

¹ Chainsweep-based efficiency of bottom trawl surveys and biomass
² estimates for flatfish, red hake, and goosefish stocks in Northwest
³ Atlantic waters of the United States

⁴ Timothy J. Miller¹, David E. Richardson², Andrew Jones², Phil Politis¹

⁵ ¹timothy.j.miller@noaa.gov, Northeast Fisheries Science Center, National Marine Fisheries Service, 166 Water
⁶ Street, Woods Hole, MA 02543, USA

⁷ ²Northeast Fisheries Science Center, National Marine Fisheries Service, Narragansett, RI USA

9 Abstract

10 Using a general hierarchical model we estimated relative efficiency of chain sweep to the rockhopper sweep
11 used by the NEFSC bottom trawl survey for from studies carried out between 2015 and 2017 aboard the
12 F/V Karen Elizabeth twin-trawl vessel. Aside from the sweeps, the rest of the trawl gear is the same. We
13 compared a set of models with different assumptions about variation of relative efficiency between paired
14 gear tows, size and diel effects on the relative efficiency, and extra-binomial variation of observations within
15 paired gear tows.

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

25 Keywords

26 hierarchical models, spline regression, gear efficiency, abundance estimation

²⁷ 1 Introduction

- ²⁸ Paired-gear studies where two gear are fished either concurrently or close together temporally and spatially
²⁹ have long been used to estimate the efficiency of one fishing gear relative to another (e.g., Gulland, 1964;
³⁰ Bourne, 1965). Of the two gears, one is often a reference gear that may be a gear currently used for annual
³¹ surveys (e.g., Munro and Sumerton 2001). Typically neither of the gears can be assumed to be fully efficient
³² and therefore the relative efficiency of gears is estimated (e.g., Miller 2013; Kotwicki et al. 2017), but there
³³ are cases where one of the gears is assumed to be at least very nearly fully efficient (e.g., Miller et al. 2019).
- ³⁴ Whether or not full efficiency of one of the gears is assumed, paired-gear studies are essential for generating
³⁵ abundance time series from fishery independent surveys when there are changes in the vessel and(or) gears
³⁶ over time due to gear failures or improved technology. These studies are also helpful for combining 2 or more
³⁷ surveys conducted using alternative gears are conducted close together in space or time.
- ³⁸ In conducting paired-gear studies it is ideal to have the two gears deployed as close together spatially and
³⁹ temporally as possible to reduce variation between the gears in densities of the species being captured.
⁴⁰ One fishing method that approaches this ideal is the twin-trawl rigging where two trawls can be fished
⁴¹ simultaneously (ICES, 1996).
- ⁴² Within the northeast US there has been a heightened focus on bottom trawl survey operations and gear
⁴³ efficiency. This focus has in part resulted from low quotas for a number of groundfish limiting fishing
⁴⁴ opportunities. To help provide clarity on the trawl operations and build trust in survey indices the New
⁴⁵ England and Mid-Atlantic Fisheries Management Councils developed a Northeast Trawl Advisory Panel.
⁴⁶ This panel is composed of members from industry, regional academics, as well as state and federal scientists.
⁴⁷ Together the group designed a set of experiments to explore the differences in efficiency between survey and
⁴⁸ trawl gear (hereafter referred to as ‘chain sweep’ experiments).
- ⁴⁹ The goal of these chain sweep experiments was to provide estimates of absolute abundance that could be
⁵⁰ used for assessments of NEUS fish stocks. For example, developing an estimate of absolute abundance allows
⁵¹ for some swept-area biomass calculations at the region scale. These estimates can then be compared to other
⁵² index-based empirical assessment methods. This is especially valuable for species where limited commercial
⁵³ harvest occurs, but assessments suggest the species is in poor status (e.g., red hake).
- ⁵⁴ Importance of biomass or (catchability) efficiency estimates for both index-based methods as well as age-
⁵⁵ structured models with low contrast.
- ⁵⁶ The basic methods we used here are based on those used by Miller (2013) to estimate size effects on relative

57 catch efficiency of the Henry B. Bigelow to the Albatross IV for a variety of commercially important species,
58 but we extend the model to consider different size effects for tows conducted during the day or night since both
59 the spring and fall bottom trawl surveys conducted in the Northeast US are 24-hour operations. We apply
60 these methods to paired gear observations for estimate relative efficiency of the chainsweep and rockhopper
61 sweep gears. We apply the estimated efficiency of the rockhopper gear to survey data to estimate spring and
62 fall abundance indices from 2009-2019 for 17 commercially important fish stocks in the Northeast US (Table
63 1).

64 Often overlooked aspects of the application of relative catch efficiency estimates is the impact on the precision
65 of abundance indices and the correlation among annual indices that the application induces. These indices
66 are typically used as measures of relative abundance in stock assessment with the precision of the indices used
67 to weight the observations within the assessment model. Furthermore, these indices are typically assumed to
68 be independent. Here we compare the precision of the calibrated and uncalibrated indices and measure the
69 correlation of calibrated indices for each stock.

70 **2 Methods**

71 **2.1 Data collection**

72 Data were collected during three field experiments carried out in 2015, 2016, and 2017, respectively, aboard
73 the F/V Karen Elizabeth, a 78ft stern trawler capable of towing two trawls simultaneously side by side.
74 However, red hake were only observed during the 2017 field experiments. One side of the twin-trawl rig
75 towed a NEFSC standard 400 x 12 cm survey bottom trawl rigged with the NEFSC standard rockhopper
76 sweep (Politis et al., 2014) (Figure 1). The other side of the twin-trawl rig towed a version the NEFSC 400 x
77 12cm survey bottom trawl modified to maximize the capture of flatfish. The trawl was modified by reducing
78 the headline flotation from 66 to 32, 20cm, spherical floats, reducing the port and starboard top wing-end
79 extensions by 50cm each and utilizing a chain sweep. The chain sweep was constructed of 1.6cm (5/8in)
80 trawl chain covered by 12.7cm diameter x 1cm thick rubber discs on every other chain link (Figure 2). Two
81 rows of 1.3cm (1/2in) tickler chains were attached to the 1.6cm trawl chain by 1.3cm shackles. To ensure
82 equivalent net geometry of each gear, 32m restrictor ropes, made of 1.4cm (9/16in) buoyant, Polytron rope,
83 were attached between each of the trawl doors and the center clump. 3.4m² Thyboron Type 4 trawl doors
84 were used to provide enough spreading force to ensure the restrictor ropes remained taut throughout each
85 tow. Each trawl used the NEFSC standard 36.6m bridles. All tows followed the NEFSC standard survey

86 towing protocols of 20 minutes at 3.0 knots. In 2015, 108 (45 day, 63 night) paired tows were conducted in
87 eastern Georges Bank and off of southern New England (Figure 3). In 2016, 117 (74 day, 43 night) paired
88 tows were conducted in western Gulf of Maine and northern edge of Georges Bank (Figure 4). In 2017, 103
89 (61 day, 42 night) paired tows were conducted in waters off of southern New England (Figure 5). Paired tows
90 were denoted as “day” and “night” by whether the sun was above or below the horizon at the time of the tow.

