

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

<sup>4</sup> Timothy J. Miller<sup>1,2</sup>      David E. Richardson<sup>3</sup>      Philip J. Politis<sup>2</sup>

<sup>5</sup> Christopher D. Roebuck<sup>4</sup>      John P. Manderson<sup>5,6</sup>

<sup>6</sup> Michael H. Martin<sup>2,7</sup>      Andrew W. Jones<sup>3</sup>

<sup>7</sup> <sup>1</sup>corresponding author: timothy.j.miller@noaa.gov

<sup>8</sup> <sup>2</sup>Northeast Fisheries Science Center, Woods Hole Laboratory, 166 Water Street, Woods Hole,  
<sup>9</sup> MA 02543 USA

<sup>10</sup> <sup>3</sup>Northeast Fisheries Science Center, Narragansett Laboratory, 28 Tarzwell Drive Narra-  
<sup>11</sup> gansett, RI 02882 USA

<sup>12</sup> <sup>4</sup>Fishing Vessel Karen Elizabeth, Narragansett, RI 02882, USA

<sup>13</sup> <sup>5</sup>Northeast Fisheries Science Center, James J. Howard Marine Sciences Laboratory, 74 Ma-  
<sup>14</sup> gruder Road Highlands, NJ 07732 USA

<sup>15</sup> <sup>6</sup>Current address: OpenOcean Research, Suite 101, 40 West Evergreen Avenue, Philadelphia,  
<sup>16</sup> PA, 19118 USA

<sup>17</sup> <sup>7</sup>Current address: Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Building 4  
<sup>18</sup> Seattle, WA 98115 USA

<sup>19</sup> **Abstract**

<sup>20</sup> 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.

<sup>36</sup> **Keywords**

<sup>37</sup> gear efficiency, biomass estimation, hierarchical generalized additive models

<sup>38</sup> **1 Introduction**

<sup>39</sup> Ecosystem monitoring surveys such as fisheries-independent trawl surveys are used to obtain  
<sup>40</sup> information on a range of species and are therefore not optimized with respect to sampling  
<sup>41</sup> design or gear for any one species (Bijleveld et al., 2012; Wang et al., 2018). Gear and  
<sup>42</sup> sampling protocols are designed to provide consistent and representative samples that allow  
<sup>43</sup> indices of abundance at size and age to be developed for a suite of species (Azarovitz, 1981;  
<sup>44</sup> Thiess et al., 2018). To provide indices of population abundance with minimal potential  
<sup>45</sup> sources of bias, survey bottom trawl gear must be configured to be towed across as wide a  
<sup>46</sup> variety of habitats as possible, including seafloor habitats with complex physical structures.

<sup>47</sup> Indices of abundance at age and size derived from fishery-independent bottom trawl surveys  
<sup>48</sup> are scaled to population size by the survey catchability ( $q$ ) parameter (Arreguín-Sánchez,  
<sup>49</sup> 1996). Catchability is typically estimated internally within stock assessment models that  
<sup>50</sup> incorporate fisheries landings, indices of abundance, and life history parameters. However,  
<sup>51</sup> the amount or quality of data and degree of contrast in the time series is often such that this  
<sup>52</sup> parameter, and therefore the population size, is difficult to estimate (Maunder and Piner,  
<sup>53</sup> 2015). In such cases, estimates of survey catchability from auxiliary data can inform the stock  
<sup>54</sup> assessment. These external estimates can be used as a direct input into the assessment model  
<sup>55</sup> (Somerton et al., 1999), can serve as a diagnostic measure of model accuracy (Miller et al.,  
<sup>56</sup> 2019), or contribute to an alternate means of providing catch advice when an assessment  
<sup>57</sup> model is not considered acceptable (Legault and McCurdy, 2017).

<sup>58</sup> Catchability can be decomposed into two components, the proportion of the population  
<sup>59</sup> available to the survey sampling frame and the efficiency of the survey gear given an indi-  
<sup>60</sup> vidual is available to the gear (Paloheimo and Dickie, 1964). Here efficiency is the fraction  
<sup>61</sup> of available fish retained by the gear, equivalent to availability-selection in Millar and Fryer  
<sup>62</sup> (1999). Estimates of these components allow relative abundance indices to be converted  
<sup>63</sup> into absolute abundance indices without a population model. As such, investigations of gear

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

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

71 Typically neither of the gears is fully efficient and therefore the relative efficiency of gears  
72 is estimated (e.g., Miller, 2013; Kotwicki et al., 2017), but there are cases where one of the  
73 gears is assumed to be very nearly fully efficient (e.g., Somerton et al., 2013; Miller et al.,  
74 2019).

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

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

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

90 can be fished simultaneously approaches this ideal (ICES, 1996), and is the data-collection  
91 platform chosen by the Trawl Advisory Panel. The Panel decided to rig one of the twin  
92 trawls as the gear used by the bottom trawl survey which uses a rockhopper sweep.  
93 The Trawl Advisory Panel decided to focus the experiments on efficiency for flatfishes, so  
94 the other trawl was rigged similarly except with a chain sweep in an attempt to eliminate  
95 any escapement of fish under the gear. The Panel thought that a chain sweep would limit  
96 escapement under the sweep better than a flat or cookie sweep or other potential sweep  
97 designs. The sweep was constructed of multiple layers of chain so as to maximize bottom  
98 contact and minimize loss but also reduce retaining debris in the net and the trawl hanging  
99 on obstructions. If the chain sweep-based gear is assumed to be fully efficient, the efficiency  
100 of the rockhopper sweep-based gear used by the bottom trawl survey can be estimated from  
101 these experiments.

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

112 The relative catch efficiency estimates provided by analyses of paired gear data have un-  
113 certainty which may not be propagated when applied to survey data to make estimates of  
114 abundance. The application to survey data also induces correlation of the annual (and sea-  
115 sonal) abundance estimates from these surveys. These indices are typically used as measures  
116 of relative abundance in stock assessment with the precision of the indices used to weight

117 the observations within the assessment model where the observations for each of the annual  
118 and seasonal indices is typically assumed to be independent of the others. Here we compare  
119 the precision of the biomass indices calibrated to the chain sweep gear to that of the and un-  
120 calibrated indices using the rockhopper sweep gear and measure the correlation of calibrated  
121 indices for each stock.

## 122 2 Methods

### 123 2.1 Data collection

124 Data were collected during three field experiments carried out in 2015, 2016, and 2017,  
125 respectively, aboard the *F/V Karen Elizabeth*, a 23.8m (78ft) stern trawler capable of towing  
126 two trawls simultaneously side by side (Figure 1). One side of the twin-trawl rig towed  
127 a NEFSC standard 400 x 12 cm survey bottom trawl rigged with the NEFSC standard  
128 rockhopper sweep (Politis et al., 2014) (Figure 2). The other side of the twin-trawl rig towed  
129 a modified version of the NEFSC 400 x 12cm survey bottom trawl with the intent of altering  
130 design characteristics of the standard survey trawl to improve bottom contact and maximize  
131 the capture of flatfish. The modifications included reducing the headline flotation from 66 to  
132 32, 20cm, spherical floats, reducing the port and starboard top wing-end extensions by 50cm  
133 each, and utilizing a chain sweep. The chain sweep was constructed of 1.6cm ( $\frac{5}{8}$ in) trawl  
134 chain covered by 12.7cm diameter x 1cm thick rubber discs on every other chain link (Figure  
135 2). Two rows of 1.3cm ( $\frac{1}{2}$ in) tickler chains were attached to the 1.6cm trawl chain by 1.3cm  
136 shackles. To ensure equivalent net geometry of each gear, 32m restrictor ropes, made of  
137 1.4cm ( $\frac{9}{16}$ in) buoyant, Polytron rope, were attached between each of the trawl doors and the  
138 center clump. 3.4m<sup>2</sup> Thyboron Type 4 trawl doors were used to provide enough spreading  
139 force to ensure the restrictor ropes remained taut throughout each tow. Each trawl used  
140 the NEFSC standard 36.6m bridles. Every tow was monitored with net mensuration sensors

<sup>141</sup> to verify bridals were held to optimal angles and identical spread. All tows followed the  
<sup>142</sup> NEFSC standard survey towing protocols of 20 minutes at 3.0 knots. Port and starboard  
<sup>143</sup> net spreads were measured separately with two sets of Simrad ITI acoustic net mensuration  
<sup>144</sup> sensors measuring from the port wing-end to the center clump and the starboard wing-end  
<sup>145</sup> to the center clump. In 2015, 108 (45 day, 63 night) paired tows were conducted in eastern  
<sup>146</sup> Georges Bank and off of southern New England (Figure 3). In 2016, 117 (74 day, 43 night)  
<sup>147</sup> paired tows were conducted in western Gulf of Maine and the northern edge of Georges  
<sup>148</sup> Bank. In 2017, 103 (61 day, 42 night) paired tows were conducted in the western Gulf of  
<sup>149</sup> Maine and off of southern New England. Paired tows were denoted as “day” and “night” by  
<sup>150</sup> whether the sun was above or below the horizon at the time of the tow.

