

Supplementary Material – Extended Methods and Results

Supplementary Material for ‘A statistical censoring approach accounts for hook competition in abundance indices from longline surveys’ by Joe Watson, Andrew M. Edwards, and Marie Auger-Méthé, published in *Canadian Journal of Fisheries and Aquatic Sciences*.

Table of Contents

S.1 Simulation study set-up and additional results	2
S.1.1 Step 1 – generate the numbers of each species attracted to the fishing gear . .	2
S.1.2 Step 2 – sample the bite times at which individual fish would take the bait given unlimited hooks	4
S.1.3 Step 3 – explicitly simulate hook competition to see which fish get caught . .	5
S.1.4 Step 4 – fit the simulated data using three alternative methods	7
S.1.5 Experiment 1	9
S.1.6 Experiment 2	11
S.1.7 Experiment 3	13
S.1.8 Experiment 4	14
S.1.9 Experiment 5	15
S.1.10 Experiment 6	15
S.1.11 Extrapolating results from experiments 1-5 to a target species with a decreas- ing abundance	18
S.1.12 (Bonus) experiment 7	21
S.1.13 Fitting details for all experiments and additional results	22
 S.2 Proof of third theoretical example	 42
 S.3 Deriving an upper bound to improve censored method’s convergence in data- poor settings	 43
 S.4 Exploratory analysis for the case study	 46
S.4.1 Assessing the linear relationship between effective skate and catch counts . .	46

S.4.2 Empirically assessing values of \hat{p}_i^*	48
S.5 Additional details on the case study models	50
S.6 Additional case study results	52
S.7 R code for fitting the censored method	61
S.7.1 R-INLA code	62
S.7.2 TMB code	63
S.7.3 Stan code via the BRMS package	64

S.1 Simulation study set-up and additional results

We generate data for the simulation study by determining the underlying numbers of each species that are attracted to the fishing gear (Step 1), determining how quickly the individual fish would take a bait (in the absence of hook competition) depending on the aggressiveness of the species (Step 2), and explicitly simulate hook competition to ascertain which individual fish actually get caught by the gear (Step 3). We then apply competing methods to the resulting simulated catch counts to infer temporal changes in the abundance of each species (Step 4).

S.1.1 Step 1 – generate the numbers of each species attracted to the fishing gear

We sample the number of individuals of each of three species groups that are attracted to the longline gear, which involves first specifying the underlying species density $\lambda_{it}(\mathbf{s})$. We want to illustrate the strengths and weaknesses of the competing methods in the simplest settings, free from confounding by external factors and spatial heterogeneity. So we do not explicitly simulate the influence of covariates on the population or on the response function that determines whether an individual fish removes bait. This involves setting $x_{itkm} = w_{itkr} = Z_{it}(\mathbf{s}_k) = 0$ in (13), and also we can set $\beta_{i0} = 0$ (it can be set to an arbitrary constant that would be absorbed by α_{it0} anyway).

Table S1: Extra notation used primarily in the Supplementary Material.

L_i	lower bound for the probability that an attracted fish of species i is captured by a baited hook, equal to the probability when only one baited hook remains
S_{tk}	soak time of fishing event k in year t
\tilde{c}_{itk}	response variable of interest, equal to c_{itk} for the CPUE and censored method, and multiplied by the hook-competition adjustment factor for the ICR method, to enable writing all models under the framework (S.23)–(S.27)
Q	fishing events with target species' catch counts lying above the Q^{th} quantile of annual observed catch counts are considered high quality when the 'quantile adjustment' is implemented
c'_{itk}	the number of species i caught during the first stage of the soak time in the k^{th} fishing event of year t
c''_{itk}	the number of species i caught during the second stage of the soak time in the k^{th} fishing event of year t
\mathbf{g}_{itk}	a vector for which each element corresponds to a bite (gobble) time of a fish of species i that was attracted to the gear during fishing event k of year t
$p(\tau)$	the proportion of baits removed τ units of time into the soak time
u_{itk}	upper bound of the catch counts of species i in year t at fishing event k
μ_{it}	expected value of C_{itk} assuming no hook competition and no spatial heterogeneity between stations

Then (13) reduces to $\log \mu_{itk} = \alpha_{it0}$, and (1) and (10) reduce to $\log \lambda_{it}(\mathbf{s}) = \alpha_{it}(\mathbf{s}) = \alpha_{it0}$, such that $\mu_{itk} = \lambda_{it}(\mathbf{s})$. By spatial homogeneity, $\mu_{itk} = \mu_{it} = \exp(\alpha_{it0})$, with μ_{it} determining the expected catch count of species i in year t during any fishing event, assuming no hook competition. These settings help constrain the number of experiments conducted. Extra notation used primarily in the Supplementary Material is summarised in Table S1.

For species group $i \in \{1 = \text{competitor}, 2 = \text{other}, 3 = \text{target}\}$, in year $t = 1, 2, \dots, 6$, and station (replicate) $k = 1, 2, \dots, K_t$, the catch count in the absence of hook competition can be sampled from

the following model based on (5) and (13):

$$[C_{itk}|e_{itk}] \sim \text{Poisson}(\mu_{it}e_{itk}) \quad (\text{S.1})$$

$$\log \mu_{it} = \begin{cases} \alpha_{i,1}, & t = 1 \\ \alpha_{i,1} + \alpha_{it}, & t = 2, 3, 4, 5, 6 \end{cases} \quad (\text{S.2})$$

$$[\log e_{1tk}, \log e_{2tk}, \log e_{3tk}] \sim \text{MVNormal}(\mathbf{0}, \mathbf{\Sigma}), \quad (\text{S.3})$$

where e_{itk} represents lognormal random noise to account for overdispersion in catch counts, α_{it} prescribe the μ_{it} (though for ease of understanding we give values of μ_{it}), and $\mathbf{\Sigma}$ describes: i) the magnitude of overdispersion in the count distributions of the three species groups; ii) the prevalence and severity of schooling behaviours exhibited by each species; iii) the correlations between the count distributions of the three species groups. The covariance matrix $\mathbf{\Sigma}$ is fixed in time and so $\exp(\alpha_{it})$ for $t > 1$ defines the relative abundance of species i in year t , compared with its abundance in year 1 (considered as the baseline level). The $\exp(\alpha_{it})$ for $t > 1$ are the inferential targets for the simulation study. We chose them since they remain identifiable when $\beta_{i0} \neq 0$ in (13) (expected in more general settings).

The catch count C_{itk} of each species in the absence of hook competition is the same as the number of individual fish of that species that are attracted to the fishing gear during the fishing event. We have not yet put in any constraints regarding the number of hooks available to the fish.

S.1.2 Step 2 – sample the bite times at which individual fish would take the bait given unlimited hooks

Next, we sample the bite times at which each fish would gobble the bait if there were unlimited hooks, denoted \mathbf{g}_{itk} , for which each element of the vector \mathbf{g}_{itk} corresponds to each of the C_{itk} individuals of species i attracted to the gear during the 5-hour soak time of the fishing event at station k during year t . So there is a vector \mathbf{g}_{itk} for each species and fishing event, and each element of the vector corresponds to the bite time of each of the individual fish that are attracted

to the gear. We can vary the aggressiveness of each species by varying the assumptions regarding their bite times.

We consider three distributions to sample from. First, is the truncated exponential distribution with rate 1 (truncated to the 5 hours). The majority of the individuals sampled from this distribution will bite the hooks early on in the soak time. Thus, the exponential distribution reflects the bite times of an ‘aggressive’ species.

Next, is a mixture distribution of a unit rate truncated exponential distribution and a uniform distribution with limits 0 and 5. The mixture weights are fixed at 0.5. This distribution reflects a large species group comprised of fish species with different levels of aggressiveness, olfactory sensitivities, and other biological characteristics. Under this distribution, some of the species are expected to bite early on in the soak time, while others are expected to bite the hooks more consistently throughout the entire soak time.

Finally, a uniform distribution with limits of 0 and 5 is considered. Individuals sampled from this distribution are expected to bite the hooks consistently throughout the entire soak time and are thus relatively ‘slow’ to reach the baited hooks on average (low ‘aggressiveness’).

S.1.3 Step 3 – explicitly simulate hook competition to see which fish get caught

Clearly, step 1 may sample more attracted individuals than the $h_{tk} = 800$ hooks that are available during every fishing event, yet step 2 has assumed unlimited hooks. So we now sequentially order each unique attracted fish in ascending order of the sampled bite times (so the more aggressive species are more likely to get to the hooks early on, whereas slower species are expected to bite later). Then, for each ordered fish, a Bernoulli random variable with known ‘capture’ probability is sampled from to determine if the fish successfully takes a bait and is captured by an available hook and is subsequently recorded in the data. If no more hooks are available then the gear is saturated and no more fish can be caught. We assume that if a fish takes a bait then it is captured

(fish of all species that steal bait without being caught could be explicitly modelled as another species group using counts of empty hooks).

We consider two situations. The first assumes that if any hooks are available then the fish will be captured (i.e. the Bernoulli capture probabilities are all 1). This setting implies the species is adept at locating remaining baited hooks, even when few remain actively fishing, e.g. observed Sablefish behaviour in hook timer experiments (Rodgveller et al., 2011).

The second situation assumes that the species' capture probabilities can decline as baits are removed. Specifically, a threshold p_i^* exists for all species. For simplicity in the experiments we set $p_i^* = p^*$. The capture probability of a fish is < 1 if the proportion of baits removed from the longline gear at the fish's specified bite time is $> p^*$. The probability equals 1 when the proportion is $\leq p^*$. Explicitly, the species-specific probabilities decrease linearly from 1 as baits are removed beyond p^* , reaching a species-specific lower bound, L_i , when one bait remains. For example, if $p^* = 0.85$ and $L_i = 0.2$, then all attracted individuals of species i with bite times before 85% of the baits have been removed are caught, but after that the probability of capture declines linearly from 1 to 0.2 as further baits are removed. Changing L_i changes the severity of the hook competition effects experienced by the species, and setting $L_i = 1$ means that hook competition only occurs when all the baits are removed (regardless of p^*). Thus, $L_i = 1$ reverts to the first situation. All attracted fish who fail their Bernoulli trial are not considered for future matches.

So, in Step 1 we generate numbers of each species that are attracted to the fishing gear from (S.1), then in Step 2 generate the bite times of each fish of each species depending on the species' aggressiveness. The bite times for all attracted fish are ordered sequentially and a Bernoulli trial is performed to see if the fish is captured by the gear in Step 3, with a probability that depends on whether the proportion of baits removed has exceeded the threshold p^* . This results in simulated catch counts, c_{itk} , for each species and each fishing event.

S.1.4 Step 4 – fit the simulated data using three alternative methods

Having generated the data, we use three methods to fit the catch count data of the target species.

The first method does not account for hook competition, and is the CPUE Poisson-lognormal model from equations (S.1)-(S.3) that was used to generate the catch counts that would occur in the absence of hook competition. The response variables are the observed catch counts of the target species. Thus, in some sense, the model has been correctly specified, albeit with hook competition not considered and with only the lower-right diagonal element of Σ estimated.

The second method uses a simple scaling to account for hook competition, and is here called the ICR (instantaneous catch rate) method. The method attempts to scale up the catch counts to account for the fish that would have been caught if all hooks had been actively fishing during the full fishing event. For example, if 5 fish of the target species were caught in an event, but only 200 hooks were returned with bait, the second method attempts to infer the number of fish of the target species that would have been caught if all 800 hooks had been actively fishing during the whole fishing event.

To do this, the ICR method multiplies the observed catch count by a competition adjustment factor, and then uses the same Poisson-lognormal model as for the CPUE method just discussed. The adjustment factor is derived and explained in Clark (2008), Webster and Stewart (2013), and Anderson et al. (2019), and it assumes that every fishing event with at least one catch is impacted by hook competition. It is applied to each fishing event, and is given by

$$F_{tk} = \begin{cases} \frac{-\log(1-p_{tk})}{p_{tk}}, & p_{tk} < 1 \\ \frac{-\log(1/800)}{799/800}, & p_{tk} = 1 \end{cases} \quad (\text{S.4})$$

where $p_{tk} = c_{.tk}/h_{tk}$ is the proportion of baits removed during fishing event k of year t . For each fishing event, the observed catch count of the target species is multiplied by the adjustment factor F_{tk} and then rounded to the nearest integer, to give a competition-adjusted integer count. The first line of (S.4) is infinite when $p_{tk} = 1$ (i.e. when all 800 baits are removed), and so we set

its value to the value when 799 of the 800 baits are removed (Anderson et al., 2019), giving the second line of (S.4) which is an adjustment factor of 6.69 for 800 hooks. The adjustment factor derivation involves assumptions on the instantaneous catch rates (ICR) of each species, and so the estimation method is an ICR-adjusted Poisson-lognormal model, hereafter called the ICR method.

The final method is our new censored method described in the main text. In many settings, the assumed censorship proportion for the target species, \hat{p}^* , is deliberately mis-specified to not equal the true value p^* that was used to generate the data, to test the influence of such mis-specification (since \hat{p}^* is not estimable within the model). For some settings with $\hat{p}^* < 1$, we use a conservative upper bound u_{3tk} for each fishing event to help improve numerical convergence (see later for derivation). No upper bound is used when $\hat{p}^* = 1$. To further improve the convergence of the model, for each year t we investigate the effect of considering all observed catch counts c_{3tk} that exceed or equal the Q^{th} quantile of the catch counts that year as ‘high-quality’, regardless of the value of p_{tk} . Thus, the censorship interval is replaced with the original count c_{3tk} for these counts. We experiment with different values of Q .

After all models have been fit, the posterior median, posterior 2.5th percentile, and the posterior 97.5th percentile of each $\exp(\alpha_{3t})$ for $t > 1$ are computed. Absolute bias is calculated as the posterior median minus the true value, and the squared error is the absolute bias squared. The coverage indicator variable is assigned value 1 uniquely for each $\exp(\alpha_{3t})$ for $t > 1$, if the 95% credible interval contains the true value, and 0 if not.

Six experiments are performed, with each experiment (except 6) testing a different combination of settings 100 times, to give 100 estimates of the statistics mentioned in the previous paragraph. To summarise the performance of each method, the 100 bias, squared error, and coverage values are aggregated, with confidence intervals produced for each aggregate measure. The confidence intervals allow for levels of statistical significance to be attributed to the comparisons being made between the competing methods. For the bias and squared error values, the median

and a robust 95% confidence intervals are computed to reduce the impact of outliers on summary measures of performance. Robust confidence intervals are computed using the median absolute deviation (MAD) measure of dispersion scaled by 1.48 to ensure consistency with standard confidence intervals (Maronna et al., 2019). For the coverage indicators, the proportion of 1's are computed along with an approximate 95% asymptotic confidence interval for the proportion computed using the normal approximation to the binomial proportion (i.e. $\pm 2\text{SE}$). For bias and median-squared error (MSE), a value close to zero is desired. For coverage, the desired value is 95%.

The different settings for each experiment are now given.

S.1.5 Experiment 1

For experiment 1 we use the following settings for a single-species scenario, with all combinations of the two choices of each of $\mu_{3,t}$ and p^* giving the four panels of results in Figure S1:

- $\mu_{it} = \begin{cases} (0, 0, 0, 0, 0, 0), & i = 1, \text{ main competitor absent} \\ (0, 0, 0, 0, 0, 0), & i = 2, \text{ no other species or sources of bait loss} \end{cases}$
- $\mu_{3t} = \begin{cases} (800, 800, 800, 800, 800, 800), & \text{target species remains constant (test type I property)} \\ (300, 400, 500, 600, 700, 800), & \text{target species increases (test type II property)} \end{cases}$
- $L_3 = 0.2$, severity of hook competition (i.e. gear saturation effects) when $p^* \neq 1$
- $\mathbf{g}_{3tk} \sim \text{U}(0, 5)$, uniform distribution of bite times
- $p^* = \begin{cases} 0.85, & \text{true breakdown point begins in fishing events once 85\% baits removed} \\ 1, & \text{the only low-quality events are when all baits are removed} \end{cases}$
- $\hat{p}^* = 1$, assumed breakdown point for fitting is not always the true value

- $\Sigma = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.8^2 \end{pmatrix}$, covariance matrix (only a single schooling species is present)
- $K_t = 30$, number of stations sampled at per year
- $Q = \begin{cases} NA, & \text{the largest catch counts can be censored} \end{cases}$

Thus, in experiment 1, there is only one species present, the target species. Very high levels of overdispersion and right-skewness are present in the target species' count distribution (indicative of frequent schooling behaviours) due to the value of 0.8^2 in Σ .

When $\mu_{3t} = 800$ for all years, the target species' relative abundance does not change. This setting explicitly evaluates type I properties. This results in around 50% of the fishing events having all 800 baits removed. The purpose is to confirm that the censored method can correctly identify the lack of trend, regardless of whether attracted individuals are always captured if any baited hooks remain available ($p^* = 1$), or the probability of capture declines as less hooks are available ($p^* = 0.85$), and even when the assumed value of $\hat{p}^* = 1$ does not equal the true value of p^* . The target species' count distribution is designed to reflect a species that often appears in group sizes that are large enough to saturate the fishing gear.

When $\mu_{3t} = (300, 400, 500, 600, 700, 800)$ we are explicitly evaluating type II properties. The high levels of overdispersion results in all 800 baits being removed during many of the fishing events, including in the early years. The purposes of this experiment are: i) to verify the result of the theoretical example given in the main text, namely, that the CPUE index will underestimate the increase in the relative abundance due to the frequent gear saturation; ii) to verify that the censored method can still recover the true trend of a schooling species, despite the saturated fishing gear. Note that no upper bound on the species catch count is assumed (i.e. $u_{itk} = \infty$ is assumed).

S.1.6 Experiment 2

In experiment 2 many species are caught, and characterized as the main competitor ($i = 1$), other ($i = 2$) and the target species ($i = 3$). The target species is non-schooling due its distribution not being overdispersed. The lack of overdispersion implies that counts of the target species have the highest possible signal-to-noise ratio which should result in the methods inferring the relative abundance trends with the greatest precision. Again, the relative abundance of the target species is fixed through time to assess the type I property and increased to assess the type II property.

However, in this experiment the target species faces fierce competition from the main competitor for the hooks. The main competitor is a schooling species with a strongly overdispersed distribution. Thus, the main competitor frequently appears in very large group sizes and can rapidly deplete the baits. Furthermore, the main competitor increases in abundance through time (either slowly or quickly) and is able to, on average, reach the baits ahead of the target species on average (when bite times have a mixed distribution). Lastly, the other species group consists of multiple species with differing behaviours and levels of overdispersion, but the mean abundance of the other species group is stationary. Specific settings are:

- $\mu_{it} = \begin{cases} (120\kappa, 140\kappa, 160\kappa, 180\kappa, 200\kappa, 280\kappa), & i = 1, \text{main competitor increases} \\ (400, 400, 400, 400, 400, 400), & i = 2, \text{'other' species constant} \end{cases}$
- $\mu_{3t} = \begin{cases} (100, 100, 100, 100, 100, 100), & \text{target species constant (test type I property)} \\ (20, 40, 60, 80, 100, 120), & \text{target species increases (test type II property)} \end{cases}$
- $\kappa = \begin{cases} 1, & \text{lower levels of increase of main competitor} \\ 2, & \text{higher levels of increase of main competitor} \end{cases}$
- $(L_1, L_2, L_3) = (1, 0.8, 0.2)$, gear saturation effects when $p^* \neq 1$
- $(\mathbf{g}_{1tk}, \mathbf{g}_{2tk}, \mathbf{g}_{3tk}) \sim$

- $$\begin{cases} (U(0, 5), U(0, 5), U(0, 5)), & \text{uniform distribution of bite times} \\ (\text{Exp}(1), 0.5\text{Exp}(1) + 0.5U(0, 5), U(0, 5)), & \text{mixed distributions of bite times} \end{cases}$$
- $$p^* = \begin{cases} 0.85, & \text{true breakdown point begins in fishing events once 85\% baits removed} \\ 1, & \text{the only low-quality events are when all baits are removed} \end{cases}$$
- $$\hat{p}^* = \begin{cases} 0.85, & \text{assumed breakdown point for fitting can equal the true value} \\ 0.95, & \text{assumed breakdown point for fitting does not equal the true value} \end{cases}$$
- $$\Sigma = \begin{pmatrix} 0.8^2 & 0 & 0 \\ 0 & 0.2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ target species is non-schooling with two schooling competitors}$$
- $K_t = 30$, number of stations sampled at per year
- $Q = NA$, no restriction on which catch counts can be censored.

The four combinations of the two choices of each of $\mu_{3,t}$ and κ give the four simulated abundances in the leftmost column of Figure 1. The four combinations of the two choices of each of the breakdown points (p^*) and the bite rate distributions (the \mathbf{g} 's) give the four columns of results of Figure 1. The CPUE and ICR methods, plus the Censored method with the two values of \hat{p}^* (which is only correctly specified when $\hat{p}^* = 0.85 = p^*$) are tested, giving the four methods in the legend in Figure 1. Hence we have four different methods tested for each of the sixteen different simulation settings, giving 64 sets of results.