91 2.2 Paired-tow analysis

92 We use the hierarchical modeling approach from Miller (2013) to estimate the relative efficiency of chain sweep
93 to the rockhopper sweep used by the NEFSC bottom trawl survey for six species from three studies carried
94 out aboard a twin trawl vessel. Aside from the sweeps the rest of the trawl gear is the same. As in Miller
95 (2013), we compared a set of models with different assumptions about variation of relative efficiency between
96 paired gear tows, size effects on the relative efficiency, and extra-binomial variation of observations within
97 paired gear tows. We began with the same 13 models considered by Miller (2013). The binomial(BI₀ to BI₄)
98 and beta-binomial (BB₀ to BB₇) models that were fitted for all species are described in Table 2 including
99 pseudo-formulas comparable to those used for fitting mixed or generalized additive models in R (R Core
100 Team, 2019; Wood, 2006). We then also included diel effects on relative catch efficiency and interactions with
101 size effects with the best performing model of the original 13 models for each species. The model framework
102 is more generalized than those in Miller (2013) in that we now allow multiple smooth effects (differing by day
103 or night) on relative catch efficiency. We implemented the models using the Template Model Builder package
104 (Kristensen et al., 2016) in R and we used the “nlnimb” optimizer to fit the models (R Core Team, 2019).

105 If the best model included cubic regression splines of length and the estimated smoothing parameter implied
106 a linear functions of length (on the transformed mean), then simple linear functions (i.e., completely smooth)
107 were assumed for further models that included diel effects on relative efficiency.

108 One less parameter (smoothing parameter) for these models.

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

116 model converged and the hessian of the maximized log-likelihood was invertible.

117 Red hake: had to assume station-specific random effects that corresponded to the population-level fixed
118 effects were uncorrelated because of convergence issues.

119 2.3 Length-weight analysis

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

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

122 We used a bias correction to ensure the expected weight $E(W_{ij}) = \alpha_i L_{ij}^{\beta_i}$. We estimated parameters by
123 maximizing the model likelihood programmed in TMB (Kristensen et al., 2016) and R (R Core Team,
124 2019). Like the relative catch efficiency, bootstrap predictions of weight at length were made by sampling
125 with replacement the length-weight observations within each annual survey and refitting the length-weight
126 relationship to each of the bootstrap data sets.

127 2.4 Biomass estimation

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

$$\tilde{N}_{hi}(L) = N_{hi}(L) \hat{\rho}_i(L) \quad (2)$$

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

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

$$\widehat{N}(L) = \sum_{h=1}^H \frac{A_h}{an_h} \sum_{i=1}^{n_h} \widetilde{N}_{hi}(L) \quad (3)$$

139 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
140 of tows that were made in stratum h . The corresponding biomass estimate is then

$$\widehat{B} = \sum_{l=1}^{n_L} \widehat{N}(L = l) \widehat{w}(L = l) \quad (4)$$

141 where $\widehat{w}(L = l)$ is the estimated weight at length from fitting length-weight observations described above.
142 Length is typically measured to the nearest cm so n_L indicates the number of 1 cm length categories that
143 were observed during the survey.

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

152 To estimate uncertainty in biomass, we used bootstrap results for the relative catch efficiency and weight at
153 length estimates along with bootstrap samples of the survey data. Bootstrap data sets for each of the annual
154 surveys respected the stratified random designs by resampling with replacement within each stratum (Smith,
155 1997). For each of the 1000 combined bootstraps, survey observations for bootstrap b were scaled with the
156 corresponding bootstrap estimates of relative chain to rockhopper sweep efficiency and predicted weight at
157 length, using Eqs. 3 and 4.

158 We also used the bootstraps to summarize other aspects of the biomass estimates. First, we used the
159 bootstraps to calculate the ratio of calibrated and uncalibrated biomass for each spring and fall annual survey
160 which is the implicit relative catch efficiency in terms of biomass. The uncalibrated biomass estimate for
161 bootstrap b uses the resampled survey data as the calibrated biomass estimate except that the bootstrap for
162 the relative catch efficiency is not used (i.e., $\widehat{\rho}_i(L) = 1$ in Eq. 2). We also used the bootstraps to compare the
163 coefficients of variation (CV) of the calibrated and uncalibrated biomass estimates. The CV for an annual

¹⁶⁴ biomass estimate for year y from either the spring or fall survey was calculated as

$$CV(\hat{B}_y) = \frac{SD(\hat{B}_y)}{\bar{\hat{B}}_y}$$

¹⁶⁵ where

$$SD(\hat{B}_y) = \sqrt{\frac{\sum_{b=1}^K (\hat{B}_{y,b} - \bar{\hat{B}}_y)^2}{K-1}},$$

¹⁶⁶

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

¹⁶⁷ and K is the number of bootstraps.

¹⁶⁸ For summer flounder it was necessary to omit one of the 1000 bootstraps of relative catch efficiency at
¹⁶⁹ length due to an extremely large value which the standard deviation and mean of the bootstraps was sensitive
¹⁷⁰ to. Finally, we calculated correlation of annual biomass estimates for years y and z using the bootstrap
¹⁷¹ estimates of biomass

$$Cor(\hat{B}_y, \hat{B}_z) = \frac{Cov(\hat{B}_y, \hat{B}_z)}{SD(\hat{B}_y) SD(\hat{B}_z)}$$

¹⁷² where the covariance is

$$Cov(\hat{B}_y, \hat{B}_z) = \frac{\sum_{b=1}^K (\hat{B}_{y,b} - \bar{\hat{B}}_y)(\hat{B}_{z,b} - \bar{\hat{B}}_z)}{K-1}.$$

¹⁷³ We summarized the relative precision of the calibrated and uncalibrated biomass estimates as the average of
¹⁷⁴ the annual ratios of the CVs for the calibrated and uncalibrated estimates

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

¹⁷⁵ We summarized the correlation of biomass estimates as the mean correlation of all annual calibrated biomass
¹⁷⁶ estimates

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

₁₇₇ **3 Results**

₁₇₈ **3.1 Paired-tow observations**

₁₇₉ In terms on paired tows and total numbers of fish, flatfish were the most well sampled species, but goosefish
₁₈₀ was observed in the most paired-tows and red hake was the most prevalent in terms of total numbers caught
₁₈₁ (Table 3). Witch flounder was the most prevalent flatfish species caught while yellowtail flounder was observed
₁₈₂ in the the most frequently observed flatfish in terms of paired tows. For all but summer flounder, and
₁₈₃ barndoor and thorny skates, only a subsample of all of the fish that were caught were measured for length,
₁₈₄ but nearly all caught winter flounder and goosefish were measured.

