Chainsweep-based efficiency of bottom trawl surveys and biomass estimates for flatfish, red hake, and goosefish stocks in Northwest Atlantic waters of the United States

Timothy J. Miller1, David E. Richardson2, Andrew Jones2, Phil Politis1

[timothy.j.miller@noaa.gov](mailto:timothy.j.miller@noaa.gov), Northeast Fisheries Science Center, National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543, USA  
Northeast Fisheries Science Center, National Marine Fisheries Service, Narragansett, RI USA

## Abstract

Using a general hierarchical model we estimated relative efficiency of chain sweep to the rockhopper sweep used by the NEFSC bottom trawl survey for from studies carried out between 2015 and 2017 aboard the F/V Karen Elizabeth twin-trawl vessel. Aside from the sweeps, the rest of the trawl gear is the same. We compared a set of models with different 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.

Using a general hierarchical model we estimated relative efficiency of chain sweep to the rockhopper sweep used by the NEFSC bottom trawl survey for winter and windowpane flounder stocks and red hake stocks from studies carried out between 2015 and 2017 aboard the F/V Karen Elizabeth twin-trawl vessel. Aside from the sweeps, the rest of the trawl gear is the same. We compared a set of models with different 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. Diel effects provided improved model performance for all three species. We used the best performing model to make annual chain sweep-based swept area biomass and abundance-at-length estimates. We estimated uncertainty in all results using bootstrap procedures for each data component.

### Keywords

hieararchical models, spline regression, gear efficiency, abundance estimation

# 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 . 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 . 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 fisheries independent bottom trawl surveys are scaled to population size by the survey catchability () parameter . 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 is difficult to estimate . In such cases, estimates of survey catchability from auxilliary data can inform the stock assessment. These external estimates can be used as a direct input into the assessment model , can serve as a diagnostic measure of model accuracy , or contribute to an alternate means of providing catch advice when an assessment model is not considered acceptable .

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 . Here efficiency is the fraction of available fish retained by the gear, equivalent to availability-selection in . 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 , species-specific gear efficiency , and availability of the stock to the survey design frame improve our understanding of catchability and therefore abundance of fish stocks.

Paired-gear studies where two gear 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 . Of the two gears, one is often a reference gear that may be a gear currently used for annual surveys . Typically neither of the gears is fully efficient and therefore the relative efficiency of gears is estimated , but there are cases where one of the gears is assumed to be at least very nearly fully efficient .

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 . These studies are also helpful for combining surveys conducted close together in space or time using alternative gears .

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 a number of groundfish limiting fishing opportunities. To help provide clarity on the trawl operations and build trust in survey indies 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 (hereafter referred to as chain sweep experiments) 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 captured. The twin-trawl rigging where two trawls can be fished simultaneously approaches this ideal , 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 the same except with a chain sweep in an attempt to eliminate any escapement of fish under the gear.

The analytical methods to estimate the efficiency of the bottom trawl gear are based on those used by to estimate size effects on relative catch efficiency of the *Henry B. Bigelow* to the *Albatrosss IV* for a variety of commercially important species, but 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 chainsweep and rockhopper sweep gears. We apply the estimated efficiency of the rockhopper gear to survey data to estimate spring and fall abundance indices from 2009-2019 for 17 commercially important fish stocks in the Northeast US (Table ).

Often overlooked aspects of the application of relative catch efficiency estimates is 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 sampling variability of the annual indices is typically assumed to be independent. Here we compare the precision of the calibrated and uncalibrated indices and measure the correlation of calibrated indices for each stock.

# Methods

## 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. However, red hake were only observed during the 2017 field experiments. 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 (Figure ). The other side of the twin-trawl rig towed a version the NEFSC 400 x 12cm survey bottom trawl modified to maximize the capture of flatfish. The trawl was modified by 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 (in) trawl chain covered by 12.7cm diameter x 1cm thick rubber discs on every other chain link (Figure ). Two rows of 1.3cm (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 (in) buoyant, Polytron rope, were attached between each of the trawl doors and the center clump. 3.4m 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. In 2015, 108 (45 day, 63 night) paired tows were conducted in eastern Georges Bank and off of southern New England (Figure ). In 2016, 117 (74 day, 43 night) paired tows were conducted in western Gulf of Maine and 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.

