

¹ Estimation of survey efficiency and biomass for
² commercially important species from industry-based
³ paired gear experiments

⁴ Timothy J. Miller¹ David E. Richardson² Philip J. Politis¹

⁵ Christopher D. Roebuck³ John P. Manderson^{4,5}

⁶ Michael H. Martin^{1,6} Andrew W. Jones²

⁷ ¹corresponding author: timothy.j.miller@noaa.gov, Northeast Fisheries Science Center,
⁸ Woods Hole Laboratory, 166 Water Street, Woods Hole, MA 02543 USA

⁹ ²Northeast Fisheries Science Center, Narragansett Laboratory, 28 Tarzwell Drive Narra-
¹⁰ gansett, RI 02882 USA

¹¹ ³Fishing Vessel Karen Elizabeth, Narragansett, RI 02882, USA

¹² ⁴Northeast Fisheries Science Center, James J. Howard Marine Sciences Laboratory, 74
¹³ Magruder Road Highlands, NJ 07732 USA

¹⁴ ⁵Current address: OpenOcean Research, Suite 101, 40 West Evergreen Avenue, Philadelphia,
¹⁵ PA, 19118 USA

¹⁶ ⁶Current address: Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Building 4
¹⁷ Seattle, WA 98115 USA

¹⁸ **Abstract**

¹⁹ 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.

³⁵ **Keywords**

³⁶ gear efficiency, biomass estimation, hierarchical generalized additive models

³⁷ **1 Introduction**

³⁸ Ecosystem monitoring surveys such as fisheries-independent trawl surveys are used to obtain
³⁹ information on a range of species and are therefore not optimized with respect to sampling
⁴⁰ design or gear for any one species (Bijleveld et al., 2012; Wang et al., 2018). Gear and
⁴¹ sampling protocols are designed to provide consistent and representative samples that allow
⁴² indices of abundance at size and age to be developed for a suite of species (Azarovitz, 1981;
⁴³ Thiess et al., 2018). To provide indices of population abundance with minimal potential
⁴⁴ sources of bias, survey bottom trawl gear must be configured to be towed across as wide a
⁴⁵ variety of habitats as possible, including seafloor habitats with complex physical structures.

⁴⁶ Indices of abundance at age and size derived from fishery-independent bottom trawl surveys
⁴⁷ are scaled to population size by the survey catchability (q) parameter (Arreguín-Sánchez,
⁴⁸ 1996). Catchability is typically estimated internally within stock assessment models that
⁴⁹ incorporate fisheries landings, indices of abundance, and life history parameters. However,
⁵⁰ the amount or quality of data and degree of contrast in the time series is often such that this
⁵¹ parameter, and therefore the population size, is difficult to estimate (Maunder and Piner,
⁵² 2015). In such cases, estimates of survey catchability from auxiliary data can inform the stock
⁵³ assessment. These external estimates can be used as a direct input into the assessment model
⁵⁴ (Somerton et al., 1999), can serve as a diagnostic measure of model accuracy (Miller et al.,
⁵⁵ 2019), or contribute to an alternate means of providing catch advice when an assessment
⁵⁶ model is not considered acceptable (Legault and McCurdy, 2017).

⁵⁷ Catchability can be decomposed into two components, the proportion of the population
⁵⁸ available to the survey sampling frame and the efficiency of the survey gear given an indi-
⁵⁹ vidual is available to the gear (Paloheimo and Dickie, 1964). Here efficiency is the fraction
⁶⁰ of available fish retained by the gear, equivalent to availability-selection in Millar and Fryer
⁶¹ (1999). Estimates of these components allow relative abundance indices to be converted
⁶² into absolute abundance indices without a population model. As such, investigations of gear

63 mensuration (Kotwicki et al., 2011), species-specific gear efficiency (Thygesen et al., 2019),
64 and availability of the stock to the survey design frame (Nichol et al., 2019) improve our
65 understanding of catchability and therefore abundance of fish stocks.

66 Paired-gear studies where two gears are fished either concurrently or close together tempo-
67 rally and spatially have long been used to estimate the efficiency of one fishing gear relative
68 to another (e.g., Gulland, 1964; Bourne, 1965). Of the two gears, one is often a reference
69 gear that may be a gear currently used for annual surveys (e.g., Munro and Somerton, 2001).

70 Typically neither of the gears is fully efficient and therefore the relative efficiency of gears
71 is estimated (e.g., Miller, 2013; Kotwicki et al., 2017), but there are cases where one of the
72 gears is assumed to be very nearly fully efficient (e.g., Somerton et al., 2013; Miller et al.,
73 2019).

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

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

86 In conducting paired-gear studies it is ideal to have the two gears deployed as close together
87 spatially and temporally as possible to reduce variation between the gears in densities of
88 the species being encountered. The twin-trawl rigging (Krag et al., 2015) where two trawls

89 can be fished simultaneously approaches this ideal (ICES, 1996), and is the data-collection
90 platform chosen by the Trawl Advisory Panel. The Panel decided to rig one of the twin
91 trawls as the the gear used by the bottom trawl survey which uses a rockhopper sweep. The
92 other trawl was rigged similarly except with a chain sweep in an attempt to eliminate any
93 escapement of fish under the gear. Assuming the chain sweep-based gear is fully efficient
94 allows the efficiency of the rockhopper sweep-based gear used by the bottom trawl survey to
95 be estimated from these experiments.

