

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

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

<sup>5</sup> Christopher D. Roebuck<sup>3</sup>      John P. Manderson<sup>4,5</sup>

<sup>6</sup> Michael H. Martin<sup>1,6</sup>      Andrew W. Jones<sup>2</sup>

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

<sup>9</sup> <sup>2</sup>Northeast Fisheries Science Center, Narragansett Laboratory, 28 Tarzwell Drive Narragansett, RI 02882 USA

<sup>11</sup> <sup>3</sup>Fishing Vessel Karen Elizabeth, Narragansett, RI 02882, USA

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

<sup>14</sup> <sup>5</sup>Current address: OpenOcean Research, Suite 101, 40 West Evergreen Avenue, Philadelphia, PA, 19118 USA

<sup>16</sup> <sup>6</sup>Current address: Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Building 4 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 relative efficiency of <sup>24</sup> 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 <sup>26</sup> a twin-trawl vessel. For 10 commercially important species, we fitted and compared a set <sup>27</sup> 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 <sup>29</sup> 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 <sup>31</sup> then used the bottom trawl survey data from surveys between 2009 and 2019 with the relative <sup>32</sup> catch efficiency estimates from the best performing models to estimate annual and seasonal <sup>33</sup> chain sweep-based swept area biomass for 17 managed stocks. We estimated uncertainty in <sup>34</sup> all results using bootstrap procedures for each data component. We also assessed the effect <sup>35</sup> of calibration on uncertainty of the biomass estimates and the degree of correlation of the <sup>36</sup> annual biomass estimates.

<sup>37</sup> **Keywords**

<sup>38</sup> gear efficiency, abundance estimation, hierarchical generalized additive models

<sup>39</sup> **1 Introduction**

<sup>40</sup> Ecosystem monitoring surveys such as fisheries-independent trawl surveys are used to obtain  
<sup>41</sup> information on a range of species and are therefore not optimized with respect to sampling  
<sup>42</sup> design or gear for any one species (Bijleveld et al., 2012; Wang et al., 2018). Gear and  
<sup>43</sup> sampling protocols are designed to provide consistent and representative samples that allow  
<sup>44</sup> indices of abundance at size and age to be developed for a suite of species (Azarovitz, 1981;  
<sup>45</sup> Thiess et al., 2018). To provide indices of population abundance with minimal potential  
<sup>46</sup> sources of bias, survey bottom trawl gear must be configured to be towed across as wide a  
<sup>47</sup> variety of habitats as possible, including seafloor habitats with complex physical structures.

<sup>48</sup> Indices of abundance at age and size derived from fisheries independent bottom trawl surveys  
<sup>49</sup> are scaled to population size by the survey catchability ( $q$ ) parameter (Arreguín-Sánchez,  
<sup>50</sup> 1996). Catchability is typically estimated internally within stock assessment models that  
<sup>51</sup> incorporate fisheries landings, indices of abundance, and life history parameters. However,  
<sup>52</sup> the amount or quality of data and degree of contrast in the time series is often such that  
<sup>53</sup> this parameter is difficult to estimate (Maunder and Piner, 2015). In such cases, estimates  
<sup>54</sup> of survey catchability from auxilliary data can inform the stock assessment. These external  
<sup>55</sup> estimates can be used as a direct input into the assessment model (Somerton et al., 1999),  
<sup>56</sup> can serve as a diagnostic measure of model accuracy (Miller et al., 2019), or contribute to  
<sup>57</sup> an alternate means of providing catch advice when an assessment model is not considered  
<sup>58</sup> acceptable (Legault and McCurdy, 2017).

<sup>59</sup> Catchability can be decomposed into two components, the proportion of the population  
<sup>60</sup> available to the survey sampling frame and the efficiency of the survey gear given an  
<sup>61</sup> individual is available to the gear (Paloheimo and Dickie, 1964). Here efficiency is the fraction  
<sup>62</sup> of available fish retained by the gear, equivalent to availability-selection in Millar and Fryer  
<sup>63</sup> (1999). Estimates of these components allow relative abundance indices to be converted into  
<sup>64</sup> absolute abundance indices without a population model. As such, investigations of gear

65 mensuration (Kotwicki et al., 2011), species-specific gear efficiency (Thygesen et al., 2019;  
66 Jones et al., 2021), and availability of the stock to the survey design frame (Nichol et al.,  
67 2019) improve our understanding of catchability and therefore abundance of fish stocks.

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

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

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

89 In conducting paired-gear studies it is ideal to have the two gears deployed as close together  
90 spatially and temporally as possible to reduce variation between the gears in densities of the

91 species being captured. The twin-trawl rigging (Krag et al., 2015) where two trawls can be  
92 fished simultaneously approaches this ideal (ICES, 1996), and is the data-collection platform  
93 chosen by the Trawl Advisory Panel. The Panel decided to rig one of the twin trawls as the  
94 the gear used by the bottom trawl survey which uses a rockhopper sweep. The other trawl  
95 was rigged similarly except with a chain sweep in an attempt to eliminate any escapement of  
96 fish under the gear. Assuming the chain sweep-based gear is fully efficient allows the efficiency  
97 of the rockhopper sweep-based gear used by the bottom trawl survey to be estimated from  
98 these experiments.

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

108 Often overlooked aspects of the application of relative catch efficiency estimates is the impact  
109 on the precision of abundance indices and the correlation among annual indices that the  
110 application induces. These indices are typically used as measures of relative abundance in  
111 stock assessment with the precision of the indices used to weight the observations within the  
112 assessment model. Furthermore, the sampling variability of the annual indices is typically  
113 assumed to be independent. Here we compare the precision of the calibrated and uncalibrated  
114 indices and measure the correlation of calibrated indices for each stock.

<sub>115</sub> **2 Methods**

<sub>116</sub> **2.1 Data collection**

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

<sup>140</sup> **2.2 Paired-tow analysis**

<sup>141</sup> We employed the hierarchical modeling approach from Miller (2013) to estimate the efficiency  
<sup>142</sup> ( $\rho$ ) of the rockhopper sweep used by the NEFSC bottom trawl survey relative to the chain  
<sup>143</sup> sweep-based gear for ten species (Summer flounder, *Paralichthys dentatus*; American plaice,  
<sup>144</sup> *Hippoglossoides platessoides*; windowpane flounder, *Scophthalmus aquosus*; winter flounder,  
<sup>145</sup> *Pseudopleuronectes americanus*; yellowtail flounder, *Limanda ferruginea*; witch flounder,  
<sup>146</sup> *Glyptocephalus cynoglossus*; red hake, *Urophycis chuss*; goosefish, *Lophius americanus*; barn-  
<sup>147</sup> door skate, *Dipturus laevis*; thorny skate, *Amblyraja radiata*) from three separate research  
<sup>148</sup> trips carried out aboard a twin trawl vessel. We first fit and compared the same set of  
<sup>149</sup> 13 models as Miller (2013) with different assumptions about variation of relative efficiency  
<sup>150</sup> between paired gear tows, size effects on the relative efficiency, and extra-binomial variation  
<sup>151</sup> of observations within paired gear tows. The binomial (BI<sub>0</sub> to BI<sub>4</sub>) and beta-binomial (BB<sub>0</sub> to  
<sup>152</sup> BB<sub>7</sub>) models that were fitted for all species are described in Table 2 including pseudo-formulas  
<sup>153</sup> analogous to those used to specify and fit mixed or generalized additive models in R (R Core  
<sup>154</sup> Team, 2019; Wood, 2006). We then also included diel effects on relative catch efficiency and  
<sup>155</sup> interactions with size effects with the best performing model of the original 13 models for  
<sup>156</sup> each species. To fit these diel effects, we generalized the modeling framework somewhat in  
<sup>157</sup> that we allow multiple (cubic regression spline) smooth effects, differing by day and night,  
<sup>158</sup> on relative catch efficiency. We implemented the models using the Template Model Builder  
<sup>159</sup> package (Kristensen et al., 2016) in R and we used the “nlnminb” optimizer to fit the models  
<sup>160</sup> by maximizing the Laplace approximation of the marginal likelihood (R Core Team, 2019).