## Paired-tow analysis

We employed the hierarchical modeling approach from to estimate the relative efficiency of chain sweep to the rockhopper sweep used by the NEFSC bottom trawl survey for six species from three studies carried out aboard a twin trawl vessel. We first fit and compared the same set of 13 models as 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 to BI) and beta-binomial (BB to BB) models that were fitted for all species are described in Table including pseudo-formulas analogous to those used to specify and fit mixed or generalized additive models in R . 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 now allow multiple (cubic regression spline) smooth effects (differing by day or night) on relative catch efficiency. We implemented the models using the Template Model Builder package in R and we used the “nlminb” optimizer to fit the models by maximizing the laplace approximation of the marginal likelihood .

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

## Length-weight analysis

We fit length-weight relationships to the length and weight observations for each survey each year. We assumed weight observation from survey , was log-normal distributed, We used a bias correction to ensure the expected weight . We estimated parameters by maximizing the model likelihood programmed in TMB [@kristensenetal16] and R [@R19] and generated predictions of weight at length Like the relative catch efficiency, bootstrap predictions of weight at length were made 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.

## Biomass estimation

For the 17 managed stocks in the Northeast US that are populations of the species where we have estimated relative efficiency, we estimated stock biomass for each spring and fall annual survey assuming 100% efficiency of the chainsweep gear by scaling the survey tow observations by the relative efficiency of the chainsweep and rockhopper sweep gears. There are single unit stocks for summer and witch flounders, American plaice, and barndoor and thorny skates, but there are three stocks of winter and yellowtail flounders, and two stocks of windowpane, red hake, and goosefish (Table ). First the tow-specific catches at length are rescaled, where is the number at length in tow from stratum and is the relative efficiency of the chain sweep to rockhopper sweep at length estimated from the twin trawl observations, that may depend on the diel characteristic of tow 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 survey.

The stratified abundance estimate is then calculated using the design-based estimator, where is the area of stratum , is the average swept area of a survey station tow, and is the number of tows that were made in stratum . The corresponding biomass estimate is then where is the predicted weight at length (Eq. ) from fitting length-weight observations described above. Length is typically measured to the nearest cm so indicates the number of 1 cm length categories that were 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 abundance 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. 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 due vessel mechanical issues 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 [@smith97]. For each of the 1000 combined bootstraps, survey observations for bootstrap were scaled with the corresponding bootstrap estimates of relative chain to rockhopper sweep efficiency and predicted weight at length, using Eqs. and .

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 uses the resampled survey data as the calibrated biomass estimate except that the bootstrap for the relative catch efficiency is not used (i.e., in Eq. ). 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 from either the spring or fall survey was calculated as

where

and is the number of bootstraps.

For summmer flounder it was necessary to omit one of the 1000 bootstraps of relative catch efficiency at length due to an extremely large value which the standard devation and mean of the bootstraps was sensitive to. Finally, we calculated correlation of annual biomass estimates for years and using the bootstrap estimates of biomass

where the covariance is

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

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

# Results

## Paired-tow observations

In terms on paired tows and total numbers of fish, flatfish were the most well sampled species, but goosefish was observed in the most paired-tows and red hake was the most prevalent in terms of total numbers caught (Table ). Witch flounder was the most prevalent flatfish species caught while yellowtail flounder was observed in the the most frequently observed flatfish in terms of paired tows. For all but summer flounder, and barndoor and thorny skates, only a subsample of all of the fish that were caught were measured for length, but nearly all caught winter flounder and goosefish were measured.

## Relative model performance

As measured by AIC, the best performing model 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 ). 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) and smooth effects of size on the beta-binomial dispersion parameter (BB,BB, and BB).

Including diel effects on relative catch efficiency improved model performance for all species except American plaice (Table ).

Higher relative catch efficiency was always during the daytime when fish typically associate with the substrate more.

winter flounder before considering day/night effects was the conditional binomial model BI (Table ). Allowing smooth size-effects on relative catch efficiency and variation in these effects among paired-tows provided primary improvements in model performance. Including diel effects on relative efficiency for the twin-trawl observations improved performance of the binomial model (BI),

The relative efficiency of the chain sweep gear to the rockhopper sweep gear is greatest at the smallest sizes of winter flounder, but is fairly constant over over sizes greater than 25 cm. The minimum relative efficency is between 1.5 and 2 during the day, but efficiencies of the two sweeps are approximately equal efficiency at night (Figure ).