₁₈₅ **3.2 Relative model performance**

₁₈₆ As measured by AIC, the best performing model for all 10 species included size effects on the relative efficiency
₁₈₇ of the chain and rockhopper sweep gears and between-pair variability in relative catch efficiency (Table 4).

₁₈₈ Extrabinomial variation (i.e., beta-binomial) in relative catch efficiency at size within pairs was also important
₁₈₉ for American plaice, yellowtail flounder, witch flounder, red hake, and thorny skate. Model convergence was
₁₉₀ an issue for all species, particularly for the most complex models with pair-specific smooth functions of length
₁₉₁ (BI_4) and smooth effects of size on the beta-binomial dispersion parameter (BB_3, BB_5 , and BB_7).

₁₉₂ Including diel effects on relative catch efficiency improved model performance for all species except American
₁₉₃ plaice (Table 5).

₁₉₄ Higher relative catch efficiency was always during the daytime when fish typically associate with the substrate
₁₉₅ more.

₁₉₆ winter flounder before considering day/night effects was the conditional binomial model BI_4 (Table 4).
₁₉₇ Allowing smooth size-effects on relative catch efficiency and variation in these effects among paired-tows
₁₉₈ provided primary improvements in model performance. Including diel effects on relative efficiency for the
₁₉₉ twin-trawl observations improved performance of the binomial model (BI_5),

₂₀₀ The relative efficiency of the chain sweep gear to the rockhopper sweep gear is greatest at the smallest sizes
₂₀₁ of winter flounder, but is fairly constant over over sizes greater than 25 cm. The minimum relative efficency
₂₀₂ is between 1.5 and 2 during the day, but efficiencies of the two sweeps are approximately equal efficiency at
₂₀₃ night (Figure 6).

204 **3.3 Bootstrap-based uncertainty**

205 All 1000 bootstrap fits of the paired tow data provided estimates of relative catch efficiency at size for summer,
206 windowpane, and yellowtail flounder, and red hake, goosefish, and thorny skate. All but 2 of the bootstraps
207 for winter flounder and 3 for barndoor skate provided estimates of relative catch efficiency. For witch flounder,
208 817 bootstraps provided estimates and only 386 provided estimates for American plaice. One bootstrap fit
209 for summer flounder was excluded due to an extremely high relative efficiency of the chainsweep gear which
210 impeded estimation of standard errors from the bootstrap fits.

211 We see that generally where data are prevalent the bootstrap and hessian-based confidence intervals are
212 similar across all species. However, sometimes substantially different perceptions of confidence ranges exist at
213 the extremes of the length range for particular species.

214 **3.4 Summer flounder**

215 **3.5 American plaice**

216 **3.6 Yellowtail flounder**

217 **3.7 Witch flounder**

218 **3.8 Goosefish**

219 **3.9 Winter flounder**

220 As measured by AIC, the best performing model for winter flounder before considering day/night effects was
221 the conditional binomial model BI₄ (Table 4). Allowing smooth size-effects on relative catch efficiency and
222 variation in these effects among paired-tows provided primary improvements in model performance. Including
223 diel effects on relative efficiency for the twin-trawl observations improved performance of the binomial model
224 (BI₅), however the model allowing the size effects on relative efficiency to differ between day and night (BI₆)
225 would not converge. The relative efficiency of the chain sweep gear to the rockhopper sweep gear is greatest at
226 the smallest sizes of winter flounder, but is fairly constant over sizes greater than 25 cm. The minimum
227 relative efficiency is between 1.5 and 2 during the day, but efficiencies of the two sweeps are approximately
228 equal efficiency at night (Figure 6).

229 Stock-specific biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey were variable.
230 Georges Bank winter flounder biomass estimates range between 1800 and 9400 mt in the spring and 3000 and
231 24,000 mt in the fall and are lower in recent years than those in the early years (Figure 7). However, we
232 note that the estimates of biomass made here were determined in the 2019 assessment to be problematic
233 because of the larger sizes that predominate in the Georges Bank stock area than other stock areas and
234 the low number of observations in the chainsweep study of these larger individuals (NEFSC, 2020). Gulf of
235 Maine winter flounder biomass estimates are constrained to the segment of the population at least 30 cm in
236 length. The spring biomass estimates have been fairly stable ranging between 900 and 2700 mt whereas the
237 fall estimates were greater at the beginning of the time series than recent years and range between 1900 and
238 4300 mt (Figure 7). Southern New England winter flounder biomass estimates are also lower in recent years
239 than the beginning of the time series for both seasons and spring and fall estimates range from 1300 to 8500
240 mt and 2100 to 47,500 mt, respectively (Figure 7).

241 The efficiency of the rockhopper gear relative to the chainsweep in terms of biomass changes from year to
242 year due primarily to corresponding changes in the estimated numbers at length (Table ??). Annual biomass
243 relative efficiency for Georges Bank winter flounder varied between 0.55 and 0.79 in the spring and 0.61 and
244 0.92 in the fall. Relative efficiencies for the Gulf of Maine stock range between 0.54 and 0.70 for the spring
245 and 0.63 and 0.88 in the fall. Relative efficiencies for the Southern New England stock range between 0.64
246 and 0.91 for the spring and 0.60 and 1.0 for the fall.

247 Because the length-weight relationship which is used with the numbers at length to estimate biomass is
248 estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the
249 biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just
250 the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at
251 length (i.e., Eqs. 3 and 4 without the relative efficiency at size). These ratios should be approximately 1.
252 The ratios for all years and seasons for all three stocks of winter flounder varied from 0.94 to 1.04 (Table ??).