<sup>161</sup> If the best model included smooth length effects and the estimated smoothing parameter  
<sup>162</sup> implied a linear functions of length (on the transformed mean), then simple linear functions  
<sup>163</sup> (i.e., completely smooth) were assumed for further models that included diel effects on relative  
<sup>164</sup> efficiency. As such, there was one less (smoothing) parameter estimated for these models.

<sup>165</sup> We compared two alternative ways of estimating uncertainty in relative catch efficiency. The

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

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

### 179 2.3 Length-weight analysis

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

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

182 We used a bias correction to ensure the expected weight  $E(W_{ij}) = \alpha_i L_{ij}^{\beta_i}$ . We estimated  
 183 parameters by maximizing the model likelihood programmed with the Template Model Builder  
 184 package (Kristensen et al., 2016) and R (R Core Team, 2019) and generated predictions of  
 185 weight at length

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

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

## 189 2.4 Biomass estimation

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

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

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

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

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

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

206 is then

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

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

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

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

226 We also used the bootstraps to summarize other aspects of the biomass estimates. First, we  
227 used the bootstraps to calculate the ratio of calibrated and uncalibrated biomass for each  
228 spring and fall annual survey which is the implicit relative catch efficiency in terms of biomass.  
229 The uncalibrated biomass estimate for bootstrap  $b$  uses the resampled survey data as the  
230 calibrated biomass estimate except that the bootstrap for the relative catch efficiency is not

<sup>231</sup> used (i.e.,  $\hat{\rho}_i(L) = 1$  in Eq. 3). We also used the bootstraps to compare the coefficients of  
<sup>232</sup> variation (CV) of the calibrated and uncalibrated biomass estimates. The CV for an annual  
<sup>233</sup> biomass estimate for year  $y$  from either the spring or fall survey was calculated as

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

<sup>234</sup> where

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

<sup>235</sup>

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

<sup>236</sup> and  $K$  is the number of bootstraps.

<sup>237</sup> For summer flounder it was necessary to omit one of the 1000 bootstraps of relative catch  
<sup>238</sup> efficiency at length due to an extremely large value which the standard deviation and mean of  
<sup>239</sup> the bootstraps was sensitive to. Finally, we calculated correlation of annual biomass estimates  
<sup>240</sup> for years  $y$  and  $z$  using the bootstrap estimates of biomass

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

<sup>241</sup> where the covariance is

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

<sup>242</sup> We summarized the relative precision of the calibrated and uncalibrated biomass estimates  
<sup>243</sup> as the average of the annual ratios of the CVs for the calibrated and uncalibrated estimates

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

<sup>244</sup> We summarized the correlation of biomass estimates as the mean correlation of all annual  
<sup>245</sup> calibrated biomass estimates

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

<sup>246</sup> All code and most data files to run the analysis and generate biomass estimates are available  
<sup>247</sup> at [https://github.com/timjmiller/chainsweep\\_paper](https://github.com/timjmiller/chainsweep_paper).

## <sup>248</sup> 3 Results

### <sup>249</sup> 3.1 Paired-tow observations

<sup>250</sup> In terms of paired tows and total numbers of fish, flatfish were the most well sampled species,  
<sup>251</sup> but goosefish was observed in the most paired-tows and red hake was the most prevalent  
<sup>252</sup> in terms of total numbers caught (Table 3). Witch flounder was the most prevalent flatfish  
<sup>253</sup> species caught while yellowtail flounder was the most frequently observed flatfish in terms of  
<sup>254</sup> paired tows. For all species but summer flounder, and barndoor and thorny skates, only a  
<sup>255</sup> subsample of all of the fish that were caught were measured for length, but nearly all winter  
<sup>256</sup> flounder and goosefish were measured.

### <sup>257</sup> 3.2 Relative catch efficiency

<sup>258</sup> As measured by AIC, the best performing models for all 10 species included size effects on  
<sup>259</sup> the relative efficiency of the chain and rockhopper sweep gears and between-pair variability  
<sup>260</sup> in relative catch efficiency (Table 4). Extrabinomial variation (i.e., beta-binomial) in relative  
<sup>261</sup> catch efficiency at size within pairs was also important for American plaice, yellowtail flounder,  
<sup>262</sup> witch flounder, red hake, and thorny skate. Model convergence was an issue for all species,

263 particularly for the most complex models with pair-specific smooth functions of length ( $BI_4$ )  
264 and smooth effects of size on the beta-binomial dispersion parameter ( $BB_3, BB_5$ , and  $BB_7$ ).

265 Including diel effects on relative catch efficiency improved model performance for all species  
266 except American plaice (Table 5). For those species with diel effects on relative catch  
267 efficiency, the ratio of the efficiencies was generally greater for daytime observations, when  
268 fish are typically more closely associate with the sea floor, than those for nighttime tows,  
269 with the exception of large winter flounder (Figure 4). The largest differences in efficiency  
270 was estimated for smaller barndoor skate. For most of the species, the difference in efficiecies  
271 between the gears was generally greater for smaller individuals.

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

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

### 283 3.3 Biomass estimation

284 Total biomass estimates calibrated to the chain sweep gear were variable across years for most  
285 stocks and without strong trend (Figure 5). However, declining trends exist for the George  
286 Bank and southern New England-Mid-Atlantic yellowtail flounder stocks and an increasing

287 trend for northern goosefish. Biomass estimates were greatest on average for northern red  
288 hake and least for Gulf of Maine winter flounder, although this biomass estimate excludes fish  
289 less than 30 cm in length. Fall and spring biomass estimates were similar in scale for most  
290 stocks, except that southern New England winter flounder and northern goosefish estimates  
291 were typically greater in the fall than the spring.

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

302 Comparing the average of estimated coefficients of variation for annual calibrated and  
303 uncalibrated biomass estimates showed large increases for summer flounder in the fall  
304 (> 50%), southern New England winter flounder in the spring (77%), Georges Bank winter  
305 flounder (more than 200% for spring and fall), northern red hake (95% for spring and 178%  
306 for fall), northern goosefish in the fall (93%), and barndoor skate (> 100% for both spring and  
307 fall) induced by the variability in the estimation of the relative catch efficiency of the gears  
308 using chain and rockhopper sweep gears (Table 6). Effects of calibration on the precision of  
309 the biomass estimates was relatively minor for other stocks.

310 We observed little correlation of annual biomass estimates induced by the relative catch  
311 efficiency estimation for most of the stocks (Table 7). However, the biomass estimates were  
312 highly correlated for George Bank winter flounder in the spring (65%) and barndoor skate

<sup>313</sup> ( $> 70\%$  on average). Estimates for Georges Bank winter flounder in the fall, both red hake  
<sup>314</sup> stocks, northern goosefish, and thorny skate were greater than 20% on average.