## Bootstrap-based uncertainty

All 1000 bootstrap fits of the paired tow data 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 chainsweep gear which impeded estimation of standard errors from the bootstrap fits.

We see that 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.

*Winter flounder*

As measured by AIC, the best performing model for winter flounder before considering day/night effects was the conditional binomial model BI (Table ). Allowing smooth size-effects on relative catch efficiency and variation in these effects among paired-tows provided primary improvements in model performance. Including diel effects on relative efficiency for the twin-trawl observations improved performance of the binomial model (BI), however the model allowing the size effects on relative efficiency to differ between day and night (BI) would not converge. The relative efficiency of the chain sweep gear to the rockhopper sweep gear is greatest at the smallest sizes of winter flounder, but is fairly constant over over sizes greater than 25 cm. The minimum relative efficency is between 1.5 and 2 during the day, but efficiencies of the two sweeps are approximately equal efficiency at night (Figure ).

Stock-specific biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey were variable. Georges Bank winter flounder biomass estimates range between 1800 and 9400 mt in the spring and 3000 and 24,000 mt in the fall and are lower in recent years than those in the early years (Figure ). However, we note that the estimates of biomass made here were determined in the 2019 assessment to be problematic because of the larger sizes that predominate in the Georges Bank stock area than other stock areas and the low number of observations in the chainsweep study of these larger individuals [@nefsc2020]. Gulf of Maine winter flounder biomass estimates are constrained to the segment of the population at least 30 cm in length. The spring biomass estimates have been fairly stable ranging between 900 and 2700 mt whereas the fall estimates were greater at the beginning of the time series than recent years and range between 1900 and 4300 mt (Figure ). Southern New England winter flounder biomass estimates are also lower in recent years than the beginning of the time series for both seasons and spring and fall estimates range from 1300 to 8500 mt and 2100 to 47,500 mt, respectively (Figure ).

The efficiency of the rockhopper gear relative to the chainsweep in terms of biomass changes from year to year due primarily to corresponding changes in the estimated numbers at length (Table ). Annual biomass relative efficiency for Georges Bank winter flounder varied between 0.55 and 0.79 in the spring and 0.61 and 0.92 in the fall. Relative efficiencies for the Gulf of Maine stock range between 0.54 and 0.70 for the spring and 0.63 and 0.88 in the fall. Relative efficiencies for the Southern New England stock range between 0.64 and 0.91 for the spring and 0.60 and 1.0 for the fall.

Because the length-weight relationship which is used with the numbers at length to estimate biomass is estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at length (i.e., Eqs. and without the relative efficiency at size). These ratios should be approximately 1. The ratios for all years and seasons for all three stocks of winter flounder varied from 0.94 to 1.04 (Table ).

## Windowpane flounder

As measured by AIC, the best performing model for windowpane flounder before considering day/night effects was the conditional binomial model BI (Table ). Allowing smooth size-effects on relative catch efficiency and variation in these effects among paired-tows provided primary improvements in model performance (Table ). Including diel effects on relative efficiency at size for the twin-trawl observations improved performance of the binomial model (BI). The relative efficiency of the chain sweep gear to the rockhopper sweep gear decreases with size of windowpane flounder. The minimum relative efficiency is between 4.5 and 21 during the day, and between 1.8 and 2.9 at night (Figure ).

Stock-specific biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey were variable. Georges Bank-Gulf of Maine windowpane flounder biomass estimates range between 3000 and 20,300 mt in the spring and 4700 and 18,300 mt in the fall and are lower in recent years than those in the early years (Figure ). Southern New England-Mid-Atlantic Bight windowpane flounder biomass estimates in the spring ranged between 7300 and 15,600 mt whereas the fall estimates ranged between 7300 and 14,700 mt (Figure ).