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

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

¹¹⁵ **2 Methods**

¹¹⁶ **2.1 Data collection**

¹¹⁷ Data were collected during three field experiments carried out in 2015, 2016, and 2017,
¹¹⁸ respectively, aboard the *F/V Karen Elizabeth*, a 23.8m (78ft) stern trawler capable of towing
¹¹⁹ two trawls simultaneously side by side (Figure 1). One side of the twin-trawl rig towed
¹²⁰ a NEFSC standard 400 x 12 cm survey bottom trawl rigged with the NEFSC standard
¹²¹ rockhopper sweep (Politis et al., 2014) (Figure 2). The other side of the twin-trawl rig
¹²² towed a version of the NEFSC 400 x 12cm survey bottom trawl modified to maximize the
¹²³ capture of flatfish. The modifications included reducing the headline flotation from 66 to 32,
¹²⁴ 20cm, spherical floats, reducing the port and starboard top wing-end extensions by 50cm
¹²⁵ each, and utilizing a chain sweep. The chain sweep was constructed of 1.6cm ($\frac{5}{8}$ in) trawl
¹²⁶ chain covered by 12.7cm diameter x 1cm thick rubber discs on every other chain link (Figure
¹²⁷ 2). Two rows of 1.3cm ($\frac{1}{2}$ in) tickler chains were attached to the 1.6cm trawl chain by 1.3cm
¹²⁸ shackles. To ensure equivalent net geometry of each gear, 32m restrictor ropes, made of 1.4cm
¹²⁹ ($\frac{9}{16}$ in) buoyant, Polytron rope, were attached between each of the trawl doors and the center
¹³⁰ clump. 3.4m^2 Thyboron Type 4 trawl doors were used to provide enough spreading force to
¹³¹ ensure the restrictor ropes remained taut throughout each tow. Each trawl used the NEFSC
¹³² standard 36.6m bridles. All tows followed the NEFSC standard survey towing protocols of
¹³³ 20 minutes at 3.0 knots. Port and starboard net spreads were measured separately with two
¹³⁴ sets of Simrad ITI acoustic net mensuration sensors measuring from the port wing-end to
¹³⁵ the center clump and the starboard wing-end to the center clump. In 2015, 108 (45 day, 63
¹³⁶ night) paired tows were conducted in eastern Georges Bank and off of southern New England
¹³⁷ (Figure 3). In 2016, 117 (74 day, 43 night) paired tows were conducted in western Gulf of
¹³⁸ Maine and the northern edge of Georges Bank. In 2017, 103 (61 day, 42 night) paired tows
¹³⁹ were conducted in the western Gulf of Maine and off of southern New England. Paired tows
¹⁴⁰ were denoted as “day” and “night” by whether the sun was above or below the horizon at

¹⁴¹ the time of the tow.

¹⁴² In order to reduce shipboard processing time and maximize the number of tows, only select
¹⁴³ taxa were enumerated and measured, rather than the full processing of all species as oc-
¹⁴⁴ curs on the trawl survey (Politis et al., 2014). All flatfish species (order Pleuronectiformes),
¹⁴⁵ thorny skate (*Amblyraja radiata*), barndoor skate (*Dipturus laevis*) and goosefish (*Lophias*
¹⁴⁶ *americanus*) collected in each net of each tow were independently sorted, weighed and mea-
¹⁴⁷ sured in all years. If the catch of a species was greater than \approx 150 individuals, a subsample
¹⁴⁸ of \approx 150 individuals was measured. Red hake (*Urophycis chuss*) were not quantified during
¹⁴⁹ the 2015 and 2016 sampling because other species were prioritized, but were fully processed
¹⁵⁰ in 2017. Winter skate (*Leucoraja ocellata*) and little skate (*L. erinacea*) were weighed in
¹⁵¹ all years and but were not separated to species nor measured. Sea scallops were weighed in
¹⁵² 2015 and 2016, but not 2017.

¹⁵³ 2.2 Paired-tow analysis

¹⁵⁴ We employed the hierarchical modeling approach from Miller (2013) to estimate the efficiency
¹⁵⁵ (ρ) of the rockhopper sweep used by the NEFSC bottom trawl survey relative to the chain
¹⁵⁶ sweep-based gear for ten species (Summer flounder, *Paralichthys dentatus*; American plaice,
¹⁵⁷ *Hippoglossoides platessoides*; windowpane flounder, *Scophthalmus aquosus*; winter flounder,
¹⁵⁸ *Pseudopleuronectes americanus*; yellowtail flounder, *Limanda ferruginea*; witch flounder,
¹⁵⁹ *Glyptocephalus cynoglossus*; red hake; goosefish; barndoor skate; thorny skate) from three
¹⁶⁰ separate research trips carried out aboard a twin trawl vessel. We first fit and compared
¹⁶¹ the same set of 13 models as Miller (2013) with different assumptions about variation of
¹⁶² relative efficiency between paired gear tows, size effects on the relative efficiency, and extra-
¹⁶³ binomial variation of observations within paired gear tows. The binomial (BI_0 to BI_4) and
¹⁶⁴ beta-binomial (BB_0 to BB_7) models that were fitted for all species are described in Table
¹⁶⁵ 2 including pseudo-formulas analogous to those used to specify and fit mixed or general-

166 sized additive models in R (R Core Team, 2019; Wood, 2006). We then also included diel
167 effects on relative catch efficiency and interactions with size effects with the best performing
168 model of the original 13 models for each species. To fit these diel effects, we generalized the
169 modeling framework somewhat in that we allowed multiple (cubic regression spline) smooth
170 effects, differing by day and night, on relative catch efficiency. We implemented the mod-
171 els using the Template Model Builder package (Kristensen et al., 2016) in R and we used
172 the “nlminb” optimizer to fit the models by maximizing the Laplace approximation of the
173 marginal likelihood (R Core Team, 2019).

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

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

189 For models BI₄, BB₆, and BB₇, there are two fixed effects parameters associated with the
190 spline coefficients that are treated as random effects for station-specific smoothers and the
191 correlation of these pairs of random effects is estimated. However, this parameter was not

192 estimable for red hake for BB₆ and assumed equal to zero.

193 2.3 Length-weight analysis

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

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

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

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

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

202 2.4 Biomass estimation

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

210 rescaled,

$$\widetilde{N}_{hi}(L) = N_{hi}(L) \hat{\rho}_i(L) \quad (3)$$

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

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

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

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

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

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

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

231 of vessel mechanical failure and indices are not available for summer flounder, SNE-MA
232 windowpane, and SNE-MA yellowtail flounder.

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

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

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

248 where

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

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

250 and K is the number of bootstraps.

251 For summer flounder it was necessary to omit one of the 1000 bootstraps of relative catch
252 efficiency at length due to an extremely large value to which the standard deviation and

253 mean of the bootstraps were sensitive. Finally, just as the uncertainty in $\rho(L)$ affects the
 254 uncertainty in the calibrated abundance at length and biomass estimates, it also induces
 255 correlation among the annual and seasonal estimates because the same estimates are applied
 256 to all of them. We calculated the correlation of annual biomass estimates for years y and z
 257 using the bootstrap estimates of biomass

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

258 where the covariance is

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

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

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

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

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

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

265 **3 Results**

266 **3.1 Paired-tow observations**

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

277 **3.2 Relative catch efficiency**

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

286 Including diel effects on relative catch efficiency improved model performance for all species
287 except American plaice (Table 5). For those species with diel effects on relative catch effi-
288 ciency, the ratio of the efficiencies was generally greater for daytime observations than those

289 for nighttime tows, with the exception of large winter flounder (Figure 4). The largest dif-
290 ferences in efficiency was estimated for smaller barndoor skate. For most of the species, the
291 differences in efficiency between the gears was generally greater for smaller individuals. The
292 large variability in the empirical estimates of the relative efficiency at size for each paired
293 tow is reflected in the variation in the posterior smooth estimates of relative efficiency at
294 size for each paired tow.

