estimated relative efficiency, we estimated stock biomass for each spring and fall annual survey assuming 100% efficiency of the chain sweep gear by scaling the survey tow observations by the relative efficiency of the chain sweep and rockhopper sweep gears. Summer and witch flounders, American plaice, and barndoor and thorny skates are managed as single unit stocks, but there are three stocks of winter and yellowtail flounders, and two stocks of windowpane, red hake, and goosefish (Table 1). First, the tow-specific catches at length are rescaled,

$$\widetilde{N}_{hi}\left(L\right) = N_{hi}\left(L\right)\widehat{\rho}_i\left(L\right) \tag{3}$$

where $N_{hi}(L)$ is the number at length L in tow i from stratum h and $\widehat{\rho}_i\left(L\right)$ is the relative efficiency of the chain sweep to rockhopper sweep at length L estimated from the twin trawl observations that may depend on the diel characteristic of tow i if that factor is in the best model fitted to the twin-trawl observations. Note that we have omitted any subscripts denoting the year or season.

The stratified abundance estimate is then calculated using the design-based estimator,

$$\widehat{N}(L) = \sum_{h=1}^{H}\frac{A_h}{an_h}\sum_{i=1}^{n_h}\widetilde{N}_{hi}(L) \tag{4}$$

where A_h is the area of stratum h, a is the average swept area of a survey station tow, and n_h is the number of tows that were made in stratum h. The corresponding biomass estimate is then

$$\widehat{B} = \sum_{l=1}^{n_L}\widehat{N}(L=l)\widehat{W}(L=l) \tag{5}$$

where $\widehat{W}(L=l)$ is the predicted weight at length (Eq. 2) from fitting length-weight observations described above. Length is typically measured to the nearest cm so n_L indicates the number of 1 cm length categories observed during the survey.

We used the same criteria for survey station selection as those currently used to estimate indices of abundance or biomass for management of each stock. For Gulf of Maine winter flounder we also restricted the size classes in each tow to those ≥ 30 cm as the biomass of the population over this threshold is currently used for management of this stock. For some stocks there were certain years where some but not all of the set of survey strata used to define indices of abundances were sampled by the bottom trawl survey. In those years, the average catch per unit area was expanded to all of the stock strata proportionally to the areas of the sampled and unsampled strata. The fall 2017 survey was extremely restricted because of vessel mechanical failure and indices are not available for summer flounder, SNE-MA windowpane, and SNE-MA yellowtail flounder.

To estimate uncertainty in biomass, we used bootstrap results for the relative catch efficiency and weight at length estimates along with bootstrap samples of the survey data. Bootstrap data sets for each of the annual surveys respected the stratified random designs by resampling with replacement within each stratum (Smith, 1997). For each of the 1000 combined bootstraps, survey observations for bootstrap b were scaled with the corresponding bootstrap estimates of relative catch efficiency and predicted weight at length, using Eqs. 4 and 5.

We also used the bootstraps to summarize other aspects of the biomass estimates. First, we used the bootstraps to calculate the ratio of calibrated and uncalibrated biomass for each spring and fall annual survey, which is the implicit relative catch efficiency in terms of biomass. The uncalibrated biomass estimate for bootstrap b uses the resampled survey data as the calibrated biomass estimate except that the bootstrap for the relative catch efficiency is not used (i.e., $\widehat{\rho}_i\left(L\right) = 1$ in Eq. 3). We also used the bootstraps to compare the coefficients of variation (CV) of the calibrated and uncalibrated biomass estimates. The CV for an annual biomass estimate for year y from either the spring or fall survey was calculated as

$$\mathrm{CV}\left(\widehat{B}_y\right) = \frac{\mathrm{SD}\left(\widehat{B}_y\right)}{\overline{\widehat{B}}_y}$$

where

$$\mathrm{SD}\left(\widehat{B}_y\right) = \sqrt{\frac{\sum_{b=1}^{K}\left(\widehat{B}_{y,b} - \overline{\widehat{B}}_y\right)^2}{K-1}},$$

$$\overline{\widehat{B}}_y = \frac{\sum_{b=1}^{K}\widehat{B}_{y,b}}{K},$$

and K is the number of bootstraps.

For summer flounder it was necessary to omit one of the 1000 bootstraps of relative catch efficiency at length due to an extremely large value to which the standard deviation and mean of the bootstraps were sensitive. Finally, just as the uncertainty in $\rho(L)$ affects the uncertainty in the calibrated abundance at length and biomass estimates, it also induces correlation among the annual and seasonal estimates because the same estimates are applied to all of them. We calculated the correlation of annual biomass estimates for years y and z using the bootstrap estimates of biomass

$$Cor\left(\widehat{B}_y, \widehat{B}_z\right) = \frac{Cov\left(\widehat{B}_y, \widehat{B}_z\right)}{\mathrm{SD}\left(\widehat{B}_y\right)\mathrm{SD}\left(\widehat{B}_z\right)}$$

where the covariance is

$$Cov\left(\widehat{B}_y, \widehat{B}_z\right) = \frac{\sum_{b=1}^{K}\left(\widehat{B}_{y,b} - \overline{\widehat{B}}_y\right)\left(\widehat{B}_{z,b} - \overline{\widehat{B}}_z\right)}{K-1}.$$

We summarized the relative precision of the calibrated and uncalibrated biomass estimates as the average of the annual ratios of the CVs for the calibrated and uncalibrated estimates

$$\frac{1}{n_y}\sum_{y=1}^{n_y}\frac{CV\left(\widehat{B}(\rho)\right)}{CV\left(\widehat{B}\right)}.$$

We summarized the correlation of biomass estimates as the mean correlation of all annual calibrated biomass estimates

$$\overline{Cor} = \frac{1}{n_y(n_y-1)/2}\sum_{y=2}^{n_y}\sum_{z=1}^{y} Cor\left(\widehat{B}_y, \widehat{B}_z\right).$$

All code and most data files to run the analysis and generate biomass estimates are available at https://github.com/timjmiller/chainsweep_paper.

3 Results

3.1 Paired-tow observations

In terms of paired tows and total numbers of fish, flatfish were the best sampled species, but goosefish was observed in the most paired-tows and red hake was one of the most prevalent in terms of total numbers caught (Table 3). Witch flounder was the most prevalent flatfish species caught while yellowtail flounder was the most frequently observed flatfish in terms of paired tows. The proportion of fish measured for length relative to the number caught varied across species. All summer flounder, barndoor skate, and thorny skate that were captured were measured. Subsampling occurred for all other species with a high proportion (>97%) measured for winter flounder and goosefish, a moderate proportion (50-97%) measured for American plaice, windowpane flounder, and yellowtail flounder, and a low proportion (<50%) measured for witch flounder and red hake.

3.2 Relative catch efficiency

As measured by AIC, the best performing models for all 10 species included size effects on the relative efficiency of the chain and rockhopper sweep gears and between-pair variability in relative catch efficiency (Table 4). Extrabinomial variation (i.e., beta-binomial) in relative catch efficiency at size within pairs was also important for American plaice, yellowtail flounder, witch flounder, red hake, and thorny skate. Model convergence was an issue for all species, particularly for the most complex models with pair-specific smooth functions of length (BI_4) and smooth effects of size on the beta-binomial dispersion parameter (BB_3,BB_5, and BB_7).