To summarize, interspecies hook competition (sometimes severe) is present between three species groups with independent distributions (i.e. no correlation is present due to the zero off-diagonal entries of Σ). The target species in this experiment never appears in large schools, and (especially in the ‘mixed’ bite time distribution setting, also called ‘Differing Aggressiveness’ in the Figures) is sensitive to the competition for hooks from the other two species groups. The level

of competition increases with time due to the increasing abundance of the main competitor. Note that a conservative upper bound, u_{itk} , is placed on the species catch counts but is not believed to impact the results (derived and discussed later).

S.1.7 Experiment 3

Experiment 3 uses the same settings as experiment 2 except for:

- $\mu_{3t} = \begin{cases} (100, 100, 100, 100, 100, 100)e^{-0.2^2/2}, & \text{target species constant (test type I property)} \\ (20, 40, 60, 80, 100, 120)e^{-0.2^2/2}, & \text{target species increases (test type II property)} \end{cases}$
- $(\mathbf{g}_{1tk}, \mathbf{g}_{2tk}, \mathbf{g}_{3tk}) \sim (\text{Exp}(1), 0.5\text{Exp}(1) + 0.5\text{U}(0, 5), \text{U}(0, 5))$, mixed distribution of bite rates
- $p^* = 0.85$, true breakdown point in simulations
- $\hat{p}^* = 0.95$, the breakdown point is always mis-specified
- $\Sigma = \begin{cases} \begin{pmatrix} 0.8^2 & 0 & 0 \\ 0 & 0.2^2 & -0.024 \\ 0 & -0.024 & 0.2^2 \end{pmatrix}, & \text{negative correlation } (\rho = -0.6) \\ \begin{pmatrix} 0.8^2 & 0 & 0 \\ 0 & 0.2^2 & 0.004 \\ 0 & 0.004 & 0.2^2 \end{pmatrix}, & \text{low correlation } (\rho = 0.2) \\ \begin{pmatrix} 0.8^2 & 0 & 0 \\ 0 & 0.2^2 & -0.012 \\ 0 & -0.012 & 0.2^2 \end{pmatrix}, & \text{medium correlation } (\rho = 0.4) \\ \begin{pmatrix} 0.8^2 & 0 & 0 \\ 0 & 0.2^2 & 0.024 \\ 0 & 0.024 & 0.2^2 \end{pmatrix}, & \text{high correlation } (\rho = 0.6) \end{cases},$
- $K_t = 100$, an increased number of stations sampled at per year

$$\bullet Q = \begin{cases} \text{NA,} & \text{no restriction on which catch counts can be censored} \\ 1, & \text{greatest annual target species' catch counts cannot be censored} \\ 0.95, & \text{greatest 5\% of annual target species' catch counts cannot be censored} \\ 0.85, & \text{greatest 15\% of annual target species' catch counts cannot be censored} \end{cases}$$

Thus, experiment 3 allows the count distribution of the target species to be overdispersed and correlated with the count distribution of the ‘other’ species group. Note that we scale the mean catch count of the target species $\mu_{3,t}$ by $e^{-0.2^2/2}$ to account for the bias due to overdispersion and to ensure the expected catch counts match experiment 2. The target species is highly sensitive to the competition for hooks from the other species, typically reaching baits later in the soak times of fishing events. More stations are sampled because: i) the extra overdispersion and hence the schooling behaviours now exhibited by the target species reduces the signal-to-noise ratio in the target species distribution; ii) the extra correlations between the target and the other species group’s distributions induces an additional stochastic dependency between the probability that a fishing event will be censored and the expected catch count of the target species. Only the results of $Q = \text{NA}$ and $Q = 0.85$ are shown since the performance for the other two values across all metrics lies in-between. The combination of settings gives the results in Figure 3. Note that a conservative upper bound, u_{itk} , is placed on the species catch counts but is not believed to impact the results (derived and discussed later).

S.1.8 Experiment 4

For experiment 4 we repeat experiment 2, but with:

- $p^* = 0.5$, much lower true breakdown point in simulations
- $\hat{p}^* = \begin{cases} 0.85, & \text{assumed censorship proportion for fitting is mis-specified} \\ 0.95, & \text{assumed censorship proportion for fitting is mis-specified} \end{cases}$

Thus we still specify $\hat{p}^* = 0.85$ and 0.95 as in experiment 2, but neither are close to the true value of $p^* = 0.5$.

S.1.9 Experiment 5

For experiment 5 we repeat experiment 3 but with:

- $(\mathbf{g}_{1tk}, \mathbf{g}_{2tk}, \mathbf{g}_{3tk}) \sim (\text{U}(0, 5), 0.5\text{Exp}(1) + 0.5\text{U}(0, 5), \text{Exp}(1))$,
the target species is now fastest
- $\hat{p}^* = \begin{cases} 0.95, & \text{assumed censorship proportion for fitting is mis-specified} \\ 1, & \text{assumed censorship proportion for fitting is mis-specified} \end{cases}$

The target species typically reaches the baits ahead of the competing species. The purposes of this experiment are to: i) empirically confirm the results of the theoretical example presented in the main text, namely, that the ICR method will not satisfy the type I setting due to the insensitivity of the target species to inter-species hook competition; ii) investigate the robustness of the censored method when the target species is not strongly affected by hook competition, nor appears in large enough numbers to saturate the fishing gear. Put differently, experiment 5 considers a setting where a target species is largely ‘immune’ to hook competition in the data collected. Note that a conservative upper bound, u_{itk} , is placed on the species catch counts when $\hat{p}^* < 1$ but is not believed to impact the results (discussed and derived later).

S.1.10 Experiment 6

Exploratory analysis of the simulated observed catch counts c_{itk} (after the bite times have been matched to the next available hook) versus p_{tk} for the target species of experiments 2-4 appears to show that the mean of c_{itk} declines rapidly after p_{tk} exceeds \hat{p}_i^* . This trend appears to be consistent across different choices of \hat{p}_i^* . Conversely, a rapidly increasing and negligible trend is seen between c_{itk} and p_{tk} for the main competitor of experiments 2-4 and the immune species of experiment 5, respectively.

Experiment 6 formally investigates if these trends between c_{itk} and p_{tk} persist across very large numbers of replications and changing settings. If so, this would imply that exploratory analysis of c_{itk} versus p_{tk} in real longline data can help the scientist: i) select a suitable value of \hat{p}_i^* as the value where the trend begins; ii) understand the statistical properties of the estimates, most notably helping the scientist to understand the levels of trust they can assign the reported levels of uncertainty based on the findings from experiments 1-5.

We simulate data (scenarios described below) with $p^* = 0.85$ and fit the following model to each simulated dataset (values of c_{3tk} and p_{tk}):

$$[c_{3tk}|e_{3tk}, p_{tk}] \sim \text{Poisson}(\mu_{3t}e_{3tk}) \quad (\text{S.5})$$

$$[\log e_{3tk}] \sim \text{Normal}(0, \sigma^2) \quad (\text{S.6})$$

$$\log(\mu_{3t}) = \alpha_{3t} + \beta I(0.8 < p_{tk} \leq 0.95) + \gamma I(0.95 < p_{tk}), \quad 1 \leq t \leq 6, \quad (\text{S.7})$$

where σ is an estimated standard deviation, $I(\cdot)$ is the indicator function (1 if the condition is satisfied, 0 if not), $\alpha_{3t} + \beta$ represents the log of the mean catch counts observed at ‘medium’ values of p_{tk} , and $\alpha_{3t} + \gamma$ the log of the mean catch counts observed at ‘high’ values. We use these discrete values (rather than continuous) here for simplicity, but allow continuous in the application to real data described in the case study.

We define these intervals ($0.8 < p_{tk} \leq 0.95$ and $0.95 < p_{tk}$) to see if we can detect a change in the mean observed catch count when $p_{tk} \approx 0.85$ (the known breakpoint is $p^* = 0.85$) to when $p_{tk} \approx 1$. These values were also chosen based on the exploratory plots. Then, the target for inference is $\exp(\gamma - \beta)$, the relative value of the mean catch counts observed at high and medium levels of p_{tk} . By plotting estimates of $\exp(\gamma - \beta)$ across a wide range of simulation settings, we aim to develop a method for choosing \hat{p}^* for real data. We repeat all combinations of the following settings 20 times:

- $\bullet \mu_{it} = \begin{cases} (120\kappa_1, 140\kappa_1, 160\kappa_1, 180\kappa_1, 200\kappa_1, 280\kappa_1), & i = 1 \\ (400, 400, 400, 400, 400, 400), & i = 2 \\ (100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3)e^{-0.2^2/2}, & i = 3 \end{cases}$

 $\bullet \mu_{it} = \begin{cases} (200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1), & i = 1 \\ (400, 400, 400, 400, 400, 400), & i = 2 \\ (100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3, 100\kappa_3)e^{-0.2^2/2}, & i = 3 \end{cases},$

 $\bullet \mu_{it} = \begin{cases} (200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1, 200\kappa_1), & i = 1 \\ (400, 400, 400, 400, 400, 400), & i = 2 \\ (20\kappa_3, 40\kappa_3, 60\kappa_3, 80\kappa_3, 100\kappa_3, 120\kappa_3)e^{-0.2^2/2}, & i = 3 \end{cases}$
- $\bullet (\kappa_1, \kappa_3) = \begin{cases} (1, 2), & \text{relatively lower levels of main competitor} \\ (2, 1), & \text{relatively higher levels of main competitor} \end{cases}$
- $\bullet (L_1, L_2, L_3) = (1, 0.8, 0.2), \quad \text{saturation effects when } p^* \neq 1$
- $\bullet p^* = \begin{cases} 0.85, & \text{true breakdown point occurs in simulation when 85\% baits removed} \end{cases}$
- $\bullet (\mathbf{g}_{1tk}, \mathbf{g}_{2tk}, \mathbf{g}_{3tk}) \sim \begin{cases} (U(0, 5), U(0, 5), U(0, 5)), & \text{all species equally-matched to reach baits} \\ (\text{Exp}(1), 0.5\text{Exp}(1) + 0.5U(0, 5), U(0, 5)), & \text{target species slow to reach baits} \\ (U(0, 5), 0.5\text{Exp}(1) + 0.5U(0, 5), \text{Exp}(1)), & \text{target species quickly reaches baits} \end{cases}$

- $\Sigma = \begin{cases} \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & -0.6\sigma_2\sigma_3 \\ 0 & -0.6\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}, & \text{negative correlation } (\rho = -0.6) \\ \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}, & \text{no correlation } (\rho = 0) \\ \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0.6\sigma_2\sigma_3 \\ 0 & 0.6\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}, & \text{high correlation } (\rho = 0.6) \end{cases}$
- $(\sigma_1, \sigma_2, \sigma_3) = \begin{cases} (0.2, 0.2, 0.2), & \text{no species strongly overdispersed} \\ (0.8, 0.2, 0.2), & \text{main competitor strongly overdispersed} \\ (0.2, 0.2, 0.8), & \text{target species strongly overdispersed} \end{cases}$
- $K_t = 100$, high number of stations sampled at per year

Finally, we produce boxplots of estimates of $\exp(\gamma - \beta)$ across all the settings, but grouped by the relative speeds of each species group to the baits (i.e. the 3 \mathbf{g}_{itk} settings) and the relative degree of schooling exhibited by the three species groups (i.e. the 3 σ_i settings). See Figure 5 in the main text for results.

S.1.11 Extrapolating results from experiments 1-5 to a target species with a decreasing abundance

Results from simulation experiments 1-5 shed light on the relative performances of the competing methods at inferring the relative abundance of a target species whose abundance decreases over time.

To see why, consider the settings of simulation experiments 1-5, but reverse the abundance trends of all species groups in time. Thus, the target species' abundance is either stationary or decreasing and the abundance of the competitor species, when present, decreases.

In these time-reversed settings, the mean catch counts of all three species groups in year 1 equal the mean catch counts in year 6 under the original settings. Suppose we parameterize the mean catch counts as before, but using notation $\gamma_{i,1}$ and γ_{it} instead of $\alpha_{i,1}$ and α_{it} for the mean catch counts. Next, let $\tilde{\mu}_{it}$ and μ_{it} denote the mean catch count in the time-reversed setting and corresponding original setting respectively. Thus, for a given simulation experimental setting, $\mu_{it} = \tilde{\mu}_{i,(7-t)}$ and hence $\mu_{i,1} = \tilde{\mu}_{i,6}$ and $\mu_{i,6} = \tilde{\mu}_{i,1}$ for all species i .

In these time-reversed settings in the absence of hook competition, the catch counts, \tilde{C}_{itk} , from the three species groups $i \in \{1 = \text{competitor}, 2 = \text{other}, 3 = \text{target}\}$, in year $t = 1, 2, \dots, 6$, and station (replicate) $k = 1, 2, \dots, K_t$ are therefore viewed as samples from the following model based on (5) and (13):

$$[\tilde{C}_{itk} | e_{itk}] \sim \text{Poisson}(\tilde{\mu}_{it} e_{itk}) \quad (\text{S.8})$$

$$\log \tilde{\mu}_{it} = \begin{cases} \gamma_{i,1}, & t = 1, \\ \gamma_{i,1} + \gamma_{it}, & t = 2, 3, 4, 5, 6, \end{cases} \quad (\text{S.9})$$

$$[\log e_{1tk}, \log e_{2tk}, \log e_{3tk}] \sim \text{MVNormal}(\mathbf{0}, \mathbf{\Sigma}). \quad (\text{S.10})$$

But this is equivalent to simply parameterizing the original experiments with year 6 as the baseline. To see why, note that we can equivalently write the model in the original time direction as:

$$\log \mu_{it} = \begin{cases} \gamma_{i,1} + \gamma_{i,(7-t)}, & t = 1, 2, 3, 4, 5, \\ \gamma_{i,1}, & t = 6. \end{cases} \quad (\text{S.11})$$

Therefore, we can write down the following parametric equality relating the relative abun-

dances in the time-reversed and original settings:

$$\exp(\gamma_{i,6}) = \frac{\exp(\gamma_{i,1} + \gamma_{i,6})}{\exp(\gamma_{i,1})} \quad (\text{S.12})$$

$$= \frac{\exp(\tilde{\mu}_{i,6})}{\exp(\tilde{\mu}_{i,1})} \quad (\text{S.13})$$

$$= \frac{\exp(\mu_{i,1})}{\exp(\mu_{i,6})} \quad (\text{S.14})$$

$$= \frac{\exp(\alpha_{i,1})}{\exp(\alpha_{i,1} + \alpha_{i,6})} \quad (\text{S.15})$$

$$= \frac{1}{\exp(\alpha_{i,6})}. \quad (\text{S.16})$$

This equality shows that we can learn the statistical properties of the competing methods at estimating year 6 relative abundance in the hypothetical time-reversed settings of experiments 1-5 by simply looking at the year 6 results on bias, MSE, and coverage from the original experiments 1-5. In particular, this is because the quantile-based equal-tailed (central) posterior credible intervals and posterior medians used in this simulation study are invariant to monotone transformations.

Thus, for any particular simulation experiment from 1-5, methods that overestimate the relative abundance of the target species in year 6, $\exp(\alpha_{i,6})$, will underestimate the relative abundance of the target species in year 6 in the corresponding time-reversed setting, $\exp(\gamma_{i,6})$, and visa-versa. Similarly, the coverage levels of each method's credible intervals for the relative abundance in year 6 in the time-reversed setting should match that seen in the corresponding original setting in year 6.

Therefore, additional simulation studies are not required for evaluating the methods' performances in settings where the target species' abundance is decreasing in time (in conjunction with a main competitor species). Bias, MSE, and coverage results seen in year 6, the year which typically differentiates the competing methods best, all carry over from the original experiments. However, the setting where the abundance trend of the target species moves in the opposite direction of the main competitor's abundance trend cannot be inferred by experiments 1-5 alone.

This setting is tested next.

S.1.12 (Bonus) experiment 7

In bonus experiment 7 the target species' abundance always decreases in time, whereas the main competitor's abundance is allowed to increase and decrease. Contained within experiment 7's settings are the type II settings of experiment 2 with the arrow of time reversed. The upper bounds on modelled target species' catch counts are removed and an improved optimization routine which was discovered after simulation experiments 1-6 were conducted (discussed in the next section) is implemented. The data are generated as follows:

$$\begin{aligned}
 \bullet \mu_{it} &= \begin{cases} (400, 400, 400, 400, 400, 400), & i = 2, \text{'other' species constant} \\ (120, 100, 80, 60, 40, 20), & i = 3, \text{target species decreases} \end{cases} \\
 \bullet \mu_{1,t} &= \begin{cases} (280\kappa, 200\kappa, 180\kappa, 160\kappa, 140\kappa, 120\kappa), & \text{main competitor decreases} \\ (120\kappa, 140\kappa, 160\kappa, 180\kappa, 200\kappa, 280\kappa), & \text{main competitor increases} \end{cases}
 \end{aligned}$$

The goals of this experiment are to empirically verify that: i) the censored method remains the superior choice of method in a decreasing abundance setting; ii) the improved optimization routine (discussed in the next section) does not negatively impact the censored method; iii) removing the conservative upper bound (discussed in the next section) from the catch counts does not negatively impact the censored method;

Goals ii) and iii) can be investigated due to the relationships between year 6 bias, MSE and coverage results from experiments 2 and experiment 7 discussed in the previous section. If we do not observe opposing signs of year 6 bias across the equivalent experiment 7 and experiment 2 settings, or, if we observe drastically different year 6 coverage levels across the two experiments, then this would suggest that the upper bound or optimization routine significantly impacts performance.

S.1.13 Fitting details for all experiments and additional results

For all experiments, we use the R-INLA package for model fitting (Rue et al., 2009; Lindgren et al., 2011; Lindgren and Rue, 2015). We place a flat prior on $\alpha_{3,1}$ and a weakly-informative normal prior with mean 0 and variance 1000 on the remaining $\alpha_{3,t}$ for $t > 1$. Lastly, we use R-INLA's default Gamma prior with shape 1 and scale 5×10^{-5} for the precision of the lognormal distribution describing the $e_{3,t,k}$ terms. To derive the posterior median and posterior quantiles used for the model comparison, we generate 5000 approximately independent Monte Carlo samples of $\alpha_{3,t}$ from the (integrated nested Laplace approximated) posterior distribution, exponentiate each sampled value, and then compute the necessary functions.

For experiment 1, Figures S1, S2, and S3 present the bias, MSE, and coverage respectively. Only the censored method satisfies both the type I and type II properties. In particular, the CPUE and ICR methods perform poorly across all three metrics when the abundance of the single species increases over time. In the type I settings, the coverage levels of the censored method fall to 75%. The reason for this is currently unknown, however we do not see this as a cause for concern given the very low MSE values.

For experiment 2, Figures 1, S4, and 2 present the bias, MSE, and coverage respectively. Results demonstrate the improved performance of the censored method, with: i) up to 100-fold reductions in MSE observed in both type I and type II settings; ii) very low bias across all settings; iii) coverage levels close to 95% in all settings. Note that the other two methods perform poorly, coverage often drops to 0% in both type I and type II settings and the bias in estimates can be extreme.

For experiment 3, conclusions are similar to experiment 2 and the censored method performs best across all metrics. Figures 3, S5, and 4 present the bias, MSE, and coverage respectively. For the censored method, coverage levels close to 95% are attained in all type I settings and they always exceed 50% in the type II settings. Coverage levels from the other two methods are close to 0% in many settings. A moderate deterioration in the performance of the censored

method occurs in settings where the target species' abundance distribution is positively correlated with the abundance distributions of competing species. Figure S5 shows an increase in MSE as the correlation increases. The quantile-modification has only a small impact on the statistical properties of the censored method but ensures convergence (Figure S6 and discussed later in this section). We show the quantile-adjusted censored method results only in the MSE figure for brevity. Values of bias and coverage change little under the quantile adjustment.

For experiment 4, conclusions are similar to experiment 2, despite the badly mis-specified \hat{p}^* value. Figures S7, S8, and S9 present the bias, MSE, and coverage respectively. They show the bias and MSE from the grossly mis-specified censored method to be at least as good as the competing methods across all settings tested, and far better in most settings. In summary, results show the censored method retains improved statistical properties when \hat{p}^* is poorly chosen.

For experiment 5, Figures S10, S11, and S12 present the bias, MSE, and coverage respectively. They confirm that the CPUE method yields values that closely follow the true abundance trends of a target species when it is unlikely to be subject to hook competition. When absolute fish numbers are low, the CPUE method method performs best with respect to MSE and coverage. However, this performance improvement is lost as fish numbers increase (increasing the proportion of baits removed) and the censored method outperforms across all metrics once again. In fact, coverage levels of both the ICR and CPUE methods fall close to 0% in both type I and type II settings when the total fish density is high. Unfortunately, however, the coverage of censored method can fall as low as 40% and MSE values can remain high. Note that the ICR method performs poorly in all settings across all metrics, in agreement with the earlier theoretical example – the ICR method is not well suited for species that can typically reach baited hooks early on in the soak times since it assumes that hook competition impacts every fishing event.