## <sup>315</sup> 4 Discussion

<sup>316</sup> The data that we used to estimate bottom trawl survey catch efficiency came from an  
<sup>317</sup> experiment using a twin trawler and many of the standard tow protocols for the NEFSC  
<sup>318</sup> survey on the *R/V Henry B. Bigelow*. The experimental net used on one side of the twin  
<sup>319</sup> trawl was the same as the standard survey trawl used on the *Bigelow* except that it contained  
<sup>320</sup> roughly half number of floats and the sweep was modified to optimize flatfish catch by  
<sup>321</sup> minimizing the ability of flatfish to pass under the net. The other side of the twin trawl was  
<sup>322</sup> essentially identical to the standard gear used on the *Bigelow*. The towing of the standard  
<sup>323</sup> survey bottom trawl on the twin trawl experiment differed in a few ways from its deployment  
<sup>324</sup> on the spring and fall bottom trawl surveys, but we believe that these differences did not  
<sup>325</sup> have a significant effect on the results. The use of larger doors and the restrictor rope served  
<sup>326</sup> to fix the net geometries which may be the biggest source of variability in comparative trawl  
<sup>327</sup> catches. This setup also allowed us to avoid many of the potential problems due to the large  
<sup>328</sup> differences in size of the *Bigelow* and the *F/V Karen Elizabeth*. We do not suspect that the  
<sup>329</sup> use of the restrictor rope would influenced flatfish behavior in front of the trawl because  
<sup>330</sup> flatfish have been shown to generally not react to trawling induced stimuli until they are in  
<sup>331</sup> very close proximity or even contacted by the fishing gear (Ryer et al., 2010). The spread  
<sup>332</sup> data indicated that the restrictor rope remained taut throughout the towing process (setting,  
<sup>333</sup> towing, hauling back), so we believe it likely that the restrictor rope was almost always at  
<sup>334</sup> least 1 m off the bottom.

<sup>335</sup> Herding is a known phenomena for flatfish and many other species when certain types of gear  
<sup>336</sup> are used (Ramm and Xiao, 1995; Somerton and Munro, 2001; Somerton et al., 2007; Rose

<sup>337</sup> et al., 2010). Somerton and Munro (2001) considered two factors of bridle herding effects on  
<sup>338</sup> efficiency. The first factor was the size of the bridle path where the bridle is off the ground  
<sup>339</sup> ( $W_{off}$ ) and the second factor, the herding efficiency ( $h$ ) was fraction of fish in the bridle  
<sup>340</sup> contact path that are moved into the path of the net path. The former is a function of gear  
<sup>341</sup> design, and controllable, whereas the latter is a function of fish behavior with regard to the  
<sup>342</sup> bridle when it is in contact with the substrate. The bridle configuration on the bottom trawl  
<sup>343</sup> survey are designed to minimize contact with the bottom (reference) and lack of abrasion of  
<sup>344</sup> painted bridles used during one of the twin trawler research trips provided evidence of little  
<sup>345</sup> or no bridle contact during the paired tow experiments used to collect the data used in this  
<sup>346</sup> study. Furthermore, studies have consistently found that herding behavior occurred during  
<sup>347</sup> the daytime (Glass and Wardle, 1989; Somerton and Munro, 2001; Ryer and Barnett, 2006;  
<sup>348</sup> Bryan et al., 2014; Ryer et al., 2010; Dean et al., 2021) with some studies indicating high  
<sup>349</sup> herding coefficients ( $h$ ) along the sections of the bridles in contact with the bottom. Studies  
<sup>350</sup> that have evaluated herding at night or in low light conditions did not find evidence for a  
<sup>351</sup> directional herding response (Glass and Wardle, 1989; Ryer and Barnett, 2006; Ryer, 2008;  
<sup>352</sup> Ryer et al., 2010). The minimal bridle contact with the substrate and the large fraction of  
<sup>353</sup> nighttime tows during the bottom trawl survey suggests flatfish herding is unlikely to be an  
<sup>354</sup> important factor in catch efficiency based on net spread-based swept area.

<sup>355</sup> On the other hand, the biomass estimates assume that the chain sweep gear is fully efficient,  
<sup>356</sup> but it is likely at least some small fraction of fish, that may depend on size, are not captured  
<sup>357</sup> by the gear. The biomass estimates also implicitly assume that the entire stock is available  
<sup>358</sup> to the bottom trawl survey, but many of these stocks extend somewhat outside of the survey  
<sup>359</sup> strata used to define the indices either throughout the year or seasonally due to migration.  
<sup>360</sup> If either of these assumptions do not hold this method of biomass estimation would be  
<sup>361</sup> negatively biased (expected value of biomass estimates would be lower than the true value).  
<sup>362</sup> However, estimation using the data from these paired-gear studies and these assumptions is  
<sup>363</sup> less biased than those made without them.

<sup>364</sup> Treating the chain sweep-based abundance estimates implicitly assumes that the chain sweep  
<sup>365</sup> provides complete capture efficiency and that the stock resides completely within the strata  
<sup>366</sup> that are used to generate the abundance indices. It is unlikely that the chain sweep gear is  
<sup>367</sup> completely efficient for all sizes of fish of a particular species over all substrate types that are  
<sup>368</sup> sampled. It is also typical for many of these stocks to extend somewhat outside of the survey  
<sup>369</sup> strata used to define the indices either throughout the year or seasonally due to migration.

<sup>370</sup> These analyses treat the amount of daylight as binary effect (day/night) on the relative catch  
<sup>371</sup> efficiency. However, behavior of the fish with respect to the gear is likely to change more  
<sup>372</sup> gradually with the amount of light. A continuous measure of light that uses the angle of  
<sup>373</sup> the sun, the depth of the tow and light attenuation with depth, might prove to be a better  
<sup>374</sup> explanatory variable for changes in relative catch efficiency and perhaps improve estimation  
<sup>375</sup> of abundance from the bottom trawl survey (Jacobson et al., 2015; Kainge et al., 2017).

<sup>376</sup> Aside from the direct impact of estimated catch efficiency of the NEFSC trawl survey gear on  
<sup>377</sup> biomass estimation. Our analyses show more subtle impacts of using independent estimates of  
<sup>378</sup> efficiency with survey tow data to generate the abundance indices. Without the independent  
<sup>379</sup> efficiency estimates, the sampling variability of each of the seasonal and annual relative  
<sup>380</sup> abundance indices is independent of the others. The bootstrapping methods we employed  
<sup>381</sup> illustrated the including estimates of catch efficiency affects the variability of the resulting  
<sup>382</sup> abundance estimates and their independence from each other. For some stocks there was a  
<sup>383</sup> substantial effect of the relative catch efficiency estimation on the precision of the biomass  
<sup>384</sup> indices. Similarly, we found high correlation of annual indices ( $> 0.6$ ) for Georges Bank  
<sup>385</sup> winter flounder and barndoor skate. In these cases, the decrease in precision and increased  
<sup>386</sup> correlation may have an impact on bias and precision of important estimates from the  
<sup>387</sup> assessment model such as stock size and fishing mortality. As such, future work should  
<sup>388</sup> evaluate effects of incorporating this information in an assessment model.