Including diel effects on relative catch efficiency improved model performance for all species except American plaice (Table 5). For those species with diel effects on relative catch efficiency, the ratio of the efficiencies was generally greater for daytime observations than those for nighttime tows, with the exception of large winter flounder (Figure 4). The largest differences in efficiency was estimated for smaller barndoor skate. For most of the species, the differences in efficiency between the gears was generally greater for smaller individuals. The large variability in the empirical estimates of the relative efficiency at size for each paired tow is reflected in the variation in the posterior smooth estimates of relative efficiency at size for each paired tow.

All 1000 bootstrap fits of the paired tow data converged with invertible hessians at the optimized log-likelihood and provided estimates of relative catch efficiency at size for summer, windowpane, and yellowtail flounder, and red hake, goosefish, and thorny skate. All but 2 of the bootstraps for winter flounder and 3 for barndoor skate provided estimates of relative catch efficiency. For witch flounder, 817 bootstraps provided estimates and only 386 provided estimates for American plaice. One bootstrap fit for summer flounder was excluded due to an extremely high relative efficiency of the chain sweep gear which impeded estimation of standard errors from the bootstrap fits.

Generally, where data are prevalent the bootstrap and hessian-based confidence intervals are similar across all species. However, sometimes substantially different perceptions of confidence ranges exist at the extremes of the length range for particular species where there are fewer data and asymptotic properties of estimators can be less applicable.

3.3 Biomass estimation

Total biomass estimates calibrated to the chain sweep gear were variable across years for most stocks and without strong trend (Figure 5). However, declining trends exist for the Georges Bank and southern New England-Mid-Atlantic yellowtail flounder stocks and an increasing trend for northern goosefish. Biomass estimates were greatest on average for northern red hake and least for Gulf of Maine winter flounder, although this biomass estimate excludes fish less than 30 cm in length. Fall and spring biomass estimates were similar in scale for most stocks, except that southern New England winter flounder and northern goosefish estimates were typically greater in the fall than the spring.

The relative catch efficiency of the rockhopper and chan sweep gears in terms of biomass varies across survey years and seasons due primarily to differences in size composition, but also variation in estimated length-weight relationship parameters (Figure 6). The efficiency of the bottom trawl survey gear was greatest for the winter flounder stocks and American plaice (0.6 to 0.9) and least for red hake, witch flounder, windowpane, and yellowtail flounder stocks (0.2 to 0.4). Precision of the estimated annual biomass efficiencies was lowest for Georges Bank winter flounder and the skate stocks. For Gulf of Maine winter flounder, southern red hake, and barndoor skate, the average fall biomass efficiencies were typically greater than in the spring although the differences were small relative to the confidence intervals.

Comparing the average of estimated coefficients of variation for annual calibrated and uncalibrated biomass estimates showed large increases for summer flounder in the fall (> 50%), southern New England winter flounder in the spring (77%), Georges Bank winter flounder (more than 200% for spring and fall), northern red hake (95% for spring and 178% for fall), northern goosefish in the fall (93%), and barndoor skate (> 100% for both spring and fall) induced by the variability in the estimation of the relative catch efficiency of the gears using chain and rockhopper sweep gears (Table 6). The effect of calibration on the precision of the biomass estimates was relatively minor for other stocks.

We observed little correlation of annual biomass estimates induced by the relative catch efficiency estimation for most of the stocks (Table 7). However, the biomass estimates were highly correlated for George Bank winter flounder in the spring (65%) and barndoor skate (> 70% on average). Estimates for Georges Bank winter flounder in the fall, both red hake stocks, northern goosefish, and thorny skate were greater than 20% on average.

4 Discussion

The data that we used to estimate bottom trawl survey catch efficiency came from an experiment using a twin trawler and many of the standard tow protocols for the NEFSC survey on the *Bigelow*. The experimental net used on one side of the twin trawl was the same as the standard survey trawl used on the *Bigelow* except that it contained roughly half the number of floats and the sweep was modified to optimize flatfish catch by minimizing the ability of flatfish to pass under the net. The other side of the twin trawl was essentially identical to the standard gear used on the *Bigelow*. The towing of the standard survey bottom trawl on the twin trawl experiment differed in a few ways from its deployment on the spring and fall bottom trawl surveys, but we believe that these differences did not have a significant effect on the results. The use of larger doors and the restrictor rope served to fix the net geometries which may be the biggest source of variability in comparative trawl catches (Jones et al., 2021). This setup also allowed us to avoid many of the potential problems due to the large differences in size of the *Bigelow* and the *F/V Karen Elizabeth*. We do not suspect that the use of the restrictor rope would influence flatfish behavior in front of the trawl because flatfish have been shown to generally not react to trawling induced stimuli until they are in very close proximity or even contacted by the fishing gear (Ryer et al., 2010). The spread data indicated that the restrictor rope remained taut throughout the towing process (setting, towing, hauling back), so we believe it likely that the restrictor rope was almost always at least 1m off the bottom. Our concerns about potential effects of the restrictor rope on species that spend more time off the ground (e.g., Atlantic cod, *Gadus morhua*) led us to exclude them from analyses.

Herding is a known phenomenon for flatfish and many other species when certain types of gear are used (Ramm and Xiao, 1995; Somerton and Munro, 2001; Somerton et al., 2007; Rose et al., 2010). Somerton and Munro (2001) considered two factors of bridle herding effects on efficiency. The first factor was the size of the bridle path where the bridle is off the ground (W_{off}) and the second factor, the herding efficiency (h) was the fraction of fish in the bridle contact path moved into the path of the net. The former is a function of gear design, and controllable, whereas the latter is a function of fish behavior with regard to the bridle when it is in contact with the substrate. The bridle configuration on the bottom trawl survey are designed to minimize contact with the bottom and lack of abrasion of painted bridles used during one of the twin trawler research trips provided evidence of little or no bridle contact during the paired tow experiments used to collect the data used in this study. Furthermore, studies have consistently found that herding behavior occurred during the daytime (Glass and Wardle, 1989; Somerton and Munro, 2001; Ryer and Barnett, 2006; Bryan et al., 2014; Ryer et al., 2010; Dean et al., 2021) with some studies indicating high herding coefficients (h) along the sections of the bridles in contact with the bottom. Studies that have evaluated herding at night or in low light conditions did not find evidence for a directional herding response (Glass and Wardle, 1989; Ryer and Barnett, 2006; Ryer, 2008; Ryer et al., 2010). The minimal bridle contact with the substrate and the large fraction of nighttime tows during the bottom trawl survey suggests flatfish herding is unlikely to be an important factor in catch efficiency based on net spread-based swept area.

On the other hand, the biomass estimates assume that the chain sweep gear is fully efficient, but it is likely at least some small fraction of fish, that may depend on size, are not captured by the gear. The biomass estimates also implicitly assume that the entire stock is available to the bottom trawl survey, but many of these stocks extend somewhat outside of the survey strata used to define the indices throughout the year and(or) seasonally due to migration. If either of these assumptions are incorrect this method of biomass estimation would be negatively biased (expected value of biomass estimates would be lower than the true value). However, estimation using the data from these paired-gear studies and these assumptions is less biased than those made without them.