The efficiency of the rockhopper gear relative to the chainsweep in terms of biomass changes from year to year due primarily to corresponding changes in the estimated numbers at length (Table ). Annual biomass relative efficiency for Georges Bank-Gulf of Maine windowpane flounder varied between 0.21 and 0.36 in the spring and 0.19 and 0.42 in the fall. Relative efficiencies for the Southern New England-Mid-Atlantic Bight stock ranged between 0.22 and 0.36 for the spring and 0.26 and 0.35 in the fall.

Because the length-weight relationship which is used with the numbers at length to estimate biomass is estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at length (i.e., Eqs. and without the relative efficiency at size). These ratios should be approximately 1. The ratios for all years and seasons for both stocks of windowpane flounder varied from 0.97 to 1.04 (Table ).

## Red hake

For red hake, the best performing model before considering day/night effects was the conditional beta-binomial model BB (Table ). The best beta-binomial model had an AIC more than 13 units lower than the best binomial model. Allowing variation in smooth size-effects on relative catch efficiency among paired-tows and extra-binomial variation withing paired-tows (overdispersion via the beta-binomial assumption) provided primary improvements in model performance. Including diel effects on relative efficiency for the twin-trawl observations improved performance of the beta-binomial model (Table ). Initially separate smooth size effects for day and night tows were considered for the beta-binomial model (BB), but the correlation of non-smoother related random effects across stations was not estimable. Those random effects were therefore assumed uncorrelated (BB). Allowing different smooth size effects of relative efficiency for day and night observations was considerd (BB), but it did not improve model performance. The relative efficiency of the chain sweep gear to the rockhopper sweep gear generally declines with increased size whether the tow occurred during day or night, but the increase in efficiency of the chainsweep was generally greater for tows occuring during the day (Figure ).

Stock-specific trends in annual biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey were generally the same. For northern red hake both the spring and fall biomass estimates increased in 2014 and have remained higher than previous years (Figure ). The scale of the biomass estimates is also similar for the spring and fall surveys. For southern red hake, the spring biomass generally declined until 2017 and then has increased for the last two years whereas the fall biomass has remained relatively stable (Figure ).

The efficiency of the rockhopper gear relative to the chainsweep in terms of biomass changes from year to year due primarily to corresponding changes in the estimated numbers at length (Table ). Annual biomass relative efficiency for northern red hake varied between 0.19 and 0.25 in the spring and 0.21 and 0.33 in the fall. Values range between 0.15 and 0.26 for the spring and 0.19 and 0.39 in the fall for southern red hake.

Because the length-weight relationship which is used with the numbers at length to estimate biomass is estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at length (i.e., Eqs. and without the relative efficiency at size). These ratios should be approximately 1. The ratios for all years and seasons for both northern and southern red hake varied from 0.96 to 1.04 (Table ).

# Discussion

The goal of these chain sweep experiments was to provide estimates of absolute abundance that could be used for assessments of Northeast US groundfish stocks. These estimates can then be compares to other index-based empirical assessment methods. These estimates are especially valuable for stocks assessed with index-based methods (e.g., GOM winter flounder, witch flounder, GB yellowtail flounder, red hake).

Compare greater or lesser smoothness within stations with . We assume the same number of knots and order (derivatives for penalties) in the cubic regression splines for the population and station-level smoothers. also implicitly assume the random effects that correspond to the null space (intercept and fixe effects of the smoothers) are uncorrelated, but correlation in these models is estimated except for red hake where we found it to be inestimable.

couple of paragraphs about diel changes in catchability. Reference other papers about this and behavior of fish with regard to the substrate changing between day and night. Chainsweep doesn’t change but rockhopper does? Particularly important when survey is conducted during day or night only. Perhaps could improve this estimation by allowing cyclic cubic spline on time of day rather than factor treatment of day/night.

Which stocks currently use index-based methods? GOM winter flounder, GB yellowtail flounder,

Treating the chain-sweep based abundance estimates implicitly assumes that the chainsweep provides complete capture efficiency and that the stock resides completely within the strata that are used to generate the abundance indices. It is unlikely that the chainsweep gear is completely efficient for all sizes of fish of a particular species over all substrate types that are sampled. It is also typical for many of these stocks to extend somewhat outside of the survey strata used to define the indices either throughout the year or seasonally due to migration.

# Acknowledgements

We thank Adam Poquette for creating the map in Figure .