253 3.10 Windowpane flounder

254 As measured by AIC, the best performing model for windowpane flounder before considering day/night effects
255 was the conditional binomial model BI₄ (Table 4). Allowing smooth size-effects on relative catch efficiency
256 and variation in these effects among paired-tows provided primary improvements in model performance (Table
257 5). Including diel effects on relative efficiency at size for the twin-trawl observations improved performance
258 of the binomial model (BI₆). The relative efficiency of the chain sweep gear to the rockhopper sweep gear

259 decreases with size of windowpane flounder. The minimum relative efficiency is between 4.5 and 21 during
260 the day, and between 1.8 and 2.9 at night (Figure 6).

261 Stock-specific biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey were variable.
262 Georges Bank-Gulf of Maine windowpane flounder biomass estimates range between 3000 and 20,300 mt in
263 the spring and 4700 and 18,300 mt in the fall and are lower in recent years than those in the early years
264 (Figure 7). Southern New England-Mid-Atlantic Bight windowpane flounder biomass estimates in the spring
265 ranged between 7300 and 15,600 mt whereas the fall estimates ranged between 7300 and 14,700 mt (Figure 7).

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

271 Because the length-weight relationship which is used with the numbers at length to estimate biomass is
272 estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the
273 biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just
274 the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at
275 length (i.e., Eqs. 3 and 4 without the relative efficiency at size). These ratios should be approximately 1. The
276 ratios for all years and seasons for both stocks of windowpane flounder varied from 0.97 to 1.04 (Table ??).

277 3.11 Red hake

278 For red hake, the best performing model before considering day/night effects was the conditional beta-binomial
279 model BB₆ (Table 4). The best beta-binomial model had an AIC more than 13 units lower than the best
280 binomial model. Allowing variation in smooth size-effects on relative catch efficiency among paired-tows and
281 extra-binomial variation withing paired-tows (overdispersion via the beta-binomial assumption) provided
282 primary improvements in model performance. Including diel effects on relative efficiency for the twin-trawl
283 observations improved performance of the beta-binomial model (Table 5). Initially separate smooth size
284 effects for day and night tows were considered for the beta-binomial model (BB₈), but the correlation of
285 non-smoother related random effects across stations was not estimable. Those random effects were therefore
286 assumed uncorrelated (BB₉). Allowing different smooth size effects of relative efficiency for day and night
287 observations was considerd (BB₁₀), but it did not improve model performance. The relative efficiency of
288 the chain sweep gear to the rockhopper sweep gear generally declines with increased size whether the tow

289 occurred during day or night, but the increase in efficiency of the chainsweep was generally greater for tows
290 occurring during the day (Figure 6).

291 Stock-specific trends in annual biomass estimates from 2009 to 2019 for the NEFSC spring and fall survey
292 were generally the same. For northern red hake both the spring and fall biomass estimates increased in 2014
293 and have remained higher than previous years (Figure 7). The scale of the biomass estimates is also similar
294 for the spring and fall surveys. For southern red hake, the spring biomass generally declined until 2017 and
295 then has increased for the last two years whereas the fall biomass has remained relatively stable (Figure 7).

296 The efficiency of the rockhopper gear relative to the chainsweep in terms of biomass changes from year to
297 year due primarily to corresponding changes in the estimated numbers at length (Table ??). Annual biomass
298 relative efficiency for northern red hake varied between 0.19 and 0.25 in the spring and 0.21 and 0.33 in the
299 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.

300 Because the length-weight relationship which is used with the numbers at length to estimate biomass is
301 estimated by survey and year there is a possibility that poor sampling in a given year could adversely affect the
302 biomass estimates. We therefore calculated the ratios of the annual uncalibrated biomass estimates using just
303 the aggregate catch data to the biomass estimates made using the numbers at length and estimated weight at
304 length (i.e., Eqs. 3 and 4 without the relative efficiency at size). These ratios should be approximately 1. The
305 ratios for all years and seasons for both northern and southern red hake varied from 0.96 to 1.04 (Table ??).

306 4 Discussion

307 Compare greater or lesser smoothness within stations with Pedersen et al. (2019). We assume the same
308 number of knots and order (derivatives for penalties) in the cubic regression splines for the population and
309 station-level smoothers. Pedersen et al. (2019) also implicitly assume the random effects that correspond to
310 the null space (intercept and fixe effects of the smoothers) are uncorrelated, but correlation in these models is
311 estimated except for red hake where we found it to be inestimable.

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

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

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

323 **Acknowledgements**

324 References

- 325 Bourne, N. 1965. A comparison of catches by 3- and 4-inch rings on offshore scallop drags. Journal of the
326 Fisheries Research Board of Canada **22**(2): 313–333.
- 327 Gulland, J. A. 1964. Variations in selection factors, and mesh differentials. Journal Du Conseil International
328 Pour L'exploration De La Mer **29**(2): 158–165.
- 329 ICES. 1996. Manual of methods of measuring the selectivity of towed fishing gears. (Eds.) Wileman, D. A.,
330 Ferro, R. S. T., Fonteyne, R., and Millar, R. B. ICES Cooperative Research Report No. 215.
- 331 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic differentiation
332 and Laplace approximation. Journal of Statistical Software **70**(5): 1–21.
- 333 Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear
334 data for U.S. Northwest Atlantic fish stocks. Canadian Journal of Fisheries and Aquatic Sciences **70**(9):
335 1306–1316, doi: 10.1139/cjfas-2013-0136.
- 336 NEFSC. 2020. Operational assessment of 14 northeast groundfish stocks, updated through 2018. Prepublication
337 Copy of the September 2019 Operational Stock Assessment Report. The report is “in preparation” for
338 publication by the NEFSC. (Oct. 3, 2019; latest revision: Jan. 7, 2020). <https://nefsc.noaa.gov/saw/2019-groundfish-docs/Prepublication-NE-Grndfsh-1-7-2020.pdf>.
- 340 Pedersen, E. J., Miller, D. L., Simpson, G. L., and Ross, N. 2019. Hierarchical generalized additive models in
341 ecology: an introduction with mgcv. PeerJ **7**: e6876, doi: 10.7717/peerj.6876.
- 342 Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. 2014. Northeast Fisheries Science Center
343 bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. U.S. Dept. Commer., Northeast Fish.
344 Sci. Cent. Ref. Doc. 14-06, 138p.
- 345 R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical
346 Computing, Vienna, Austria.
- 347 Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance.
348 Canadian Journal of Fisheries and Aquatic Sciences **54**(3): 616–630, doi: 10.1139/f96-303.
- 349 Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall, Boca Raton,
350 Florida, 392 pp.