These analyses treat the amount of daylight as binary effect (day/night) on the relative catch efficiency. However, behavior of the fish with respect to the gear is likely to change more gradually with the amount of light. A continuous measure of light that uses the angle of the sun, the depth of the tow and light attenuation with depth, might prove to be a better explanatory variable for changes in relative catch efficiency and perhaps improve estimation of abundance from the bottom trawl survey (Jacobson et al., 2015; Kainge et al., 2017).

Aside from the direct impact of estimated catch efficiency of the NEFSC trawl survey gear on biomass estimation, our analyses show more subtle impacts of using independent estimates of efficiency with survey tow data to generate the abundance indices. Without the independent efficiency estimates, the sampling variability of each of the seasonal and annual relative abundance indices is independent of the others. The bootstrapping methods we employed illustrated that including estimates of catch efficiency affects the variability of the resulting abundance estimates and their independence from each other. For some stocks there was a substantial effect of the relative catch efficiency estimation on the precision of the biomass indices. Similarly, we found high correlation of annual indices (> 0.6) for Georges Bank winter flounder and barndoor skate. In these cases, the decrease in precision and increased correlation may have an impact on bias and precision of important estimates from the assessment model such as stock size and fishing mortality. As such, future work should evaluate the effects of incorporating this information in an assessment model.

The estimates of absolute abundance produced using the sweep comparison experiments have already been informative to assessments and management of many stocks in the Northeast U.S. Chain sweep-based abundance estimates have been used directly in the age-structured assessment model for summer flounder and northern and southern goosefish stocks (NEFSC, 2019, 2020c). Abundance estimates for southern New England-Mid-Atlantic winter flounder, both Cape Cod-Gulf of Maine and southern New England- Mid-Atlantic yellowtail flounder stocks, and American plaice were used to validate the abundance estimates produced by the assessment models (NEFSC, 2020b). The abundance estimates have also been used directly in assessments for witch flounder, Gulf of Maine winter flounder, Georges Bank yellowtail flounder, northern and southern red hake stocks, which are all assessed using simpler index-based assessment methods (Legault and McCurdy, 2017; NEFSC, 2020a,b). These estimates can be especially valuable for index-based methods where the scale of the stock is assumed known. The abundance estimates have also been used in a supporting fashion for fall-back assessments of both Gulf of Maine-Georges Bank and southern New England-Mid-Atlantic windowpane stocks (NEFSC, 2020b).

Typically, research surveys provide only a relative index of abundance rather than an absolute estimate of abundance. Stock assessment models then integrate these observations with time series of catch and other data sources to determine the scale of the population. However, various factors can make for imprecise and inaccurate scaling of population levels including inaccurate catch data (Cadigan, 2016), time-varying catchability (Wilberg et al., 2009), low fishing mortality rates over the time series (Adams et al., 2015), and uncertain and time-varying natural mortality (Stock et al., 2021). In these cases, external information such as those produced by studies such as ours, can be particularly useful in estimating the size of of the stock, the status of the stock relative to optimal levels and ultimately making catch advice for commercially important fish stocks.

Acknowledgements

The quality of this manuscript has been improved by comments from Jessica Blaylock. Tor Bendiksen and John Knight assisted with the design of the chain sweep. Jeff Pessutti, Giovanni Gianesen, Dominique St. Amand, and Calvin Alexander helped collect paired-tow data for these analyses. We thank Adam Poquette for creating the map in Figure 3. We thank Calvin Alexander also for the photo of the twin trawl gear in Figure 2.

CRediT authorship contribution statement

Timothy J. Miller: Conceptualization, Methodology, Writing - original draft, Formal analysis, Visualization. **David E. Richardson:** Conceptualization, Methodology, Writing - original draft. **Philip J. Politis:** Investigation, Methodology, Writing - review and editing. **Christopher D. Roebuck:** Project administration, Conceptualization, Funding acquisition, Investigation, Resources. **John P. Manderson:** Conceptualization, Investigation, Writing - review and editing. **Michael H. Martin:** Project administration, Conceptualization, Data curation, Writing - review and editing. **Andrew W. Jones:** Visualization, Writing - review and editing.

References

Adams, C. F., Miller, T. J., Manderson, J. P., Richardson, D. E., and Smith, B. E. 2015. Butterfish 2014 Assessment. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-06; 110 p.

Arreguín-Sánchez, F. 1996. Catchability: a key parameter for fish stock asseessment. Reviews in Fish Biology and Fisheries **6**: 221–242, doi: 10.1007/BF00182344.

Azarovitz, T. S. 1981. A brief historical review of the Woods Hole Laboratory trawl survey time series. *In* Bottom Trawl Surveys. *Edited by* W. G. Doubleday and D. Rivard. Can. Spec. Pub. Fish. Aqua. Sci. No. 58. pp. 62–67.

Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekinga, A., Kraan, C., van der Veer, H. W., and Piersma, T. 2012. Designing a benthic monitoring programme with multiple conflicting objectives. Methods in Ecology and Evolution **3**(3): 526–536, doi: 10.1111/j.2041-210X.2012.00192.x.

Bourne, N. 1965. A comparison of catches by 3- and 4-inch rings on offshore scallop drags. Journal of the Fisheries Research Board of Canada **22**(2): 313–333.

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. Fisheries Research **154**: 120–126, doi: 10.1016/j.fishres.2014.02.007.

Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including underreported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences **73**(2): 296–308, doi: 10.1139/cjfas-2015-0047.

Dean, M. J., Hoffman, W. S., Buchan, N. C., Cadrin, S. X., and Grabowski, J. H. 2021. The influence of trawl efficiency assumptions on survey-based population metrics. ICES Journal of Marine Science **78**(8): 2858–2874.

Glass, C. W. and Wardle, C. S. 1989. Comparison of the reactions of fish to a trawl gear, at high and low light intensities. Fisheries Research **7**(3): 249–266, doi: 10.1016/0165-7836(89)90059-3.

Gulland, J. A. 1964. Variations in selection factors, and mesh differentials. Journal Du Conseil International Pour L'exploration De La Mer **29**(2): 158–165.

ICES. 1996. Manual of methods of measuring the selectivity of towed fishing gears. (Eds.) Wileman, D. A., Ferro, R. S. T., Fonteyne, R., and Millar, R. B. ICES Cooperative Research Report No. 215.

Jacobson, L. D., Hendrickson, L. C., and Tang, J. 2015. Solar zenith angles for biological research and an expected catch model for diel vertical migration patterns that affect stock size estimates for longfin inshore squid (*Doryteuthis pealeii*). Canadian Journal of Fisheries and Aquatic Sciences **72**(9): 1329–1338, doi: 10.1139/cjfas-2014-0436.

Jones, A. W., Miller, T. J., Politis, P. J., Richardson, D. E., Mercer, A. M., Pol, M. V., and Roebuck, C. D. 2021. Experimental assessment of the effect of net wing spread on relative catch efficiency of four flatfishes by a four seam bottom trawl. Fisheries Research **244**: 106106, doi: 10.1016/j.fishres.2021.106106.

Kainge, P., van der Plas, A. K., Bartholomae, C. H., and Wieland, K. 2017. Effects of environmental variables on survey catch rates and distribution by size of shallow- and deep-water cape hakes, merluccius capensis and merluccius paradoxus off namibia. Fisheries Oceanography **26**(6): 680–692, doi: 10.1111/fog.12227.

Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2013. Combining bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl efficiency parameters for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences **70**(2): 208–219, doi: 10.1139/cjfas-2012-0321.

Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. 2017. Selectivity ratio: A useful tool for comparing size selectivity of multiple survey gears. Fisheries Research **191**: 76–86, doi: 10.1016/j.fishres.2017.02.012.

Kotwicki, S., Martin, M. H., and Laman, E. A. 2011. Improving area swept estimates from bottom trawl surveys. Fisheries Research **110**(1): 198–206, doi: 10.1016/j.fishres.2011.04.007.

Krag, L. A., Herrmann, B., Karlsen, J. D., and Mieske, B. 2015. Species selectivity in different sized topless trawl designs: Does size matter? Fisheries Research **172**: 243–249, doi: 10.1016/j.fishres.2015.07.010.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software **70**(5): 1–21.

Legault, C. M. and McCurdy, Q. M. 2017. Stock assessment of Georges Bank yellowtail flounder for 2017. Transboundary Resources Assessment Committee (TRAC) Reference Document 2017/03 available at: https://repository.library.noaa.gov/view/noaa/27556.

Maunder, M. N. and Piner, K. R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science **72**(1): 7–18, doi: 10.1093/icesjms/fsu015.

Millar, R. B. and Fryer, R. J. 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. Reviews in Fish Biology and Fisheries **9**(1): 89–116, doi: 10.1023/A:1008838220001.

Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for U.S. Northwest Atlantic fish stocks. Canadian Journal of Fisheries and Aquatic Sciences **70**(9): 1306–1316, doi: 10.1139/cjfas-2013-0136.

Miller, T. J., Hart, D. R., Hopkins, K., Vine, N. H., Taylor, R., York, A. D., and Gallager, S. M. 2019. Estimation of the capture efficiency and abundance of Atlantic sea scallops (*Placopecten magellanicus*) from paired photographic–dredge tows using hierarchical models. Canadian Journal of Fisheries and Aquatic Sciences **76**(6): 847–855, doi: 10.1139/cjfas-2018-0024.

Munro, P. T. and Somerton, D. A. 2001. Maximum likelihood and non-parametric methods for estimating trawl footrope selectivity. ICES Journal of Marine Science **58**(1): 220–229, doi: 10.1006/jmsc.2000.1004.

NEFSC. 2019. 66th Northeast Regional Stock Assessment Workshop (66th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 19-08; 1170 p. Available from: http://www.nefsc.noaa.gov/publications/.

NEFSC. 2020a. Final report of red hake stock structure working group. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 20-07; 185 p.

NEFSC. 2020b. Operational assessment of 14 northeast groundfish stocks, updated through 2018. Prepublication Copy of the September 2019 Operational Stock Assessment Report. The report is "in preparation" for publication by the NEFSC. (Oct. 3, 2019; latest revision: Jan. 7, 2020).

NEFSC. 2020c. Operational assessment of the black sea bass, scup, bluefish, and monkfish stocks, updated through 2018. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 20-01; 160 p.

Nichol, D. G., Kotwicki, S., Wilderbuer, T. K., Lauth, R. R., and Ianelli, J. N. 2019. Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey and its effect on estimates of survey biomass. Fisheries Research **211**: 319–330, doi: 10.1016/j.fishres.2018.11.017.

Paloheimo, J. E. and Dickie, L. M. 1964. Abundance and fishing success. Rapports et Procès-Verbaux des Réunions. Conseil Internationale pour l'Exploration de la Mer **155**: 152–163.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Canadian Journal of Fisheries and Aquatic Sciences **55**(12): 2672–2690, doi: 10.1139/f98-151.

Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. 2014. Northeast Fisheries Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. U.S. Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 14-06, 138p.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramm, D. C. and Xiao, Y. 1995. Herding in groundfish and effective pathwidth of trawls. Fisheries Research **24**(3): 243–259, doi: 10.1016/0165-7836(95)00373-I.

Rose, C. S., Gauvin, J. R., and Hammond, C. F. 2010. Effective herding of flatfish by cables with minimal seafloor contact. Fishery Bulletin **108**(2): 136–144.

Ryer, C. H. 2008. A review of flatfish behavior relative to trawls. Fisheries Research **90**(1): 138–146, doi: 10.1016/j.fishres.2007.10.005.

Ryer, C. H. and Barnett, L. A. K. 2006. Influence of illumination and temperature upon flatfish reactivity and herding behavior: Potential implications for trawl capture efficiency. Fisheries Research **81**(2): 242–250, doi: 10.1016/j.fishres.2006.07.001.

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 elevated sweeps. Fishery Bulletin **108**(2): 145–154.

Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. Canadian Journal of Fisheries and Aquatic Sciences **54**(3): 616–630, doi: 10.1139/f96-303.

Somerton, D., Ianelli, J., Walsh, S., Smith, S., Godø, O. R., and Ramm, D. 1999. Incorporating experimentally derived estimates of survey trawl efficiency into the stock assessment process: a discussion. ICES Journal of Marine Science **56**(3): 299–302, doi: 10.1006/jmsc.1999.0443.

Somerton, D. A. and Munro, P. 2001. Bridle efficiency of a survey trawl for flatfish. Fishery Bulletin **99**(4): 641–652.

Somerton, D. A., Munro, P. T., and Weinberg, K. L. 2007. Whole-gear efficiency of a benthic survey trawl for flatfish. Fishery Bulletin **105**(2): 278–291.

Somerton, D. A., Weinberg, K. L., and Goodman, S. E. 2013. Catchability of snow crab (*Chionoecetes opilio*) by the eastern Bering Sea bottom trawl survey estimated using a catch comparison experiment. Canadian Journal of Fisheries and Aquatic Sciences **70**(12): 1699–1708, doi: 10.1139/cjfas-2013-0100.

Stock, B. C., Xu, H., Miller, T. J., Thorson, J. T., and Nye, J. A. 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for Southern New England-Mid Atlantic yellowtail flounder. Fisheries Research **237**: 105873, doi: 10.1016/j.fishres.2021.105873.

Thiess, M. E., Benoit, H., Clark, D. S., Fong, K., Mello, L. G. S., Mowbray, F., Pepin, P., Cadigan, N. G., Miller, T. J., Thirkell, D., and Wheeland, L. 2018. Proceedings of the National Comparative Trawl Workshop, November 28-30, 2017, Nanaimo, B. C. Canadian Technical Report of Fisheries and Aquatic Sciences 3254, x + 40 p.

Thygesen, U. H., Kristensen, K., Jansen, T., and Beyer, J. E. 2019. Intercalibration of survey methods using paired fishing operations and log-Gaussian Cox processes. ICES Journal of Marine Science **76**(4): 1189–1199, doi: 10.1093/icesjms/fsy191.

Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. 2018. Evaluation of alternative stratifications for a stratified random fishery-independent survey. Fisheries Research **207**: 150–159, doi: 10.1016/j.fishres.2018.06.019.

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2009. Incorporating time-varying catchability into population dynamic stock assessment models. Reviews in Fisheries Science **18**(1): 7–24, doi: 10.1080/10641260903294647.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall, Boca Raton, Florida, 392 pp.

Fig. 1. Diagram of twin-trawl gear configuration. One of the two nets is rigged with a rockhopper sweep (8) and the other is rigged with a chain sweep (7) and for both a restrictor rope (5) is used to obtain consistent net spread. The other importand components are the side wires (1), middle wire (2), doors (3), the clump weight (4), and the acoustic mensuration system (6). The side where the rockhopper and chainsweep gears gears were deployed varied throughout the experimental tows of each.