<sup>151</sup> In order to reduce shipboard processing time and maximize the number of tows, only se-  
<sup>152</sup> lect taxa were enumerated and measured for total length, rather than the full processing  
<sup>153</sup> of all species as occurs on the trawl survey (Politis et al., 2014). All flatfish species (order  
<sup>154</sup> Pleuronectiformes), thorny skate (*Amblyraja radiata*), barndoor skate (*Dipturus laevis*) and  
<sup>155</sup> goosefish (*Lophias americanus*) collected in each net of each tow were independently sorted,  
<sup>156</sup> weighed and measured in all years. If the catch of a species was greater than  $\approx$ 150 individ-  
<sup>157</sup> uals, a subsample of  $\approx$ 150 individuals was measured. Red hake (*Urophycis chuss*) were not  
<sup>158</sup> quantified during the 2015 and 2016 sampling because other species were prioritized, but  
<sup>159</sup> were fully processed in 2017. Winter skate (*Leucoraja ocellata*) and little skate (*L. erinacea*)  
<sup>160</sup> were weighed in all years and but were not separated to species nor measured. Sea scallops  
<sup>161</sup> were weighed in 2015 and 2016, but not 2017.

## <sup>162</sup> 2.2 Paired-tow analysis

<sup>163</sup> We employed the hierarchical modeling approach from Miller (2013) to estimate the efficiency  
<sup>164</sup> ( $\rho$ ) of the rockhopper sweep used by the NEFSC bottom trawl survey relative to the chain  
<sup>165</sup> sweep-based gear for ten species (Summer flounder, *Paralichthys dentatus*; American plaice,

<sup>166</sup> *Hippoglossoides platessoides*; windowpane flounder, *Scophthalmus aquosus*; winter flounder,  
<sup>167</sup> *Pseudopleuronectes americanus*; yellowtail flounder, *Limanda ferruginea*; witch flounder,  
<sup>168</sup> *Glyptocephalus cynoglossus*; red hake; goosefish; barndoor skate; thorny skate) from the  
<sup>169</sup> data collected during the three trips carried out aboard the *F/V Karen Elizabeth*. We first  
<sup>170</sup> fit and compared the same set of 13 models as Miller (2013) with different assumptions  
<sup>171</sup> about variation of relative efficiency between paired gear tows, size effects on the relative  
<sup>172</sup> efficiency, and extra-binomial variation of observations within paired gear tows. The binomial  
<sup>173</sup> ( $BI_0$  to  $BI_4$ ) and beta-binomial ( $BB_0$  to  $BB_7$ ) models that were fitted for all species are  
<sup>174</sup> described in Table 2 including pseudo-formulas analogous to those used to specify and fit  
<sup>175</sup> mixed or generalized additive models in R (R Core Team, 2019; Wood, 2006). We then  
<sup>176</sup> also included diel effects on relative catch efficiency and interactions with size effects with  
<sup>177</sup> the best performing model of the original 13 models for each species. To fit these diel  
<sup>178</sup> effects, we generalized the modeling framework somewhat in that we allowed multiple (cubic  
<sup>179</sup> regression spline) smooth effects, differing by day and night, on relative catch efficiency.  
<sup>180</sup> We implemented the models using the Template Model Builder package (Kristensen et al.,  
<sup>181</sup> 2016) in R and we used the “nlminb” optimizer to fit the models by maximizing the Laplace  
<sup>182</sup> approximation of the marginal likelihood (R Core Team, 2019).

<sup>183</sup> We assessed convergence of the optimization for each model in two ways. The first criterion  
<sup>184</sup> was whether the optimization using nlminb completed without error. The errors for these  
<sup>185</sup> models was due to entering the parameter space where the gradient was not defined. Note  
<sup>186</sup> that TMB uses automatic differentiation to provide a gradient function for use in optimiza-  
<sup>187</sup> tion. The second convergence criterion was whether the flag returned by nlminb indicated  
<sup>188</sup> false convergence which is associated with overparameterization of the model and that a sim-  
<sup>189</sup> pler model (e.g., no random effects or smoother) is warranted. Models that did not satisfy  
<sup>190</sup> these criteria were not considered further for relative performance based on AIC. If the best  
<sup>191</sup> performing model included smooth length effects and the estimated smoothing parameter  
<sup>192</sup> implied a linear functions of length (on the transformed mean), then simple linear func-

193 tions (i.e., completely smooth) were assumed for further models that included diel effects on  
194 relative efficiency. As such, there was one less (smoothing) parameter estimated for these  
195 models.

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

206 For models BI<sub>4</sub>, BB<sub>6</sub>, and BB<sub>7</sub>, there are two fixed effects parameters associated with the  
207 spline coefficients that are treated as random effects for station-specific smoothers and the  
208 correlation of these pairs of random effects is estimated. However, this parameter was not  
209 estimable for red hake for BB<sub>6</sub> and assumed equal to zero.

### 210 2.3 Length-weight analysis

211 We will use the relative catch efficiency at length to rescale the abundance at length from  
212 the surveys. To generate a rescaled biomass estimate, we convert the numbers at length to  
213 biomass at length using estimates of weight at length and then sum across lengths. We fit  
214 length-weight relationships to the length and weight observations for each survey each year.  
215 We assumed weight observation  $j$  from survey  $i$ , was log-normal distributed,

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

216 Because the expectation of a log-normal random variable is a function of the mean of the  
 217 normal distribution and  $\sigma_i^2$ , we used a bias correction to ensure the expected weight  $E(W_{ij}) =$   
 218  $\alpha_i L_{ij}^{\beta_i}$ . We estimated parameters by maximizing the model likelihood programmed with the  
 219 Template Model Builder package and R and generated predictions of weight at length

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

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

## 223 2.4 Biomass estimation

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

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

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

<sup>237</sup> The stratified abundance estimate is then calculated using the design-based estimator,

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

<sup>238</sup> where  $A_h$  is the area of stratum  $h$ ,  $a$  is the average swept area of a survey station tow, and  
<sup>239</sup>  $n_h$  is the number of tows that were made in stratum  $h$ . The corresponding biomass estimate  
<sup>240</sup> is then

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

<sup>241</sup> where  $\widehat{W}(L = l)$  is the predicted weight at length (Eq. 2) from fitting length-weight obser-  
<sup>242</sup> vations described above. Length is typically measured to the nearest cm so  $n_L$  indicates the  
<sup>243</sup> number of 1 cm length categories observed during the survey.

<sup>244</sup> We used the same criteria for survey station selection as those currently used to estimate  
<sup>245</sup> indices of abundance or biomass for management of each stock. For Gulf of Maine winter  
<sup>246</sup> flounder we also restricted the size classes in each tow to those  $\geq 30$  cm as the biomass of  
<sup>247</sup> the population over this threshold is currently used for management of this stock. For some  
<sup>248</sup> stocks there were certain years where some but not all of the set of survey strata used to  
<sup>249</sup> define indices of abundances were sampled by the bottom trawl survey. In those years, the  
<sup>250</sup> average catch per unit area was expanded to all of the stock strata proportionally to the areas  
<sup>251</sup> of the sampled and unsampled strata. The fall 2017 survey was extremely restricted because  
<sup>252</sup> of vessel mechanical failure and indices are not available for summer flounder, SNE-MA  
<sup>253</sup> windowpane, and SNE-MA yellowtail flounder.