Fig. 1. Diagram of the standard Northeast Fisheries Science Center rockhopper sweep center and wing sections.

ROCKHOPPER CENTER SECTION
Section = 890cm, 100lbs Lead

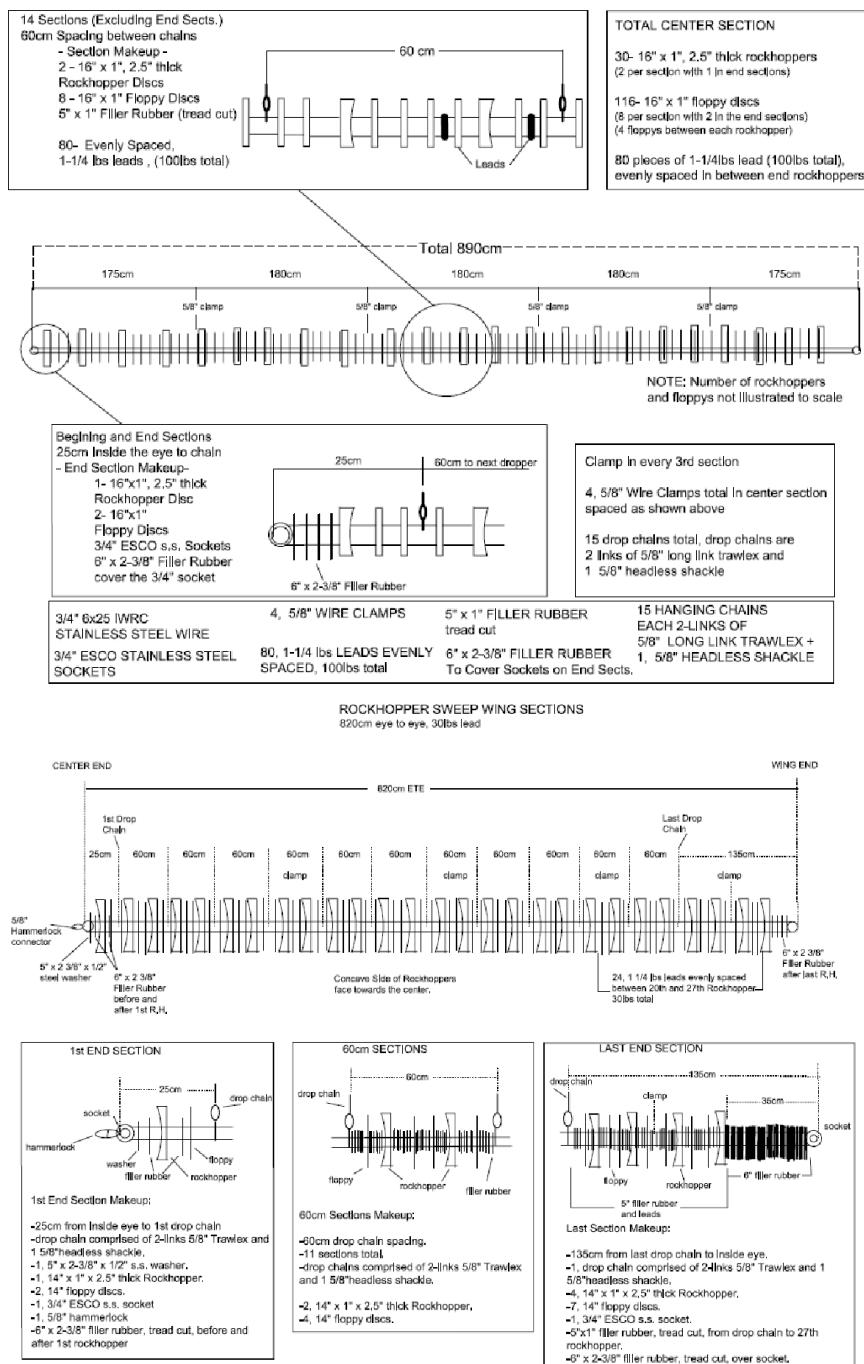


Fig. 2. Diagram of the chain sweep designed maximize bottom contact and flatfish capture.