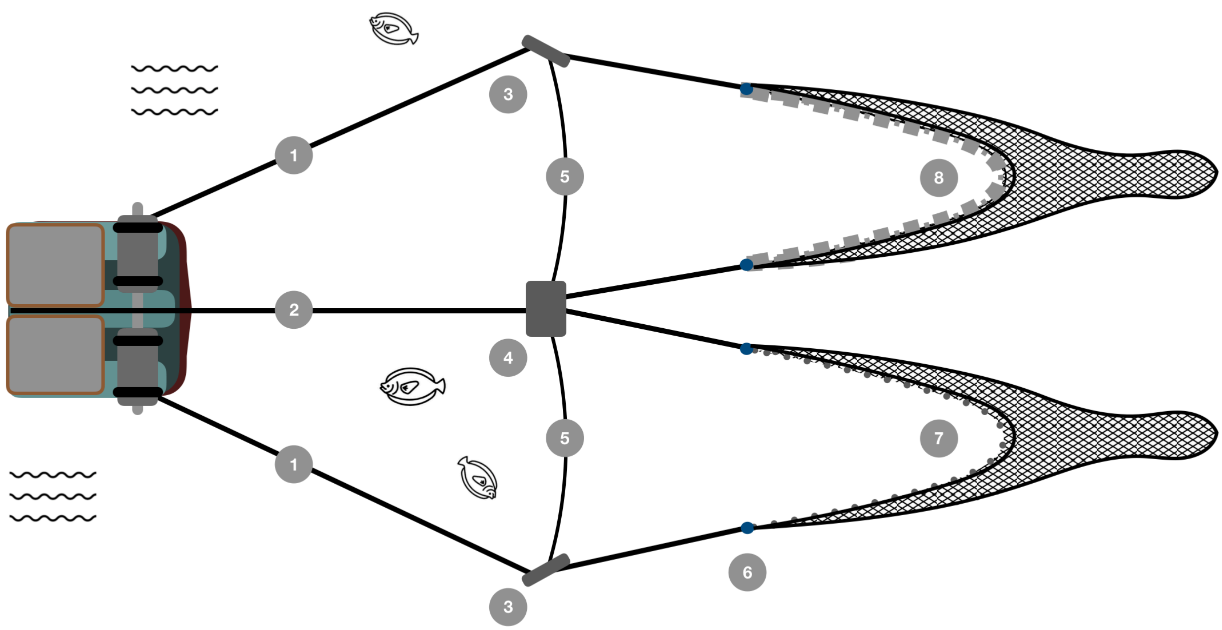

Fig. 2. The *F/V Karen Elizabeth* twin-trawl vessel rigged with rockhopper sweep gear on the right and chain sweep gear on the left.

Fig. 3. Annual locations of stations where the F/V Karen Elizabeth conducted twin-trawl sets with the standard bottom trawl gear and the gear with a chain sweep instead of the rockhopper sweep.

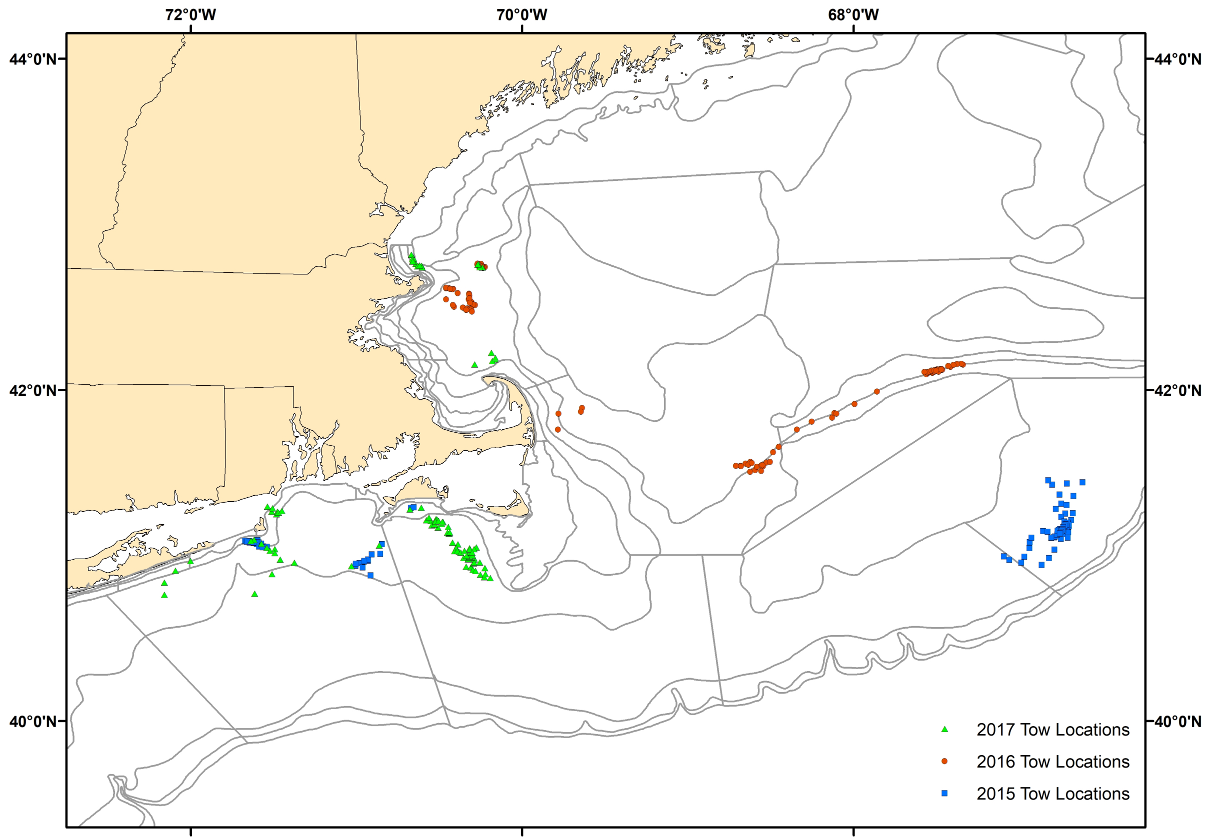

Fig. 4. Relative efficiency of gears using chain and rockhopper sweeps from the best performing model for each species (Table 5). Blue and red denote results for day and night data, respectively, and thick and thin lines represent overall and paired-tow specific estimates of relative catch efficiency, respectively. There was no diel effect in the best model for American plaice. Points represent empirical estimates of relative efficiency for paired observation by length and paired tow. Polygons and dashed lines represent hessian-based and bootstrap-based 95% confidence intervals, respectively.