For experiment 6, cross-correlating the results shown in Figure 5 with those from experiments 1-5 suggests a simple empirical test for assessing if the censored method is suitable for modelling a given species. The censored method performed very strongly for the target species

of experiments 2-4, for which Figure 5 presents a decreasing abundance trend with p_{tk} . The censored method also performed well for experiment 1, for which Figure 5 presents a strong increasing abundance trend with p_{tk} for cases where the target species are responsible for frequently saturating the fishing gear. Finally, the censored method performed adequately for the ‘immune’ species of experiment 5, for which Figure 5 presents a weak abundance trend with p_{tk} for the target species in this experiment.

Thus, the censored method should perform very well for species exhibiting strong declines in their observed catch counts c_{itk} at ‘high’ values of p_{tk} . This negative trend should remain after controlling for the effects of exogenous factors (e.g. year). A statistically significant decline in mean catch counts exceeding 10% appears to be a conservative decision rule; whilst the censored method was found to perform very well in many settings that fail this decision rule, the rule appears to filter out problematic ‘immune’ species seen in experiment 5. By comparing the ‘best-performing’ \hat{p}_i^* values in experiments 2-4 with the exploratory analysis between c_{itk} and p_{tk} , we recommend that for species exhibiting a 10% decline, a value of \hat{p}_i^* be chosen close to the value of p_{tk} where the negative trend first begins.

Similarly, the censored method should perform well for species exhibiting large increases in their observed catch counts c_{itk} at ‘high’ values of p_{tk} . A statistically significant increase in mean catch counts exceeding 20% appears to be a conservative decision rule for confidently classifying a schooling species responsible for frequently saturating the longline fishing gear. Results from experiment 1 suggest setting $\hat{p}_i^* = 1$, however uncertainty levels may be somewhat underestimated.

For species exhibiting neither a strong increase or decrease in mean catch count at ‘high’ values of p_{tk} , results from experiment 5 suggest the censored method should be cautiously used with $\hat{p}_i^* = 1$. The user should be aware that the species could be largely immune to hook competition in the dataset. Hence, levels of uncertainty could be underestimated (due to the lower coverage levels seen in immune species). Furthermore, setting $\hat{p}_i^* < 1$ for an immune species can hurt the

performance of the censored method.

For experiments 2 and 4 there were a handful of simulation iterations where the censored method failed to converge. When this happened, we simply repeated the simulation iteration (i.e. generated new data and re-fit all models) and ignored the convergence failure. In experiment 3, the convergence issues were much more prevalent due to the lower signal-to-noise ratio and were a target of research. Thus, for experiment 3, we recorded in the output instances where a model was reported by R-INLA as failing to converge.

Across all experiments, between 30-60% of fishing events returned <5% of the baits, and between 20-50% of events returned 0% (see Figure S6). Interestingly, during experiment 3, convergence issues were more common when the all-species density was lowest. Over 40% of the iterations failed to converge in the most challenging settings if the censored method was used without adjustment. Implementing the quantile adjustment improved things. This adjustment considers all catch counts c_{itk} exceeding the Q^{th} annual quantile as high quality, regardless of the value of p_{tk} . Using $Q = 0.95$ led to almost 100% convergence across all simulation settings, with $Q = 0.85$ ensuring 100% convergence throughout. We hypothesize the improved convergence is due to an improved ability to identify the variance of the overdispersion terms, σ^2 , since the right-tail of the catch count distribution is (artificially) observed. The minimal impact that the quantile adjustment has on the statistical properties of the censored method in the simulation experiments suggests this could be a default approach for eliminating convergence issues in practice.

After conducting the simulation experiments 1-6, we discovered that convergence could be further improved by starting the numerical optimization with both initial values and Laplace approximations started at the final converged values found in the CPUE model. We have not repeated simulation experiments 1-5 with this change implemented due to the simulation experiments taking a very long time to fit, but note that the 40% estimate should be viewed as a highly conservative upper bound – 8 out of the 11 models in the case study fit without convergence issues and without requiring the use of any upper bound $u_{3,t,k}$ (including the one described later),

or requiring the quantile adjustment (i.e. $Q \leq 1$ for $\hat{p}_i^* < 1$).

Finally, results from experiment 7 confirm that the censored method outperforms the other methods in settings where the target species' abundance is decreasing, regardless of the direction in the competing species' abundance trend (see Figures S13, S14, and S15). The censored method consistently has the lowest bias and MSE and the coverage levels are almost always close to the targetted 95%. Furthermore, the coverage levels of credible intervals for year 6 relative abundance between equivalent experiment 2 and experiment 7 settings match and the biases in year 6 abundance are all of opposite sign, as dictated by theory. Combined, this suggests that implementing the new optimization routine and removing the upper bound in catch counts does not grossly impact inference. Note that we ended up with 130 replicates instead of the targetted 100 (we ran the experiment in parallel jobs of 10 iterations each and anticipated greater numbers of convergence failures than witnessed).

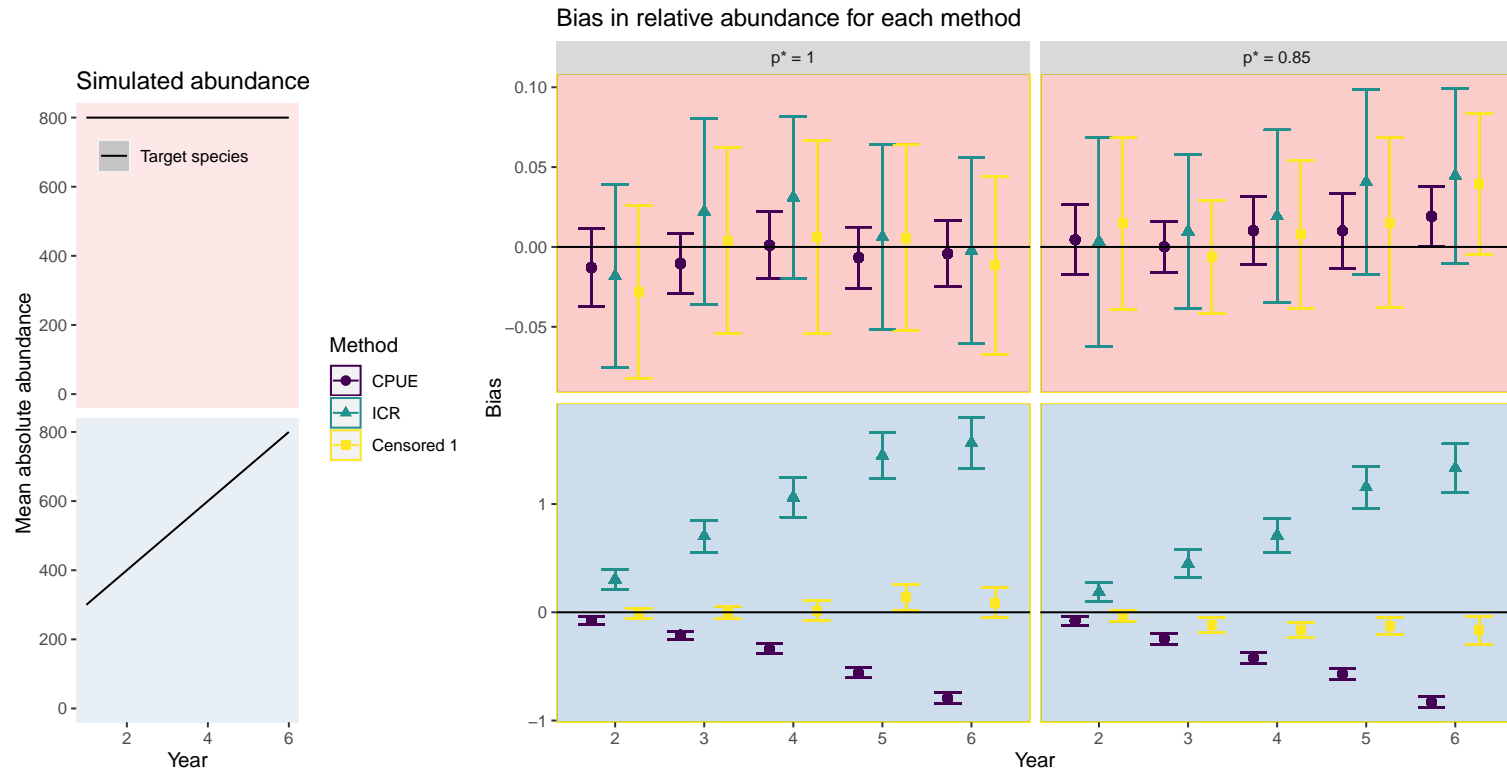


Figure S1: Absolute bias in the relative indices across three methods for experiment 1. The leftmost column shows the mean abundance in the target species across time, such that the first row is testing the type I properties and the second row the type II. Columns specify if hook competition occurs only when all baits are removed ($p^* = 1$) or begins when 85% are removed ($p^* = 0.85$). Plotted are the median and robust 95% confidence intervals calculated from the 100 repetitions. The ‘Censored 1’ results correspond to the censored method with $\hat{p}^* = 1$. The bias from the censored method remains low in all settings, unlike the other methods which become severely biased in the type II setting.

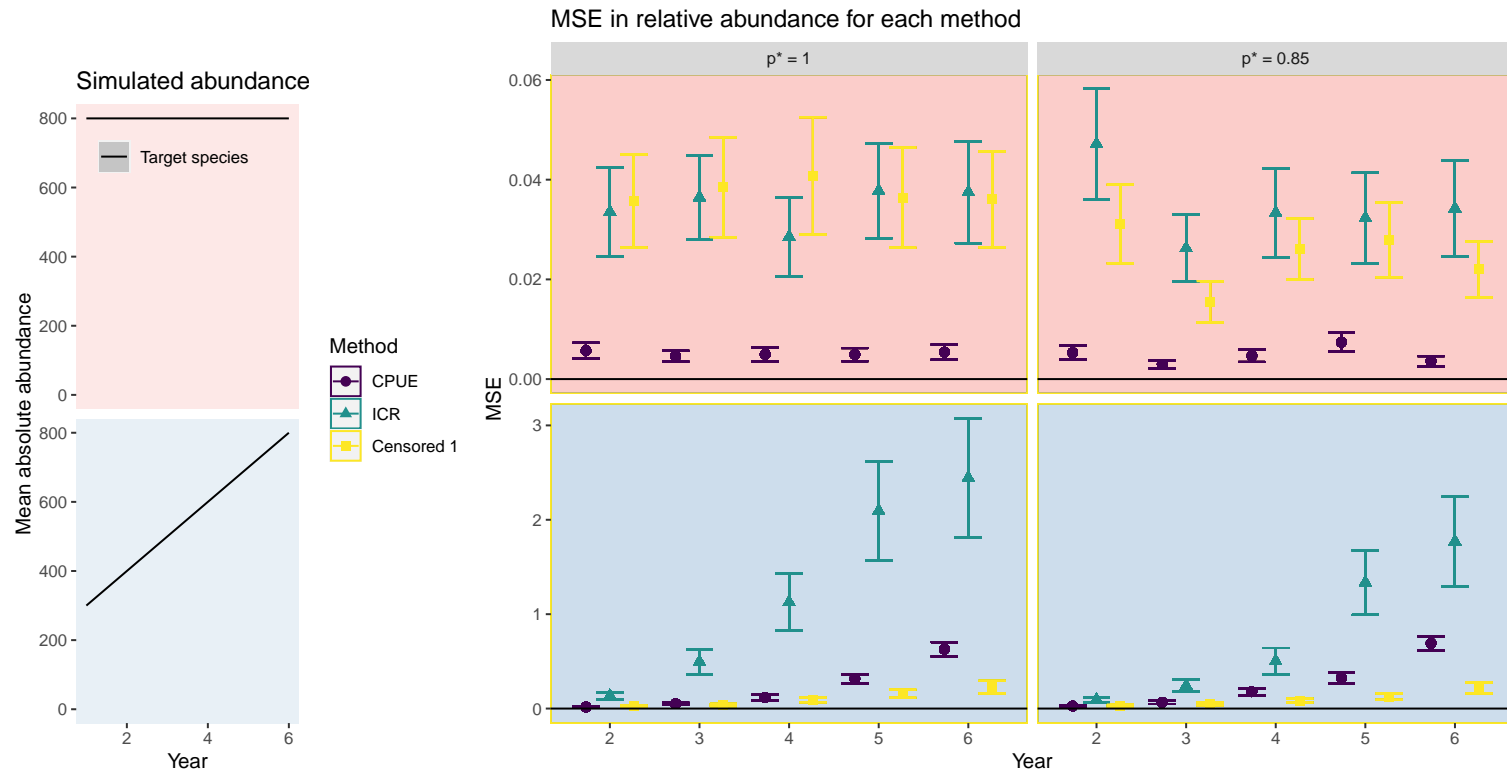


Figure S2: Median-squared error results for experiment 1. Rows and columns as for Figure S1. Plotted are the median and robust 95% confidence intervals calculated from the 100 repetitions. The errors from the censored method remain low in type I settings (though not as low as the CPUE method) and are far smaller than the CPUE and ICR methods in the type II settings.

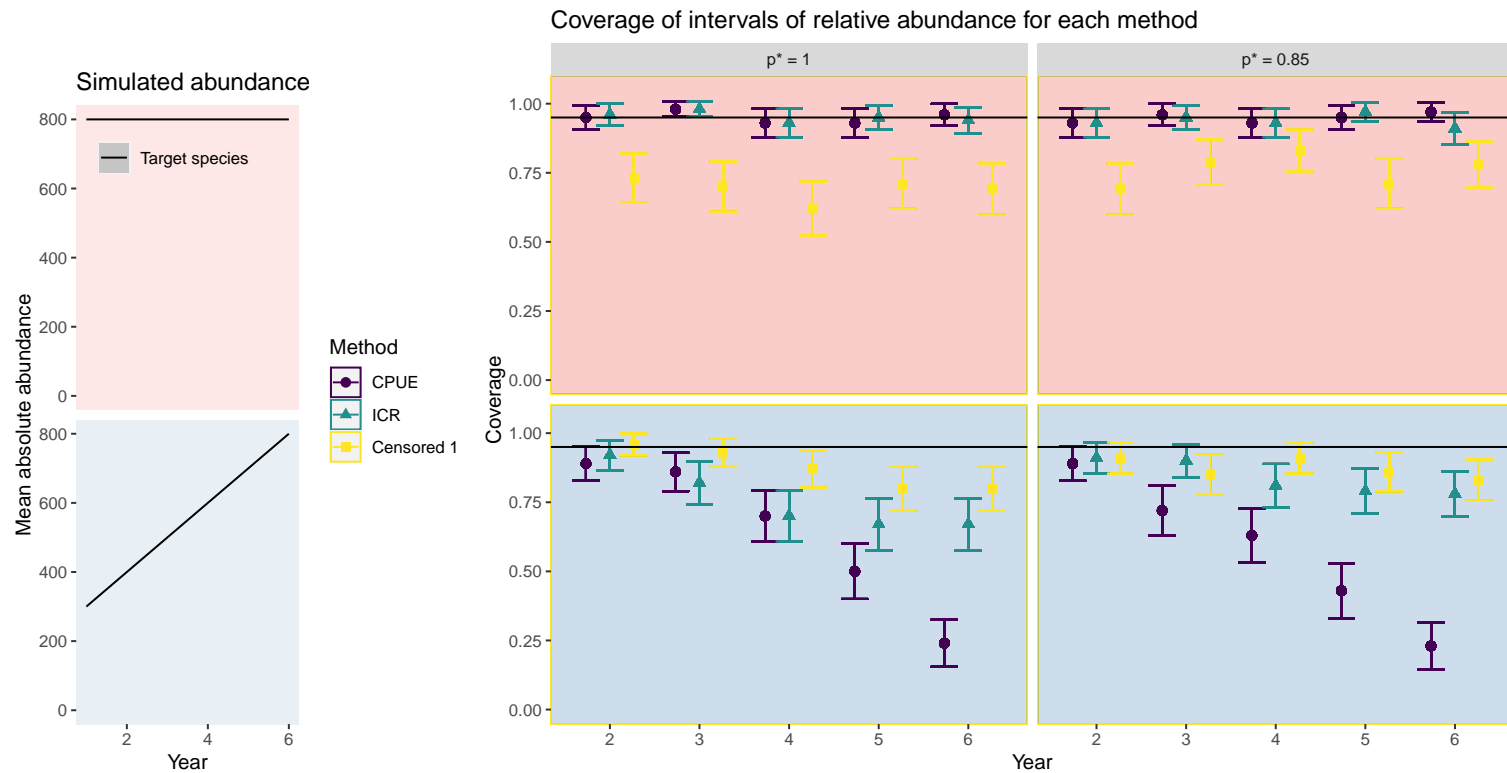


Figure S3: Coverage results for experiment 1. Rows and columns as for Figure S1. Plotted are the mean and approximate 95% confidence intervals across the 100 repetitions. The coverage levels of the censored method remain high in all settings, however, can fall to ~70% in the type I setting.

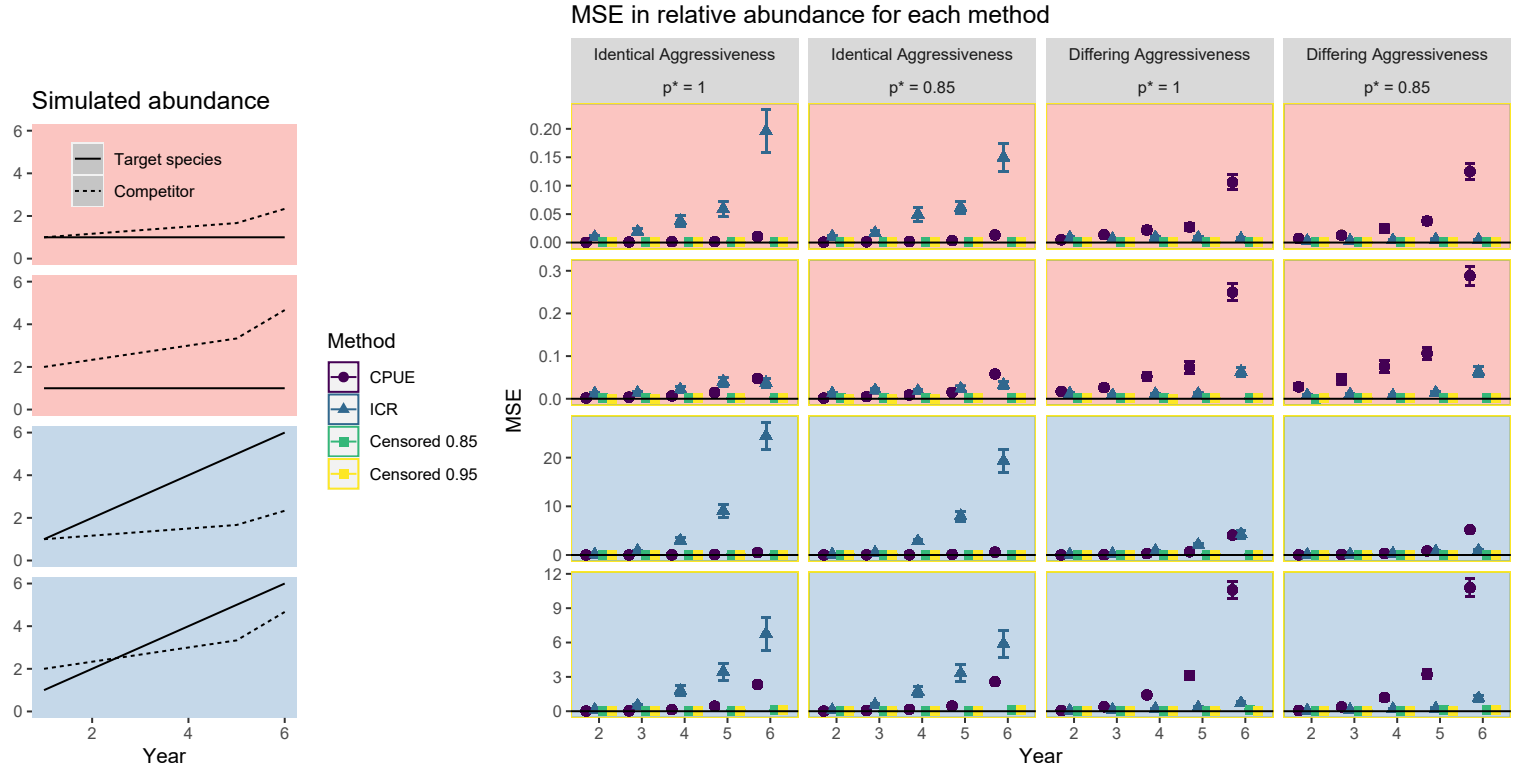


Figure S4: The median-squared error of the relative abundance indices for experiment 2. Columns and rows as in Figure 1. Plotted are the median and robust 95% confidence intervals computed from the 100 repetitions. The ‘Censored 0.85’ and ‘Censored 0.95’ methods assume $\hat{p}^* = 0.85$ and $\hat{p}^* = 0.95$, respectively. The censored method performs better than the others across all setting.