<sup>254</sup> To estimate uncertainty in biomass, we used bootstrap results for the relative catch efficiency  
<sup>255</sup> and weight at length estimates along with bootstrap samples of the survey data. Bootstrap  
<sup>256</sup> data sets for each of the annual surveys respected the stratified random designs by resam-  
<sup>257</sup> pling with replacement within each stratum (Smith, 1997). For each of the 1000 combined  
<sup>258</sup> bootstraps, survey observations for bootstrap  $b$  were scaled with the corresponding bootstrap  
<sup>259</sup> estimates of relative catch efficiency and predicted weight at length, using Eqs. 4 and 5.

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

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

269 where

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

270

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

271 and  $K$  is the number of bootstraps.

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

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

279 where the covariance is

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

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

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

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

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

284 All code and most data files to run the analysis and generate biomass estimates are available  
285 at [https://github.com/timjmiller/chainsweep\\_paper](https://github.com/timjmiller/chainsweep_paper).

### 286 3 Results

#### 287 3.1 Paired-tow observations

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

295 measured for winter flounder and goosefish, a moderate proportion (50-97%) measured for  
296 American plaice, windowpane flounder, and yellowtail flounder, and a low proportion (<50%)  
297 measured for witch flounder and red hake.

## 298 **3.2 Relative catch efficiency**

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

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

316 All 1000 bootstrap fits of the paired tow data converged with invertible hessians at the  
317 optimized log-likelihood and provided estimates of relative catch efficiency at size for summer,  
318 windowpane, and yellowtail flounder, and red hake, goosefish, and thorny skate. All but 2  
319 of the bootstraps for winter flounder and 3 for barndoor skate provided estimates of relative

320 catch efficiency. For witch flounder, 817 bootstraps provided estimates and only 386 provided  
321 estimates for American plaice. One bootstrap fit for summer flounder was excluded due to  
322 an extremely high relative efficiency of the chain sweep gear which impeded estimation of  
323 standard errors from the bootstrap fits.

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

### 328 3.3 Biomass estimation

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

337 The relative catch efficiency of the rockhopper and chan sweep gears in terms of biomass  
338 varies across survey years and seasons due primarily to differences in size composition, but  
339 also variation in estimated length-weight relationship parameters (Figure 6). The efficiency of  
340 the bottom trawl survey gear was greatest for the winter flounder stocks and American plaice  
341 (0.6 to 0.9) and least for red hake, witch flounder, windowpane, and yellowtail flounder stocks  
342 (0.2 to 0.4). Precision of the estimated annual biomass efficiencies was lowest for Georges  
343 Bank winter flounder and the skate stocks. For Gulf of Maine winter flounder, southern red  
344 hake, and barndoor skate, the average fall biomass efficiencies were typically greater than in

345 the spring although the differences were small relative to the confidence intervals.

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

354 We observed little correlation of annual biomass estimates induced by the relative catch  
355 efficiency estimation for most of the stocks (Table 7). However, the biomass estimates were  
356 highly correlated for George Bank winter flounder in the spring (65%) and barndoor skate  
357 (> 70% on average). Estimates for Georges Bank winter flounder in the fall, both red hake  
358 stocks, northern goosefish, and thorny skate were greater than 20% on average.

## 359 4 Discussion

360 The data that we used to estimate bottom trawl survey catch efficiency came from an  
361 experiment using a twin trawler and many of the standard tow protocols for the NEFSC  
362 survey on the *Bigelow*. The experimental net used on one side of the twin trawl was the  
363 same as the standard survey trawl used on the *Bigelow* except that it contained roughly half  
364 the number of floats and the sweep was modified to optimize flatfish catch by minimizing  
365 the ability of flatfish to pass under the net. The other side of the twin trawl was essentially  
366 identical to the standard gear used on the *Bigelow*. The towing of the standard survey  
367 bottom trawl on the twin trawl experiment differed in a few ways from its deployment on  
368 the spring and fall bottom trawl surveys, but we believe that these differences did not have  
369 a significant effect on the results. The use of larger doors and the restrictor rope served to

370 fix the net geometries which may be the biggest source of variability in comparative trawl  
371 catches (Jones et al., 2021). This setup also allowed us to avoid many of the potential  
372 problems due to the large differences in size of the *Bigelow* and the *F/V Karen Elizabeth*.  
373 We do not suspect that the use of the restrictor rope would influence flatfish behavior in  
374 front of the trawl because flatfish have been shown to generally not react to trawling induced  
375 stimuli until they are in very close proximity or even contacted by the fishing gear (Ryer  
376 et al., 2010). The spread data indicated that the restrictor rope remained taut throughout  
377 the towing process (setting, towing, hauling back), so we believe it likely that the restrictor  
378 rope was almost always at least 1m off the bottom. Our concerns about potential effects of  
379 the restrictor rope on species that spend more time off the ground (e.g., Atlantic cod, *Gadus*  
380 *morhua*) led us to exclude them from analyses.

381 Herding is a known phenomenon for flatfish and many other species when certain types of  
382 gear are used (Ramm and Xiao, 1995; Somerton and Munro, 2001; Somerton et al., 2007;  
383 Rose et al., 2010). Somerton and Munro (2001) considered two factors of bridle herding  
384 effects on efficiency. The first factor was the size of the bridle path where the bridle is off the  
385 ground ( $W_{\text{off}}$ ) and the second factor, the herding efficiency ( $h$ ) was the fraction of fish in the  
386 bridle contact path moved into the path of the net. The former is a function of gear design,  
387 and controllable, whereas the latter is a function of fish behavior with regard to the bridle  
388 when it is in contact with the substrate. The bridle configuration on the bottom trawl survey  
389 is designed to minimize contact with the bottom and lack of abrasion of painted bridles used  
390 during one of the twin trawler research trips provided evidence of little or no bridle contact  
391 during the paired tow experiments used to collect the data used in this study. Furthermore,  
392 studies have consistently found that herding behavior occurred during the daytime (Glass  
393 and Wardle, 1989; Somerton and Munro, 2001; Ryer and Barnett, 2006; Bryan et al., 2014;  
394 Ryer et al., 2010; Dean et al., 2021) with some studies indicating high herding coefficients  
395 ( $h$ ) along the sections of the bridles in contact with the bottom. Studies that have evaluated  
396 herding at night or in low light conditions did not find evidence for a directional herding

<sup>397</sup> response (Glass and Wardle, 1989; Ryer and Barnett, 2006; Ryer, 2008; Ryer et al., 2010).

<sup>398</sup> The minimal bridle contact with the substrate and the large fraction of nighttime tows during

<sup>399</sup> the bottom trawl survey suggests flatfish herding is unlikely to be an important factor in

<sup>400</sup> catch efficiency for net spread-based swept area.

<sup>401</sup> On the other hand, the biomass estimates assume that the chain sweep gear is fully efficient,

<sup>402</sup> but it is likely at least some small fraction of fish, that may depend on size, are not captured

<sup>403</sup> by the gear. Trawl configurations such as that described by Munro and Somerton (2002)

<sup>404</sup> could be used to quantify the efficiency of the chain sweep gear were considered by the Trawl

<sup>405</sup> Advisory Panel, but were not used due to concerns that the attached underbag could alter

<sup>406</sup> the performance of the standard survey gear and inhibit the utility in calibrating the survey

<sup>407</sup> indices. The biomass estimates also implicitly assume that the entire stock is available to the

<sup>408</sup> bottom trawl survey, but many of these stocks extend somewhat outside of the survey strata

<sup>409</sup> used to define the indices throughout the year and(or) seasonally due to migration. If either

<sup>410</sup> of these assumptions are incorrect this method of biomass estimation would be negatively

<sup>411</sup> biased (expected value of biomass estimates would be lower than the true value). However,

<sup>412</sup> estimation using the data from these paired-gear studies and these assumptions is less biased

<sup>413</sup> than those made without them.