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

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

307 **3.3 Biomass estimation**

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

³¹⁴ stocks, except that southern New England winter flounder and northern goosefish estimates
³¹⁵ were typically greater in the fall than the spring.

³¹⁶ The relative catch efficiency of the rockhopper and chan sweep gears in terms of biomass
³¹⁷ varies across survey years and seasons due primarily to differences in size composition, but
³¹⁸ also variation in estimated length-weight relationship parameters (Figure 6). The efficiency of
³¹⁹ the bottom trawl survey gear was greatest for the winter flounder stocks and American plaice
³²⁰ (0.6 to 0.9) and least for red hake, witch flounder, windowpane, and yellowtail flounder stocks
³²¹ (0.2 to 0.4). Precision of the estimated annual biomass efficiencies was lowest for Georges
³²² Bank winter flounder and the skate stocks. For Gulf of Maine winter flounder, southern red
³²³ hake, and barndoor skate, the average fall biomass efficiencies were typically greater than in
³²⁴ the spring although the differences were small relative to the confidence intervals.

³²⁵ Comparing the average of estimated coefficients of variation for annual calibrated and uncal-
³²⁶ ibrated biomass estimates showed large increases for summer flounder in the fall (> 50%),
³²⁷ southern New England winter flounder in the spring (77%), Georges Bank winter flounder
³²⁸ (more than 200% for spring and fall), northern red hake (95% for spring and 178% for fall),
³²⁹ northern goosefish in the fall (93%), and barndoor skate (> 100% for both spring and fall)
³³⁰ induced by the variability in the estimation of the relative catch efficiency of the gears using
³³¹ chain and rockhopper sweep gears (Table 6). The effect of calibration on the precision of
³³² the biomass estimates was relatively minor for other stocks.

³³³ We observed little correlation of annual biomass estimates induced by the relative catch
³³⁴ efficiency estimation for most of the stocks (Table 7). However, the biomass estimates were
³³⁵ highly correlated for George Bank winter flounder in the spring (65%) and barndoor skate
³³⁶ (> 70% on average). Estimates for Georges Bank winter flounder in the fall, both red hake
³³⁷ stocks, northern goosefish, and thorny skate were greater than 20% on average.

³³⁸ **4 Discussion**

³³⁹ The data that we used to estimate bottom trawl survey catch efficiency came from an
³⁴⁰ experiment using a twin trawler and many of the standard tow protocols for the NEFSC
³⁴¹ survey on the *Bigelow*. The experimental net used on one side of the twin trawl was the
³⁴² same as the standard survey trawl used on the *Bigelow* except that it contained roughly half
³⁴³ the number of floats and the sweep was modified to optimize flatfish catch by minimizing
³⁴⁴ the ability of flatfish to pass under the net. The other side of the twin trawl was essentially
³⁴⁵ identical to the standard gear used on the *Bigelow*. The towing of the standard survey
³⁴⁶ bottom trawl on the twin trawl experiment differed in a few ways from its deployment on
³⁴⁷ the spring and fall bottom trawl surveys, but we believe that these differences did not have
³⁴⁸ a significant effect on the results. The use of larger doors and the restrictor rope served to
³⁴⁹ fix the net geometries which may be the biggest source of variability in comparative trawl
³⁵⁰ catches (Jones et al., 2021). This setup also allowed us to avoid many of the potential
³⁵¹ problems due to the large differences in size of the *Bigelow* and the *F/V Karen Elizabeth*.
³⁵² We do not suspect that the use of the restrictor rope would influence flatfish behavior in
³⁵³ front of the trawl because flatfish have been shown to generally not react to trawling induced
³⁵⁴ stimuli until they are in very close proximity or even contacted by the fishing gear (Ryer
³⁵⁵ et al., 2010). The spread data indicated that the restrictor rope remained taut throughout
³⁵⁶ the towing process (setting, towing, hauling back), so we believe it likely that the restrictor
³⁵⁷ rope was almost always at least 1m off the bottom. Our concerns about potential effects of
³⁵⁸ the restrictor rope on species that spend more time off the ground (e.g., Atlantic cod, *Gadus*
³⁵⁹ *morhua*) led us to exclude them from analyses.

³⁶⁰ Herding is a known phenomenon for flatfish and many other species when certain types of
³⁶¹ gear are used (Ramm and Xiao, 1995; Somerton and Munro, 2001; Somerton et al., 2007;
³⁶² Rose et al., 2010). Somerton and Munro (2001) considered two factors of bridle herding
³⁶³ effects on efficiency. The first factor was the size of the bridle path where the bridle is off

364 the ground (W_{off}) and the second factor, the herding efficiency (h) was the fraction of fish
365 in the bridle contact path moved into the path of the net. The former is a function of gear
366 design, and controllable, whereas the latter is a function of fish behavior with regard to
367 the bridle when it is in contact with the substrate. The bridle configuration on the bottom
368 trawl survey are designed to minimize contact with the bottom and lack of abrasion of
369 painted bridles used during one of the twin trawler research trips provided evidence of little
370 or no bridle contact during the paired tow experiments used to collect the data used in this
371 study. Furthermore, studies have consistently found that herding behavior occurred during
372 the daytime (Glass and Wardle, 1989; Somerton and Munro, 2001; Ryer and Barnett, 2006;
373 Bryan et al., 2014; Ryer et al., 2010; Dean et al., 2021) with some studies indicating high
374 herding coefficients (h) along the sections of the bridles in contact with the bottom. Studies
375 that have evaluated herding at night or in low light conditions did not find evidence for a
376 directional herding response (Glass and Wardle, 1989; Ryer and Barnett, 2006; Ryer, 2008;
377 Ryer et al., 2010). The minimal bridle contact with the substrate and the large fraction of
378 nighttime tows during the bottom trawl survey suggests flatfish herding is unlikely to be an
379 important factor in catch efficiency based on net spread-based swept area.