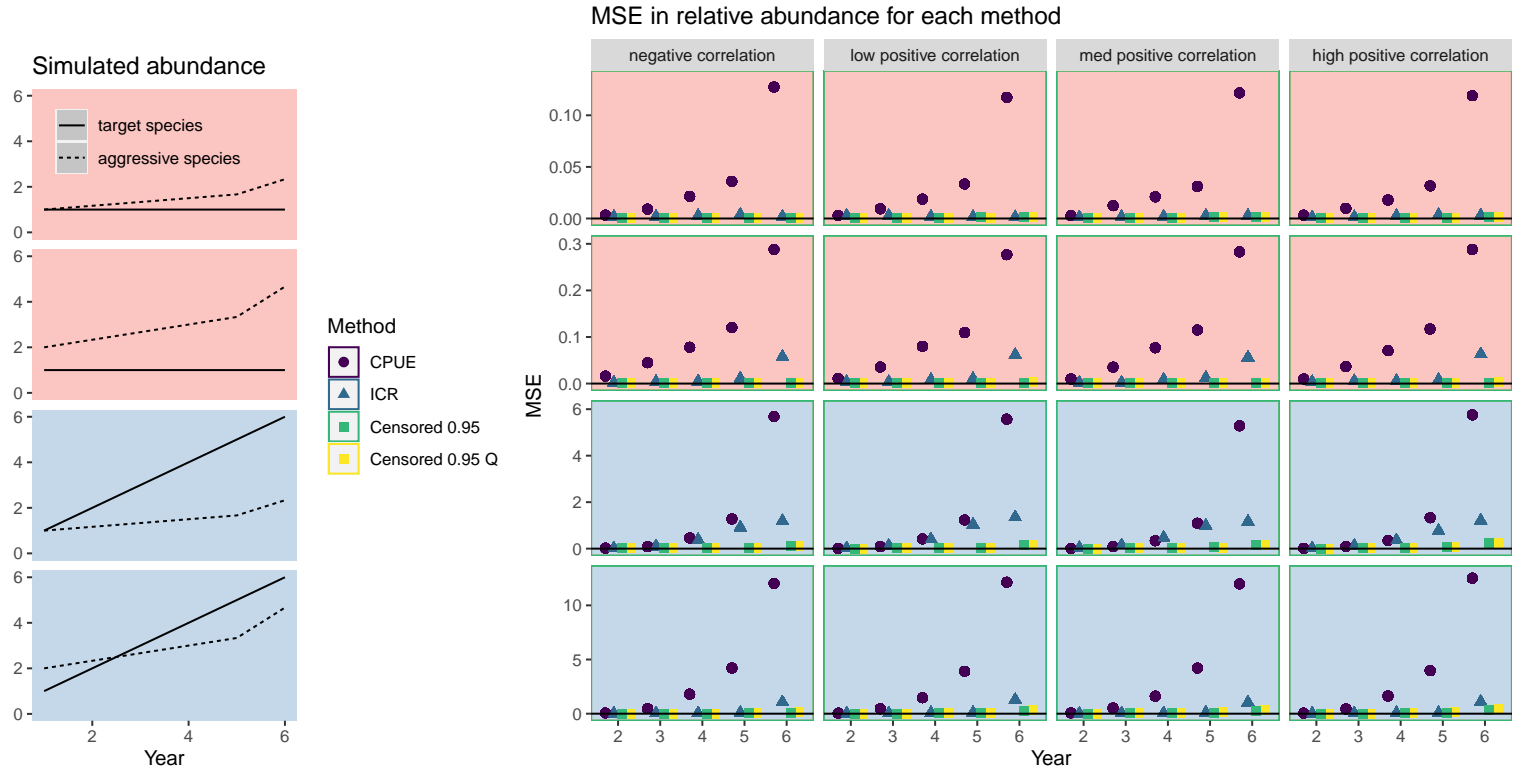


Figure S5: The median-squared error of the relative abundance indices for experiment 3. Rows, columns, and methods as in Figure 3. Also included are results from the ‘Censored 0.95 Q’ estimator, denoting the censored method where all observed counts of the target species \geq the 85th observed annual quantile are considered ‘high-quality’, regardless of p_{tk} . Plotted are the median values. The approximate 95% intervals are omitted since they are typically smaller than the plotting character. The censored methods better estimate the relative abundance across all settings compared to the CPUE and ICR-base methods. Little difference between the two censored methods can be seen. Errors for the CPUE method are really high in all settings and errors for the ICR method are highest in the settings with the greatest levels of bait removal by all sources.

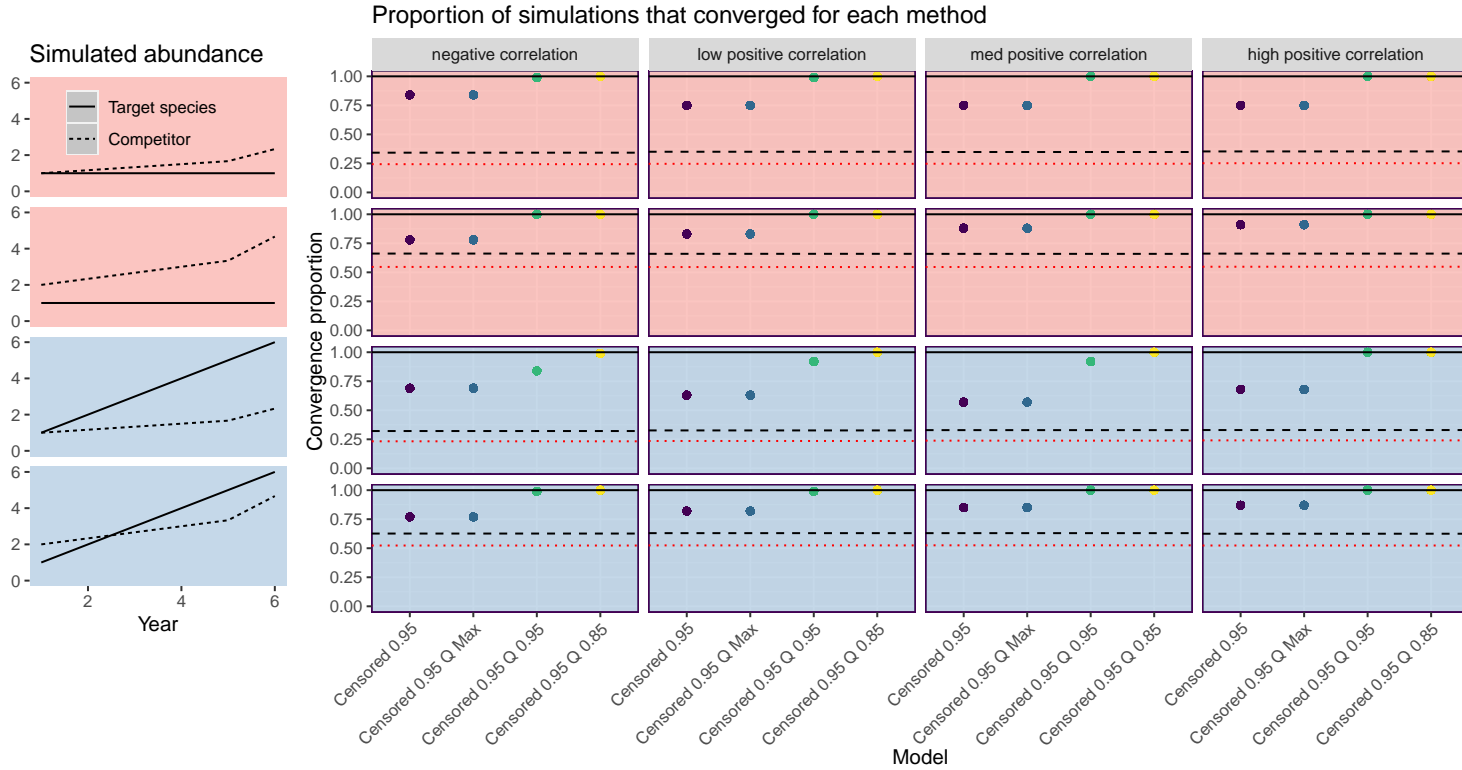


Figure S6: The proportion of the 100 simulations that successfully converged for the four variants of the censored method tested in experiment 3 with $p^* = 0.85$ throughout. We omit the CPUE and ICR methods as they always converge. Columns and rows as in Figure S5. Plotted as points are the convergence proportions. Also plotted as a dashed black line and a dotted red line are the proportions of fishing events across the 6 years that experienced at least 95% and 100% bait removal. ‘Censored Q X’ denotes the censored method where all observed counts of the target species \geq the X^{th} observed annual quantile are considered ‘high-quality’, regardless of p_{tk} . The ‘Censored 0.95 Q 0.85’ method converged in all settings but the ‘Censored 0.95’ method without the quantile adjustment failed to converge up to 45% of the time in certain settings. Over 60% of fishing events in rows 2 and 4 return 5% or fewer of the baits, and over 50% of events return no baits. In rows 1 and 3 these numbers are 20% and 30%. Note that convergence is now improved in light of the improved optimization approach discovered after conducting this simulation study.

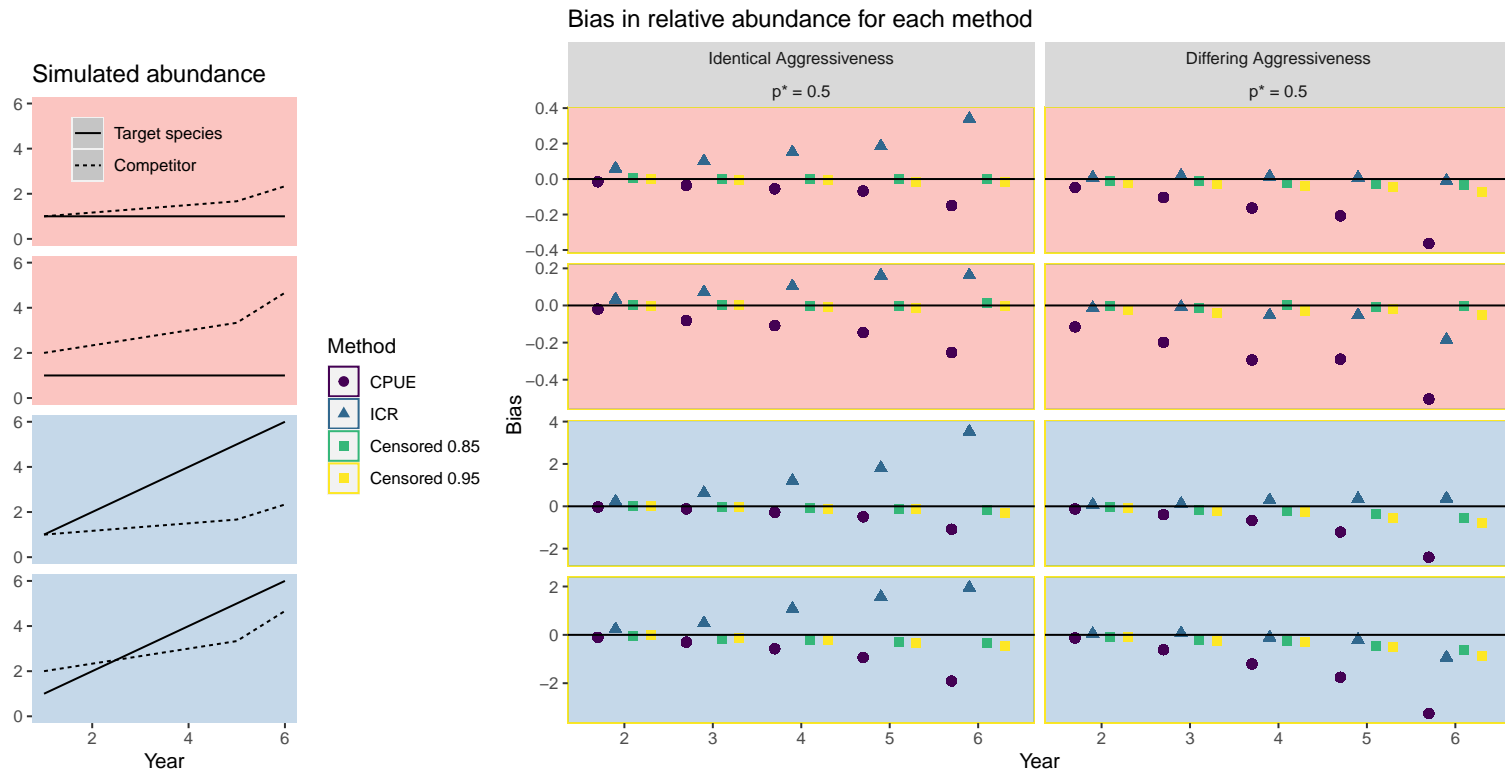


Figure S7: Absolute bias results for experiment 4. Rows are defined as in Figure 1, with the two columns testing the alternative bite-time scenarios. Plotted are the median values. The approximate 95% intervals are omitted since they are typically smaller than the plotting character. The censored method typically outperforms the other two.

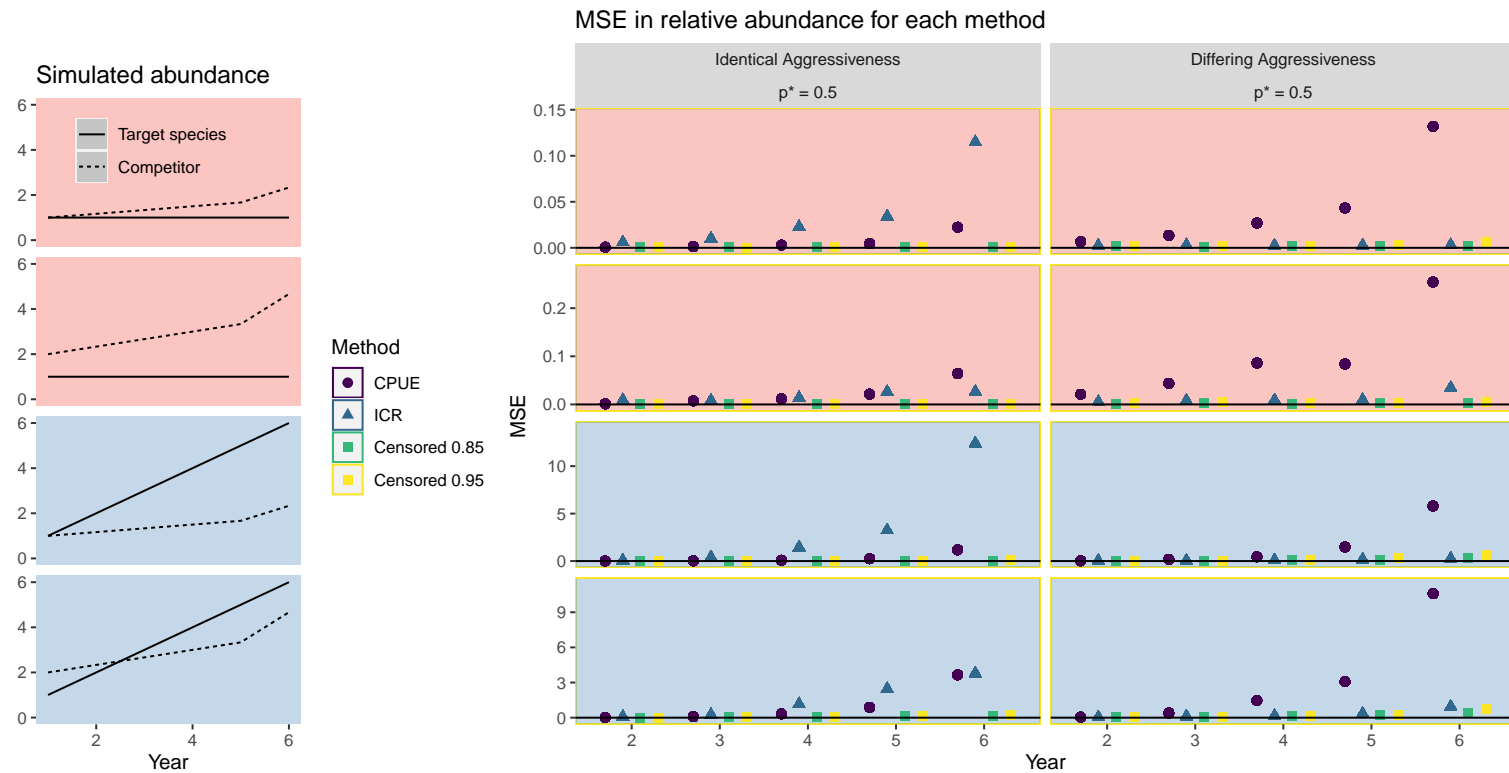


Figure S8: Median-squared error results for experiment 4. Details as for Figure S7. The MSE is always lowest for the censored method by year 6.

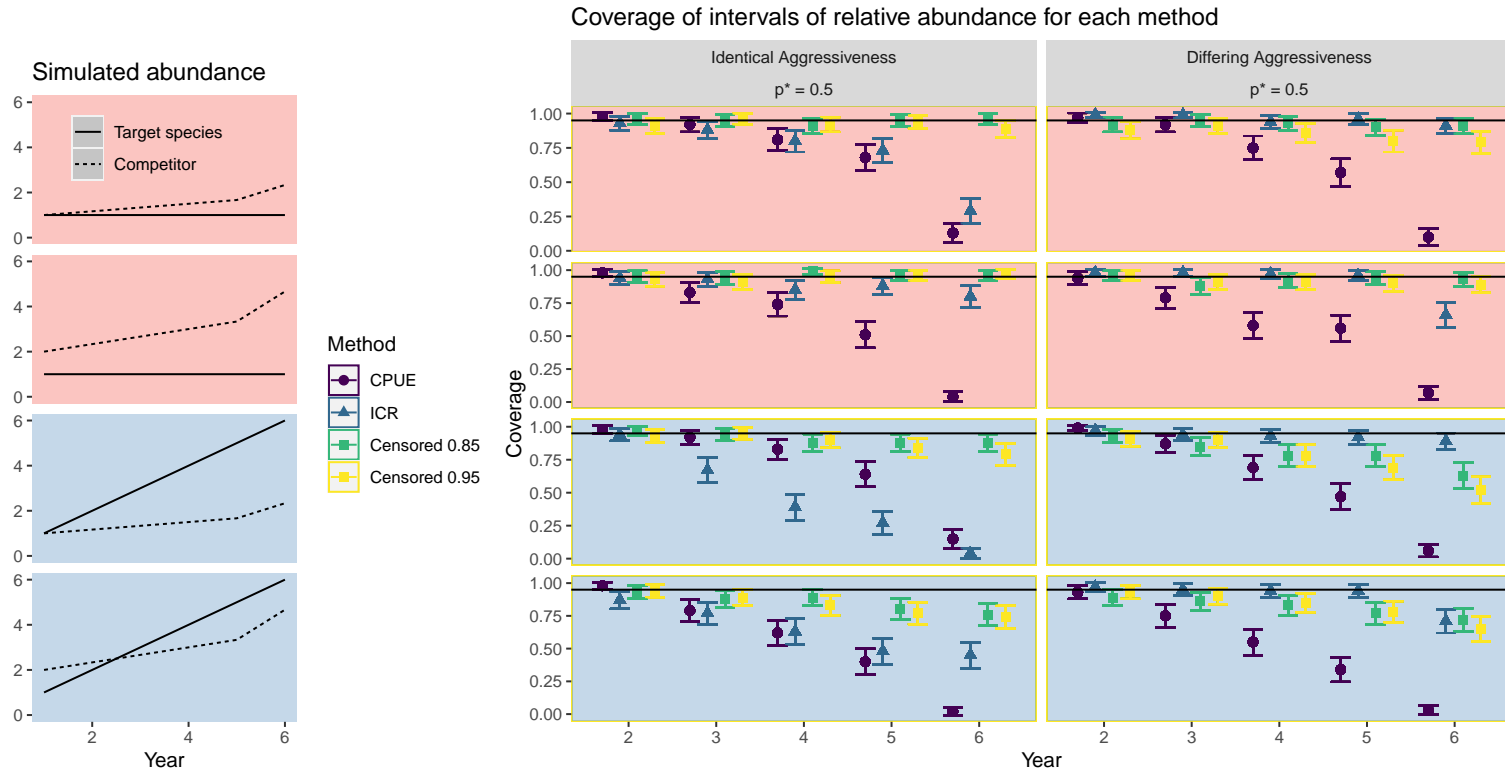


Figure S9: Nominal coverage of 95% credible intervals for experiment 4, for which $p^* = 0.5$ is fixed (way below the assumed $\hat{p}^* = 0.85$ and 0.95 values used for the ‘Censored 0.85’ and ‘Censored 0.95’ methods). Details as for Figure S7. The censored method attains the best coverage levels in all settings and close to 95% in the first two rows, satisfying the type I property but not always the type II property.