<sup>389</sup> The estimates of absolute abundance produced using the sweep comparison experiments have

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

Typically, research surveys provide only a relative index of abundance rather than an absolute estimate of abundance. Stock assessment models then integrate these observations with time series of catch and other data sources to determine the scale of the population. However, various factors can make for imprecise and inaccurate scaling of population levels including inaccurate catch data (Cadigan, 2016), time-varying catchability (Wilberg et al., 2009), low fishing mortality rates over the time series (Adams et al., 2020), and uncertain and time-varying natural mortality (Stock et al., 2021). In these cases, external information such as those produced by studies such as ours, can be particularly useful in estimating the size of the stock, the status of the stock relative to optimal levels and ultimately making catch advice for commercially important fish stocks.

## <sup>414</sup> Acknowledgements

<sup>415</sup> The quality of this manuscript has been improved by comments from Jessica Blaylock? We

<sup>416</sup> thank Adam Poquette for creating the map in Figure 3.

417 **References**

- 418 Adams, C. F., Miller, T. J., Manderson, J. P., Richardson, D. E., and Smith, B. E. 2020.  
419 Butterfish 2014 Assessment. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-06;  
420 110 p.
- 421 Arreguín-Sánchez, F. 1996. Catchability: a key parameter for fish stock assessment. Reviews  
422 in Fish Biology and Fisheries **6**: 221–242, doi: 10.1007/BF00182344.
- 423 Azarovitz, T. S. 1981. A brief historical review of the Woods Hole Laboratory trawl survey  
424 time series. In Bottom Trawl Surveys. Edited by W. G. Doubleday and D. Rivard. Can.  
425 Spec. Pub. Fish. Aqua. Sci. No. 58. pp. 62–67.
- 426 Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekkinga, A., Kraan, C., van der Veer, H. W.,  
427 and Piersma, T. 2012. Designing a benthic monitoring programme with multiple conflict-  
428 ing objectives. Methods in Ecology and Evolution **3**(3): 526–536, doi: 10.1111/j.2041-  
429 210X.2012.00192.x.
- 430 Bourne, N. 1965. A comparison of catches by 3- and 4-inch rings on offshore scallop drags.  
431 Journal of the Fisheries Research Board of Canada **22**(2): 313–333.
- 432 Bryan, D. R., Bosley, K. L., Hicks, A. C., Haltuch, M. A., and Wakefield, W. W. 2014.  
433 Quantitative video analysis of flatfish herding behavior and impact on effective area swept  
434 of a survey trawl. Fisheries Research **154**: 120–126, doi: 10.1016/j.fishres.2014.02.007.
- 435 Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including  
436 under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries  
437 and Aquatic Sciences **73**(2): 296–308, doi: 10.1139/cjfas-2015-0047.
- 438 Dean, M. J., Hoffman, W. S., Buchan, N. C., Cadrian, S. X., and Grabowski, J. H. 2021. The  
439 influence of trawl efficiency assumptions on survey-based population metrics. ICES Journal  
440 of Marine Science **78**(8): 2858–2874.

- 441 Glass, C. W. and Wardle, C. S. 1989. Comparison of the reactions of fish to a trawl gear,  
442 at high and low light intensities. *Fisheries Research* **7**(3): 249–266, doi: 10.1016/0165-  
443 7836(89)90059-3.
- 444 Gulland, J. A. 1964. Variations in selection factors, and mesh differentials. *Journal Du Conseil  
445 International Pour L'exploration De La Mer* **29**(2): 158–165.
- 446 ICES. 1996. Manual of methods of measuring the selectivity of towed fishing gears. (Eds.)  
447 Wileman, D. A., Ferro, R. S. T., Fonteyne, R., and Millar, R. B. ICES Cooperative Research  
448 Report No. 215.
- 449 Jacobson, L. D., Hendrickson, L. C., and Tang, J. 2015. Solar zenith angles for biological  
450 research and an expected catch model for diel vertical migration patterns that affect stock  
451 size estimates for longfin inshore squid (*Doryteuthis pealeii*). *Canadian Journal of Fisheries  
452 and Aquatic Sciences* **72**(9): 1329–1338, doi: 10.1139/cjfas-2014-0436.
- 453 Jones, A. W., Miller, T. J., Politis, P. J., Richardson, D. E., Mercer, A. M., Pol, M. V., and  
454 Roebuck, C. D. 2021. Experimental assessment of the effect of net wing spread on relative  
455 catch efficiency of four flatfishes by a four seam bottom trawl. *Fisheries Research* **244**:  
456 106106, doi: 10.1016/j.fishres.2021.106106.
- 457 Kainge, P., van der Plas, A. K., Bartholomae, C. H., and Wieland, K. 2017. Effects of  
458 environmental variables on survey catch rates and distribution by size of shallow- and  
459 deep-water cape hakes, *merluccius capensis* and *merluccius paradoxus* off namibia. *Fisheries  
460 Oceanography* **26**(6): 680–692, doi: 10.1111/fog.12227.
- 461 Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2013. Combining  
462 bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl  
463 efficiency parameters for semipelagic species. *Canadian Journal of Fisheries and Aquatic  
464 Sciences* **70**(2): 208–219, doi: 10.1139/cjfas-2012-0321.
- 465 Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. 2017. Selectivity ratio: A

- 466 useful tool for comparing size selectivity of multiple survey gears. *Fisheries Research* **191**:  
467 76–86, doi: 10.1016/j.fishres.2017.02.012.
- 468 Kotwicki, S., Martin, M. H., and Laman, E. A. 2011. Improving area swept estimates from bot-  
469 tom trawl surveys. *Fisheries Research* **110**(1): 198–206, doi: 10.1016/j.fishres.2011.04.007.
- 470 Krag, L. A., Herrmann, B., Karlsen, J. D., and Mieske, B. 2015. Species selectivity in different  
471 sized topless trawl designs: Does size matter? *Fisheries Research* **172**: 243–249, doi:  
472 10.1016/j.fishres.2015.07.010.
- 473 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic  
474 differentiation and Laplace approximation. *Journal of Statistical Software* **70**(5): 1–21.
- 475 Legault, C. M. and McCurdy, Q. M. 2017. Stock assessment of Georges Bank yellowtail  
476 flounder for 2017. Transboundary Resources Assessment Committee (TRAC) Reference  
477 Document 2017/03 available at: <https://repository.library.noaa.gov/view/noaa/27556>.
- 478 Maunder, M. N. and Piner, K. R. 2015. Contemporary fisheries stock assessment: many issues  
479 still remain. *ICES Journal of Marine Science* **72**(1): 7–18, doi: 10.1093/icesjms/fsu015.
- 480 Millar, R. B. and Fryer, R. J. 1999. Estimating the size-selection curves of towed gears,  
481 traps, nets and hooks. *Reviews in Fish Biology and Fisheries* **9**(1): 89–116, doi:  
482 10.1023/A:1008838220001.
- 483 Miller, T. J. 2013. A comparison of hierarchical models for relative catch efficiency based on  
484 paired-gear data for U.S. Northwest Atlantic fish stocks. *Canadian Journal of Fisheries  
485 and Aquatic Sciences* **70**(9): 1306–1316, doi: 10.1139/cjfas-2013-0136.
- 486 Miller, T. J., Hart, D. R., Hopkins, K., Vine, N. H., Taylor, R., York, A. D., and Gallager,  
487 S. M. 2019. Estimation of the capture efficiency and abundance of Atlantic sea scallops  
488 (*Placopecten magellanicus*) from paired photographic-dredge tows using hierarchical models.  
489 *Canadian Journal of Fisheries and Aquatic Sciences* **76**(6): 847–855, doi: 10.1139/cjfas-  
490 2018-0024.