380 On the other hand, the biomass estimates assume that the chain sweep gear is fully efficient,
381 but it is likely at least some small fraction of fish, that may depend on size, are not captured
382 by the gear. The biomass estimates also implicitly assume that the entire stock is available
383 to the bottom trawl survey, but many of these stocks extend somewhat outside of the survey
384 strata used to define the indices throughout the year and(or) seasonally due to migration.
385 If either of these assumptions are incorrect this method of biomass estimation would be
386 negatively biased (expected value of biomass estimates would be lower than the true value).
387 However, estimation using the data from these paired-gear studies and these assumptions is
388 less biased than those made without them.
389 These analyses treat the amount of daylight as binary effect (day/night) on the relative catch
390 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-

⁴¹⁸ based assessment methods (Legault and McCurdy, 2017; NEFSC, 2020a,b). These estimates
⁴¹⁹ can be especially valuable for index-based methods where the scale of the stock is assumed
⁴²⁰ known. The abundance estimates have also been used in a supporting fashion for fall-back
⁴²¹ assessments of both Gulf of Maine-Georges Bank and southern New England-Mid-Atlantic
⁴²² windowpane stocks (NEFSC, 2020b).

⁴²³ Typically, research surveys provide only a relative index of abundance rather than an absolute
⁴²⁴ estimate of abundance. Stock assessment models then integrate these observations with time
⁴²⁵ series of catch and other data sources to determine the scale of the population. However,
⁴²⁶ various factors can make for imprecise and inaccurate scaling of population levels including
⁴²⁷ inaccurate catch data (Cadigan, 2016), time-varying catchability (Wilberg et al., 2009), low
⁴²⁸ fishing mortality rates over the time series (Adams et al., 2015), and uncertain and time-
⁴²⁹ varying natural mortality (Stock et al., 2021). In these cases, external information such as
⁴³⁰ those produced by studies such as ours, can be particularly useful in estimating the size of
⁴³¹ of the stock, the status of the stock relative to optimal levels and ultimately making catch
⁴³² advice for commercially important fish stocks.

⁴³³ Acknowledgements

⁴³⁴ This work was funded by NOAA Fisheries Northeast Fisheries Science Center. The quality
⁴³⁵ of this manuscript has been improved by comments from Jessica Blaylock. Tor Bendiksen
⁴³⁶ and John Knight assisted with the design of the chain sweep. Jeff Pessutti, Giovanni Giane-
⁴³⁷ sen, Dominique St. Amand, and Calvin Alexander helped collect paired-tow data for these
⁴³⁸ analyses. We thank Adam Poquette for creating the map in Figure 3. We thank Calvin
⁴³⁹ Alexander also for the photo of the twin trawl gear in Figure 2.

⁴⁴⁰ **CRediT authorship contribution statement**

⁴⁴¹ **Timothy J. Miller:** Conceptualization, Methodology, Writing - original draft, Formal
⁴⁴² analysis, Visualization. **David E. Richardson:** Conceptualization, Methodology, Writing
⁴⁴³ - original draft. **Philip J. Politis:** Investigation, Methodology, Writing - review and editing.
⁴⁴⁴ **Christopher D. Roebuck:** Project administration, Conceptualization, Funding acquisi-
⁴⁴⁵ tion, Investigation, Resources. **John P. Manderson:** Conceptualization, Investigation,
⁴⁴⁶ Writing - review and editing. **Michael H. Martin:** Project administration, Conceptual-
⁴⁴⁷ ization, Data curation, Writing - review and editing. **Andrew W. Jones:** Visualization,
⁴⁴⁸ Writing - review and editing.

⁴⁴⁹ **References**

- ⁴⁵⁰ Adams, C. F., Miller, T. J., Manderson, J. P., Richardson, D. E., and Smith, B. E. 2015.
⁴⁵¹ Butterfish 2014 Assessment. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-06;
⁴⁵² 110 p.
- ⁴⁵³ Arreguín-Sánchez, F. 1996. Catchability: a key parameter for fish stock assessment. Reviews
⁴⁵⁴ in Fish Biology and Fisheries **6**: 221–242, doi: 10.1007/BF00182344.
- ⁴⁵⁵ Azarovitz, T. S. 1981. A brief historical review of the Woods Hole Laboratory trawl survey
⁴⁵⁶ time series. In Bottom Trawl Surveys. Edited by W. G. Doubleday and D. Rivard. Can.
⁴⁵⁷ Spec. Pub. Fish. Aqua. Sci. No. 58. pp. 62–67.
- ⁴⁵⁸ Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekking, A., Kraan, C., van der Veer, H. W.,
⁴⁵⁹ and Piersma, T. 2012. Designing a benthic monitoring programme with multiple conflict-
⁴⁶⁰ ing objectives. Methods in Ecology and Evolution **3**(3): 526–536, doi: 10.1111/j.2041-
⁴⁶¹ 210X.2012.00192.x.
- ⁴⁶² Bourne, N. 1965. A comparison of catches by 3- and 4-inch rings on offshore scallop drags.
⁴⁶³ Journal of the Fisheries Research Board of Canada **22**(2): 313–333.
- ⁴⁶⁴ Bryan, D. R., Bosley, K. L., Hicks, A. C., Haltuch, M. A., and Wakefield, W. W. 2014.
⁴⁶⁵ Quantitative video analysis of flatfish herding behavior and impact on effective area swept
⁴⁶⁶ of a survey trawl. Fisheries Research **154**: 120–126, doi: 10.1016/j.fishres.2014.02.007.
- ⁴⁶⁷ Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-
⁴⁶⁸ reported catches and variable natural mortality rates. Canadian Journal of Fisheries and
⁴⁶⁹ Aquatic Sciences **73**(2): 296–308, doi: 10.1139/cjfas-2015-0047.
- ⁴⁷⁰ Dean, M. J., Hoffman, W. S., Buchan, N. C., Cadrin, S. X., and Grabowski, J. H. 2021.
⁴⁷¹ The influence of trawl efficiency assumptions on survey-based population metrics. ICES
⁴⁷² Journal of Marine Science **78**(8): 2858–2874.