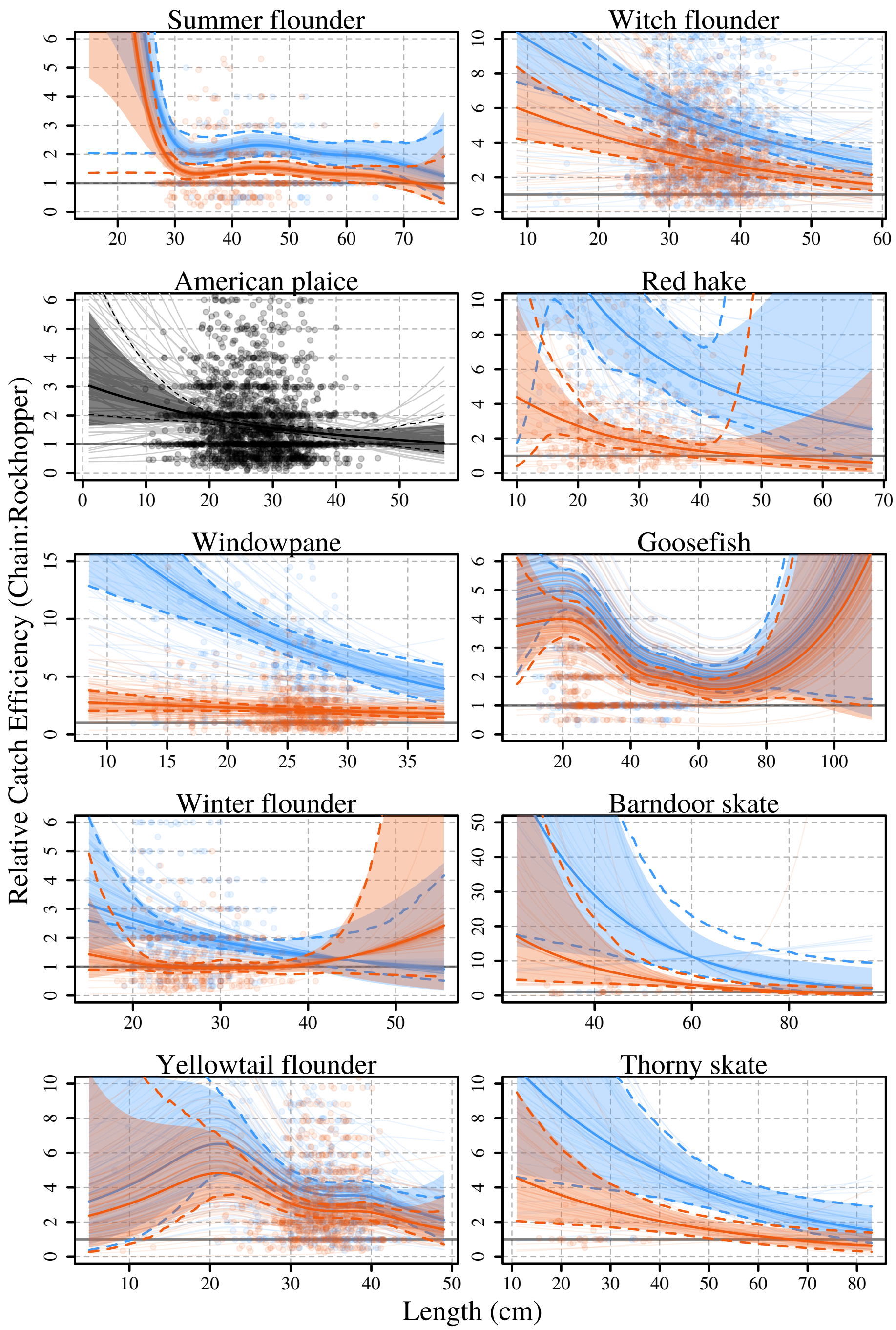

Fig. 5. Annual spring (blue) and fall (red) biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps are from the best performing model for each species (Table 5).

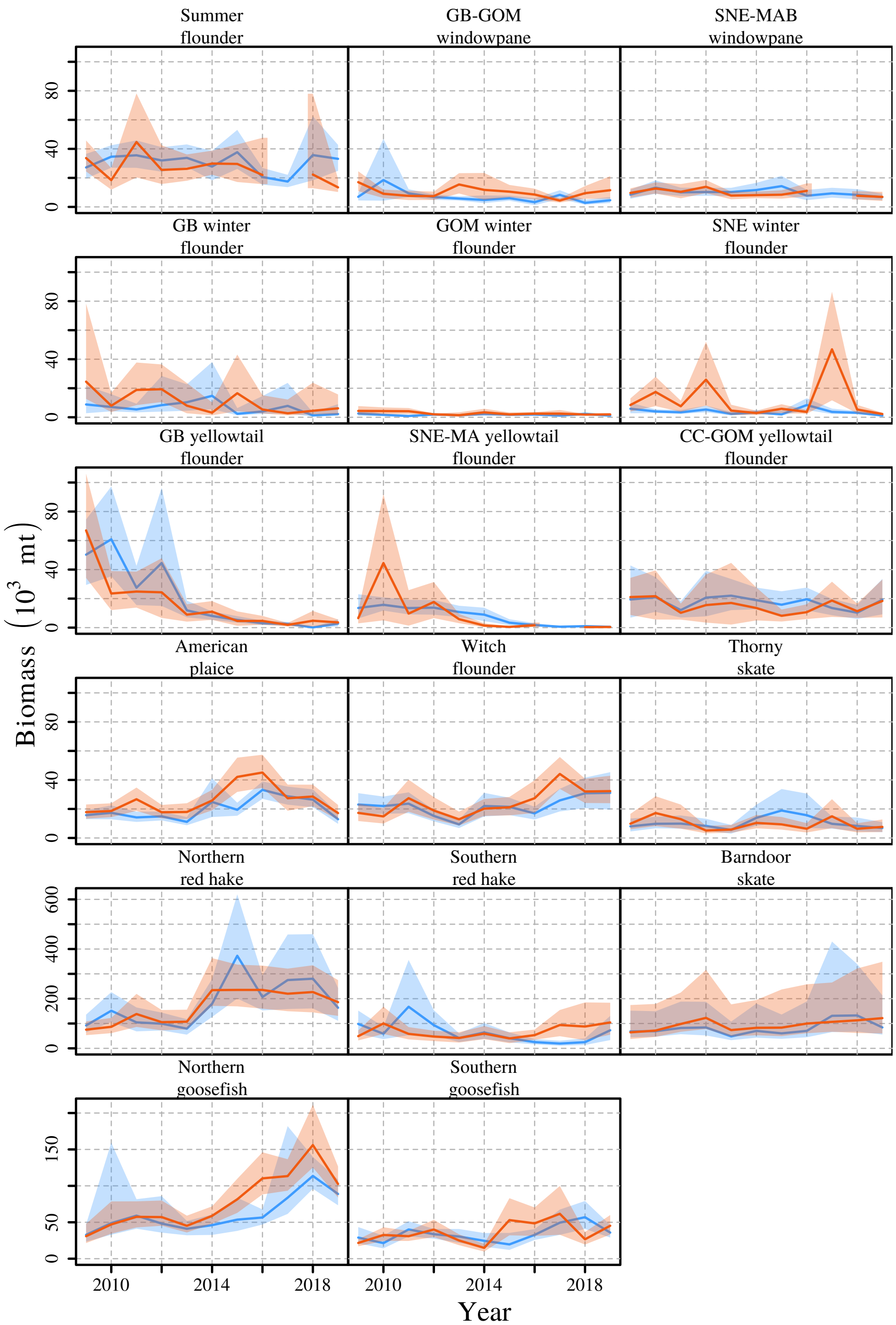

Fig. 6. Implied catch efficiency of annual spring (blue) and fall (red) bottom trawl survey biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps are from the best performing model for each species (Table 5).