<sup>414</sup> Diurnal differences in behavior including hiding from predators at higher light levels and

<sup>415</sup> occurring higher in the water column at night, possibly for feeding, have been observed for

<sup>416</sup> flatfish species (Hempel, 1964; Verheijen and de Groot, 1967; Burrows et al., 1994; Burrows

<sup>417</sup> and Gibson, 1995; Ellis et al., 1997; Hurst and Duffy, 2005). Diurnal differences in reaction

<sup>418</sup> to trawl gear have also been observed, with flatfish attempting to avoid gear by staying near

<sup>419</sup> the bottom during daytime and moving off bottom during nighttime (Ryer and Barnett,

<sup>420</sup> 2006). It is not possible to distinguish these behaviors from paired tow studies alone, but

<sup>421</sup> the greater relative catch efficiencies of the chain sweep gear we found during daytime, at

<sup>422</sup> least for smaller sizes, are consistent with these previous studies. We did not find evidence

<sup>423</sup> for differences in relative catch efficiency for American plaice, but findings of differences in

<sup>424</sup> catch efficiency in previous studies have been mixed with possibly regional differences for  
<sup>425</sup> this species (Beamish, 1966; Walsh, 1991; Casey and Myers, 1998).

<sup>426</sup> Our analyses treat the amount of daylight as binary effect (day/night) on the relative catch  
<sup>427</sup> efficiency. However, behavior of the fish with respect to the gear is likely to change more  
<sup>428</sup> gradually with the amount of light. A continuous measure of light that uses the angle of  
<sup>429</sup> the sun, the depth of the tow and light attenuation with depth, might prove to be a better  
<sup>430</sup> explanatory variable for changes in relative catch efficiency and perhaps improve estimation  
<sup>431</sup> of abundance from the bottom trawl survey (Jacobson et al., 2015; Kainge et al., 2017).

<sup>432</sup> Aside from the direct impact of estimated catch efficiency of the NEFSC trawl survey gear  
<sup>433</sup> on biomass estimation, our analyses show more subtle impacts of using efficiency estimates  
<sup>434</sup> with survey tow data to generate the abundance indices. Excluding the efficiency estimates,  
<sup>435</sup> the sampling variability of each of the seasonal and annual relative abundance indices is  
<sup>436</sup> independent of the others. The bootstrapping methods we employed illustrated that includ-  
<sup>437</sup> ing estimates of catch efficiency affects the variability of the resulting abundance estimates  
<sup>438</sup> and their independence from each other. For some stocks there was a substantial effect of  
<sup>439</sup> the relative catch efficiency estimation on the precision of the biomass indices. Similarly,  
<sup>440</sup> we found high correlation of annual indices ( $> 0.6$ ) for Georges Bank winter flounder and  
<sup>441</sup> barndoor skate. Decreased precision or increased correlation likely imply less informative-  
<sup>442</sup> ness in assessments based on integrated age-structured models than treating uncalibrated  
<sup>443</sup> indices independently. Assuming calibrated estimates as independent that are in fact highly  
<sup>444</sup> correlated could therefore cause biased estimation and inferences for important assessment  
<sup>445</sup> output such as stock size and fishing mortality. As such, future work should evaluate the  
<sup>446</sup> effects of incorporating this information in an assessment model.

<sup>447</sup> The estimates of absolute abundance and biomass produced using the sweep comparison  
<sup>448</sup> experiments have already been informative to assessments and management of many stocks  
<sup>449</sup> in the Northeast U.S. These estimates have been used directly in the age-structured as-

450 sessment model for summer flounder and northern and southern goosefish stocks (NEFSC,  
451 2019, 2020c). Estimates for southern New England-Mid-Atlantic winter flounder, both Cape  
452 Cod-Gulf of Maine and southern New England- Mid-Atlantic yellowtail flounder stocks, and  
453 American plaice were used to validate the abundance estimates produced by the assessment  
454 models (NEFSC, 2020b). These estimates have also been used directly in assessments for  
455 witch flounder, Gulf of Maine winter flounder, Georges Bank yellowtail flounder, northern  
456 and southern red hake stocks, which are all assessed using simpler index-based assessment  
457 methods (Legault and McCurdy, 2017; NEFSC, 2020a,b). The estimates can be especially  
458 valuable for index-based methods where the scale of the stock is assumed known. Esti-  
459 mates have also been used in a supporting fashion for fall-back assessments of both Gulf of  
460 Maine-Georges Bank and southern New England-Mid-Atlantic windowpane stocks (NEFSC,  
461 2020b).

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

## 472 Acknowledgements

473 This work was funded by NOAA Fisheries Northeast Fisheries Science Center. The quality of  
474 this manuscript has been improved by comments from Jessica Blaylock and two anonymous

475 reviewers. Tor Bendiksen and John Knight assisted with the design of the chain sweep. Jeff  
476 Pessutti, Giovanni Ganesen, Dominique St. Amand, and Calvin Alexander helped collect  
477 paired-tow data for these analyses. We thank Adam Poquette for creating the map in Figure  
478 3. We thank Calvin Alexander also for the photo of the twin trawl gear in Figure 2.

479 **CRediT authorship contribution statement**

480 **Timothy J. Miller:** Conceptualization, Methodology, Writing - original draft, Formal  
481 analysis, Visualization. **David E. Richardson:** Conceptualization, Methodology, Writing  
482 - original draft. **Philip J. Politis:** Investigation, Methodology, Writing - review and editing.  
483 **Christopher D. Roebuck:** Project administration, Conceptualization, Funding acquisi-  
484 tion, Investigation, Resources. **John P. Manderson:** Conceptualization, Investigation,  
485 Writing - review and editing. **Michael H. Martin:** Project administration, Conceptual-  
486 ization, Data curation, Writing - review and editing. **Andrew W. Jones:** Visualization,  
487 Writing - review and editing.

488 **References**

- 489 Adams, C. F., Miller, T. J., Manderson, J. P., Richardson, D. E., and Smith, B. E. 2015.  
490 Butterfish 2014 Assessment. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-06;  
491 110 p.
- 492 Arreguín-Sánchez, F. 1996. Catchability: a key parameter for fish stock asseessment. Reviews  
493 in Fish Biology and Fisheries **6**: 221–242, doi: 10.1007/BF00182344.
- 494 Azarowitz, T. S. 1981. A brief historical review of the Woods Hole Laboratory trawl survey  
495 time series. In Bottom Trawl Surveys. Edited by W. G. Doubleday and D. Rivard. Can.  
496 Spec. Pub. Fish. Aqua. Sci. No. 58. pp. 62–67.
- 497 Beamish, F. W. H. 1966. Vertical migration by demersal fish in the northwest atlantic.  
498 Journal of the Fisheries Research Board of Canada **23**(1): 109–139, doi: 10.1139/f66-009.
- 499 Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekkinga, A., Kraan, C., van der Veer, H. W.,  
500 and Piersma, T. 2012. Designing a benthic monitoring programme with multiple conflict-  
501 ing objectives. Methods in Ecology and Evolution **3**(3): 526–536, doi: 10.1111/j.2041-  
502 210X.2012.00192.x.
- 503 Bourne, N. 1965. A comparison of catches by 3- and 4-inch rings on offshore scallop drags.  
504 Journal of the Fisheries Research Board of Canada **22**(2): 313–333.
- 505 Bryan, D. R., Bosley, K. L., Hicks, A. C., Haltuch, M. A., and Wakefield, W. W. 2014.  
506 Quantitative video analysis of flatfish herding behavior and impact on effective area swept  
507 of a survey trawl. Fisheries Research **154**: 120–126, doi: 10.1016/j.fishres.2014.02.007.
- 508 Burrows, M. T. and Gibson, R. N. 1995. The effects of food, predation risk and endogenous  
509 rhythmicity on the behaviour of juvenile plaice, *Pleuronectes platessa* L. Animal Behaviour  
510 **50**(1): 41–52, doi: 10.1006/anbe.1995.0219.