- 473 Glass, C. W. and Wardle, C. S. 1989. Comparison of the reactions of fish to a trawl gear,
474 at high and low light intensities. *Fisheries Research* **7**(3): 249–266, doi: 10.1016/0165-
475 7836(89)90059-3.
- 476 Gulland, J. A. 1964. Variations in selection factors, and mesh differentials. *Journal Du Con-*
477 *seil International Pour L'exploration De La Mer* **29**(2): 158–165.
- 478 ICES. 1996. Manual of methods of measuring the selectivity of towed fishing gears. (Eds.)
479 Wileman, D. A., Ferro, R. S. T., Fonteyne, R., and Millar, R. B. ICES Cooperative
480 Research Report No. 215.
- 481 Jacobson, L. D., Hendrickson, L. C., and Tang, J. 2015. Solar zenith angles for biological
482 research and an expected catch model for diel vertical migration patterns that affect stock
483 size estimates for longfin inshore squid (*Doryteuthis pealeii*). *Canadian Journal of Fisheries*
484 and Aquatic Sciences
- 485 Jones, A. W., Miller, T. J., Politis, P. J., Richardson, D. E., Mercer, A. M., Pol, M. V., and
486 Roebuck, C. D. 2021. Experimental assessment of the effect of net wing spread on relative
487 catch efficiency of four flatfishes by a four seam bottom trawl. *Fisheries Research* **244**:
488 106106, doi: 10.1016/j.fishres.2021.106106.
- 489 Kainge, P., van der Plas, A. K., Bartholomae, C. H., and Wieland, K. 2017. Effects of
490 environmental variables on survey catch rates and distribution by size of shallow- and deep-
491 water cape hakes, *merluccius capensis* and *merluccius paradoxus* off namibia. *Fisheries*
492 *Oceanography* **26**(6): 680–692, doi: 10.1111/fog.12227.
- 493 Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2013. Combining
494 bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl
495 efficiency parameters for semipelagic species. *Canadian Journal of Fisheries and Aquatic*
496 *Sciences* **70**(2): 208–219, doi: 10.1139/cjfas-2012-0321.

- 497 Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. 2017. Selectivity ratio: A
498 useful tool for comparing size selectivity of multiple survey gears. *Fisheries Research* **191**:
499 76–86, doi: 10.1016/j.fishres.2017.02.012.
- 500 Kotwicki, S., Martin, M. H., and Laman, E. A. 2011. Improving area swept es-
501 timates from bottom trawl surveys. *Fisheries Research* **110**(1): 198–206, doi:
502 10.1016/j.fishres.2011.04.007.
- 503 Krag, L. A., Herrmann, B., Karlsen, J. D., and Mieske, B. 2015. Species selectivity in
504 different sized topless trawl designs: Does size matter? *Fisheries Research* **172**: 243–249,
505 doi: 10.1016/j.fishres.2015.07.010.
- 506 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic
507 differentiation and Laplace approximation. *Journal of Statistical Software* **70**(5): 1–21.
- 508 Legault, C. M. and McCurdy, Q. M. 2017. Stock assessment of Georges Bank yellowtail
509 flounder for 2017. Transboundary Resources Assessment Committee (TRAC) Reference
510 Document 2017/03 available at: <https://repository.library.noaa.gov/view/noaa/27556>.
- 511 Maunder, M. N. and Piner, K. R. 2015. Contemporary fisheries stock assessment: many issues
512 still remain. *ICES Journal of Marine Science* **72**(1): 7–18, doi: 10.1093/icesjms/fsu015.
- 513 Millar, R. B. and Fryer, R. J. 1999. Estimating the size-selection curves of towed gears,
514 traps, nets and hooks. *Reviews in Fish Biology and Fisheries* **9**(1): 89–116, doi:
515 10.1023/A:1008838220001.
- 516 Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on
517 paired-gear data for U.S. Northwest Atlantic fish stocks. *Canadian Journal of Fisheries
518 and Aquatic Sciences* **70**(9): 1306–1316, doi: 10.1139/cjfas-2013-0136.
- 519 Miller, T. J., Hart, D. R., Hopkins, K., Vine, N. H., Taylor, R., York, A. D., and Gal-
520 lager, S. M. 2019. Estimation of the capture efficiency and abundance of Atlantic sea

- 521 scallops (*Placopecten magellanicus*) from paired photographic–dredge tows using hierar-
522 chical models. Canadian Journal of Fisheries and Aquatic Sciences **76**(6): 847–855, doi:
523 10.1139/cjfas-2018-0024.
- 524 Munro, P. T. and Somerton, D. A. 2001. Maximum likelihood and non-parametric methods
525 for estimating trawl footrope selectivity. ICES Journal of Marine Science **58**(1): 220–229,
526 doi: 10.1006/jmsc.2000.1004.
- 527 NEFSC. 2019. 66th Northeast Regional Stock Assessment Workshop (66th SAW) Assessment
528 Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 19-08; 1170 p. Available from:
529 <http://www.nefsc.noaa.gov/publications/>.
- 530 NEFSC. 2020a. Final report of red hake stock structure working group. US Dept Commer,
531 Northeast Fish Sci Cent Ref Doc. 20-07; 185 p.
- 532 NEFSC. 2020b. Operational assessment of 14 northeast groundfish stocks, updated through
533 2018. Prepublication Copy of the September 2019 Operational Stock Assessment Report.
534 The report is “in preparation” for publication by the NEFSC. (Oct. 3, 2019; latest revision:
535 Jan. 7, 2020).
- 536 NEFSC. 2020c. Operational assessment of the black sea bass, scup, bluefish, and monkfish
537 stocks, updated through 2018. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 20-01;
538 160 p.
- 539 Nichol, D. G., Kotwicki, S., Wilderbuer, T. K., Lauth, R. R., and Ianelli, J. N. 2019.
540 Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey
541 and its effect on estimates of survey biomass. Fisheries Research **211**: 319–330, doi:
542 10.1016/j.fishres.2018.11.017.
- 543 Paloheimo, J. E. and Dickie, L. M. 1964. Abundance and fishing success. Rapports et Procès-
544 Verbaux des Réunions. Conseil Internationale pour l’Exploration de la Mer **155**: 152–163.

- 545 Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an
546 application. Canadian Journal of Fisheries and Aquatic Sciences **55**(12): 2672–2690, doi:
547 10.1139/f98-151.
- 548 Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. 2014. Northeast Fisheries
549 Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. U.S.
550 Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 14-06, 138p.
- 551 R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Founda-
552 tion for Statistical Computing, Vienna, Austria.
- 553 Ramm, D. C. and Xiao, Y. 1995. Herding in groundfish and effective pathwidth of trawls.
554 Fisheries Research **24**(3): 243–259, doi: 10.1016/0165-7836(95)00373-I.
- 555 Rose, C. S., Gauvin, J. R., and Hammond, C. F. 2010. Effective herding of flatfish by cables
556 with minimal seafloor contact. Fishery Bulletin **108**(2): 136–144.
- 557 Ryer, C. H. 2008. A review of flatfish behavior relative to trawls. Fisheries Research **90**(1):
558 138–146, doi: 10.1016/j.fishres.2007.10.005.
- 559 Ryer, C. H. and Barnett, L. A. K. 2006. Influence of illumination and temperature upon
560 flatfish reactivity and herding behavior: Potential implications for trawl capture efficiency.
561 Fisheries Research **81**(2): 242–250, doi: 10.1016/j.fishres.2006.07.001.
- 562 Ryer, C. H., Rose, C. S., and Iseri, P. J. 2010. Flatfish herding behavior in response to
563 trawl sweeps: a comparison of diel responses to conventional sweeps and elevated sweeps.
564 Fishery Bulletin **108**(2): 145–154.
- 565 Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean
566 abundance. Canadian Journal of Fisheries and Aquatic Sciences **54**(3): 616–630, doi:
567 10.1139/f96-303.