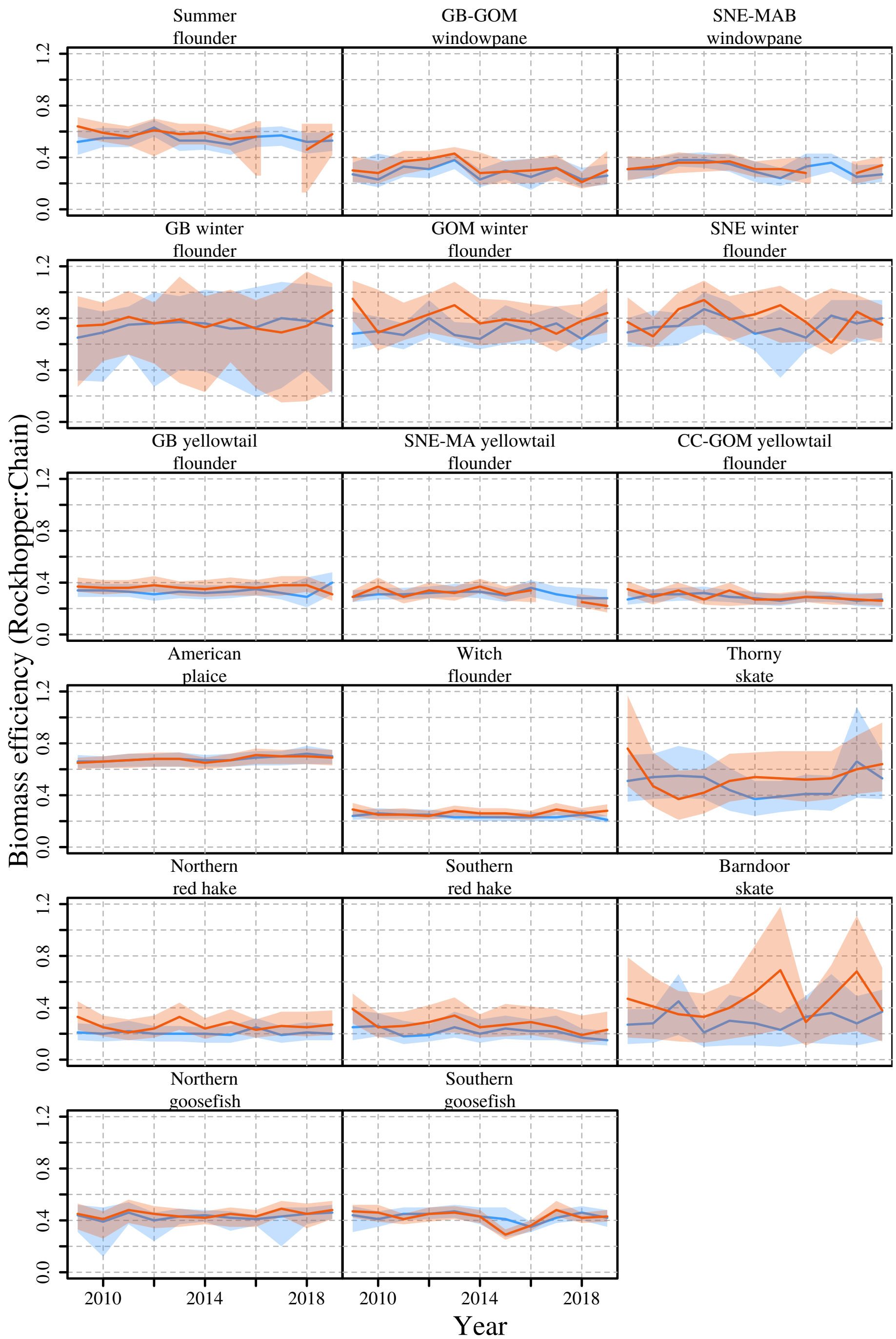

Table 1. Managed stocks associated with the species for which relative catch efficiency was estimated.

Stock
Summer flounder
American Plaice
Georges Bank-Gulf of Maine (GB-GOM) windowpane
Southern New England-Mid-Atlantic Bight (SNE-MAB) windowpane
Georges Bank (GB) winter flounder
Gulf of Maine (GOM) winter flounder
Southern New England (SNE) winter flounder
GB yellowtail flounder
Southern New England-Mid-Atlantic (SNE-MA) yellowtail flounder
Cape Cod-Gulf of Maine (CC-GOM) yellowtail flounder
Witch flounder
Northern red hake
Southern red hake
Northern goosefish
Southern goosefish
Barndoor skate
Thorny skate

Table 2. Description of relative catch efficiency (ρ) and beta-binomial dispersion (ϕ) parameterizations for binomial and beta-binomial models and number of marginal likelihood parameters (n_p) for the 13 base models from Miller (2013) and fit to paired chain sweep and rockhoppersweep tow data for each species.

Model	$\log(\rho)$	$\log(\phi)$	n_p	Description
BI_0	~ 1	–	1	population-level mean for all observations
BI_1	$\sim 1 + 1\|\text{pair}$	–	2	population- and random station-level ρ
BI_2	$\sim$ s(length)	–	3	population-level smooth size effect on ρ
BI_3	$\sim$ s(length) + 1|pair	–	4	population-level smooth size effect and random station-level intercept for ρ
BI_4	$\sim$ s(length) + s(length)|pair	–	7	population-level and random station-level smooth size effects for ρ
BB_0	~ 1	~ 1	2	population-level ρ and ϕ
BB_1	$\sim 1 + 1\|\text{pair}$	~ 1	3	population-level and random station-level intercept for ρ and population-level ϕ
BB_2	$\sim$ s(length)	~ 1	4	population-level smooth size effect on ρ and population-level ϕ
BB_3	$\sim$ s(length)	$\sim$ s(length)	6	population-level smooth size effect on ρ and ϕ
BB_4	$\sim$ s(length) + 1|pair	~ 1	5	population-level smooth size effect and random station-level intercept for ρ and population-level ϕ
BB_5	$\sim$ s(length) + 1|pair	$\sim$ s(length)	7	population-level smooth size effect on ρ and ϕ and random station-level intercepts for ρ
BB_6	$\sim$ s(length) + s(length)|pair	~ 1	8	population-level and random station-level smooth size effects on ρ and population-level ϕ
BB_7	$\sim$ s(length) + s(length)|pair	$\sim$ s(length)	10	population-level and random station-level smooth size effects on ρ and population-level smooth size effects on ϕ

Table 3. Number of paired tows where fish were captured and the number of fish captured and measured for lengths for each species in total and by day or night.