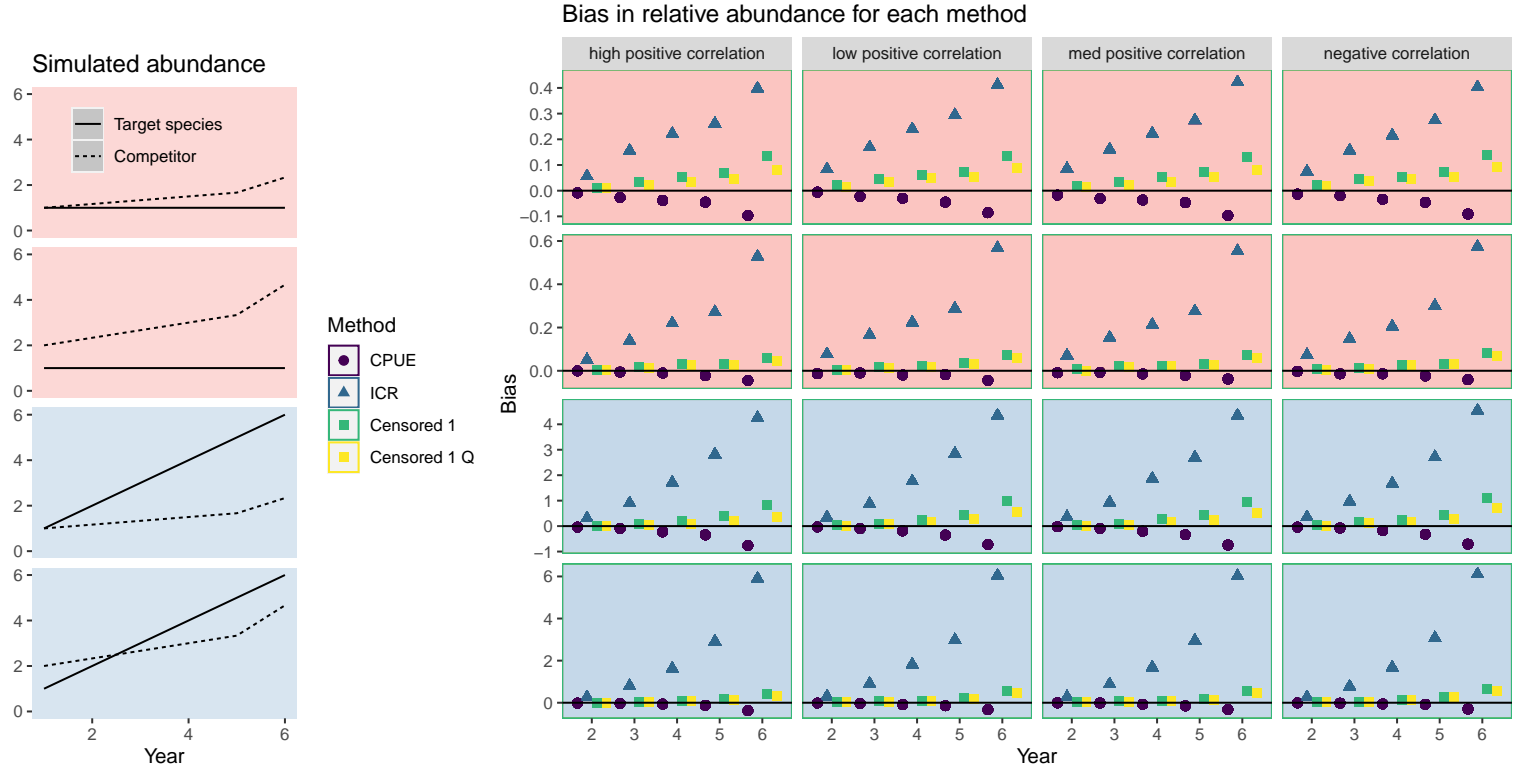


Figure S10: Absolute bias results for experiment 5. Rows, columns, and methods as in Figure 3, with the extra methods ‘Censored 1’ and ‘Censored 1 Q’ which assume $\hat{p}^* = 1$. ‘Censored 1 Q’ considers all observed counts of the target species \geq the 85th observed annual quantile as ‘high-quality’, regardless of p_{tk} . The bias from the censored method remains low in all settings and is at least as good as the CPUE method method, which is expected to perform well here due to the insensitivity of the target species to hook competition. The approximate 95% intervals are omitted since they are typically smaller than the plotting character.

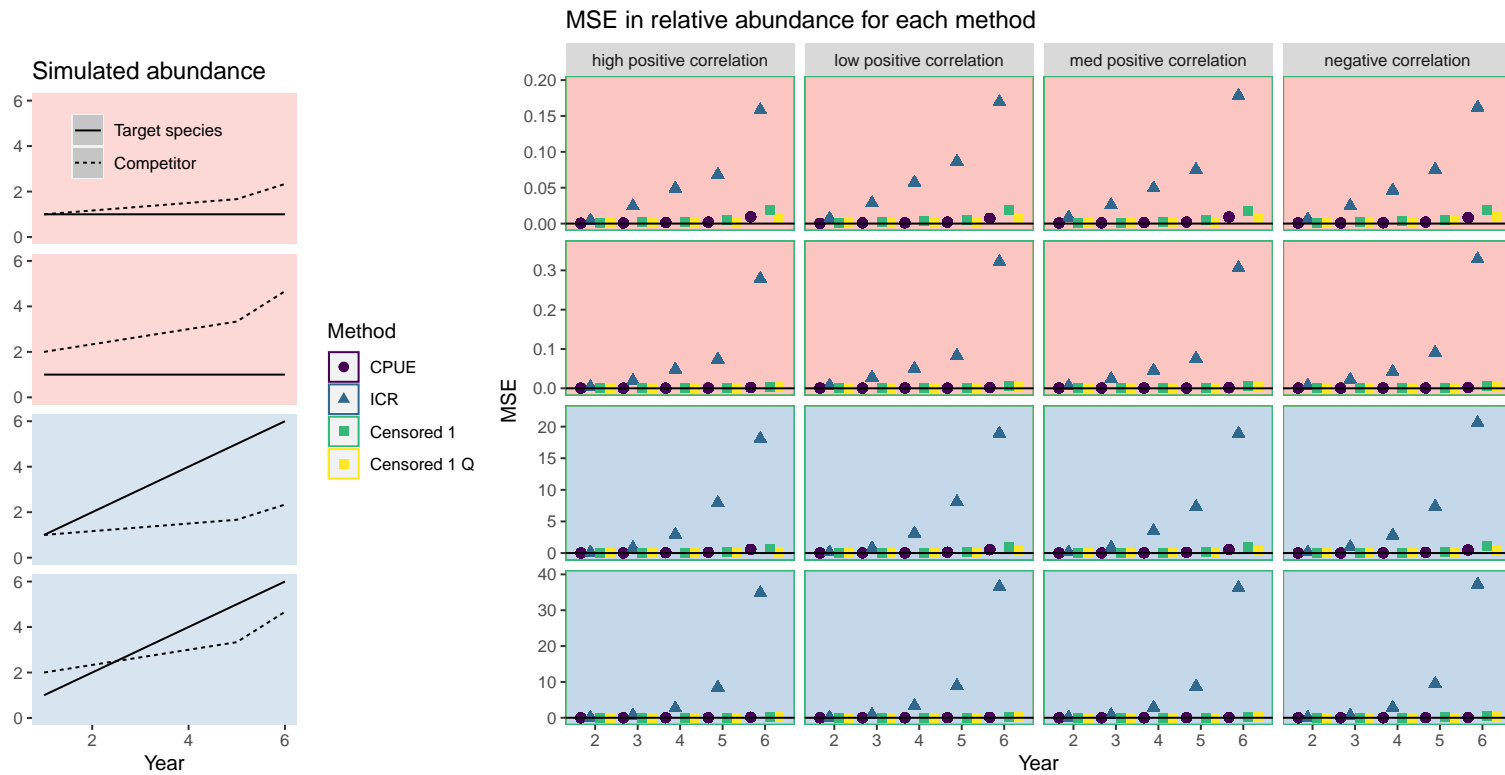


Figure S11: Median-squared error results for experiment 5. Details as for Figure S10. Values for the censored method remain competitively low in all settings (compared with the CPUE method), but are highest in the negative correlation setting. The errors from the ICR method are extremely high. The approximate 95% intervals are omitted since they are typically smaller than the plotting character.

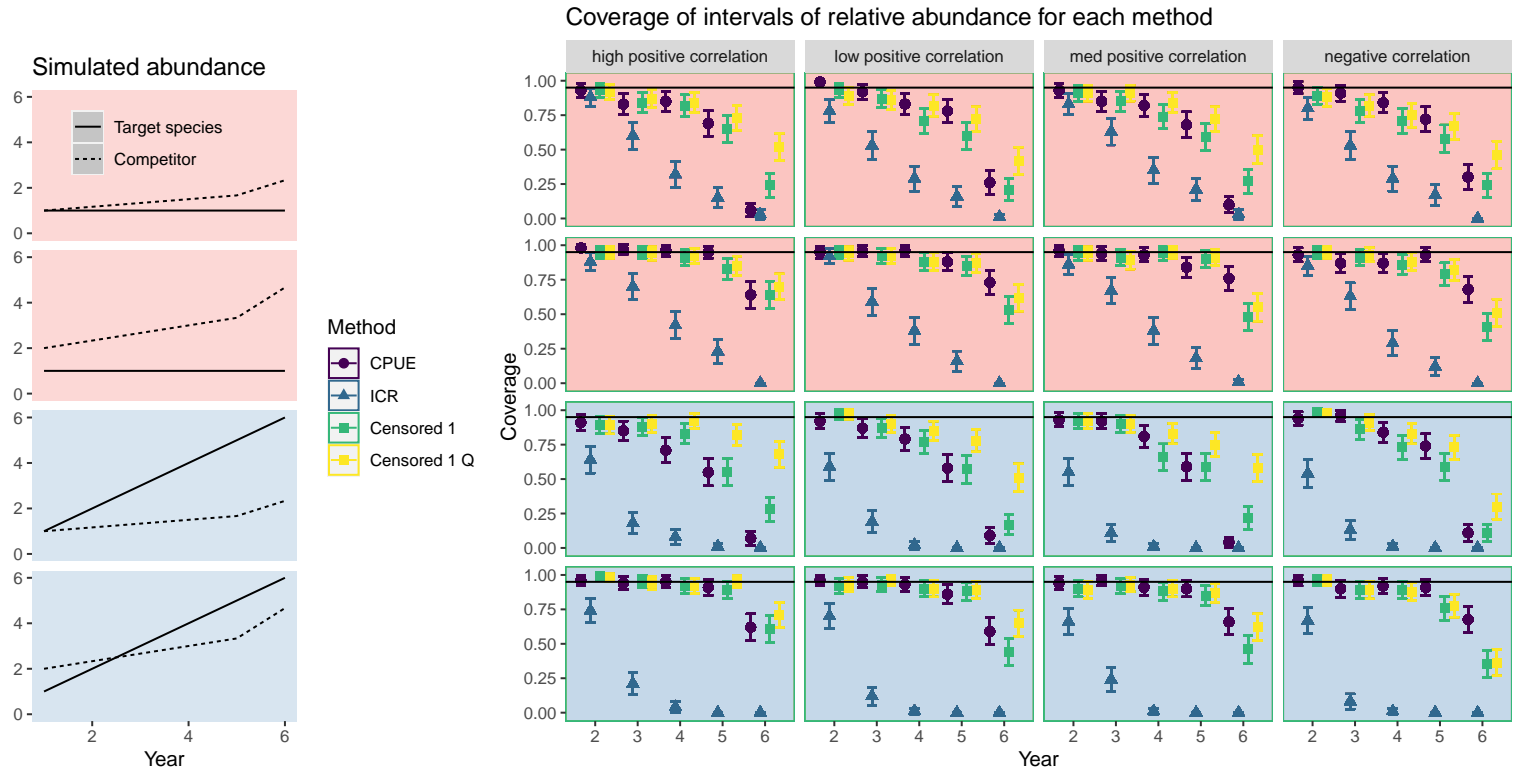


Figure S12: Coverage results for experiment 5. Details as for Figure S10. The coverage levels for the censored method are higher than the competing methods in most settings, but can fall as low as $\sim 20\%$. In general, the censored method's coverage is lowest in the negative correlation settings, and improved when $Q = 0.85$ is specified.

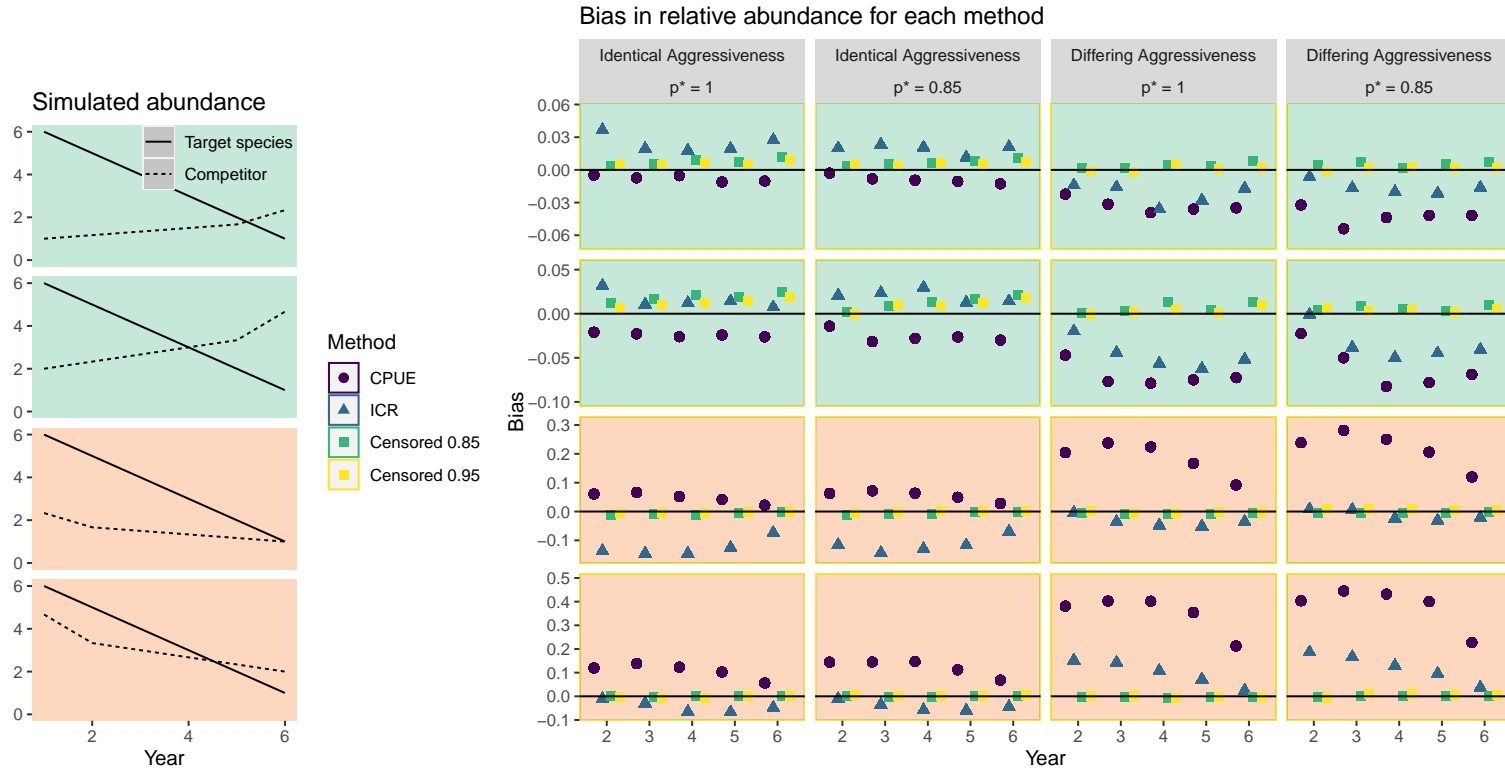


Figure S13: Absolute bias results for experiment 7. Columns and methods as in Figure 1. Leftmost column shows the relative abundances for the target species and the main competitor in the simulations. The target species' abundance is always decreasing and the competitor's abundance either increases (top two rows) or decreases (bottom two rows). Rows are coloured differently to reflect the fact that they no longer differentiate type I and type II settings. The 'other' species group's abundance remains constant. Plotted are the median values of bias. The approximate 95% intervals are omitted since they are typically smaller than the plotting character. The censored method is the only one to attain low bias in all settings. Notice that the signs of all the bias values in the bottom two rows in year 6 are opposite to those seen in Figure 1, as suggested by theory.

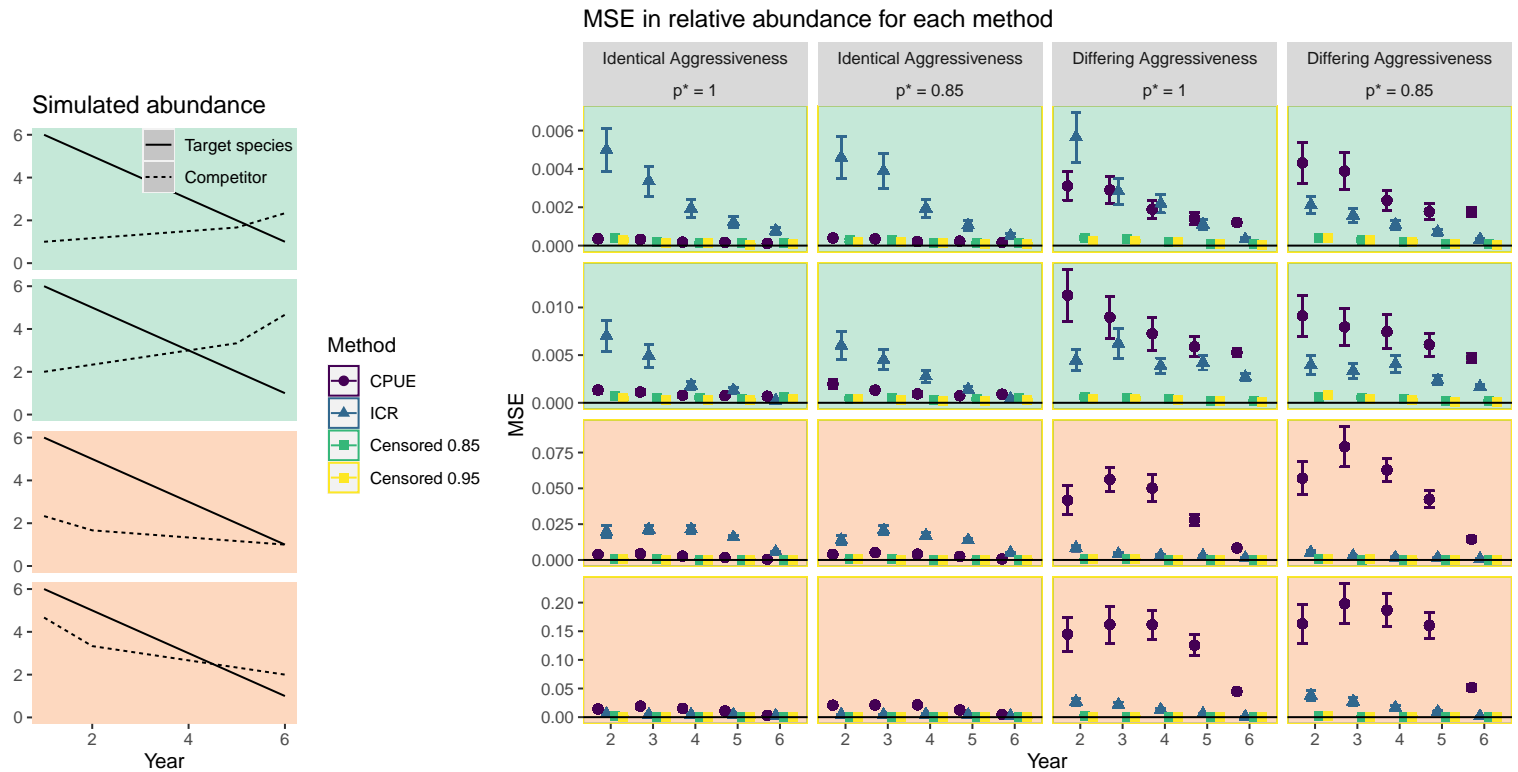


Figure S14: Median-squared error results for experiment 7. Details as for Figure S13. The approximate 95% intervals are omitted since they are typically smaller than the plotting character. MSE values for the censored method are the lowest in every setting tested.