- 568 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.
- 572 Somerton, D. A. and Munro, P. 2001. Bridle efficiency of a survey trawl for flatfish. Fishery Bulletin **99**(4): 641–652.
- 574 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.
- 576 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.
- 580 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.
- 584 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.
- 588 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.
- 591 Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. 2018. Evaluation of alternative

- 592 stratifications for a stratified random fishery-independent survey. *Fisheries Research* **207**:
593 150–159, doi: 10.1016/j.fishres.2018.06.019.
- 594 Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2009. Incorporating time-
595 varying catchability into population dynamic stock assessment models. *Reviews in Fish-
596 eries Science* **18**(1): 7–24, doi: 10.1080/10641260903294647.
- 597 Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall,
598 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 important 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 were deployed varied throughout the experimental tows of each.

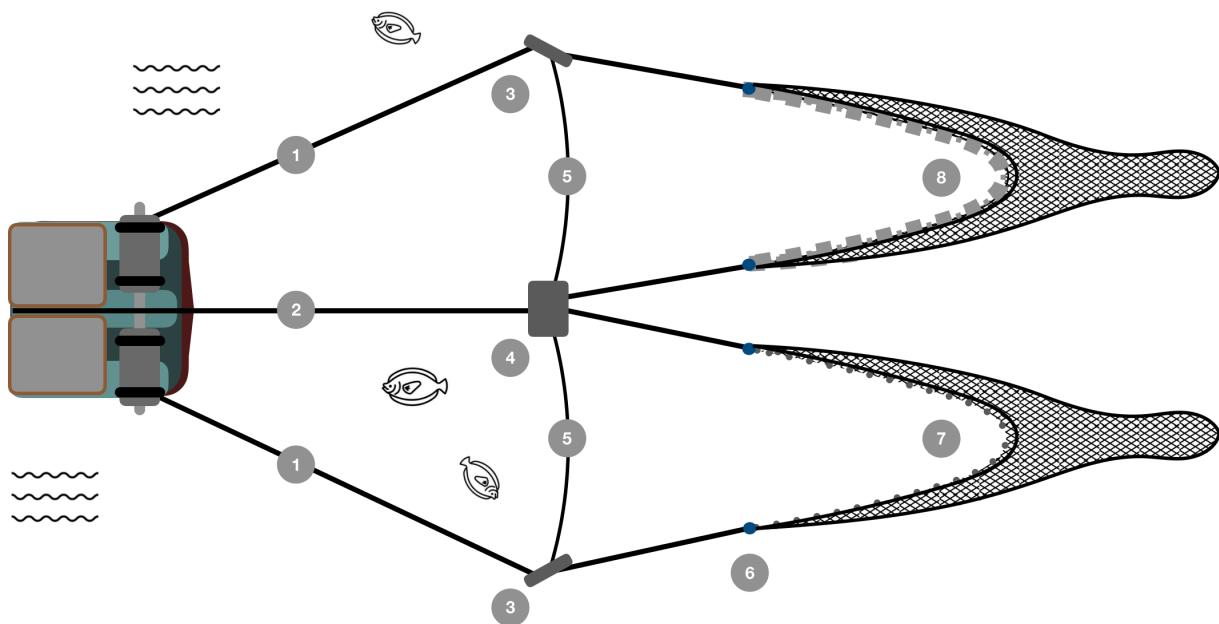


Fig. 2. The *F/V Karen Elizabeth* twin-trawl vessel rigged with rockhopper sweep gear on the right and chain sweep gear on the left.



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.