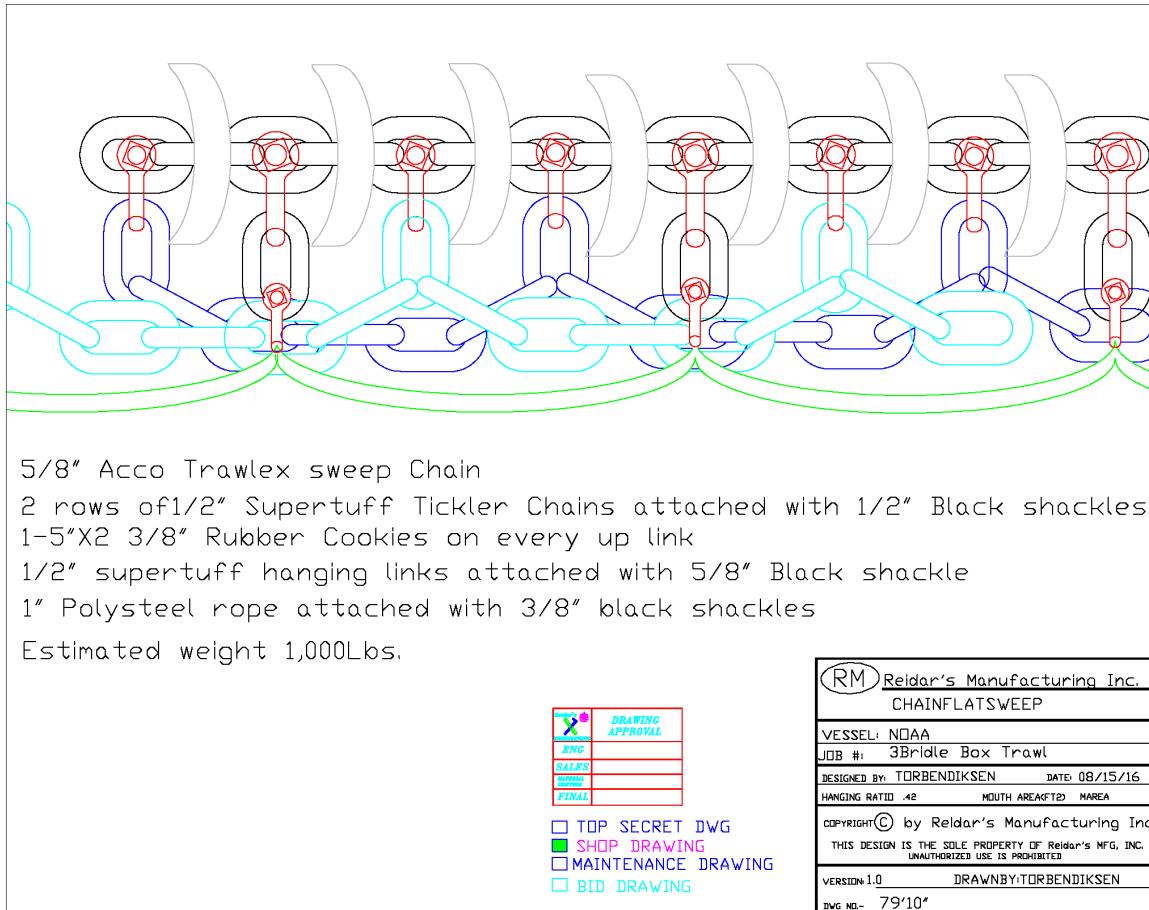


Fig. 3. Locations of stations in 2015 where the F/V Karen Elizabeth conducted twin-trawl sets with the standard bottom trawl gear and the gear with a chain sweep instead of the rockhopper sweep.

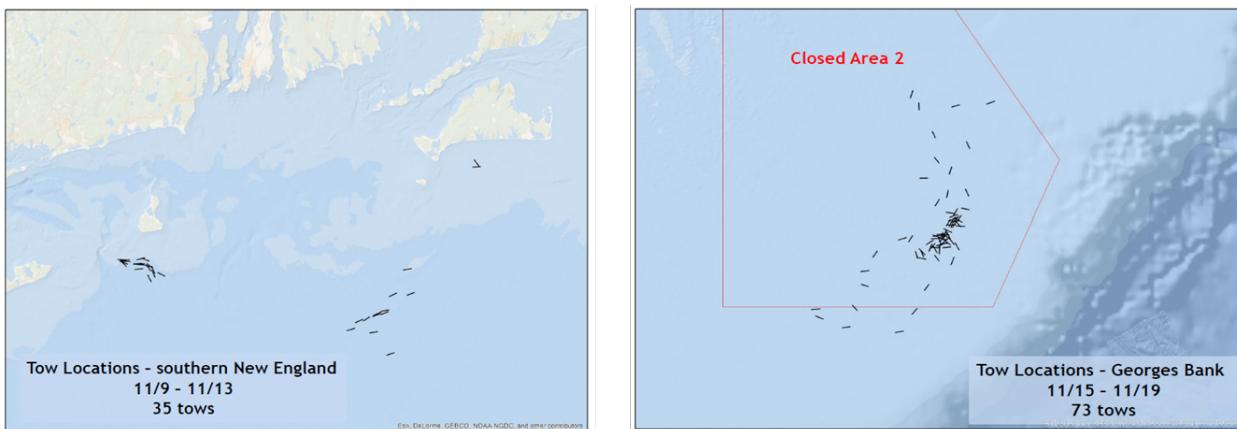


Fig. 4. Locations of stations in 2016 where the F/V Karen Elizabeth conducted twin-trawl sets with the standard bottom trawl gear and the gear with a chain sweep instead of the rockhopper sweep.

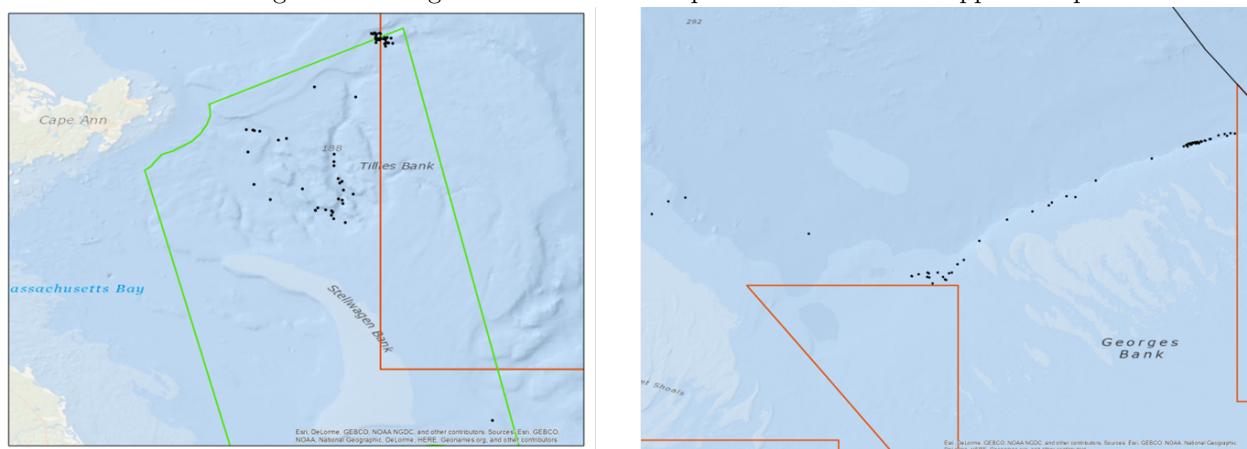


Fig. 5. Locations of stations in 2017 where the F/V Karen Elizabeth conducted twin-trawl sets with the standard bottom trawl gear and the gear with a chain sweep instead of the rockhopper sweep.