- 511 Burrows, M. T., Gibson, R. N., and Maclean, A. 1994. Effects of endogenous rhythms and  
512 light conditions on foraging and predator-avoidance behaviour in juvenile plaice. Journal  
513 of Fish Biology **45**(SA): 171–180, doi: 10.1111/j.1095-8649.1994.tb01091.x.
- 514 Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-  
515 reported catches and variable natural mortality rates. Canadian Journal of Fisheries and  
516 Aquatic Sciences **73**(2): 296–308, doi: 10.1139/cjfas-2015-0047.
- 517 Casey, J. M. and Myers, R. A. 1998. Diel variation in trawl catchability: is it as clear as day  
518 and night? Canadian Journal of Fisheries and Aquatic Sciences **55**(10): 2329–2340, doi:  
519 10.1139/f98-120.
- 520 Dean, M. J., Hoffman, W. S., Buchan, N. C., Cadrin, S. X., and Grabowski, J. H. 2021.  
521 The influence of trawl efficiency assumptions on survey-based population metrics. ICES  
522 Journal of Marine Science **78**(8): 2858–2874.
- 523 Ellis, T., Howell, B. R., and Hughes, R. N. 1997. The cryptic responses of hatchery-reared  
524 sole to a natural sand substratum. Journal of Fish Biology **51**(2): 389–401.
- 525 Glass, C. W. and Wardle, C. S. 1989. Comparison of the reactions of fish to a trawl gear,  
526 at high and low light intensities. Fisheries Research **7**(3): 249–266, doi: 10.1016/0165-  
527 7836(89)90059-3.
- 528 Gulland, J. A. 1964. Variations in selection factors, and mesh differentials. Journal Du Con-  
529 seil International Pour L'exploration De La Mer **29**(2): 158–165.
- 530 Hempel, G. 1964. Diurnal variations in catch, feeding and swimming activity of plaice (*Pleu-*  
531 *ronectes platessa* L.). Rapports et Procès-Verbaux des Réunions. Conseil Permanent In-  
532 ternationale pour l'Exploration de la Mer **155**: 58–64.
- 533 Hurst, T. P. and Duffy, T. A. 2005. Activity patterns in northern rock sole are mediated

- 534 by temperature and feeding history. *Journal of Experimental Marine Biology and Ecology*  
535 **325**(2): 201–213, doi: 10.1016/j.jembe.2005.05.003.
- 536 ICES. 1996. Manual of methods of measuring the selectivity of towed fishing gears. (Eds.)  
537 Wileman, D. A., Ferro, R. S. T., Fonteyne, R., and Millar, R. B. ICES Cooperative  
538 Research Report No. 215.
- 539 Jacobson, L. D., Hendrickson, L. C., and Tang, J. 2015. Solar zenith angles for biological  
540 research and an expected catch model for diel vertical migration patterns that affect stock  
541 size estimates for longfin inshore squid (*Doryteuthis pealeii*). *Canadian Journal of Fisheries*  
542 and Aquatic Sciences
- 72**
- (9): 1329–1338, doi: 10.1139/cjfas-2014-0436.
- 543 Jones, A. W., Miller, T. J., Politis, P. J., Richardson, D. E., Mercer, A. M., Pol, M. V., and  
544 Roebuck, C. D. 2021. Experimental assessment of the effect of net wing spread on relative  
545 catch efficiency of four flatfishes by a four seam bottom trawl. *Fisheries Research* **244**:  
546 106106, doi: 10.1016/j.fishres.2021.106106.
- 547 Kainge, P., van der Plas, A. K., Bartholomae, C. H., and Wieland, K. 2017. Effects of  
548 environmental variables on survey catch rates and distribution by size of shallow- and deep-  
549 water Cape hakes, *Merluccius capensis* and *Merluccius paradoxus* off Namibia. *Fisheries*  
550 *Oceanography* **26**(6): 680–692, doi: 10.1111/fog.12227.
- 551 Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2013. Combining  
552 bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl  
553 efficiency parameters for semipelagic species. *Canadian Journal of Fisheries and Aquatic*  
554 *Sciences* **70**(2): 208–219, doi: 10.1139/cjfas-2012-0321.
- 555 Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. 2017. Selectivity ratio: A  
556 useful tool for comparing size selectivity of multiple survey gears. *Fisheries Research* **191**:  
557 76–86, doi: 10.1016/j.fishres.2017.02.012.

- 558 Kotwicki, S., Martin, M. H., and Laman, E. A. 2011. Improving area swept es-  
559 timates from bottom trawl surveys. *Fisheries Research* **110**(1): 198–206, doi:  
560 10.1016/j.fishres.2011.04.007.
- 561 Krag, L. A., Herrmann, B., Karlsen, J. D., and Mieske, B. 2015. Species selectivity in  
562 different sized topless trawl designs: Does size matter? *Fisheries Research* **172**: 243–249,  
563 doi: 10.1016/j.fishres.2015.07.010.
- 564 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic  
565 differentiation and Laplace approximation. *Journal of Statistical Software* **70**(5): 1–21.
- 566 Legault, C. M. and McCurdy, Q. M. 2017. Stock assessment of Georges Bank yellowtail  
567 flounder for 2017. Transboundary Resources Assessment Committee (TRAC) Reference  
568 Document 2017/03 available at: <https://repository.library.noaa.gov/view/noaa/27556>.
- 569 Maunder, M. N. and Piner, K. R. 2015. Contemporary fisheries stock assessment: many issues  
570 still remain. *ICES Journal of Marine Science* **72**(1): 7–18, doi: 10.1093/icesjms/fsu015.
- 571 Millar, R. B. and Fryer, R. J. 1999. Estimating the size-selection curves of towed gears,  
572 traps, nets and hooks. *Reviews in Fish Biology and Fisheries* **9**(1): 89–116, doi:  
573 10.1023/A:1008838220001.
- 574 Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on  
575 paired-gear data for U.S. Northwest Atlantic fish stocks. *Canadian Journal of Fisheries  
576 and Aquatic Sciences* **70**(9): 1306–1316, doi: 10.1139/cjfas-2013-0136.
- 577 Miller, T. J., Hart, D. R., Hopkins, K., Vine, N. H., Taylor, R., York, A. D., and Gal-  
578 lager, S. M. 2019. Estimation of the capture efficiency and abundance of Atlantic sea  
579 scallops (*Placopecten magellanicus*) from paired photographic–dredge tows using hierar-  
580 chical models. *Canadian Journal of Fisheries and Aquatic Sciences* **76**(6): 847–855, doi:  
581 10.1139/cjfas-2018-0024.

- 582 Munro, P. T. and Somerton, D. A. 2001. Maximum likelihood and non-parametric methods  
583 for estimating trawl footrope selectivity. ICES Journal of Marine Science **58**(1): 220–229,  
584 doi: 10.1006/jmsc.2000.1004.
- 585 Munro, P. T. and Somerton, D. A. 2002. Estimating net efficiency of a survey trawl for  
586 flatfishes. Fisheries Research **55**(1–3): 267–279, doi: 10.1016/S0165-7836(01)00280-6.
- 587 NEFSC. 2019. 66th Northeast Regional Stock Assessment Workshop (66th SAW) Assessment  
588 Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 19-08; 1170 p. Available from:  
589 <http://www.nefsc.noaa.gov/publications/>.
- 590 NEFSC. 2020a. Final report of red hake stock structure working group. US Dept Commer,  
591 Northeast Fish Sci Cent Ref Doc. 20-07; 185 p.
- 592 NEFSC. 2020b. Operational assessment of 14 northeast groundfish stocks, updated through  
593 2018. Prepublication Copy of the September 2019 Operational Stock Assessment Report.  
594 The report is “in preparation” for publication by the NEFSC. (Oct. 3, 2019; latest revision:  
595 Jan. 7, 2020).
- 596 NEFSC. 2020c. Operational assessment of the black sea bass, scup, bluefish, and monkfish  
597 stocks, updated through 2018. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 20-01;  
598 160 p.
- 599 Nichol, D. G., Kotwicki, S., Wilderbuer, T. K., Lauth, R. R., and Ianelli, J. N. 2019.  
600 Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey  
601 and its effect on estimates of survey biomass. Fisheries Research **211**: 319–330, doi:  
602 10.1016/j.fishres.2018.11.017.
- 603 Paloheimo, J. E. and Dickie, L. M. 1964. Abundance and fishing success. Rapports et Procès-  
604 Verbaux des Réunions. Conseil Internationale pour l’Exploration de la Mer **155**: 152–163.