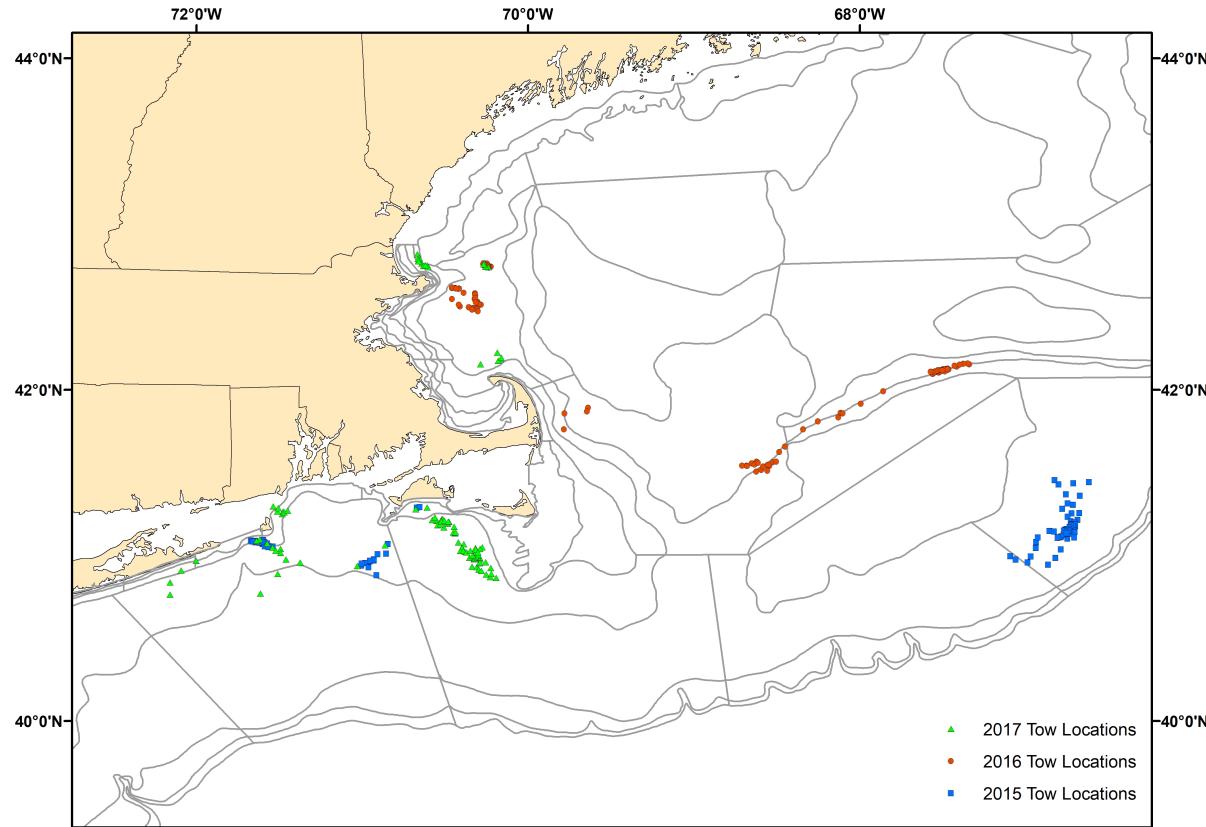


Fig. 4. Relative efficiency of gears using chain and rockhopper sweeps from the best performing model for each species (Table 5). Blue and red denote results for day and night data, respectively, and thick and thin lines represent overall and paired-tow specific estimates of relative catch efficiency, 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.

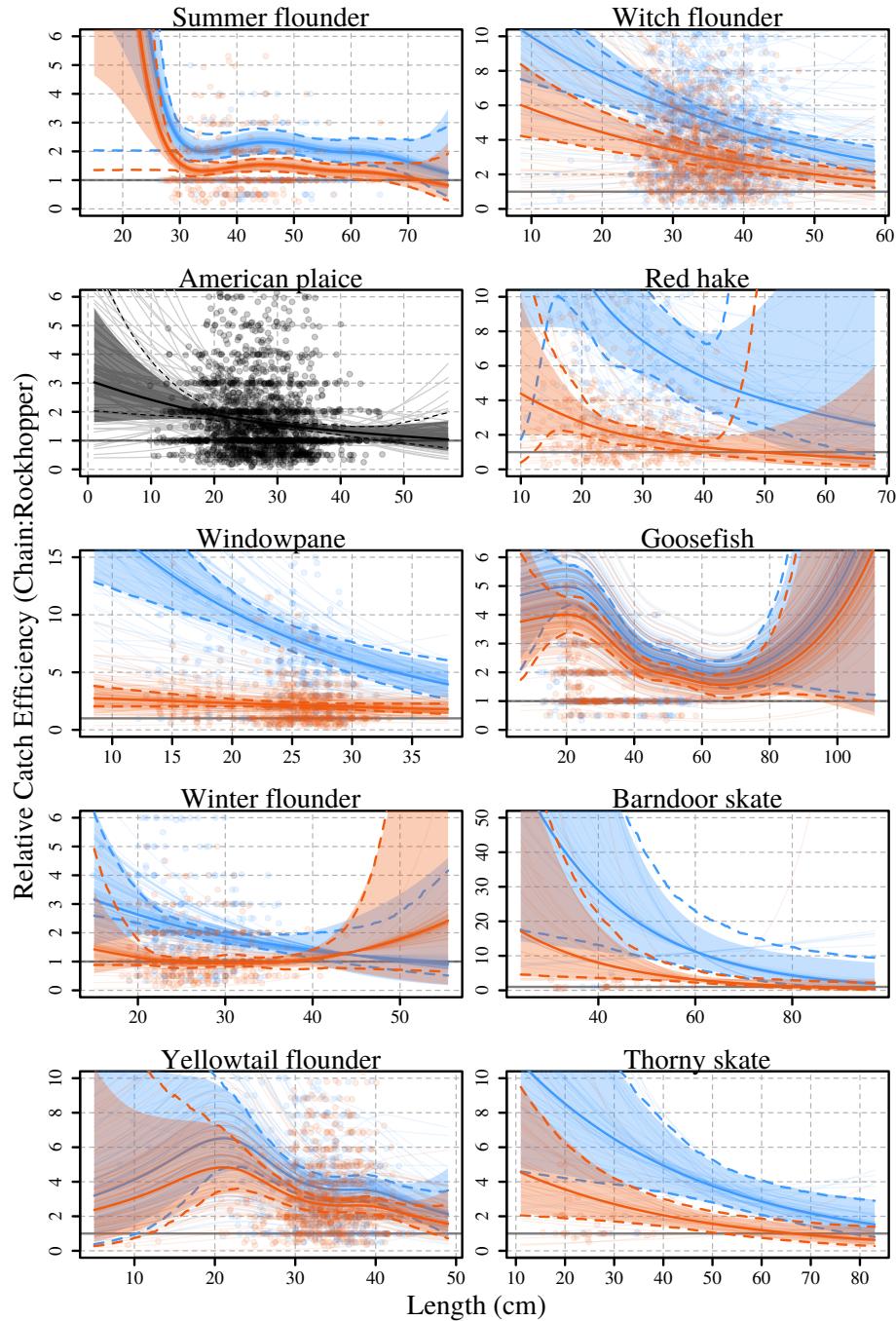


Fig. 5. Annual spring (blue) and fall (red) biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps are from the best performing model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