- 491 Munro, P. T. and Somerton, D. A. 2001. Maximum likelihood and non-parametric methods  
492 for estimating trawl footrope selectivity. ICES Journal of Marine Science **58**(1): 220–229,  
493 doi: 10.1006/jmsc.2000.1004.
- 494 NEFSC. 2019. 66th Northeast Regional Stock Assessment Workshop (66th SAW) Assessment  
495 Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 19-08; 1170 p. Available from:  
496 <http://www.nefsc.noaa.gov/publications/>.
- 497 NEFSC. 2020a. Final report of red hake stock structure working group. US Dept Commer,  
498 Northeast Fish Sci Cent Ref Doc. 20-07; 185 p.
- 499 NEFSC. 2020b. Operational assessment of 14 northeast groundfish stocks, updated through  
500 2018. Prepublication Copy of the September 2019 Operational Stock Assessment Report.  
501 The report is “in preparation” for publication by the NEFSC. (Oct. 3, 2019; latest revision:  
502 Jan. 7, 2020).
- 503 NEFSC. 2020c. Operational assessment of the black sea bass, scup, bluefish, and monkfish  
504 stocks, updated through 2018. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 20-01;  
505 160 p.
- 506 Nichol, D. G., Kotwicki, S., Wilderbuer, T. K., Lauth, R. R., and Ianelli, J. N. 2019.  
507 Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey  
508 and its effect on estimates of survey biomass. Fisheries Research **211**: 319–330, doi:  
509 10.1016/j.fishres.2018.11.017.
- 510 Paloheimo, J. E. and Dickie, L. M. 1964. Abundance and fishing success. Rapports et  
511 Procès-Verbaux des Réunions. Conseil Internationale pour l’Exploration de la Mer **155**:  
512 152–163.
- 513 Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an  
514 application. Canadian Journal of Fisheries and Aquatic Sciences **55**(12): 2672–2690, doi:  
515 10.1139/f98-151.

- 516 Politis, P. J., Galbraith, J. K., Kostovick, P., and Brown, R. W. 2014. Northeast Fisheries  
517 Science Center bottom trawl survey protocols for the NOAA Ship Henry B. Bigelow. U.S.  
518 Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 14-06, 138p.
- 519 R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation  
520 for Statistical Computing, Vienna, Austria.
- 521 Ramm, D. C. and Xiao, Y. 1995. Herding in groundfish and effective pathwidth of trawls.  
522 Fisheries Research **24**(3): 243–259, doi: 10.1016/0165-7836(95)00373-I.
- 523 Rose, C. S., Gauvin, J. R., and Hammond, C. F. 2010. Effective herding of flatfish by cables  
524 with minimal seafloor contact. Fishery Bulletin **108**(2): 136–144.
- 525 Ryer, C. H. 2008. A review of flatfish behavior relative to trawls. Fisheries Research **90**(1):  
526 138–146, doi: 10.1016/j.fishres.2007.10.005.
- 527 Ryer, C. H. and Barnett, L. A. K. 2006. Influence of illumination and temperature upon  
528 flatfish reactivity and herding behavior: Potential implications for trawl capture efficiency.  
529 Fisheries Research **81**(2): 242–250, doi: 10.1016/j.fishres.2006.07.001.
- 530 Ryer, C. H., Rose, C. S., and Iseri, P. J. 2010. Flatfish herding behavior in response to trawl  
531 sweeps: a comparison of diel responses to conventional sweeps and elevated sweeps. Fishery  
532 Bulletin **108**(2): 145–154.
- 533 Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean  
534 abundance. Canadian Journal of Fisheries and Aquatic Sciences **54**(3): 616–630, doi:  
535 10.1139/f96-303.
- 536 Somerton, D., Ianelli, J., Walsh, S., Smith, S., Godø, O. R., and Ramm, D. 1999. Incorporating  
537 experimentally derived estimates of survey trawl efficiency into the stock assessment process:  
538 a discussion. ICES Journal of Marine Science **56**(3): 299–302, doi: 10.1006/jmsc.1999.0443.

- 539 Somerton, D. A. and Munro, P. 2001. Bridle efficiency of a survey trawl for flatfish. Fishery  
540 Bulletin **99**(4): 641–652.
- 541 Somerton, D. A., Munro, P. T., and Weinberg, K. L. 2007. Whole-gear efficiency of a benthic  
542 survey trawl for flatfish. Fishery Bulletin **105**(2): 278–291.
- 543 Somerton, D. A., Weinberg, K. L., and Goodman, S. E. 2013. Catchability of snow crab  
544 (*Chionoecetes opilio*) by the eastern Bering Sea bottom trawl survey estimated using a  
545 catch comparison experiment. Canadian Journal of Fisheries and Aquatic Sciences **70**(12):  
546 1699–1708, doi: 10.1139/cjfas-2013-0100.
- 547 Stock, B. C., Xu, H., Miller, T. J., Thorson, J. T., and Nye, J. A. 2021. Implementing two-  
548 dimensional autocorrelation in either survival or natural mortality improves a state-space  
549 assessment model for Southern New England-Mid Atlantic yellowtail flounder. Fisheries  
550 Research **237**: 105873, doi: 10.1016/j.fishres.2021.105873.
- 551 Thiess, M. E., Benoit, H., Clark, D. S., Fong, K., Mello, L. G. S., Mowbray, F., Pepin, P.,  
552 Cadigan, N. G., Miller, T. J., Thirkell, D., and Wheeland, L. 2018. Proceedings of the  
553 National Comparative Trawl Workshop, November 28-30, 2017, Nanaimo, B. C. Canadian  
554 Technical Report of Fisheries and Aquatic Sciences 3254, x + 40 p.
- 555 Thygesen, U. H., Kristensen, K., Jansen, T., and Beyer, J. E. 2019. Intercalibration of survey  
556 methods using paired fishing operations and log-Gaussian Cox processes. ICES Journal of  
557 Marine Science **76**(4): 1189–1199, doi: 10.1093/icesjms/fsy191.
- 558 Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. 2018. Evaluation of alternative  
559 stratifications for a stratified random fishery-independent survey. Fisheries Research **207**:  
560 150–159, doi: 10.1016/j.fishres.2018.06.019.
- 561 Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. 2009. Incorporating time-  
562 varying catchability into population dynamic stock assessment models. Reviews in Fisheries  
563 Science **18**(1): 7–24, doi: 10.1080/10641260903294647.

- <sup>564</sup> Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall,  
<sup>565</sup> Boca Raton, Florida, 392 pp.