- 605 Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an  
606 application. Canadian Journal of Fisheries and Aquatic Sciences **55**(12): 2672–2690, doi:  
607 10.1139/f98-151.
- 608 Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. 2014. Northeast Fisheries  
609 Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. U.S.  
610 Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 14-06, 138p.
- 611 R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Founda-  
612 tion for Statistical Computing, Vienna, Austria.
- 613 Ramm, D. C. and Xiao, Y. 1995. Herding in groundfish and effective pathwidth of trawls.  
614 Fisheries Research **24**(3): 243–259, doi: 10.1016/0165-7836(95)00373-I.
- 615 Rose, C. S., Gauvin, J. R., and Hammond, C. F. 2010. Effective herding of flatfish by cables  
616 with minimal seafloor contact. Fishery Bulletin **108**(2): 136–144.
- 617 Ryer, C. H. 2008. A review of flatfish behavior relative to trawls. Fisheries Research **90**(1):  
618 138–146, doi: 10.1016/j.fishres.2007.10.005.
- 619 Ryer, C. H. and Barnett, L. A. K. 2006. Influence of illumination and temperature upon  
620 flatfish reactivity and herding behavior: Potential implications for trawl capture efficiency.  
621 Fisheries Research **81**(2): 242–250, doi: 10.1016/j.fishres.2006.07.001.
- 622 Ryer, C. H., Rose, C. S., and Iseri, P. J. 2010. Flatfish herding behavior in response to  
623 trawl sweeps: a comparison of diel responses to conventional sweeps and elevated sweeps.  
624 Fishery Bulletin **108**(2): 145–154.
- 625 Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean  
626 abundance. Canadian Journal of Fisheries and Aquatic Sciences **54**(3): 616–630, doi:  
627 10.1139/f96-303.

- 628 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.
- 632 Somerton, D. A. and Munro, P. 2001. Bridle efficiency of a survey trawl for flatfish. Fishery Bulletin **99**(4): 641–652.
- 634 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.
- 636 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.
- 640 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.
- 644 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.
- 648 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.
- 651 Verheijen, F. J. and de Groot, S. J. 1967. Diurnal activity pattern of plaice and flounder (Pleuronectidae) in aquaria. Netherlands Journal of Sea Research **3**(3): 383–390.

- 653 Walsh, S. J. 1991. Diel variation in availability and vulnerability of fish to a survey trawl.
- 654 Journal of Applied Ichthyology 7(3): 147–159, doi: 10.1111/j.1439-0426.1991.tb00521.x.
- 655 Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. 2018. Evaluation of alternative
- 656 stratifications for a stratified random fishery-independent survey. Fisheries Research 207:
- 657 150–159, doi: 10.1016/j.fishres.2018.06.019.
- 658 Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2009. Incorporating time-
- 659 varying catchability into population dynamic stock assessment models. Reviews in Fish-
- 660 eries Science 18(1): 7–24, doi: 10.1080/10641260903294647.
- 661 Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall,
- 662 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 chain sweep gears were deployed varied throughout the experimental tows of each.

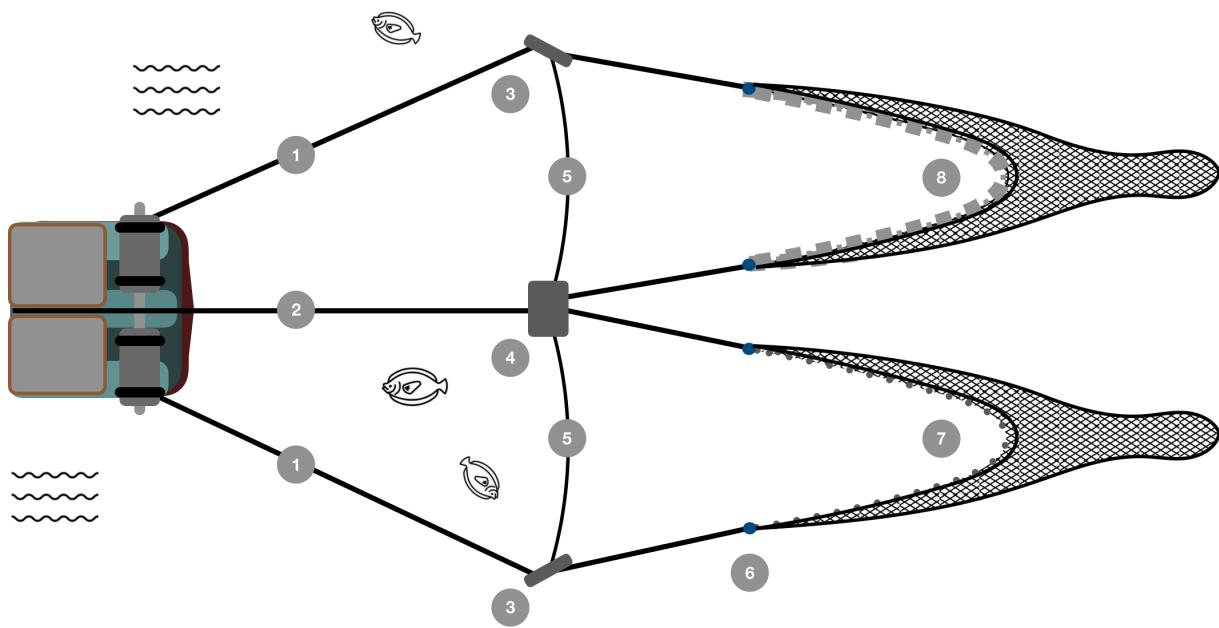


Fig. 2. The *F/V Karen Elizabeth* twin-trawl vessel rigged with rockhopper sweep gear on the right and chain sweep gear on the left.



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.