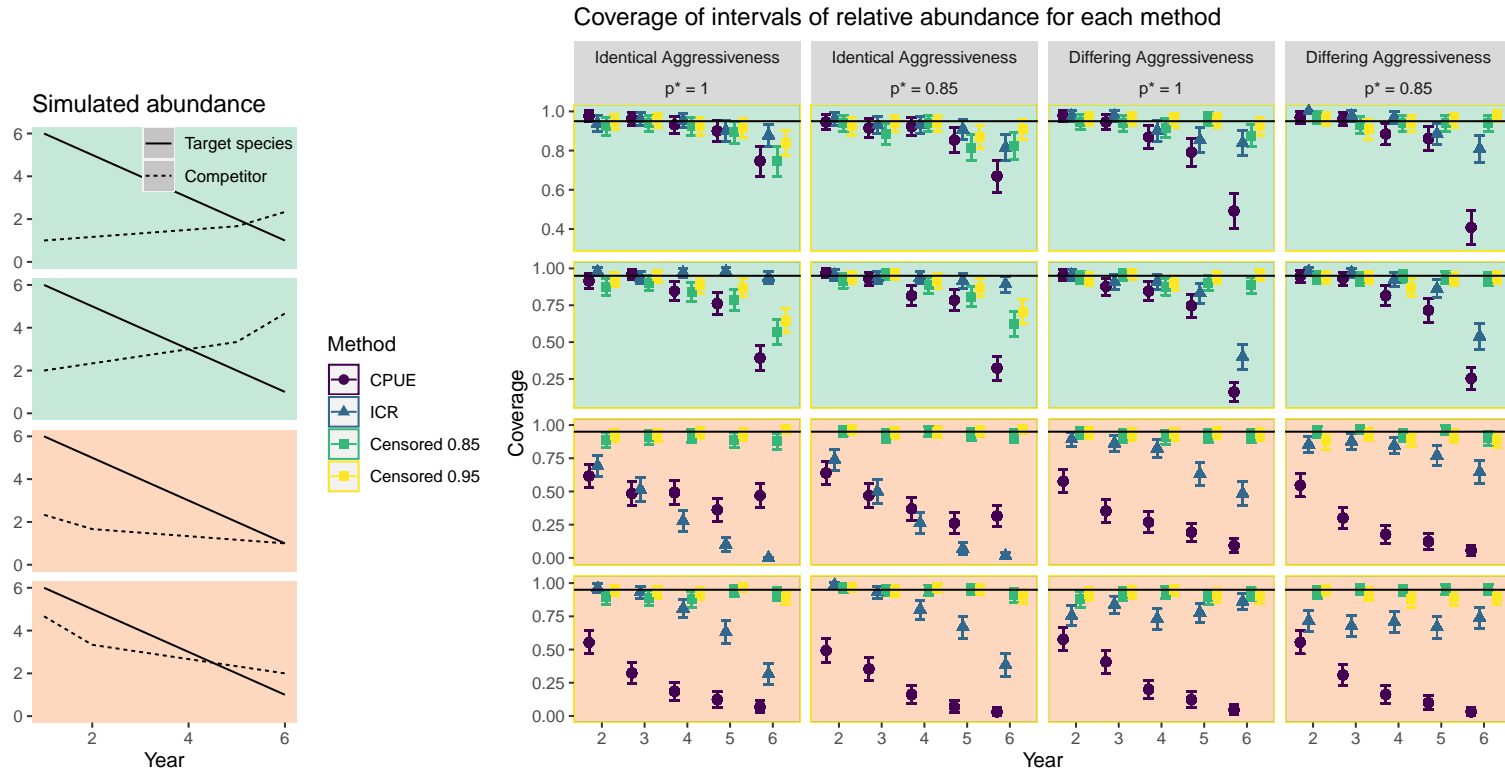


Figure S15: Coverage results for experiment 7. Details as for Figure S13. The coverage levels for the censored method are close to 95% in all settings and years, with the exception of two years (see leftmost results in row 2), where coverage falls to around 60%. Conversely, coverage levels from the other two methods are 0% in some settings. Notice that all the year 6-abundance coverage levels are close to 95% in the bottom two rows, in agreement with Figure 2 and theory.

S.2 Proof of third theoretical example

We prove the result of the third theoretical example presented in the introduction of the main text. This considered two species over two years. The abundance of species A (denoted Λ_{At}) was assumed stationary ($\Lambda_{A,1} = \Lambda_{A,2}$), but species B increased in abundance in year 2 ($\Lambda_{B,1} < \Lambda_{B,2}$). Species A was assumed to always reach the baits ahead of species B. There were 1000 baited hooks deployed in a single fishing event each year and 990 fish of species A were caught in both years. For species B, 0 and 9 fish were caught in years 1 and 2 respectively. We drop the k subscript due to the single fishing event each year.

Let c_{At} denote the observed catch of species A in year t and let F_t be the multiplicative competition adjustment factor from (S.4) but without depending on k , such that

$$F_t = \begin{cases} \frac{-\log(1-p_t)}{p_t}, & p_t < 1 \\ \frac{-\log(1/800)}{799/800}, & p_t = 1, \end{cases} \quad (\text{S.17})$$

where p_t is the proportion of the baits removed during the single fishing event of year t .

To estimate the relative abundance of species i in year t , the ICR method scales the catch per unit effort $c_{At}/1000$ by F_t as:

$$\hat{\Lambda}_{At}(\mathbf{s}_t) \propto \frac{c_{At}}{1000} F_t. \quad (\text{S.18})$$

Knowing that species A can always reach baits ahead of species B suggests that estimates of relative abundance for species A should not change between the years. However, the multiplicative competition adjustment factor depends on the observed catch count of species B (through the dependency between c_{Bt} and F_t).

Substituting in the values 990/1000 and 999/1000 for p_t , we see that the multiplicative scale factor F_t approximately equals 4.65 and 6.91 in years 1 and 2. Since c_{At} is fixed across the years, an increase of $6.91/4.65 \approx 49\%$ in the relative abundance of species A is estimated, yet its true

abundance did not change.

Thus, the ICR method's estimator of relative abundance is highly sensitive to changes in the abundances of competing species, which can easily occur in real-world settings.

S.3 Deriving an upper bound to improve censored method's convergence in data-poor settings

For some of the simulation experiments with $\hat{p}^* < 1$, and for three of the case study species, we placed an upper bound on the catch counts of the target species, denoted u_{itk} , to improve the convergence of the censored method. We now derive the upper bound, argue why it is likely conservative when used in general settings, and then empirically confirm this claim for the simulated data of experiments 2-5. Note that this upper bound will likely be needed less frequently in light of the subsequent improvements in the optimization routine.

Let a collection of longline fishing events in year t be indexed by k , each with h_{tk} baited hooks deployed and with assumed censorship threshold values $\hat{p}_i^* < 1$ for each species. First, assume that attracted individuals of species i do not have any difficulties locating a baited hook when a proportion less than \hat{p}_i^* of them have had their baits removed. We refer to this period of the soak time as stage 1. We do not need to make any further assumptions about the arrival rates of the fish to the baits during this stage. We denote the total catch counts of species i after soak time stage 1 as c'_{itk} .

Next, at a given moment in time, $0 < \tau < S_{tk}$, during the soak time S_{tk} , denote the current proportion of baits removed as $p(\tau)$. We refer to the period of the soak time where $p(\tau) \geq \hat{p}_i^*$, when it occurs, as stage 2. During stage 2, we assume that each species arrives at the fishing gear at times which are independent exponentially distributed with constant instantaneous rates ρ_{itk} . Then, the probability of capture is assumed to decrease linearly with time as

$$1 - \frac{p(\tau) - \hat{p}_i^*}{1 - \hat{p}_i^*}. \quad (\text{S.19})$$

The assumed exponentially-distributed bait removal times during the stage 2 are chosen so that the fishing process in stage 2 satisfies the assumptions required of the ICR scale factor adjustment F_{tk} .

Let $p_{tk} > \hat{p}_i^*$. If we knew the number of fish of the target species that were caught during stage 2, denoted c''_{itk} , then an ICR-based estimate for the number of fish of species i that could have been caught if more hooks been available during the stage 2 soak time would be:

$$\left[c''_{itk} \frac{-\log \left(1 - \frac{p_{tk} - p_i^*}{1 - p_i^*} \right)}{\left(\frac{p_{tk} - p_i^*}{1 - p_i^*} \right)} \right]. \quad (\text{S.20})$$

Thus, under the above assumptions, the ICR-based estimate for the total number of fish caught during the entire soak time would equal:

$$c'_{itk} + \left[c''_{itk} \frac{-\log \left(1 - \frac{p_{tk} - p_i^*}{1 - p_i^*} \right)}{\left(\frac{p_{tk} - p_i^*}{1 - p_i^*} \right)} \right]. \quad (\text{S.21})$$

Clearly, without hook timers, we will be unable to determine c'_{itk} or c''_{itk} . However, since we are only requiring u_{itk} to be an upper bound, we replace these numbers with known quantities: i) c_{itk} , the observed catch count of species i after the soak time S_{tk} ; ii) the number of hooks observed with their bait removed by *all sources* in the second stage, $h_{tk} (p_{tk} - p_i^*)$. Thus, we define a conservative upper bound for each species i during fishing events where c_{tk} (the observed number of bait removals by all causes during fishing event k of year t) satisfies $c_{tk} > h_{tk} p_i^*$:

$$u_{itk} = c_{itk} + \left[h_{tk} (p_{tk} - \hat{p}_i^*) \frac{-\log \left(1 - \frac{p_{tk} - \hat{p}_i^*}{1 - \hat{p}_i^*} \right)}{\left(\frac{p_{tk} - \hat{p}_i^*}{1 - \hat{p}_i^*} \right)} \right]. \quad (\text{S.22})$$

In words, our upper bound for the censored observations is the observed species-specific catch count c_{itk} plus the observed number of baited hooks removed in stage 2 by all sources, $h_{tk} (p_{tk} - p_i^*)$, scaled by the competition adjustment factor F_{tk} evaluated only for the stage 2 hooks. This value is then rounded up to the nearest integer.

Note that we are not assuming that the above model describes the true bait removal process in stage 2. Instead, we are trying to utilize a set of assumptions that justify a reasonable upper bound to improve the convergence of the censored method in the most challenging of settings. Given that we are assuming the instantaneous catch rates of each species is non-decreasing throughout the soak time of stage 2 and that all bait removals in stage 2 were due to species i , this should ensure that u_{itk} is a conservative upper bound in many settings where multiple species are competing for large numbers of hooks.

We empirically evaluate this claim. The upper bound is always used in simulation experiments 2-4. For these settings, the upper bound lies above the true number of attracted individuals C_{itk} in 99.9% of cases and exceeds it by a mean number of 350 fish when $\hat{p}^* = p^*$. Thus, u_{itk} is a successful (and highly conservative) upper bound which should have only a minor impact on the results of the simulation study, compared with if the simulation had been run without any upper-bound at all. Thus, we are confident that the results from experiments 2-4 with the upper bound used generalize to the setting where no upper bound is used, without us needing to re-run the computationally-costly simulation experiments without the upper-bound and using the new optimization routine. Furthermore, results from experiment 7 empirically support this claim, as explained earlier.

An upper bound was considered in experiment 5 when $\hat{p}_i^* < 1$. However, in this experiment, results showed that performance was optimized with $\hat{p}_i^* = 1$ when no upper bound was used and so we do not repeat the empirical evaluation for experiment 5. The upper bound was not used in experiment 1.

S.4 Exploratory analysis for the case study

S.4.1 Assessing the linear relationship between effective skate and catch counts

The geographic extent of the IPHC data used in the case study is shown in Figure S16, along with the locations of the stations on the 10×10 nautical mile square grid; this is IPHC Regulatory Area 2B.

We investigate the hypothesis that the measure of effective skate (Yamanaka et al., 2008) is a suitable proxy for the integral of the response surface (equation 11) for all 11 species analyzed in the case study. Despite the IPHC data being a survey with fixed protocols deployed, in some years, only the first 20 hooks of each skate were enumerated. The effective skate values have been scaled to account for this, but we remain unsure about the validity of this correction. For effective skate to be a valid proxy implies that the expected catch count of a species scales linearly with effective skate, holding all other factors constant. We test this now.

Let i denote the species and let E_{tk} denote the effective skate value corresponding to the fishing gear deployed in year t at station k . We fit the following negative binomial GAMM to the observed species catch counts (c_{itk}) and perform exploratory analysis to empirically assess this assumption:

$$[c_{itk}|Z_{ik}] \sim \text{NegBin}(\mu_{itk}, \theta)$$

$$\log(\mu_{itk}) = \beta_{it} + g_i(\log E_{tk}) + Z_{ik}$$

$$[Z_{ik}] \sim N(0, \sigma_i) \quad (\text{IID station random effects}),$$

where $g_i(\log E_{tk})$ denotes (possibly nonlinear) effects of effective skate, to be estimated with cubic spline smooths, with the degree of ‘wiggleness’ chosen automatically. We use the `mgcv` package to perform the model fitting (Wood, 2004, 2011; Wood et al., 2016).

Thus, the above model attempts to control for the changes in abundance through time and

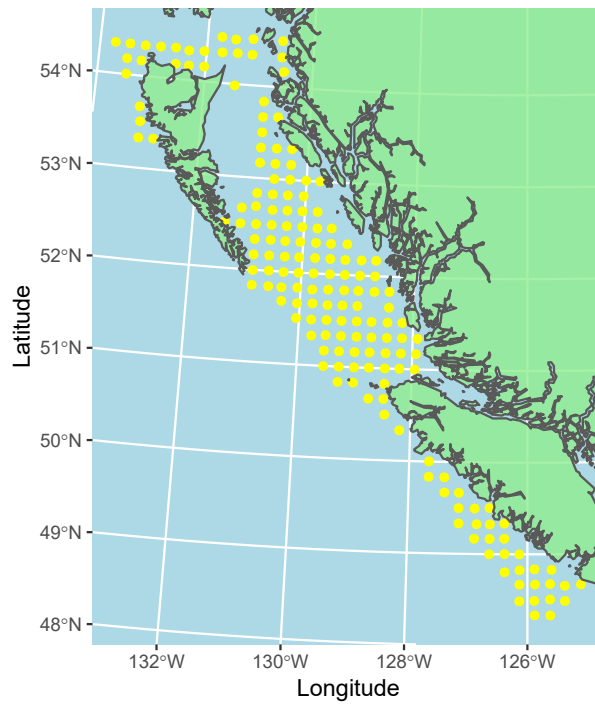


Figure S16: The locations of the 171 International Pacific Halibut Commission stations that were fished in Canadian waters off British Columbia between 1998-2021.

the station-station variability with random effects. Then, lastly, any nonlinear effects of effective skate are estimated in the $g_i(\cdot)$ terms. Note that no adjustments are made for hook competition and hence the model is only exploratory in nature. The negative binomial model is chosen to account for overdispersion in the catch counts, whilst being available in the computationally-fast mgcv R package, unlike the Poisson lognormal used elsewhere in this work.

We plot the mean and 95% confidence intervals for the $g_i(\cdot)$ terms for all 11 species. Figure S17 fails to show any strong departures from a linear trend between the mean catch counts and the value of effective skate for 6 of the species, with the 95% confidence intervals for these species covering a line with unit slope. However, significant departures are detected for 5 species, most severely for Arrowtooth Flounder, Pacific Cod, and Sablefish. Overall, we conclude that the linearity assumption between effective skate and the mean catch counts appears to be a reasonable approach to account for heterogeneous fishing effort across the fishing events for 6 of the 11 species. Future work should investigate how to better control for heterogeneous observer effort in the other 5 species. We continue to use effective skate as an offset value to account for heterogeneous fishing effort when estimating the relative abundances of all 11 species, however, the estimated indices should be treated with caution.

S.4.2 Empirically assessing values of \hat{p}_i^*

Next, we investigate suitable choices for the censorship proportion, \hat{p}_i^* , for all 11 species. To do this, we once again attempt to control for the temporal effects on the mean abundances of both species. Based on the results of the previous subsection, we remove the cubic spline smoother of effective skate and assume a linear relationship.

To detect this hook saturation effect, we include a cubic spline of the proportion of baits

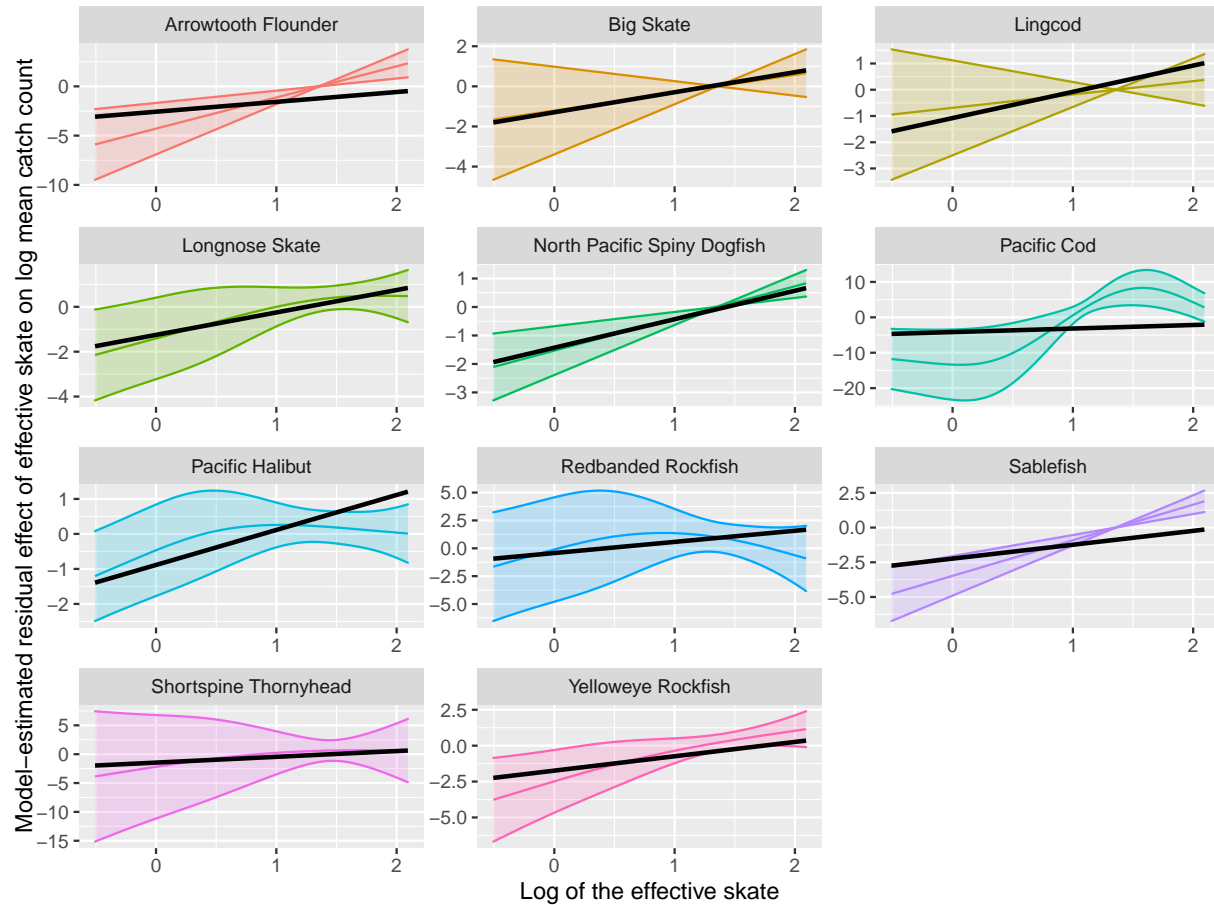


Figure S17: A plot showing the estimated nonlinear effects of log effective skate on the log mean catch counts of all 11 species, after controlling for temporal effects. 6 out of 11 of the 95% confidence intervals cover the unit slope line across all values of log effective skate. This indicates that the effective skate is a useful proxy for the heterogeneous fishing effort experienced across the fishing events for these species. Future work should investigate how to control for heterogeneous observer effort in the other 5 species.

removed during each fishing event (p_{tk}) inside the model as follows:

$$[c_{itk}|Z_{ik}] \sim \text{NegBin}(\mu_{itk}E_{tk}, \theta)$$

$$\log(\mu_{itk}) = \beta_{it} + h_i(p_{tk}) + Z_{ik}$$

$$[Z_{ik}] \sim N(0, \sigma_i) \quad (\text{IID station random effects}).$$

Thus, the above model once again attempts to control for the changes in abundance through time and the station-station variability with random effects. Lastly, any residual nonlinear effects of hook competition should appear in the $h_i(\cdot)$ terms.

We plot the mean and 95% confidence intervals for exponentially transformed $h_i(\cdot)$ terms for all 11 species. Figure 6 shows that all species experience a statistically-significant decline in the mean catch count exceeding 10%, with the exceptions of: i) dogfish which experiences an statistically significant increase in mean catch count exceeding 65%; ii) Shortspine Thornyhead which experiences a non-significant decline of only 1%. The plots show that statistically significant declines in mean catch counts are first seen around: $p_{tk} = 0.7$ for Big Skate; $p_{tk} = 0.85$ for Lingcod, Redbanded Rockfish, and Yelloweye Rockfish; $p_{tk} = 0.95$ for the remaining 5 species.