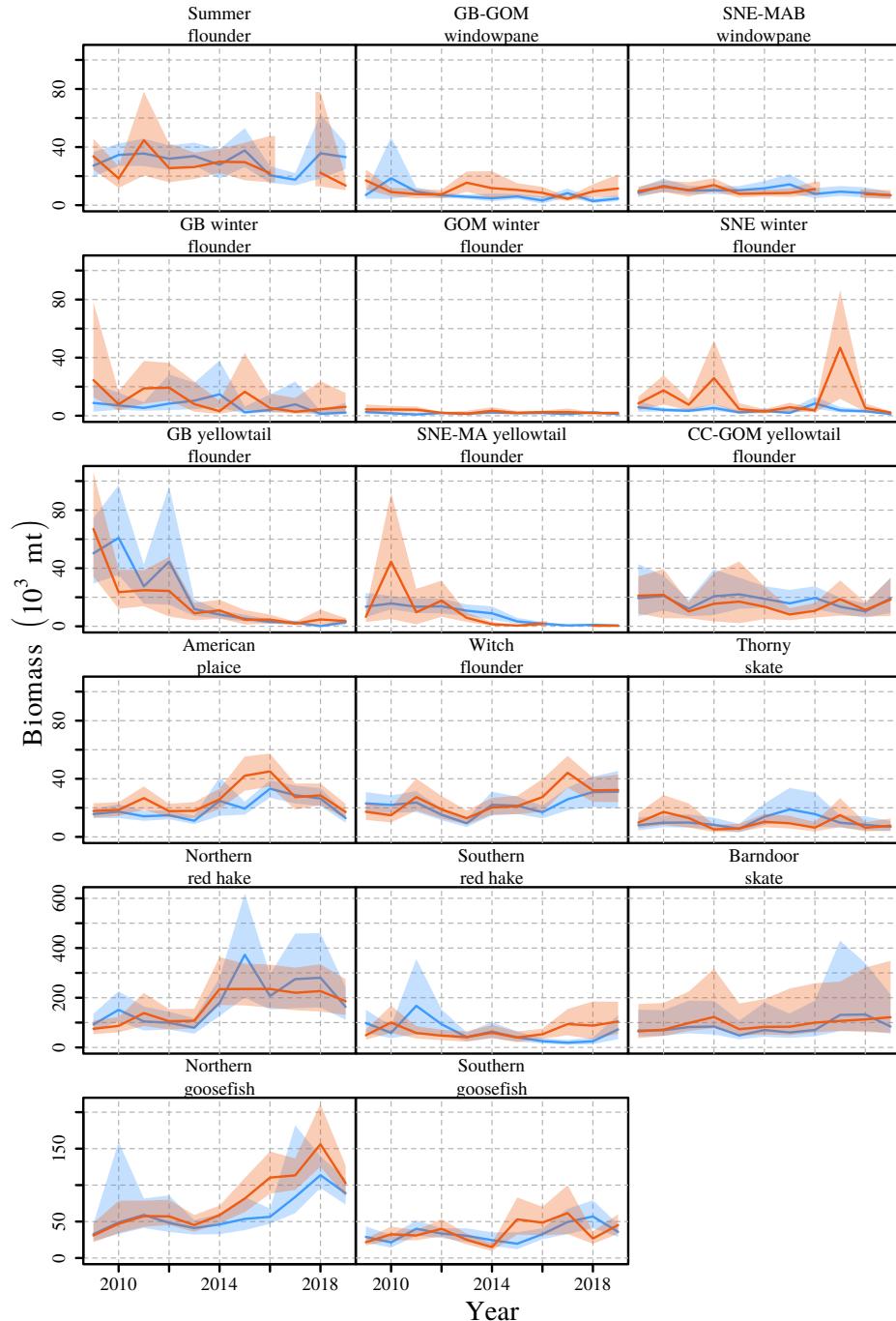


Fig. 6. Implied catch efficiency of annual spring (blue) and fall (red) bottom trawl survey biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and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

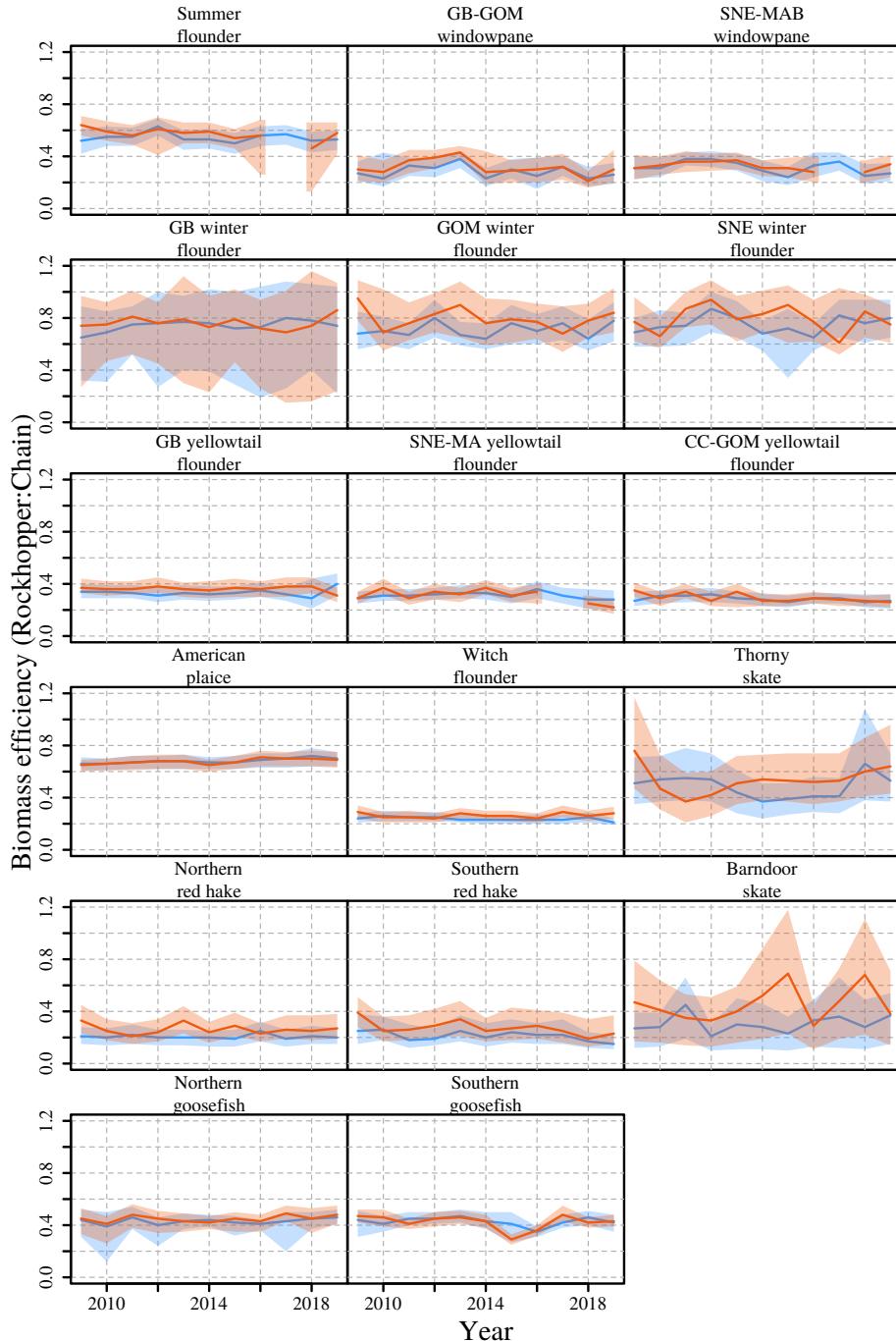


Table 1. Managed stocks associated with the species for which relative catch efficiency was estimated.

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	$\sim 1 + 1 pair$	—	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	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