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

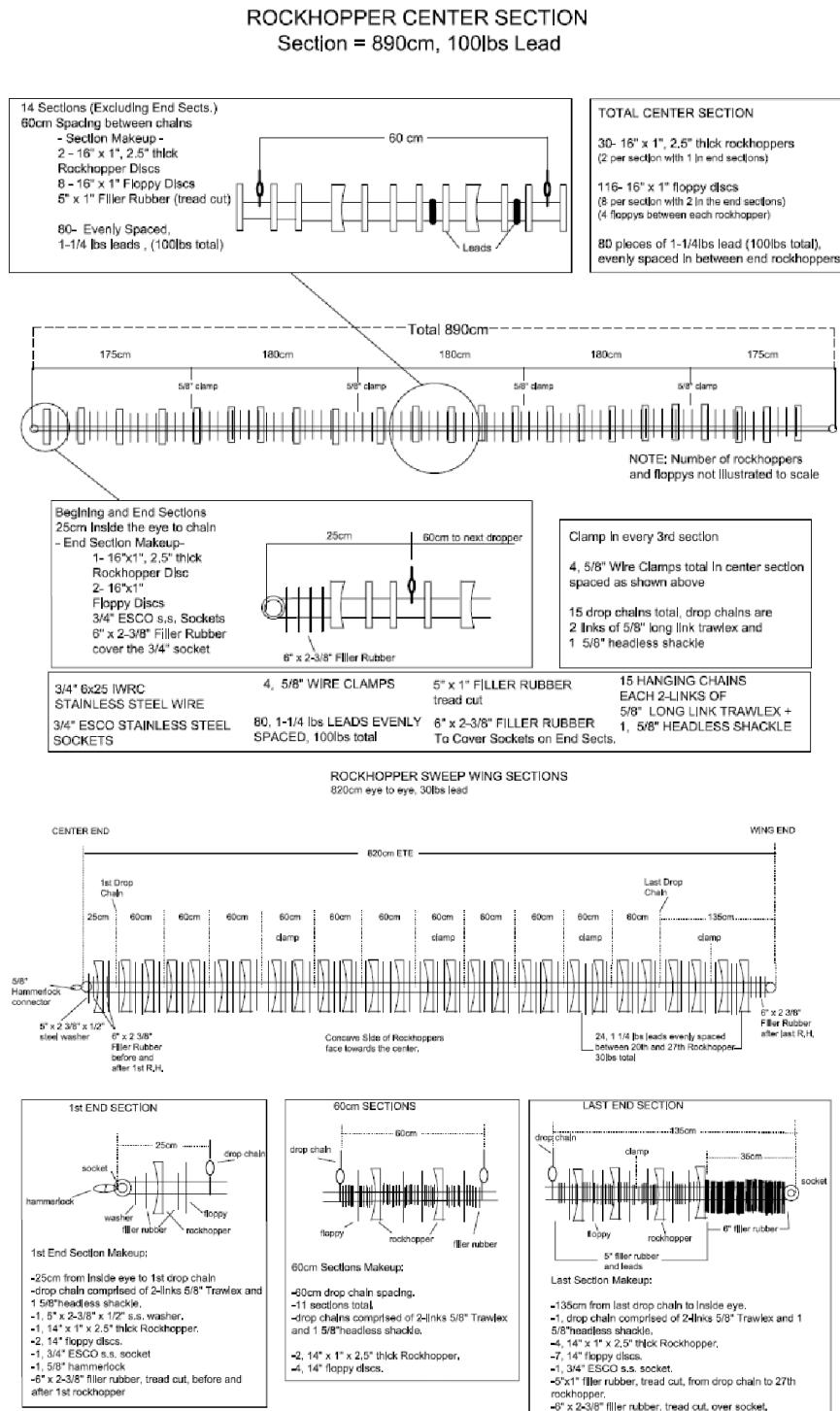


Fig. 2. Diagram of the chain sweep designed maximize bottom contact and flatfish capture.

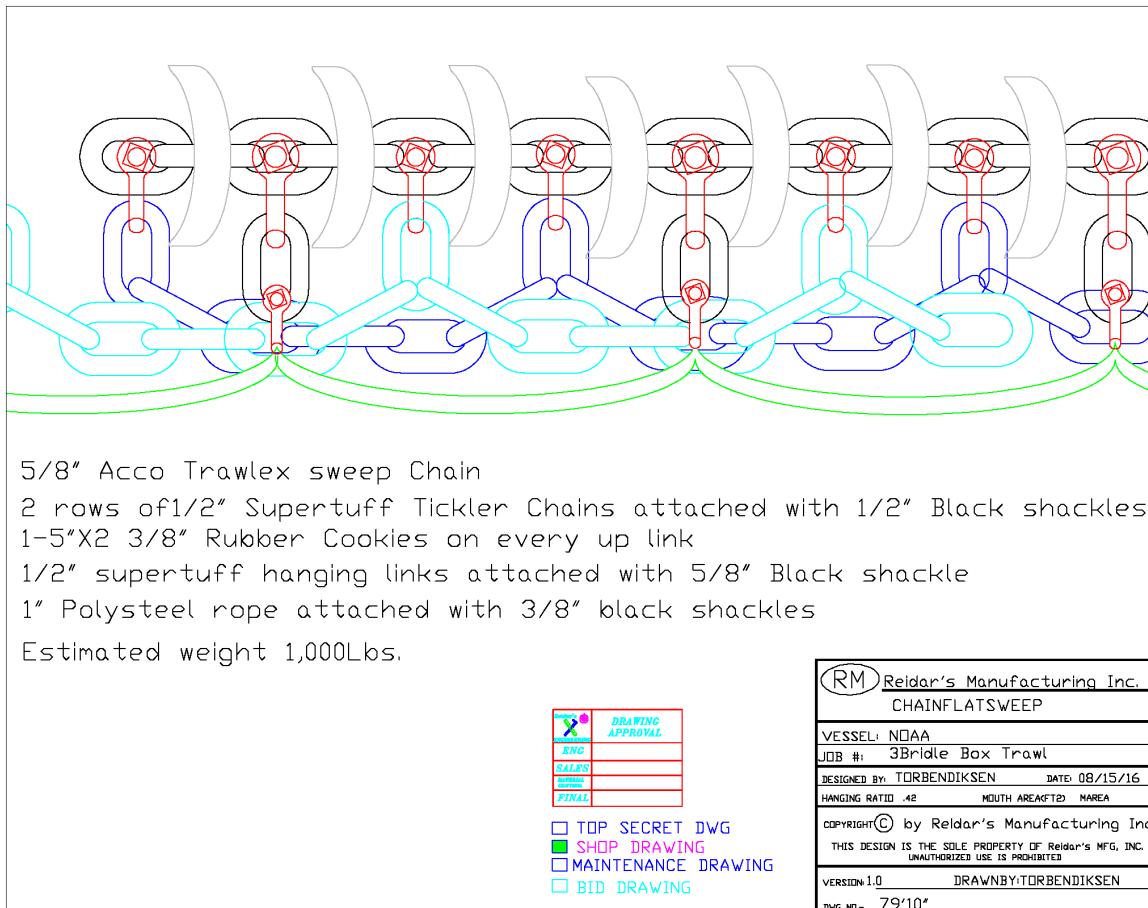


Fig. 3. Annual locations of stations during where the F/V Karen Elizabeth conducted twin-trawl sets with the standard bottom trawl gear and the gear with a chain sweep instead of the rockhopper sweep.