Based on advice given in the main text from simulation experiment 6, values of: $\hat{p}_i^* = 1$ for dogfish and Shortspine Rockfish, $\hat{p}_i^* = 0.7$ for Big Skate, $\hat{p}_i^* = 0.85$ for Lingcod, Redbanded Rockfish, and Yelloweye Rockfish, and $\hat{p}_i^* = 0.95$ for all other species should be used. Furthermore, based on the decision rule given in the main text, we should treat the censored method's indices with caution for Shortspine Thornyhead until we can determine that neither fish is 'immune' to hook competition from other species.

S.5 Additional details on the case study models

We assume that doubling the value of effective skate of a fishing event doubles the expected integral of the response function attributable to the fishing event. By denoting \tilde{c}_{itk} to be the observed raw catch count for the CPUE and censored method, and the hook-competition adjusted

value for the ICR-adjusted method (i.e. equal to $c_{itk}F_{tk}$), we can formally write all three methods under the same framework as:

$$\pi(\tilde{c}_{itk}|Z_{ik}, e_{itk}) \sim \begin{cases} f(\tilde{c}_{itk}; \mu_{itk}) & p_{tk} < \hat{p}_i^* \\ 1 - F(\tilde{c}_{itk} - 1; \mu_{itk}) & p_{tk} \geq \hat{p}_i^*, \end{cases} \quad (\text{S.23})$$

$$\mu_{itk} = \lambda_{itk} E_{tk} e_{itk} \quad (\text{S.24})$$

$$\log \lambda_{itk} = \alpha_{it} + Z_{ik} \quad (\text{S.25})$$

$$[\log e_{itk}] \sim N(0, \sigma_e) \quad \text{IID overdispersion effects} \quad (\text{S.26})$$

$$[Z_{ik}] \sim N(0, \sigma_Z) \quad \text{IID station-station effects} \quad (\text{S.27})$$

where E_{tk} is the effective skate number of fishing event k in year t and σ_e and σ_Z are standard deviations to be estimated. By setting $\hat{p}_i^* > 1$ for the CPUE and ICR-adjusted methods (which do not require values of \hat{p}_i^*), (S.23) reduces to just its first row, correctly specifying those models. The temporal trends are captured by the unique intercept each year, α_{it} . These are unstructured fixed effects to allow for flexible changes in abundance to be inferred and avoid smoothing over the temporal variation. Normalized estimates (each series divided by its geometric mean) of $\exp(\alpha_{it})$ are used to represent the relative abundance trends in Figure 7 for three species, and in Figures S22-S24 for all species, characterised by their \hat{p}_i^* values.

All models were fit using the sdmTMB and TMB software packages (Anderson et al., 2022; Kristensen et al., 2016) within R (R Core Team, 2022). Models converged for 8 out of the 11 species without requiring any upper bounds u_{itk} be placed on the catch counts or any other modification. To attain convergence for Redbanded Rockfish, Yelloweye Rockfish, and Big Skate, the upper bound derived earlier in the Supplementary Material was used and the upper 5% of catch counts each year were considered high quality, regardless of the values of p_{tk} . Based on the results of the simulation study, we do not anticipate there to be substantial declines in performance caused by these modifications made to improve convergence.

S.6 Additional case study results

Figures S18, 8, S19, and S20 help us to understand the magnitudes of the differences between Figures S21, S22, S23, and S24. They plot the estimated percentage change in relative abundance estimate from the censored method compared to the CPUE method for each year, plotted with the proportion of fishing events with $p_{tk} \geq \hat{p}_i^*$ on the x-axis. In addition to assessing the absolute magnitudes of change, the differing sensitivities of each species to hook competition can also be inferred by assessing the relative magnitude of change witnessed for each species. Of note are Pacific Halibut and dogfish which are both estimated to be only slightly impacted by hook competition. Conversely, Lingcod, Sablefish, and Pacific Cod are all estimated to be acutely sensitive to hook competition by the censored method.

Figure S25 shows that the rank correlations between the CPUE and ICR methods' indices are consistently high (≈ 0.93) when averaged across all 11 species, whereas the rank correlations between the censored method's indices and the CPUE and ICR methods' indices are lower (≈ 0.80). Similar results are seen with respect to the percentage overlap of 95% confidence intervals for abundance (Figure S25). All of the intervals overlap for the CPUE and ICR methods, unlike the censored method's intervals which overlap with around 80-85% of the CPUE and ICR methods' intervals.

Figure S25 also shows that the estimates of uncertainty are smallest for the CPUE method and largest for the ICR-based method, with the censored method in-between. Furthermore, the MAD and 'Percentage Unchanged' results of Figure S25 show that relative abundance estimates from the ICR method are most variable and those from the CPUE method are least variable, while the 'Range' results indicate that the ICR method infers the greatest fluctuations in abundance.

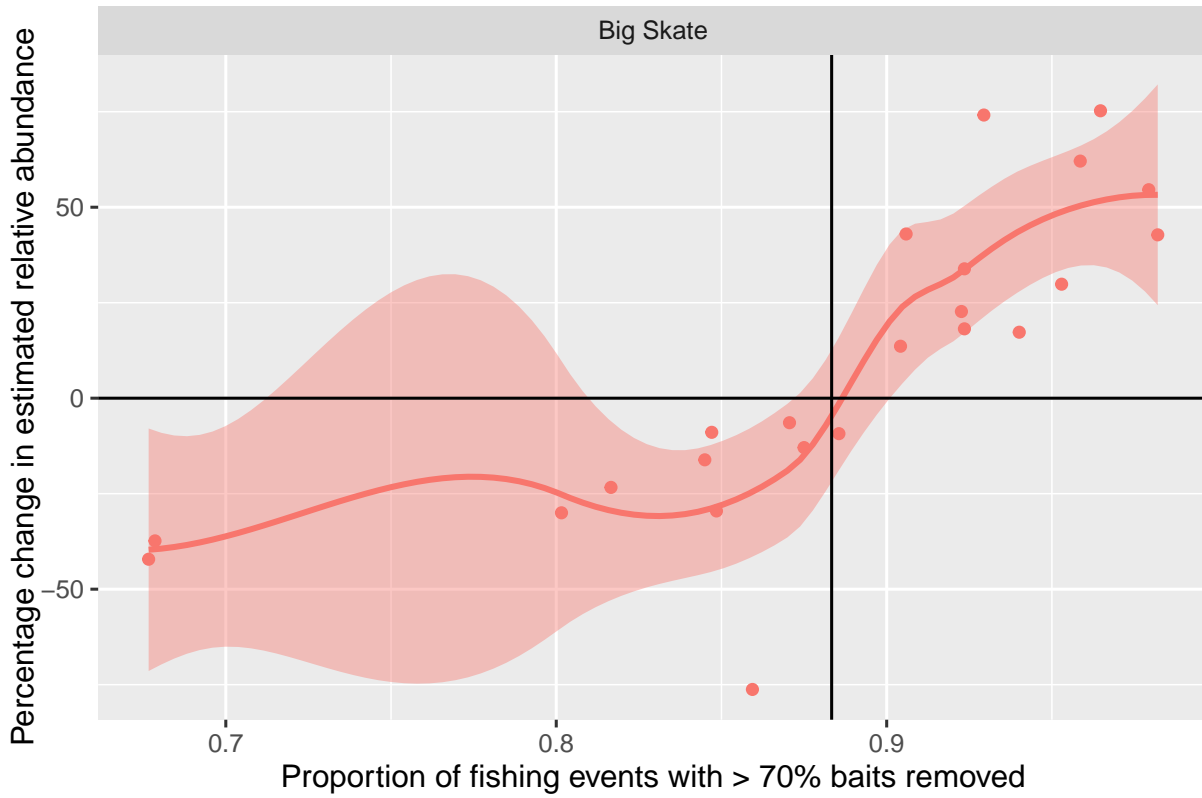


Figure S18: As for Figure 8 but for the one species with $\hat{p}_i^* = 0.7$, with vertical lines showing the average annual proportion of fishing events experiencing at least 70% bait removal. The general conclusions are the same as for Figure 8. Despite almost 100% of the fishing events being censored at the 70% level in some years, the percentage change in the estimated relative abundance in these years barely exceeds 50%. Thus, given the high prevalence of censorship for Big Skate, it appears that it is relatively insensitive to hook competition.

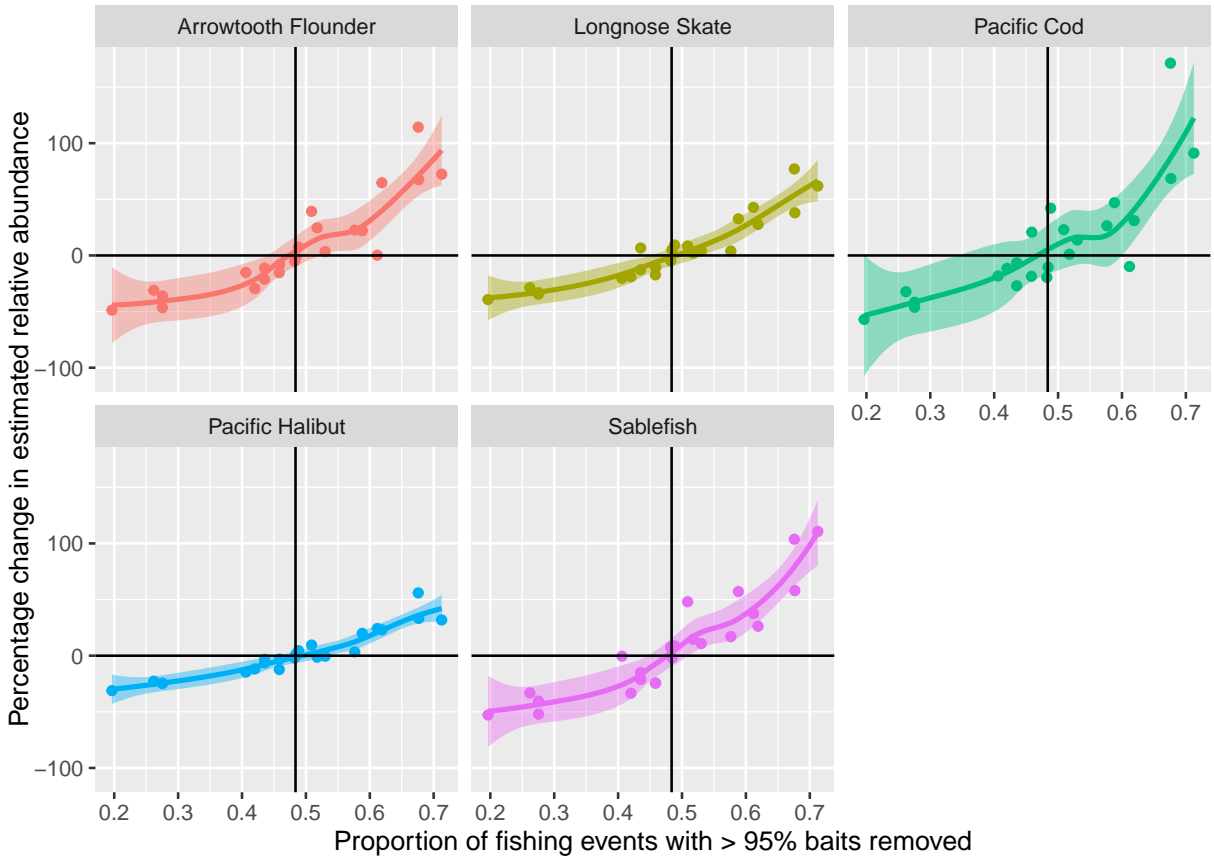


Figure S19: As for Figure 8 but for the five species with $\hat{p}_i^* = 0.95$, with vertical lines showing the average annual proportion of fishing events experiencing at least 95% bait removal. The general conclusions are the same as for Figure 8. Here, the censored method infers that Pacific Halibut is the least affected by hook competition, with Pacific Cod the most affected.

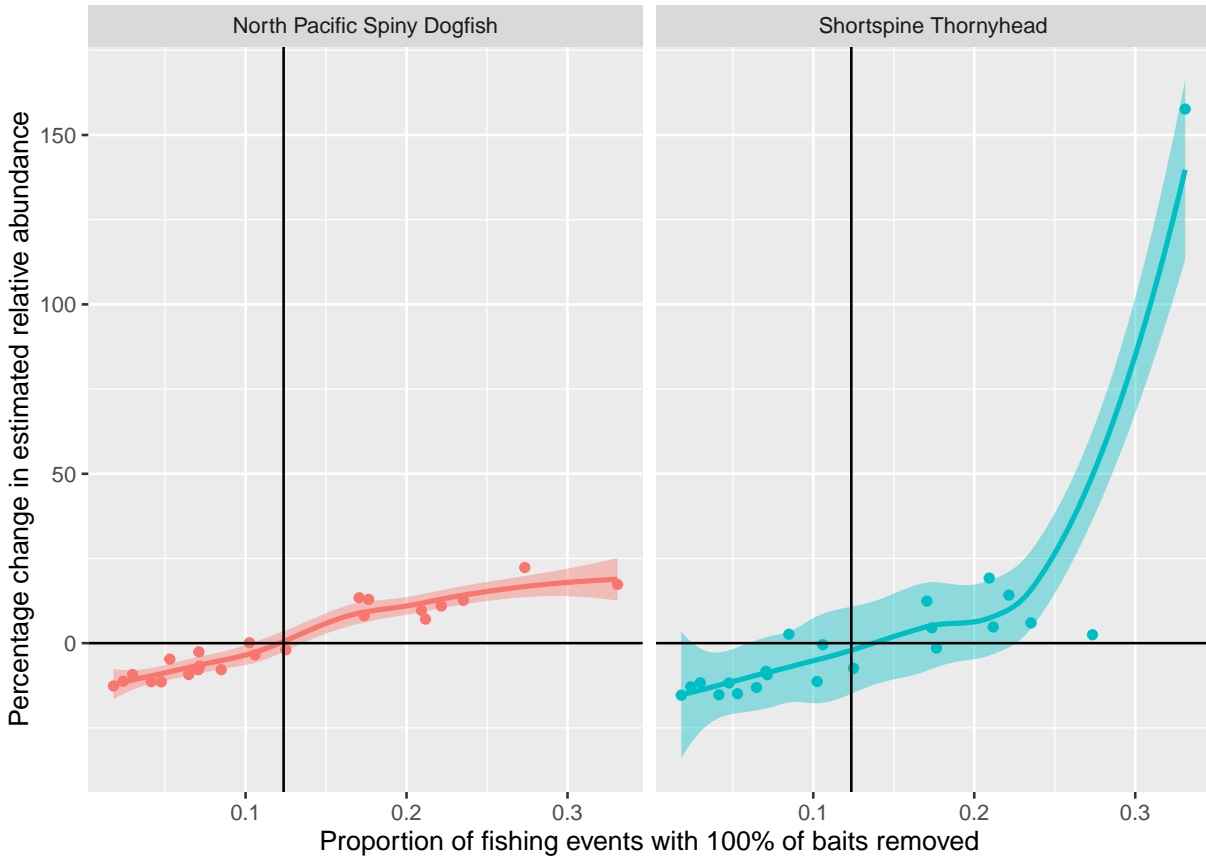


Figure S20: As for Figures 8 and S19 but for the two species with $\hat{p}_i^* = 1$, with vertical lines showing the average annual proportion of fishing events experiencing 100% bait removal. Again, the general conclusions are the same as for Figure 8. Here, the censored method infers Shortspine Thornyhead to be more severely affected by hook competition than dogfish, although this conclusion may be unduly influenced by a single outlier seen in the top right.

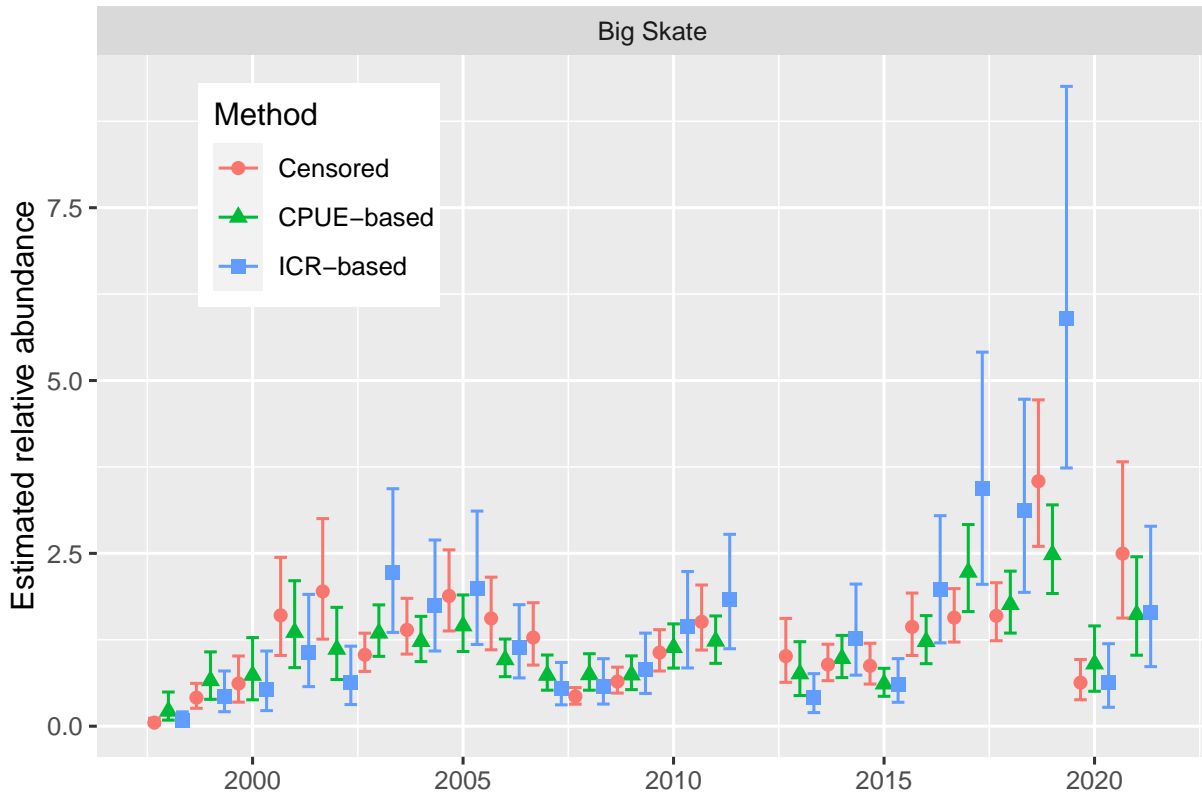


Figure S21: Plots showing the estimated relative abundances of the one species for which $\hat{p}_i^* = 0.7$ was judged appropriate. Shown are the mean and 95% confidence intervals from three model-based methods. The points and intervals have been ‘jittered’ along the x-axis for readability. The censored and CPUE methods’ estimates closely agree whereas the ICR method’s estimates differ greatly.