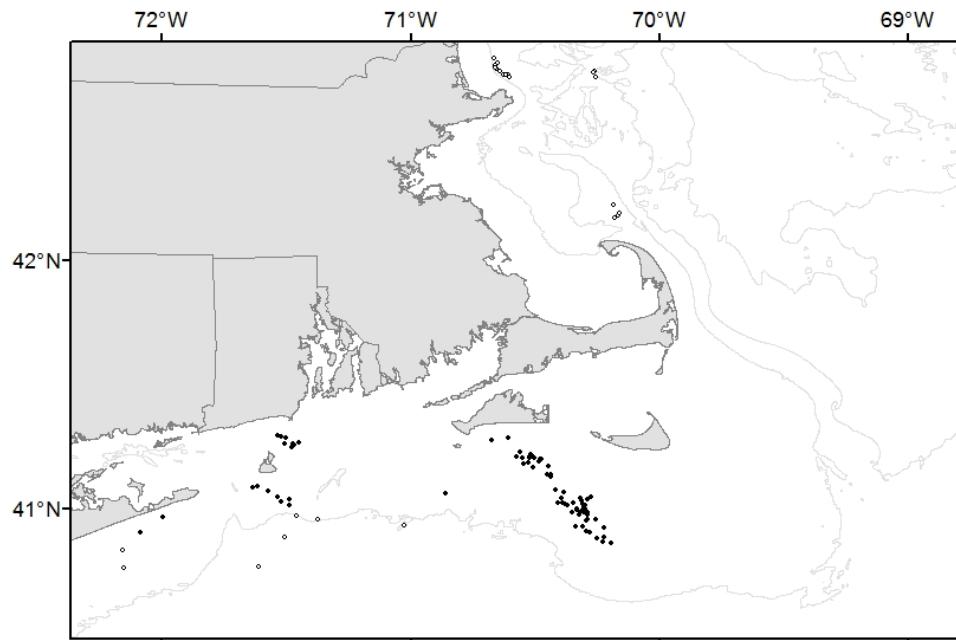


Fig. 6. Relative efficiency of gears using chain and rockhopper sweeps from the best performing model for each species (Table 5). Blue and red denote results for day and night data, respectively, and thick and thin lines represent overall and paired-tow specific estimates of relative catch efficiency, respectively. 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.

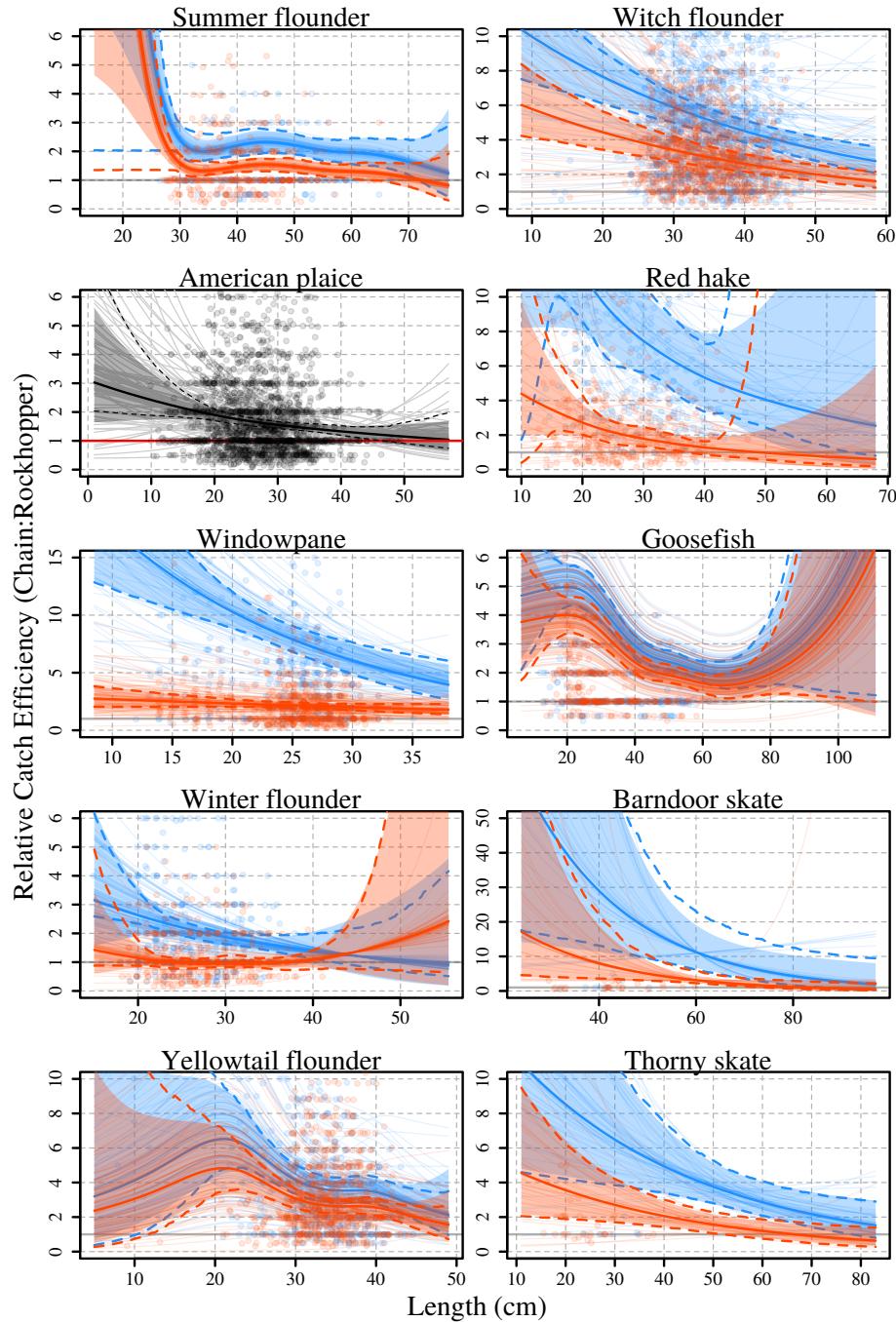


Table 1. Managed stocks associated with the species for which relative catch efficiency was estimated.

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

Fig. 7. Annual spring (blue) and fall (red) biomass estimates for each managed stock assuming 100% efficiency for chainsweep 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).