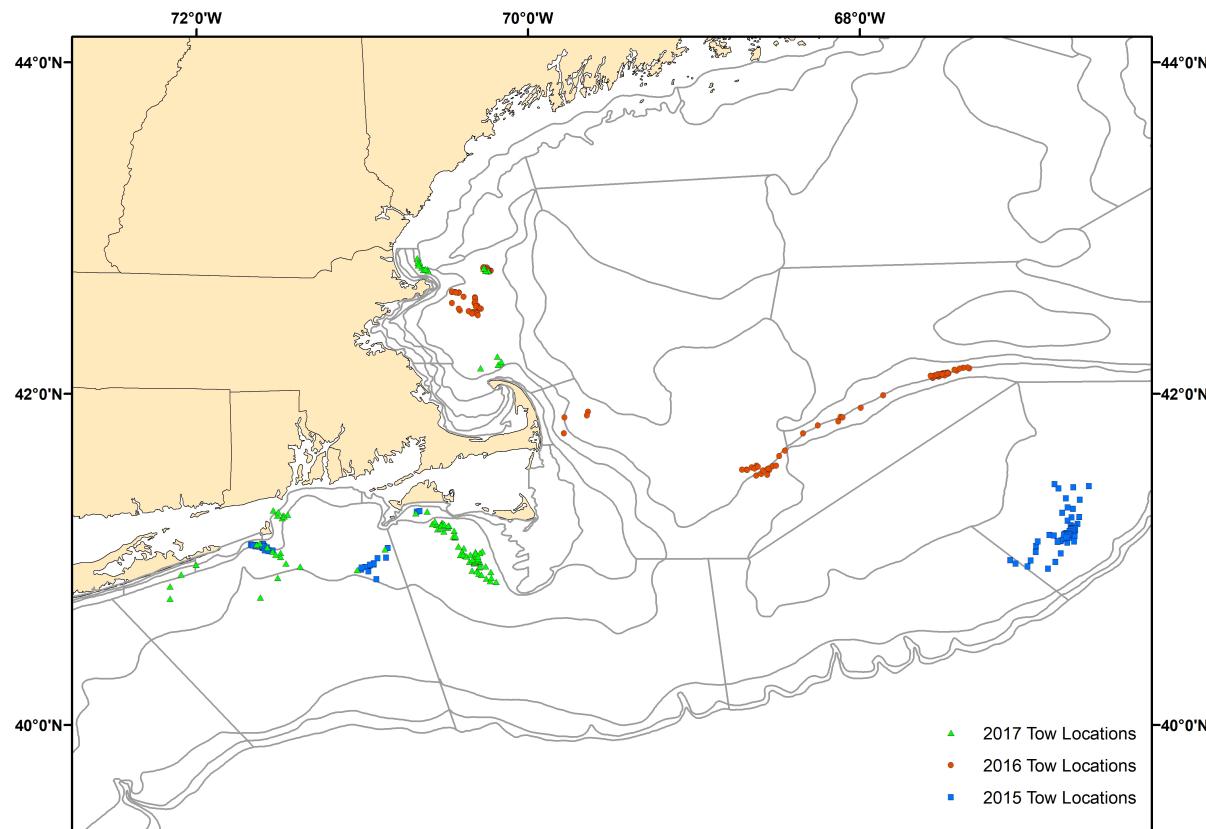


Fig. 4. Relative efficiency of gears using chain and rockhopper sweeps from the best performing model for each species (Table 5). Blue and red denote results for day and night data, respectively, and thick and thin lines represent overall and paired-tow specific estimates of relative catch efficiency,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

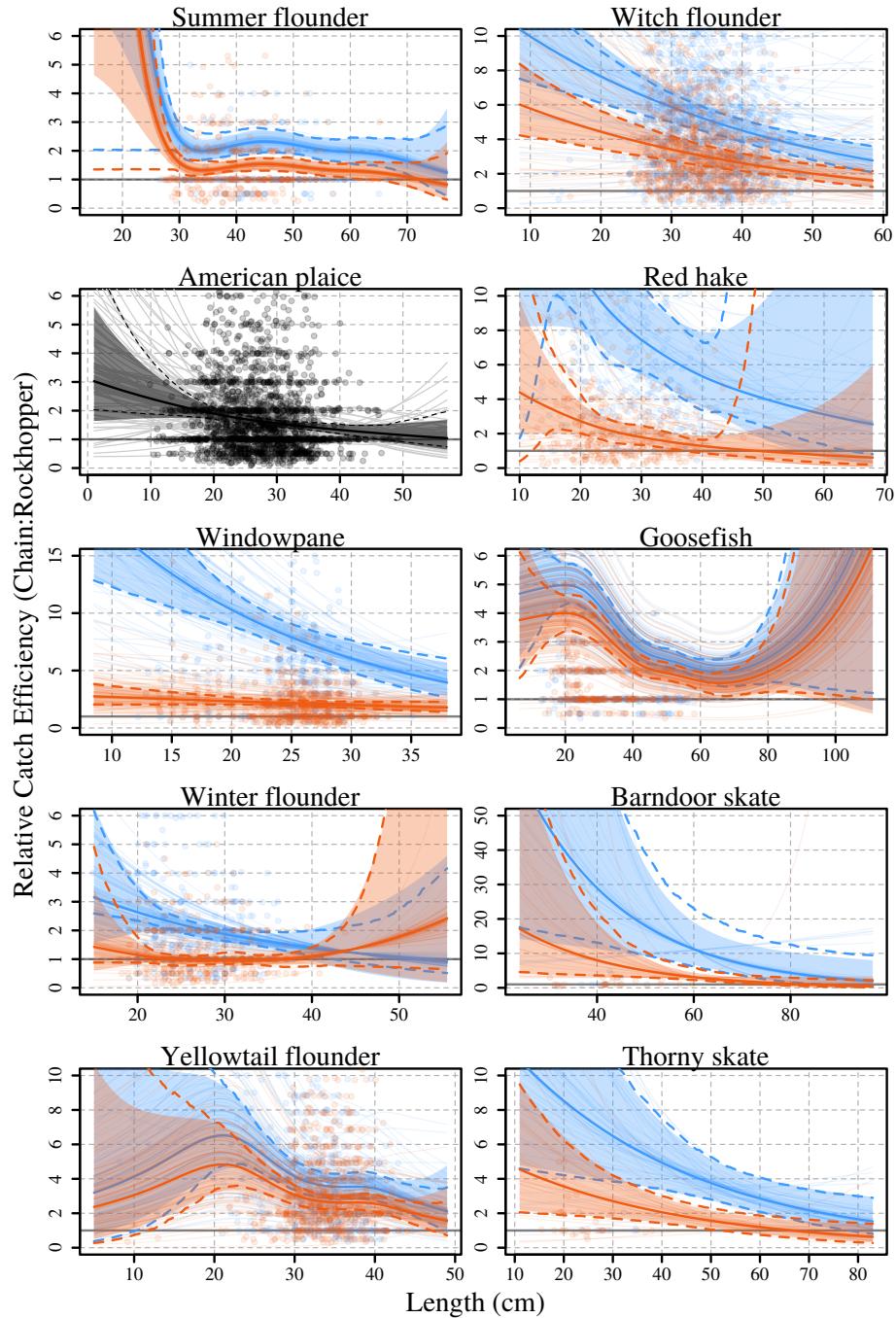


Fig. 5. Annual spring (blue) and fall (red) biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps are from the best performing model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