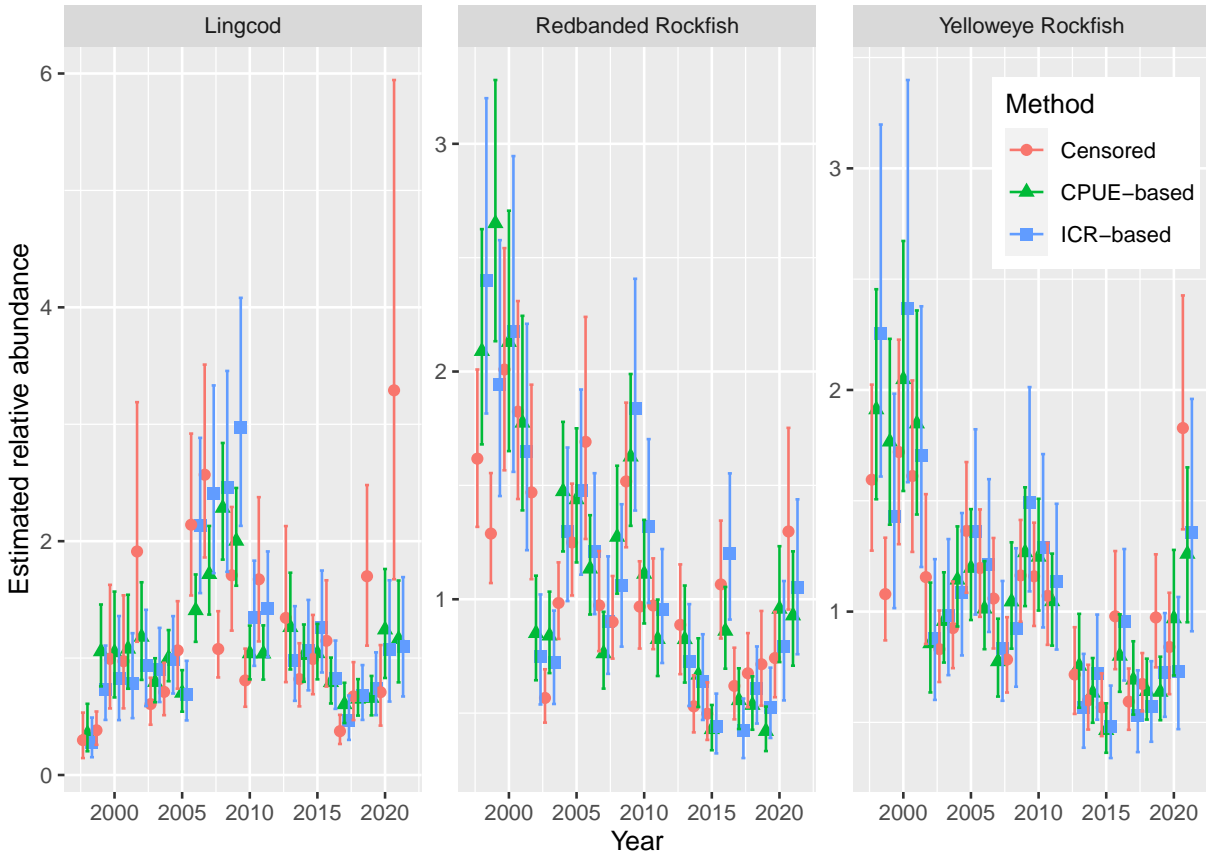


Figure S22: Plots showing the estimated relative abundances of the three species for which $\hat{p}_i^* = 0.85$ was judged appropriate. Shown are the mean and 95% confidence intervals from three model-based methods. The points and intervals have been ‘jittered’ along the x-axis for readability. The ICR and CPUE methods closely agree for both species. Conversely the censored method’s estimates differ greatly for Lingcod.

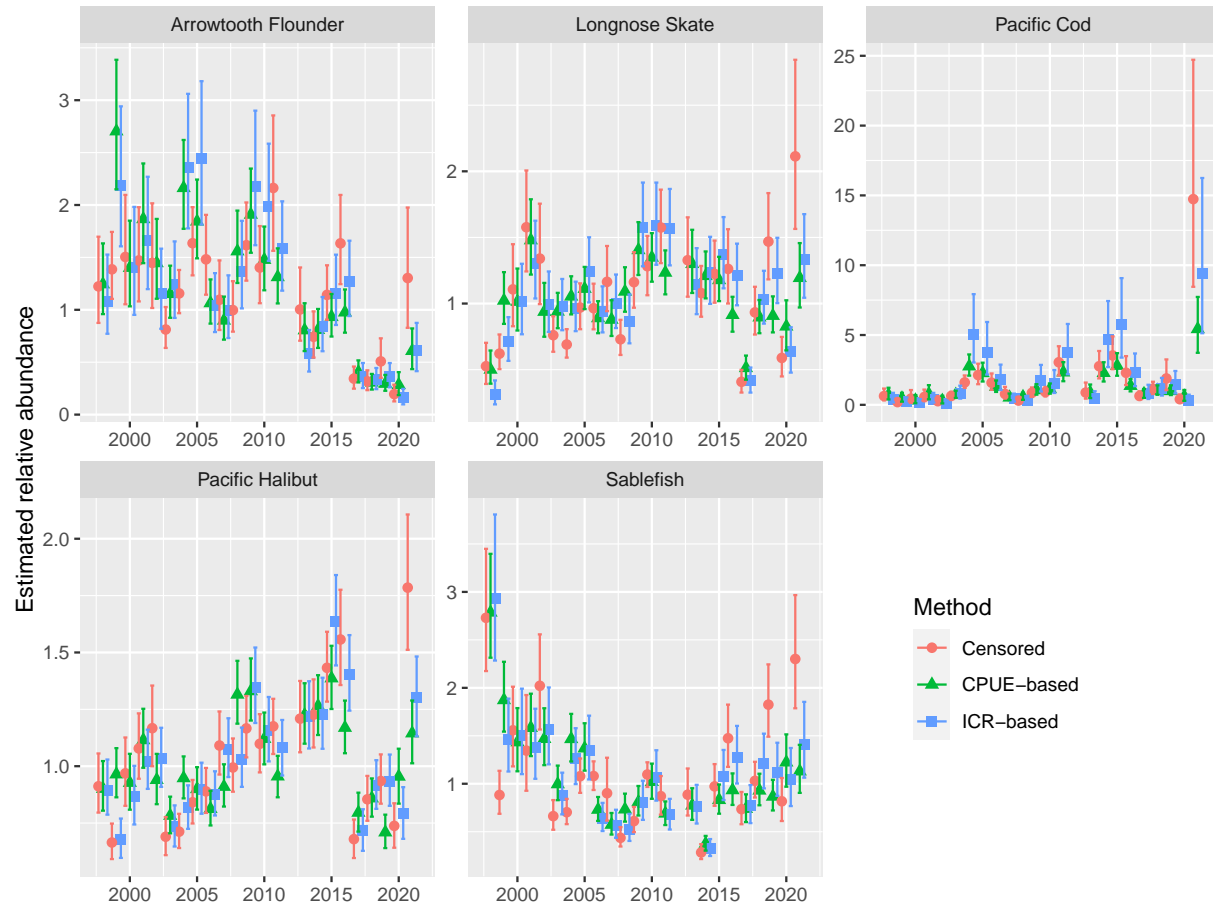


Figure S23: Plots showing the estimated relative abundances of the five species for which $\hat{p}_i^* = 0.95$ was judged appropriate. Shown are the mean and 95% confidence intervals from three model-based methods. The points and intervals have been ‘jittered’ along the x-axis for readability. The ICR and CPUE methods closely agree for all species. Conversely the censored method’s estimates differ greatly, most notably for Sablefish and Arrowtooth Flounder.

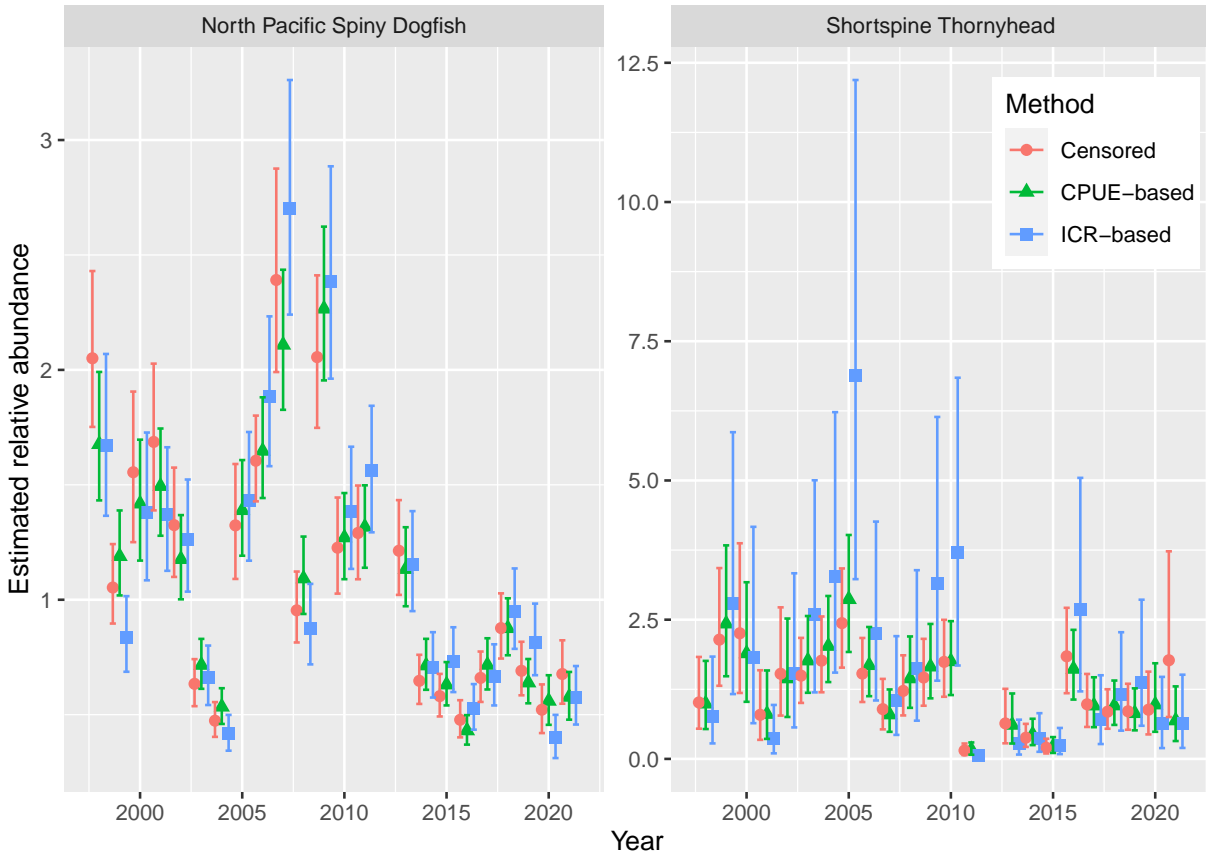


Figure S24: Plots showing the estimated relative abundances of the two species for which $\hat{p}_i^* = 1$ was judged appropriate. Shown are the mean and 95% confidence intervals from three model-based methods. The points and intervals have been ‘jittered’ along the x-axis for readability. The CPUE and censored methods’ estimates closely agree for both species. Conversely the ICR method’s estimates differ greatly for Shortspine Thornyhead.

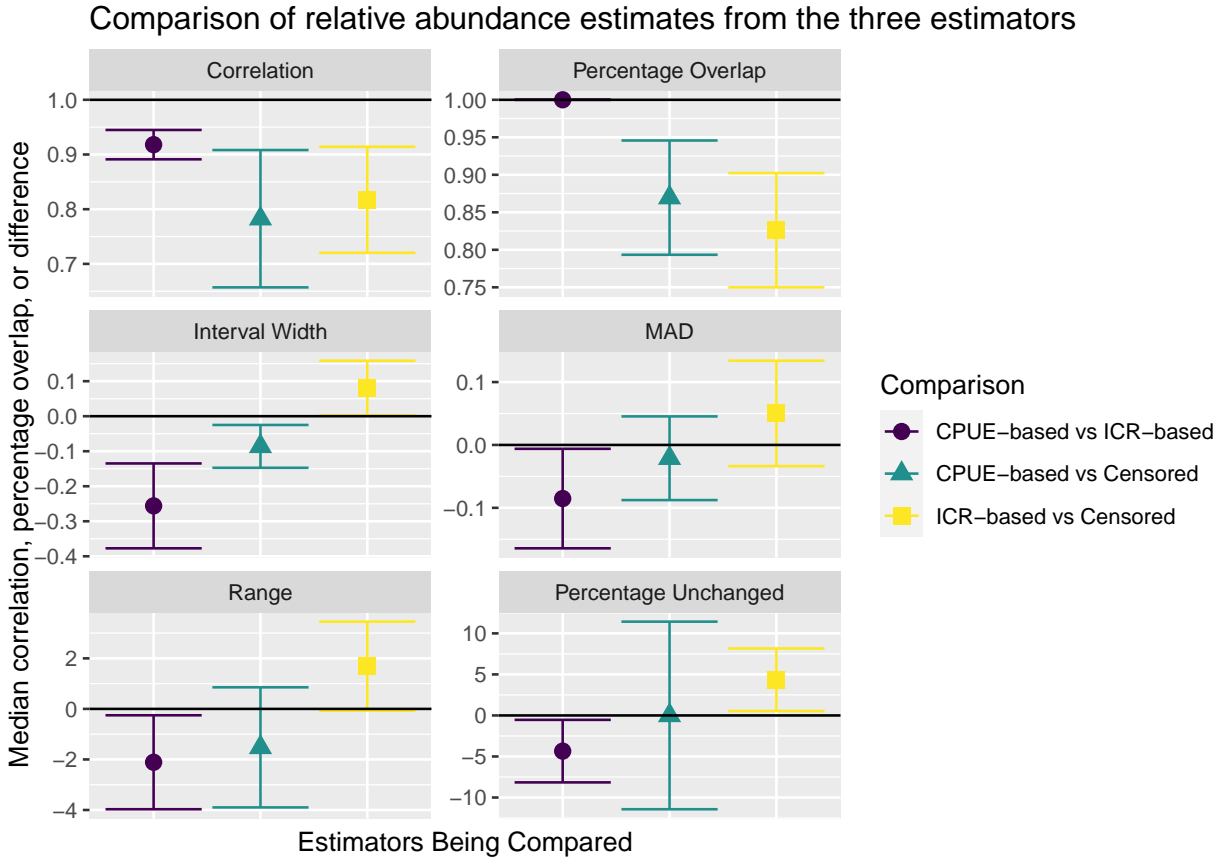


Figure S25: Plots showing empirical comparisons of the relative abundance estimates from the three model-based indices averaged across all 11 species. Confidence intervals are computed assuming the 11 species are a representative sample of all species. Shown are: Spearman's rank correlations between the intervals, percentage overlap between the competing 95% confidence intervals, differences in the widths of the 95% confidence intervals, differences in the median absolute deviations about the median (MAD) values, differences in the range of abundance estimates, and differences in the percentage of 95% confidence intervals covering 1 (i.e. abundance is not significantly different from the mean abundance). The censored method's estimates are least correlated with and its intervals overlap least with those from the other two methods. The CPUE intervals are narrowest and the ICR intervals are widest. The CPUE estimates of relative abundance vary the least (smallest range and smallest MAD).

S.7 R code for fitting the censored method

We provide R code (below and in `cjfas-2022-0159supplb.zip`) to simulate censored longline count data of a single species Y_i from the following Poisson regression model:

$$\log \lambda_i = a + bx_i$$

$$N_i \sim \text{Poisson}(\lambda_i)$$

$$Y_i = \begin{cases} N_i & \text{for } i = 1, \dots, 100; \text{ i.e. data are uncensored} \\ \min(N_i, 3) & \text{for } i = 101, \dots, 200; \text{ i.e. data are right-censored at 3} \\ \min(N_i, 2) & \text{for } i = 201, \dots, 300; \text{ i.e. data are right-censored at 2} \end{cases}$$

where x_i are values of an observed covariate (for $i = 1, \dots, 300$), a and b are parameters relating $\log \lambda_i$ to x_i , λ_i is the mean number of attracted fish, and the Y_i are the observations. The first 100 observations are uncensored and collected by trawling (i.e. $Y_i = N_i$ for $i = 1, \dots, 100$), the second 100 observations are right-censored at 3 (i.e. collected by longline fishing gear with only 3 hooks deployed), and the final 100 observations are right-censored at 2 (i.e. collected with longline fishing gear with only 2 hooks deployed). Note that $p^* = 1$ and identical integrated response functions across the three fishing gears are assumed for this simulated example.

The goal is then to account for the different fishing gear and hence levels of censorship to infer the parameters a and b . In R, we simulate the data as follows:

```

# simulate.R - simulate some data
set.seed(29072021)
n <- 300
a <- 0
b <- 1
x <- rnorm(n, sd = 0.3)      # random observed covariates
lambda <- exp(a + b*x)
N <- rpois(n, lambda = lambda)
# Censor the data:
y <- c(N[1:100],
      ifelse(N[101:200] >= 3, 3, N[101:200]),
      ifelse(N[201:300] >= 2, 2, N[201:300]))
rm(N) # To guarantee that the uncensored data not observed
      later

```

Next, we show how R-INLA, sdmTMB/TMB, and brms/stan can be used to correctly model the censored fishing data.

S.7.1 R-INLA code

To fit the censored model in R-INLA, we run the following R code:

```

# r-inla.R - fit the censored method using R-INLA
source("simulate.R")
library(INLA)
# version 22.04.16 or higher.

# Define the lower bounds for the 300 censorship intervals.
# Note that for values of low set to Inf, INLA assumes the
# data are perfectly observed (no censorship).
low <- c(rep(Inf, 100),
        ifelse(y[101:200] == 3, 3, Inf),
        ifelse(y[201:300] == 2, 2, Inf))

# Define the upper bounds, which are always infinite.
high <- rep(Inf, n)

# What do this data look like?
cbind(low, y, high)

# Fit the model
INLA_mod <- inla(inla.mdata(cbind(y, low, high)) ~ 1 + x,

```

```

        family = "cenpoisson2",
        data = data.frame(y, low, high, x))

# View the results
summary(INLA_mod)

# If interval-censoring is desired change the upper bound
# Suppose the upper bound is 10
high <- c(y[1:100],
          ifelse(y[101:200] == 3, 10, y[101:200]),
          ifelse(y[201:300] == 2, 10, y[201:300]))

# Refit the model
INLA_mod <- inla(inla.mdata(cbind(y, low, high)) ~ 1 + x,
                family = "cenpoisson2",
                data = data.frame(y, low, high, x))

# View the results
summary(INLA_mod)
# Doesn't change much

```

S.7.2 TMB code

To fit the censored model in sdmTMB and hence TMB, we run the following R code:

```

# tmb.R - fit the censored method using TMB
source("simulate.R")
library(sdmTMB)
# Install from https://github.com/pbs-assess/sdmTMB

# Define the lower and upper bounds. Note that equality
# in bounds is equivalent to no censoring
low <- y

high <- c(y[1:100],
          ifelse(y[101:200] == 3, NA, y[101:200]),
          ifelse(y[201:300] == 2, NA, y[201:300]))

# Fit the model (note that an arbitrary mesh needs to be made)
# The mesh is not used in the computation with spatial='off'
sdmTMB_mod <- sdmTMB(
  data = data.frame(x=x, y=y),

```

```

formula = y ~ x,
mesh = make_mesh(data.frame(x=1:300,y=1:300),
                    xy_cols = c('x','y'),
                    n_knots = 3),
family = censored_poisson(link = "log"),
experimental = list(lwr = low, upr = high),
spatial = 'off',
spatiotemporal = 'off'
)

# View the results
summary(sdmTMB_mod)

# If interval-censoring is desired change the upper bound
high <- c(y[1:100],
          ifelse(y[101:200] == 3, 10, y[101:200]),
          ifelse(y[201:300] == 2, 10, y[201:300]))

# Refit the model
sdmTMB_mod <- sdmTMB(
  data = data.frame(x=x,y=y),
  formula = y ~ x,
  mesh = make_mesh(data.frame(x=1:300,y=1:300),
                    xy_cols = c('x','y'),
                    n_knots = 3),
  family = censored_poisson(link = "log"),
  experimental = list(lwr = low, upr = high),
  spatial = 'off',
  spatiotemporal = 'off'
)

# View the results
summary(sdmTMB_mod)
# Doesn't change much

```

S.7.3 Stan code via the BRMS package

To fit the censored model in brms and hence stan, we run the following R code:

```

# stan.R - fit the censored method using Stan
source("simulate.R")
library(rstan)

```



```

library(brms)

# Need to create a variable with values 'none', 'left',
# 'right', or 'interval'

# We will use 'right' and 'none' here
y_BRMS <- c(y[1:100],
            ifelse(y[101:200] == 3, 2, y[101:200]),
            ifelse(y[201:300] == 2, 1, y[201:300]))
# NOTICE THAT FOR BRMS, WE NEED TO SUBTRACT 1 FROM THE
# RESPONSE AS THE CENSORSHIP INTERVAL IS CONSIDERED
# OPEN ON THE LEFT AND DATA HAVE TO BE INTEGERS

censored_txt <- rep('none', 300)
censored_txt[101:200] <- ifelse(y[101:200] == 3, 'right',
                                'none')
censored_txt[201:300] <- ifelse(y[201:300] == 2, 'right',
                                'none')

brms_mod <-
  brm(y_BRMS | cens(censored_txt) ~ x,
      data = data.frame(y_BRMS, x, censored_txt),
      family='Poisson')
summary(brms_mod)
# Compare the estimated intercept and linear value with the
# known values of a
# and b.

# To make an interval censored value, simply define an upper
# bound. We use 10 here for demonstrative purposes.
censored_txt2 <- rep('none', 300)
censored_txt2[101:200] <- ifelse(y[101:200] == 3, 'interval',
                                'none')
censored_txt2[201:300] <- ifelse(y[201:300] == 2, 'interval',
                                'none')

# Define right interval endpoint as 10
U_BRMS2 <- y
U_BRMS2[censored_txt2 == 'interval'] <- 10

brms_mod2 <-

```

```
brm(y_BRMS | cens(censored_txt2,U_BRMS2) ~ x,  
     data = data.frame(y_BRMS, x, censored_txt2, U_BRMS2),  
     family='Poisson')  
summary(brms_mod2)  
# Notice that the results are almost identical as before
```

References

- Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Am. Stat. Assoc.*, **99**(467), 673–686.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc., Ser. B*, **73**(1), 3–36.
- Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *J. Am. Stat. Assoc.*, **111**, 1548–1575.