Economists counting fish: modeling profit-maximizing behavior to improve stock assessments

Y. Allen Chen[a*], Cole C. Monnahan[b], Owen S. Hamel[a], Alan C. Haynie[c]

[a] Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, NOAA Fisheries, 2725 Montlake Blvd. E., Seattle, WA 98112

[b] Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, NOAA Fisheries, 7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115

[c] International Council for the Exploration of the Sea, H. C. Andersens Boulevard 44-46, 1553 Copenhagen V, Denmark

[*] Corresponding author, allen.chen@noaa.gov

Abstract

Stock assessments are an integral part of contemporary fisheries management, estimating the health of wild fish populations and ensuring sustainable marine resource management. Fishery-independent data (e.g., scientific surveys) usually provide unbiased information not influenced by fisher targeting, although fishery-dependent data can be more widely available. One reason the use of fishery-dependent data in assessments is treated with care is that the data do not account for the spatial site selection and economic behavior of fishers. By understanding and explicitly modeling how fishers make tradeoffs, economic models can correct for selection if fishers systematically choose areas with greater expected catches or revenues. We develop and test an approach to correct abundance indices taking into account the sampling process of fishers in a simulation framework. Corrected indices can supplement fishery-independent surveys and improve the accuracy of spawning biomass estimates in some scenarios, but are unreliable in others. We demonstrate the potential for economic models to serve as a tool to improve certain stock assessments and the management based upon them.

Keywords

Selection bias, economics, fishery-dependent data, CPUE index standardization

1. Introduction

Stock assessments estimate the status and sustainable levels of catch in exploited stocks and are a vital tool for sustainable marine resource management (Hilborn & Walters 1992). When stock assessments use fishery-dependent observations to infer trends in abundance, it is important to account for effects other than true changes in the population trend, such as gear or spatial changes, or the site selection and incentives of fishers (Eales & Wilen 1986), through a process known as index standardization (e.g., Maunder & Punt 2004; Walters 2003). Fishers do not sample randomly across fish habitat, but instead choose areas to target different species, densities, and sizes of fish. Standardizing fishery-dependent indices based on catch rates to correct for these factors is an ongoing challenge to producing accurate and reliable stock assessments and the associated advice used by fisheries managers. Observed catches are influenced by the incentives and catch patterns of individual fishers.

There have been a number of studies that examine standardization of catch per unit effort (CPUE) indices, for example by allowing correlation between observations (Bishop et al. 2004), taking into account changes in fishing power (Allen & Punsly 1984), or incorporating targeting strategies that vary spatially (Hoyle & Okamoto 2013). More recently, geostatistical approaches have shown improved properties and this class of models has proliferated (Thorson et al. 2015, 2020), including work integrating data from both commercial fisheries and fishery-independent surveys (Rufener et al. 2021, Rufener et al. 2023). However, little attention has been devoted to index standardization methods that use economic drivers, which offer a mechanistic understanding of the sampling process, resulting from the decisions of fishers, that generates catch observations and fishery-dependent data. This information could be used to correct for bias in the sample of data collected from fishers.

Economic models that explicitly characterize how fishers make tradeoffs can correct for selection that occurs if, for example, fishers choose to systematically fish at areas with greater expected catches, as we would expect. In this paper we examine a form of selection bias that results from fisher targeting, due to site selection across space. However, more broadly we are referring to bias that occurs due to the selection of data, such that the sample observed is conditioned on factors such as the preferences and choices of fishers, and results in a sample that is potentially different from the population. The study of self-selection in economics is well-developed and robust, following the seminal work of Heckman (1979). An example of correcting for selection bias in a polychotomous choice setting is provided by Dahl (2002), examining how migrants choose where to live based on their expected wages. In a fisheries extension, Chen et al. (2023) modelled how fishers choose areas in the Bering Sea pollock fishery, suggesting a method that estimates average catches at areas, corrected for selection bias.

Chen et al. (2023) were primarily interested in understanding the spatial tradeoffs and preferences of fishers, which allows policymakers and researchers to evaluate policies such as spatial management strategies. An intermediate step necessary for developing that inference was to obtain unbiased estimates of fisher catches across space. In order to estimate the marginal utility from catch, you must have estimates of expected catch, including at locations where fishing did not occur (in order to compare those alternatives). If these data are biased, this results in biased estimates of the marginal utility of catch. We hypothesize that the same process to correct for fisher-induced selection could potentially be used to create corrected fishery-dependent indices of abundance that better reflect stock abundance.

Our methods use a structural model of fisher behavior, where fishers optimize over a given set of preferences, are statistically testable for selection, and can incorporate any number of relevant factors or covariates into the fisher’s profit function. We calibrate a model of a managed fishery where fishers are allowed to choose where to fish across discrete locations, with private information about the stochastic nature of the catch at each location before they choose where to fish. Conversely, we model a manager that observes catches, and fisher characteristics, but does not know whether observed catches are larger or smaller than average. Therefore, not only do fishers have information about expected catches the manager does not, but the catches the manager observes are driven by the choices and selection of fishers (and the fisher's private information).

We believe this approach provides an opportunity to improve existing index standardization methods by explicitly incorporating structural models of fisher behavior (and not just economic covariates), and provide a mechanistic understanding of their site selection process, which to our knowledge is novel in fisheries literature. The specific assessment modeling framework we examine in this paper is Stock Synthesis (SS3; Methot & Wetzel 2013), where relative abundance indices can be constructed from fishery-dependent data and are treated separately from age- or length-compositions, although the latter data is also used. Broadly, the form of our bias occurs due to the selection of data such that the observed sample is conditioned on factors such as fisher preferences. We detail the model and test it in a simulation framework using scenarios that vary in how fish are distributed spatially and how fisher decisions are made. We find that site selection biases observed catches and estimates of interannual abundance indices as long as fishers are not selecting locations at random, and our methods can create unbiased indices under some circumstances, but not when there is a spatial trend in abundance.