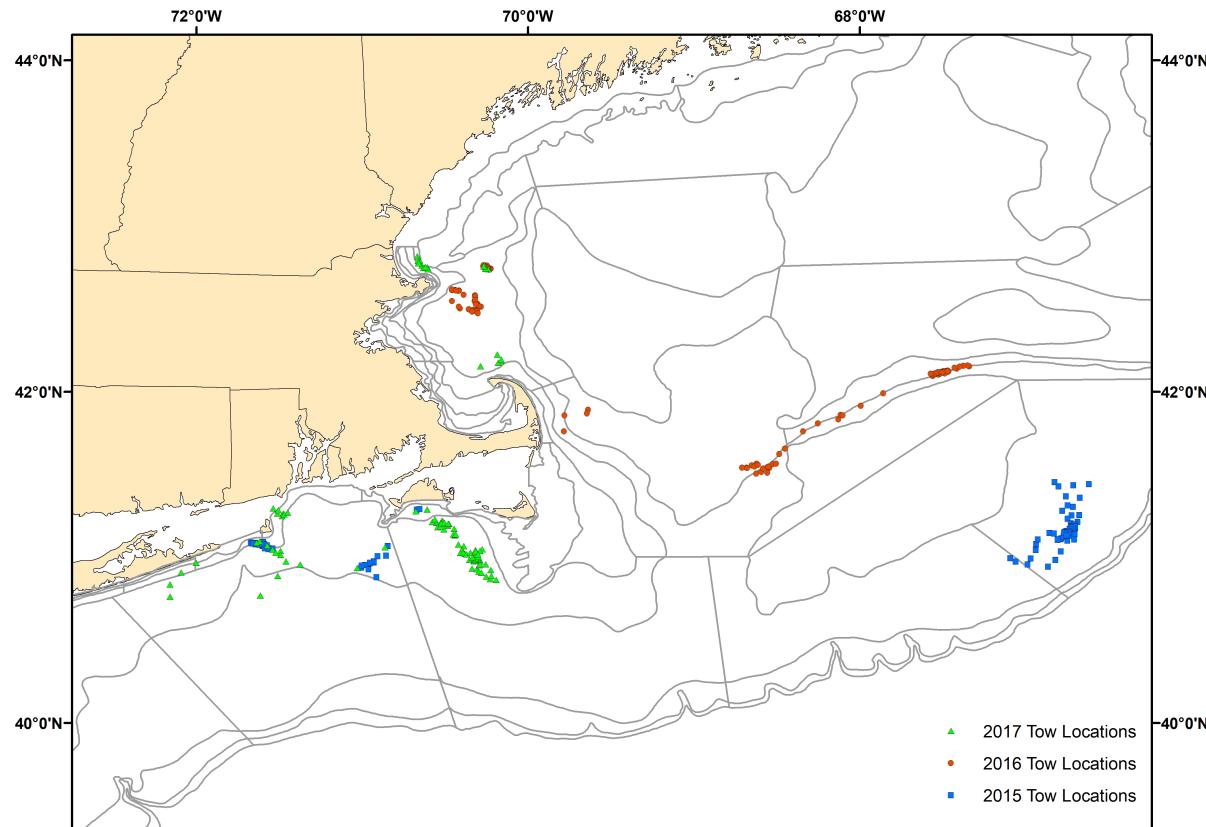


Fig. 4. Relative efficiency of gears using chain and rockhopper sweeps from the best performing model for each species (Table 5). Blue and red denote results for day and night data, respectively, and thick and thin lines represent overall and paired-tow specific estimates of relative catch efficiency,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

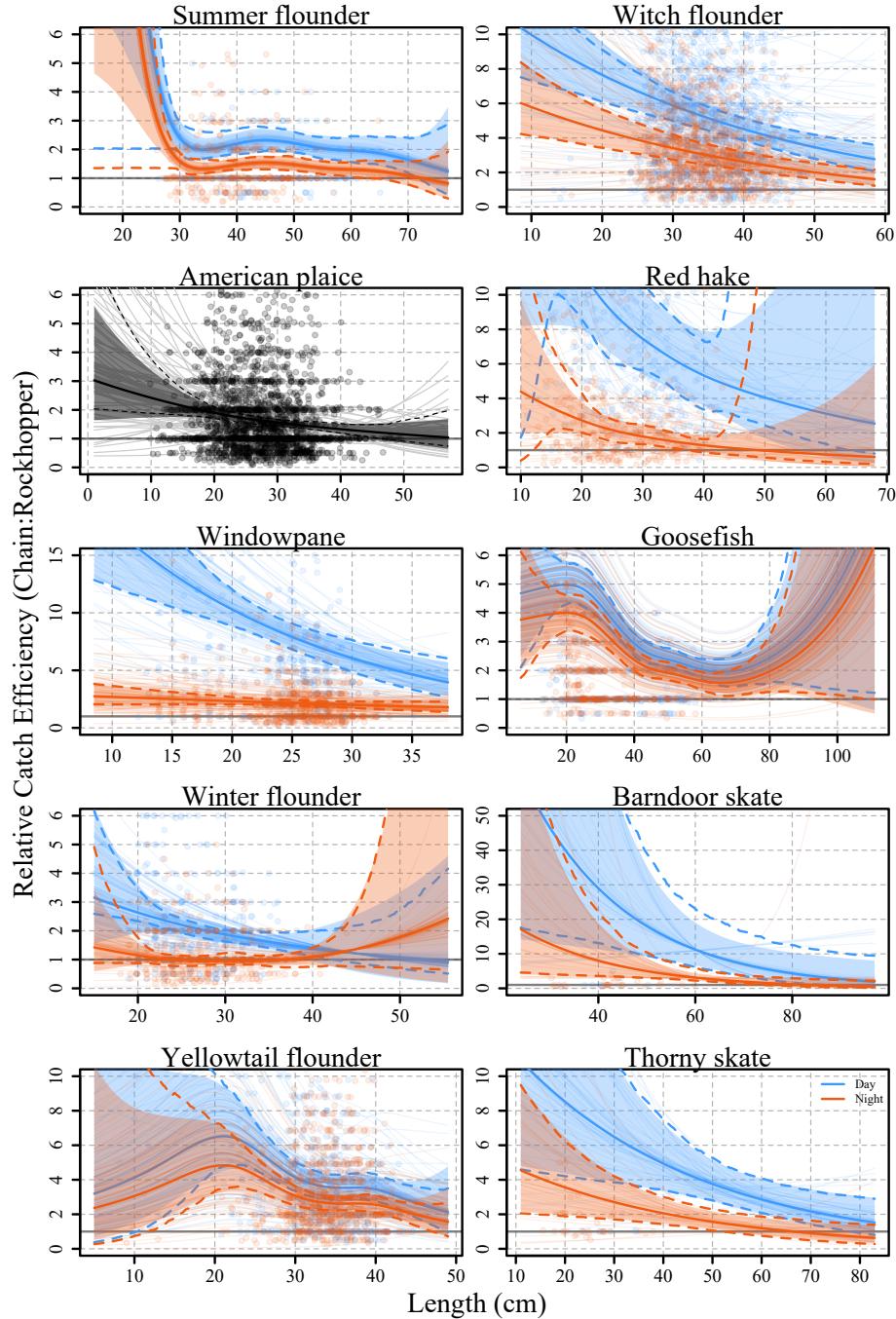


Fig. 5. Annual spring (blue) and fall (red) biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps are from the best performing model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

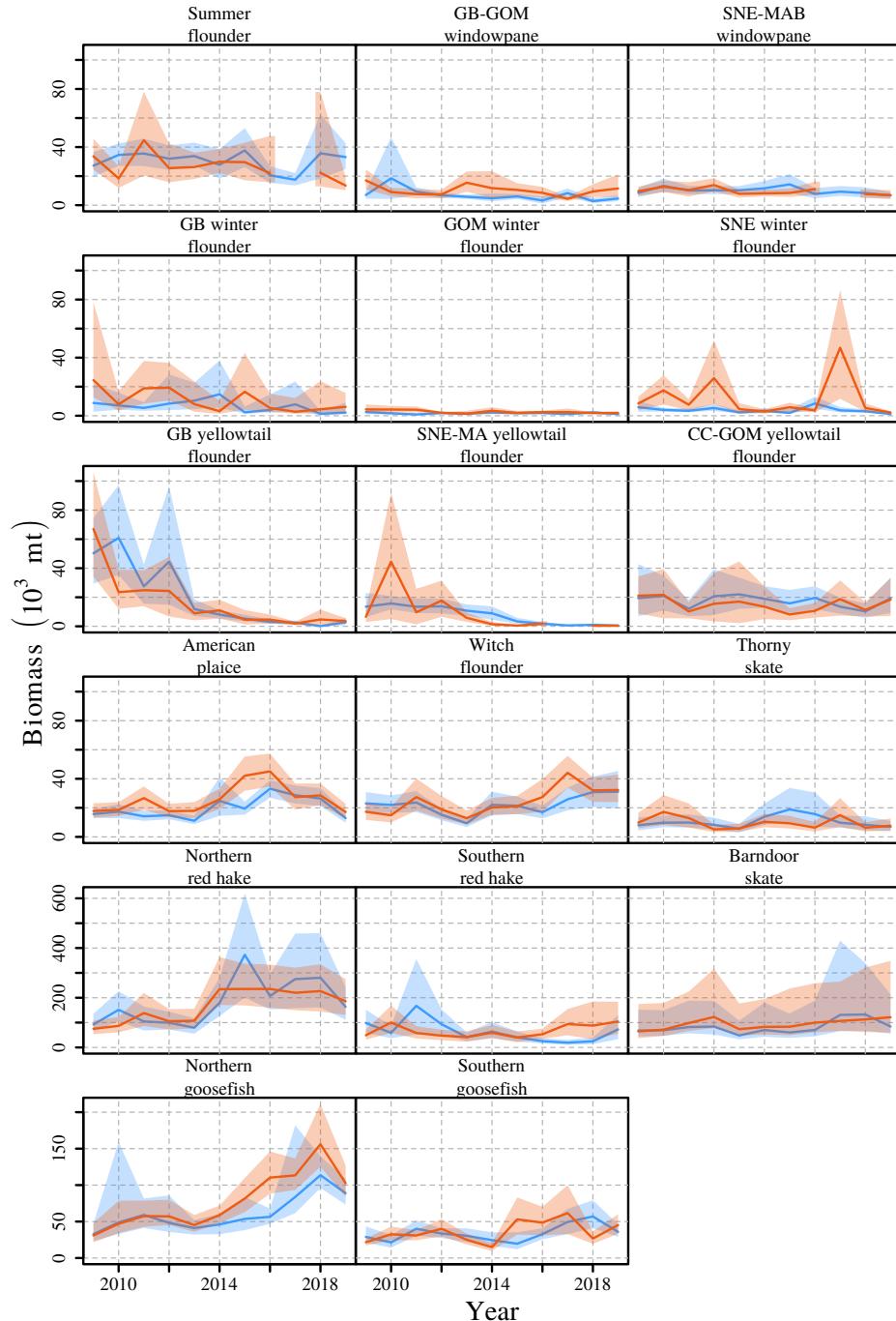


Fig. 6. Implied catch efficiency of annual spring (blue) and fall (red) bottom trawl survey biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and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