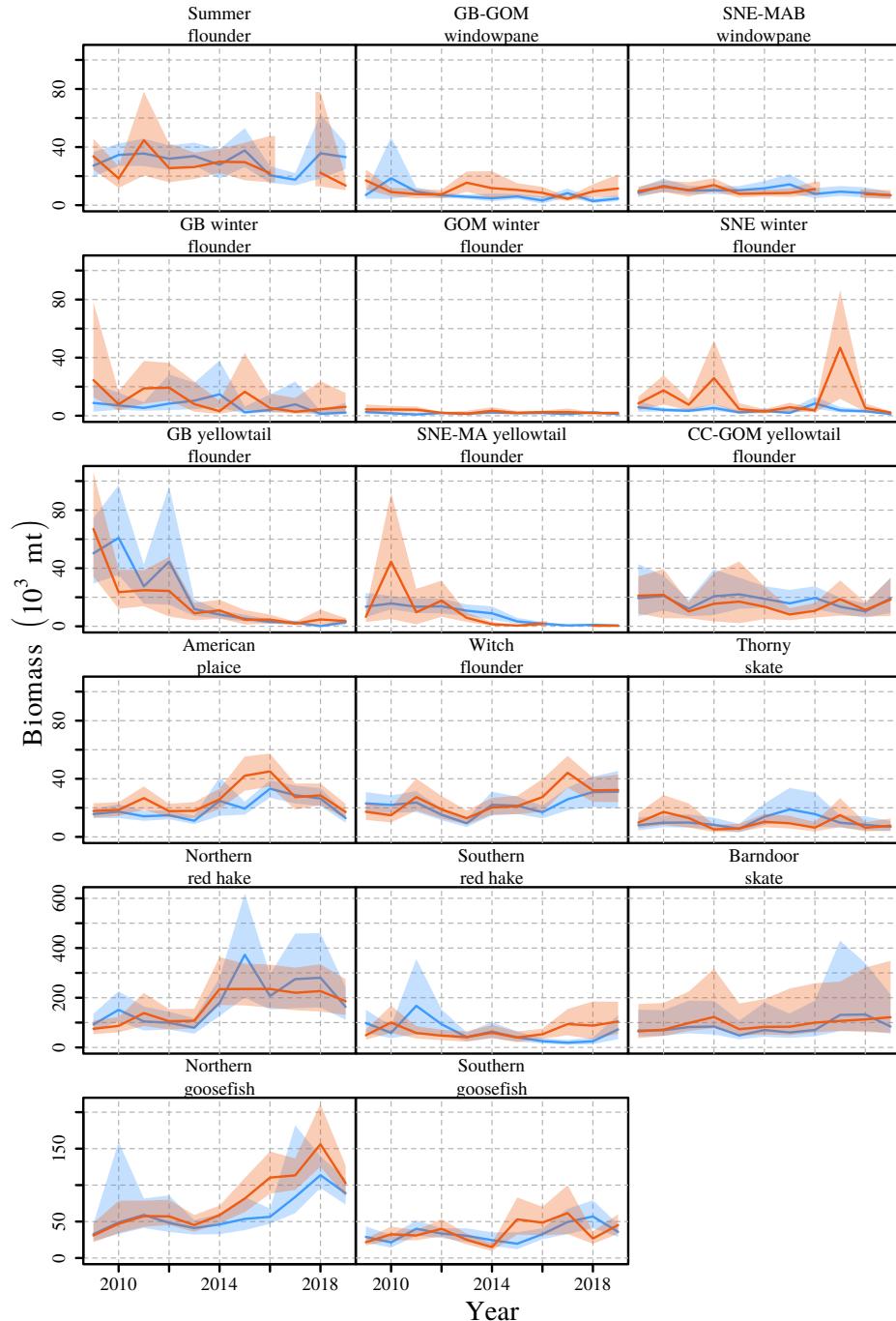


Fig. 6. Implied efficiency of annual spring (blue) and fall (red) bottom trawl survey biomass estimates for each managed stock assuming 100% efficiency for chain sweep gear with shaded polygons representing bootstrap-based 95% confidence intervals. Relative catch efficiency at size estimates and bootstraps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

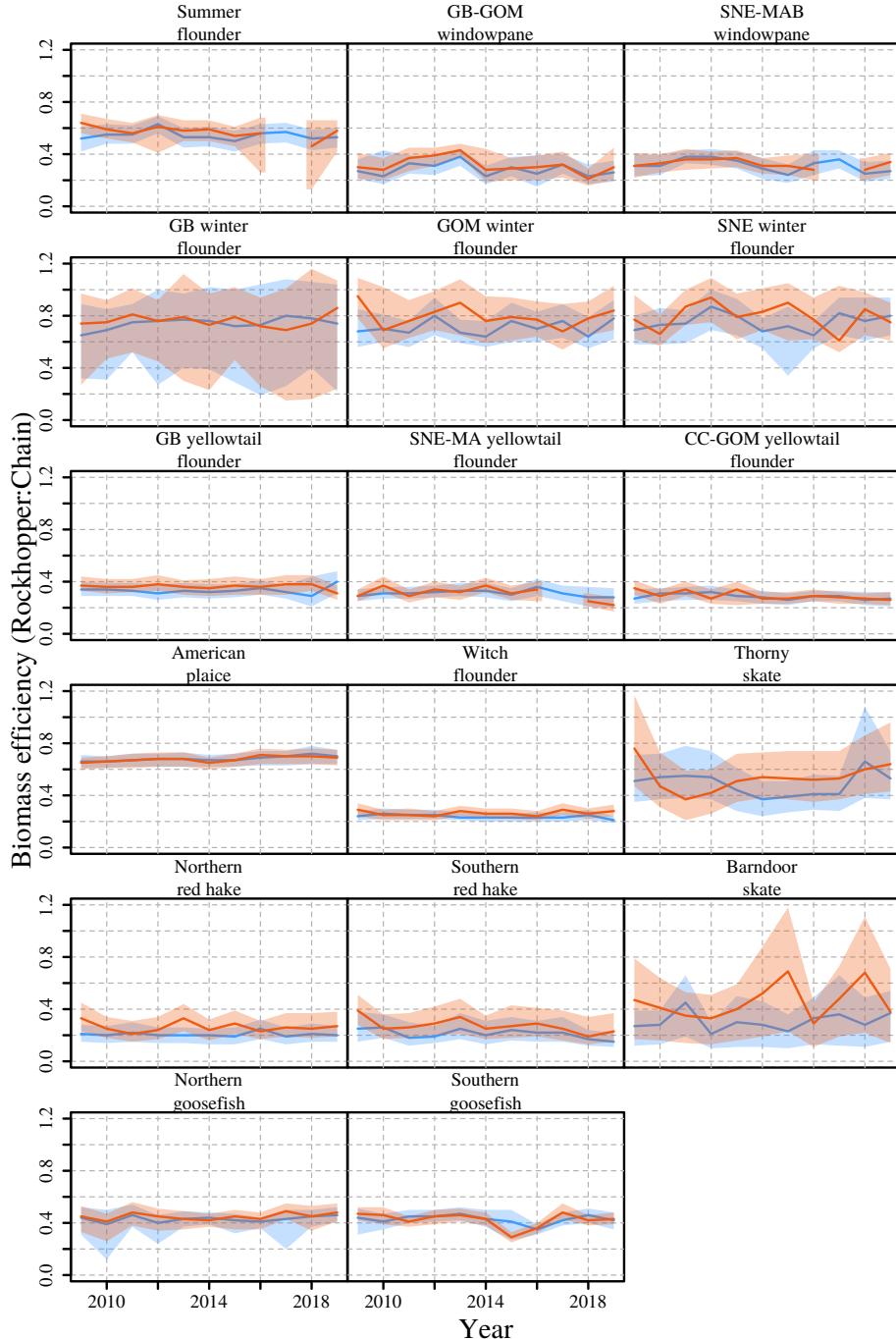


Table 1. Managed stocks associated with the species for which relative catch efficiency was estimated.

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		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