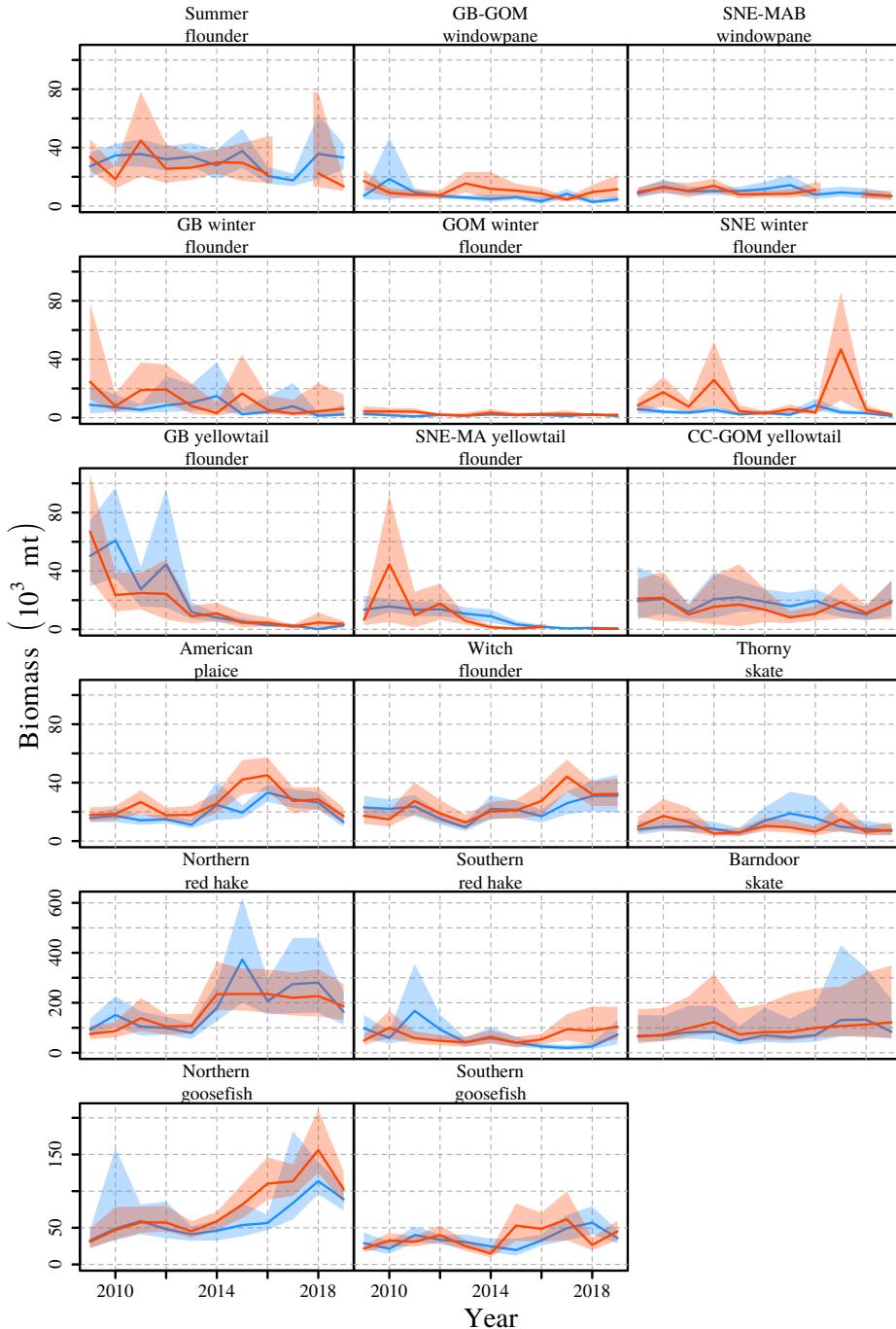


Table 2. Description of relative catch efficiency (ρ) and beta-binomial dispersion (ϕ) parameterizations for binomial and beta-binomial models and number of marginal likelihood parameters (n_p) for the 13 base models from Miller (2013) and fit to paired chainsweep and rockhoppersweep tow data for each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

Model	$\log(\rho)$	$\log(\phi)$	n_p	Description
BI ₀	~ 1	–	1	population-level mean for all observations
BI ₁	$\sim 1 + 1 pair$	–	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	$\sim 1 + 1 pair$	~ 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	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	$\sim s(\text{length}) + 1 \text{pair}$	–	4	22.92
	BI _{3a}	$\sim dn + s(\text{length}) + 1 \text{pair}$	–	5	0
	BI _{3b}	$\sim dn * s(\text{length}) + 1 \text{pair}$	–	7	1.74
American plaice					
	BB ₇	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	$\sim s(\text{length})$	10	0
	BB _{7a}	$\sim dn + s(\text{length}) + s(\text{length}) \text{pair}$	$\sim s(\text{length})$	11	1.43
	BB _{7b}	$\sim dn * s(\text{length}) + s(\text{length}) \text{pair}$	$\sim s(\text{length})$	13	2.95
Windowpane					
	BI ₄	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	–	7	152.1
	BI _{4a}	$\sim dn + \text{length} + s(\text{length}) \text{pair}$	–	7	4.06
	BI _{4b}	$\sim dn * \text{length} + s(\text{length}) \text{pair}$	–	8	0
Winter flounder					
	BI ₄	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	–	7	50.68
	BI _{4a}	$\sim dn + s(\text{length}) + \text{length} \text{pair}$	–	7	0.3
	BI _{4b}	$\sim dn * s(\text{length}) + \text{length} \text{pair}$	–	9	0
Yellowtail flounder					
	BB ₆	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	8	3.84
	BB _{6a}	$\sim dn + s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	9	0
	BB _{6b}	$\sim dn * s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	11	3.48
Witch flounder					
	BB ₆	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	8	19.68
	BB _{6a}	$\sim dn + \text{length} + s(\text{length}) \text{pair}$	~ 1	8	0
	BB _{6b}	$\sim dn * \text{length} + s(\text{length}) \text{pair}$	~ 1	9	1.52
Red hake					
	BB ₆	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	8	32.35
	BB _{6a}	$\sim dn + s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	8	0
	BB _{6b}	$\sim dn * s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	10	3.18
Goosefish					
	BI ₃	$\sim s(\text{length}) + 1 \text{pair}$	–	4	5.44
	BI _{3a}	$\sim dn + s(\text{length}) + 1 \text{pair}$	–	5	0
	BI _{3b}	$\sim dn * s(\text{length}) + 1 \text{pair}$	–	7	6.8
Barndoor skate					
	BI ₄	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	–	7	15.57
	BI _{4a}	$\sim dn + \text{length} + \text{length} \text{pair}$	–	5	0
	BI _{4b}	$\sim dn * \text{length} + \text{length} \text{pair}$	–	6	1.83
Thorny skate					
	BB ₆	$\sim s(\text{length}) + s(\text{length}) \text{pair}$	~ 1	8	15.51
	BB _{6a}	$\sim dn + \text{length} + \text{length} \text{pair}$	~ 1	7	0
	BB _{6b}	$\sim dn * \text{length} + \text{length} \text{pair}$	~ 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