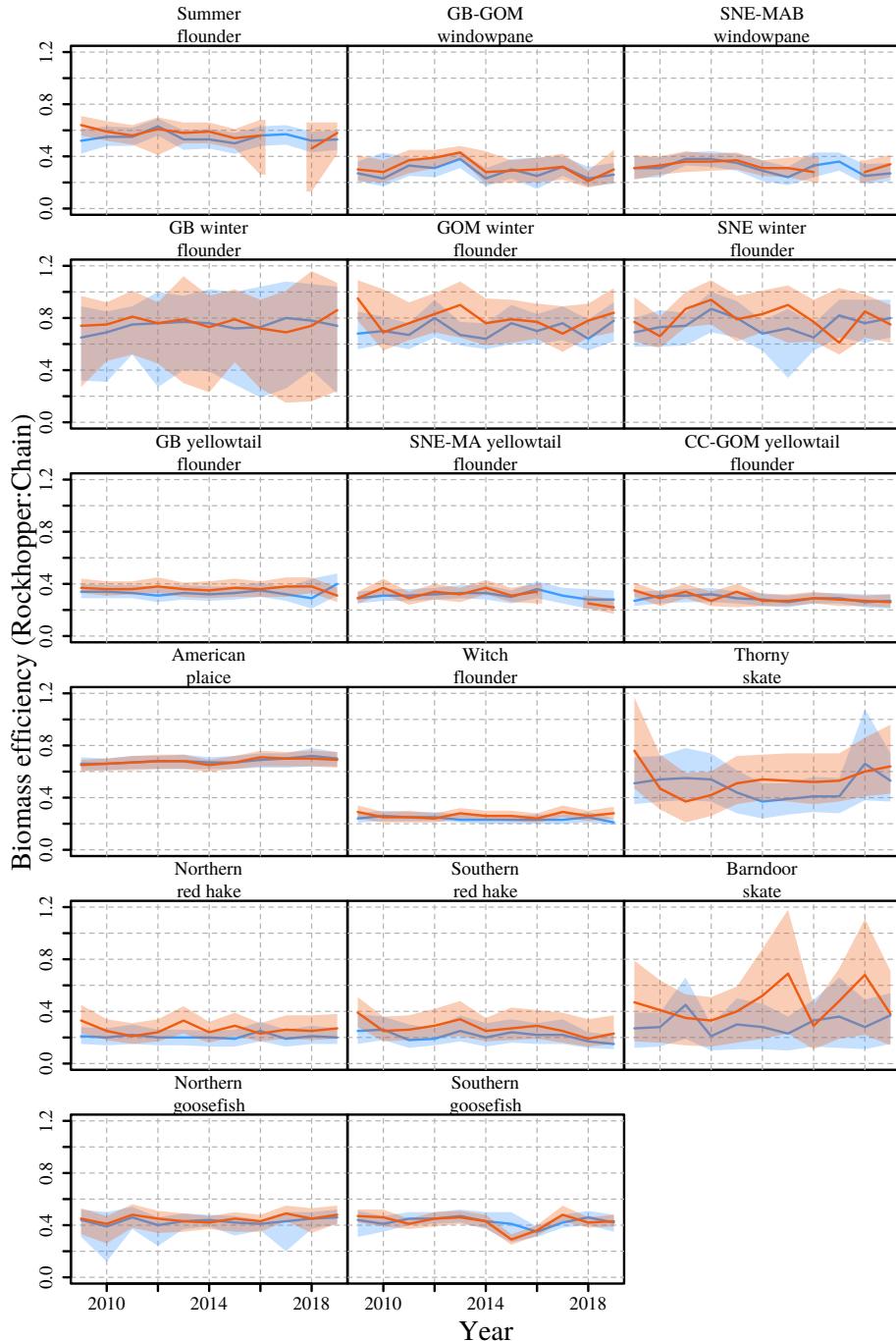


Table 1. Managed stocks associated with the species for which relative catch efficiency was estimated.

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 ( $\rho$ ) and beta-binomial dispersion ( $\phi$ ) 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 <sub>0</sub>	$\sim 1$	—	1	population-level mean for all observations
BI <sub>1</sub>	$\sim 1 + 1 pair$	—	2	population- and random station-level $\rho$
BI <sub>2</sub>	$\sim s(length)$	—	3	population-level smooth size effect on $\rho$
BI <sub>3</sub>	$\sim s(length) + 1 pair$	—	4	population-level smooth size effect and random station-level intercept for $\rho$
BI <sub>4</sub>	$\sim s(length) + s(length) pair$	—	7	population-level and random station-level smooth size effects for $\rho$
BB <sub>0</sub>	$\sim 1$	$\sim 1$	2	population-level $\rho$ and $\phi$
BB <sub>1</sub>	$\sim 1 + 1 pair$	$\sim 1$	3	population-level and random station-level intercept for $\rho$ and population-level $\phi$
BB <sub>2</sub>	$\sim s(length)$	$\sim 1$	4	population-level smooth size effect on $\rho$ and population-level $\phi$
BB <sub>3</sub>	$\sim s(length)$	$\sim s(length)$	6	population-level smooth size effect on $\rho$ and $\phi$
BB <sub>4</sub>	$\sim s(length) + 1 pair$	$\sim 1$	5	population-level smooth size effect and random station-level intercept for $\rho$ and population-level $\phi$
BB <sub>5</sub>	$\sim s(length) + 1 pair$	$\sim s(length)$	7	population-level smooth size effect on $\rho$ and $\phi$ and random station-level intercepts for $\rho$
BB <sub>6</sub>	$\sim s(length) + s(length) pair$	$\sim 1$	8	population-level and random station-level smooth size effects on $\rho$ and population-level $\phi$
BB <sub>7</sub>	$\sim s(length) + s(length) pair$	$\sim s(length)$	10	population-level and random station-level smooth size effects on $\rho$ and population-level smooth size effects on $\phi$

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 <sub>0</sub>	BI <sub>1</sub>	BI <sub>2</sub>	BI <sub>3</sub>	BI <sub>4</sub>	BB <sub>0</sub>	BB <sub>1</sub>	BB <sub>2</sub>	BB <sub>3</sub>	BB <sub>4</sub>	BB <sub>5</sub>	BB <sub>6</sub>	BB <sub>7</sub>
Summer flounder	27.96	13.53	8.9	<b>0</b>		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	<b>0</b>	
Windowpane	1045.06	38.51	1029.72	17.03	<b>0</b>	585.7	32.22	572.73		15.27			
Winter flounder	216.47	15.73	200.33	3.02	<b>0</b>	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	<b>0</b>	3.33	
Witch flounder	1424.17	212.64	1372.66		35.33	881.28	142.53	844.47		81.37		<b>0</b>	
Red hake	1884.51	295.85	1697.48	170.75		627.33	166.43	590.92		95.8	59.31	<b>0</b>	0.83
Goosefish	227.67	87.23	80.37	<b>0</b>		219.13		76.54					
Barndoor skate	36.51	10.01	31.34	2.72	<b>0</b>	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	<b>0</b>	

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 ( $\Delta\text{AIC}$ ) from the best of the three models (**0**) by species.

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