Species	Paired Tows			Captured	Both Gears Measured			Chainsweep Measured			Rockhopper Measured		
	Total	Day	Night	Total	Total	Day	Night	Total	Day	Night	Total	Day	Night
Summer flounder	141	75	66	4,154	4,154	1,770	2,384	2,616	1,195	1,421	1,538	575	963
American plaice	134	84	50	31,983	19,245	13,619	5,626	10,982	7,775	3,207	8,263	5,844	2,419
Windowpane	195	100	95	15,310	13,014	6,221	6,793	9,854	5,443	4,411	3,160	778	2,382
Winter flounder	171	97	74	6,586	6,449	3,605	2,844	3,805	2,385	1,420	2,644	1,220	1,424
Yellowtail flounder	192	101	91	18,545	14,134	6,849	7,285	10,065	5,297	4,768	4,069	1,552	2,517
Witch flounder	132	83	49	57,133	23,927	13,899	10,028	14,899	9,271	5,628	9,028	4,628	4,400
Red hake	73	40	33	47,275	12,585	6,614	5,971	8,587	4,908	3,679	3,998	1,706	2,292
Goosefish	302	165	137	8,798	8,541	3,985	4,556	6,409	3,053	3,356	2,132	932	1,200
Barndoor skate	62	33	29	502	502	219	283	397	198	199	105	21	84
Thorny skate	90	56	34	907	907	399	508	648	311	337	259	88	171

Table 4. Difference in AIC for each of the 13 models described in Table 2 from the best model (**0**) by species.

	BI_0	BI_1	BI_2	BI_3	BI_4	BB_0	BB_1	BB_2	BB_3	BB_4	BB_5	BB_6	BB_7
Summer flounder	27.96	13.53	8.9	**0**		28.64	15.45	10.59					
American plaice	821.11	546.54	743.34	494.92	415.63	179.48	71.76	141.44		37.06		0.71	**0**
Windowpane	1045.06	38.51	1029.72	17.03	**0**	585.7	32.22	572.73		15.27			
Winter flounder	216.47	15.73	200.33	3.02	**0**	163.31	16.63	151.66	151.01	4.21	6.78	1.41	
Yellowtail flounder	727.15	97.93	727.36	51.84	10.96	394.94	70.2	391.13	371.13	31.85		**0**	3.33
Witch flounder	1424.17	212.64	1372.66		35.33	881.28	142.53	844.47		81.37		**0**	
Red hake	1884.51	295.85		170.75		627.33	166.43	590.92		95.8	59.31	**0**	0.83
Goosefish	227.67	87.23	80.37	**0**		219.13		76.54					
Barndoor skate	36.51	10.01	31.34	2.72	**0**	36.23	11.99	29.03		4.6			
Thorny skate	39.04	8.57	32.65	3.44	1.15	22.38	5.84	18.66		1.38	5.19	**0**	

Table 5. Best performing models from Table 4 and extended models that include diel effects on relative catch efficiency for each species with the number of parameters for each model (n_p) and the differences in AIC (ΔAIC) from the best of the three models (**0**) by species.

	Model	$\log(\rho)$	$\log(\phi)$	n_p	ΔAIC
Summer flounder					
	BI_3	~ s(length) + 1|pair	–	4	22.92
	BI_{3a}	~ dn + s(length) + 1|pair	–	5	**0**
	BI_{3b}	~ dn * s(length) + 1|pair	–	7	1.74
American plaice					
	BB_7	~ s(length) + s(length)|pair	~ s(length)	10	**0**
	BB_{7a}	~ dn + s(length) + s(length)|pair	~ s(length)	11	1.43
	BB_{7b}	~ dn * s(length) + s(length)|pair	~ s(length)	13	2.95
Windowpane					
	BI_4	~ s(length) + s(length)|pair	–	7	152.1
	BI_{4a}	~ dn + length + s(length)|pair	–	7	4.06
	BI_{4b}	~ dn * length + s(length)|pair	–	8	**0**
Winter flounder					
	BI_4	~ s(length) + s(length)|pair	–	7	50.68
	BI_{4a}	~ dn + s(length) + length|pair	–	7	0.3
	BI_{4b}	~ dn * s(length) + length|pair	–	9	**0**
Yellowtail flounder					
	BB_6	~ s(length) + s(length)|pair	~ 1	8	3.84
	BB_{6a}	~ dn + s(length) + s(length)|pair	~ 1	9	**0**
	BB_{6b}	~ dn * s(length) + s(length)|pair	~ 1	11	3.48
Witch flounder					
	BB_6	~ s(length) + s(length)|pair	~ 1	8	19.68
	BB_{6a}	~ dn + length + s(length)|pair	~ 1	8	**0**
	BB_{6b}	~ dn * length + s(length)|pair	~ 1	9	1.52
Red hake					
	BB_6	~ s(length) + s(length)|pair	~ 1	8	32.35
	BB_{6a}	~ dn + s(length) + s(length)|pair	~ 1	8	**0**
	BB_{6b}	~ dn * s(length) + s(length)|pair	~ 1	10	3.18
Goosefish					
	BI_3	~ s(length) + 1|pair	–	4	5.44
	BI_{3a}	~ dn + s(length) + 1|pair	–	5	**0**
	BI_{3b}	~ dn * s(length) + 1|pair	–	7	6.8
Barndoor skate					
	BI_4	~ s(length) + s(length)|pair	–	7	15.57
	BI_{4a}	~ dn + length + length|pair	–	5	**0**
	BI_{4b}	~ dn * length + length|pair	–	6	1.83
Thorny skate					
	BB_6	~ s(length) + s(length)|pair	~ 1	8	15.51
	BB_{6a}	~ dn + length + length|pair	~ 1	7	**0**
	BB_{6b}	~ dn * length + length|pair	~ 1	8	1.38

Table 6. Average of annual (2009-2019) ratios of coefficients of variation for calibrated and uncalibrated biomass indices for each stock by seasonal survey. Coefficients of variation are based on bootstrap resampling of paired tow observations, survey station data and associated length and weight observations. Annual indices for fall 2017 were not available for summer flounder, SNE-MA windowpane, and SNE-MA yellowtail flounder.

Stock	Average CV Ratio Calibrated:Uncalibrated	
	Spring	Fall
Summer flounder	1.13	1.51
American plaice	1.07	1.02
GB-GOM windowpane	1.03	1.07
SNE-MAB windowpane	1.06	0.90
GB winter flounder	3.19	3.89
GOM winter flounder	1.05	1.07
SNE winter flounder	1.77	0.99
GB yellowtail flounder	1.06	0.98
SNE-MA yellowtail flounder	1.05	0.99
CC-GOM yellowtail flounder	1.01	1.02
Witch flounder	1.12	1.11
Northern red hake	1.95	2.78
Southern red hake	1.28	1.28
Northern goosefish	1.93	1.34
Southern goosefish	1.18	1.04
Barndoor skate	2.47	2.78
Thorny skate	1.14	1.20

Table 7. Average correlation of annual (2009-2019) calibrated biomass indices for each stock by seasonal survey. Annual indices for fall 2017 were not available for SNE-MA windowpane and SNE-MA yellowtail flounder.

Stock	Spring	Fall
Summer flounder	0.16	0.14
American plaice	0.09	0.06
GB-GOM windowpane	0.06	0.04
SNE-MAB windowpane	0.06	0.05
GB winter flounder	0.65	0.45
GOM winter flounder	0.05	0.05
SNE winter flounder	0.07	0.03
GB yellowtail flounder	0.05	0.04
SNE-MA yellowtail flounder	0.07	0.02
CC-GOM yellowtail flounder	0.05	0.04
Witch flounder	0.10	0.10
Northern red hake	0.42	0.34
Southern red hake	0.25	0.21
Northern goosefish	0.21	0.30
Southern goosefish	0.10	0.07
Barndoor skate	0.74	0.81
Thorny skate	0.29	0.25