We focus on a single, widely-used stock assessment model framework, SS3, and assume the fisher preferences result in biases in catch data, though not in compositional data. Therefore, we focus primarily on the construction of indices of catch per unit effort. Age and size compositions of catch are rarely expected to match population compositions, and this difference is typically addressed through size and/or age selectivity estimation. Our effort focuses on fishers who seek to maximize catch, and the estimation of catch per unit effort indices, and not size-selectivity, although we examine possible ramifications of size-dependent pricing and targeting in Appendix A.

Following development and application of the simulation framework, we apply the economic model to the simulated data to recover corrected fishery-dependent indices of abundance. We find that corrected indices can better approximate the true interannual abundance trend, even when site selection preferences remain constant across years. Finally, we supply the corrected indices to a stock assessment to quantify the improvement to management advice.

2. Methods

To investigate the potential of using fishery-dependent indices of abundance corrected for selection bias, we conduct simulations using a biological operating model (OM) (Figure 1). This operating model generates age-structured population dynamics for each simulation replicate. Given a population trajectory we then simulate annual spatial sampling patterns with a fishing model that encodes fisher behavior. The fishing model and simulated sampling patterns generate fishery-dependent data, which can then be corrected with an econometric model of fisher behavior that recognizes that fishers trade off the expected catch and revenue in an area with the costs of fishing in different areas. The corrected data are then used to construct indices of relative abundance, which are included in an estimating model to evaluate if they improve stock assessment estimates.

We compare assessments with the addition of uncorrected versus corrected indices of abundance, as well as indices from a fishery constructed with random rather than preferential fishing area selection (emulating an additional annual fishery-independent survey). Note that we do not complete the loop in the simulation (Figure 1), as the estimating model is not used to dictate catch limits or fishing mortality rates in the operating model, but rather the fishing mortality rate *F* is fixed for each year within the operating model (as well as catches at length and age). This still allows us to evaluate how correcting the fishery-dependent abundance index improves the assessment, without conducting a full closed-loop simulation analysis.

2.1 The "true" biological operating model

We use the R package ss3sim (Anderson et al. 2014) to configure a single-sex, age-structured fishery based on a "cod-like species", calibrated using North Sea cod. This package allows us to simulate a "true" abundance trend in its operating model, and then test the effect of different fishery sampling patterns in the estimating model using the assessment software SS3. The cod-like species is a default model included in the ss3sim package, and is a similar model to those investigated in other fisheries applications (e.g., Johnson et al. 2015, Ono et al. 2015). Key life-history and modeling parameter calibrations are described in Table 1, and we refer the reader to previous studies cited above for more information. The modified ss3sim package, including the fishery model, econometric correction, and code to reproduce the results in this paper, is available on the corresponding author's Github.[1]

We explore a linearly increasing fishing mortality rate starting in year 62 and ramping up to F_{MSY} (~0.17; Figure 2). Natural mortality does not vary with time nor age and is fixed at a constant 0.2. Fishery and survey size selectivity is specified as the flexible double normal curve

[1] https://github.com/allen-chen-noaa-gov/ss3simfork. The fishery model code also uses routines from the FishSET package (Pfeiffer et al. 2024).

recommended in SS3 applications (Privitera-Johnson et al. 2022), but are calibrated to be logistic in our case (see the parameters in Table 1). Growth uses the von Bertalanffy equation, while spawner-recruitment is specified with Beverton-Holt, both as parameterized in Table 1. Finally, our process error takes the form of independent lognormal random deviations around median recruitment. The choices in the operating model produce the true abundance trend and total catches, from which simulated data are generated (described below) and used in the estimation and fishing model.

2.2 Simulated fishing model

Notably, the ss3sim framework is not spatially explicit. To characterize a fishing model that allows fishers to choose different areas, we must distribute total abundance in year *y* available to the fishery A_y^f across a defined spatial grid. We assign each grid cell *j* a relative abundance, parameterized by $\tilde{\beta}_{jy}$ whose sum is equal to a scalar with constant value over all years, and we multiply these $\tilde{\beta}_{jy}$ by a scaling factor equal to "true abundance" A_y^t from the operating model for that year, where t denotes the truth, relative to average abundance μ_A, where A_y^t is assumed to be exploitable true biomass and $\mu_A = \frac{\sum_y^Y A_y^t}{Y}$. Thus, the only factor changing the observed relative abundance β_{jy} across years defining catches is a factor scaling operating model abundance A_y^t in year *y*, from all *J* areas:

(1a)
$$A_y^f = \sum_{j=1}^{J} \left(\frac{A_y^t}{\mu_A}\right) \tilde{\beta}_{jy}$$

(1b)
$$\beta_{jy} = \left(\frac{A_y^t}{\mu_A}\right) \tilde{\beta}_{jy}$$

Notably, the sum of relative abundances $\sum_{j=1}^{J} \tilde{\beta}_{jy}$ is always constant across *y*, such that the only change across years in the observed β_{jy} we use to define catches, comes from operating model abundance A_y^t and $\tilde{\beta}_{jy}$ (equation 1b), and not the spatial distribution of abundance. Intuitively, if there are more fish in the sea, catches should be bigger. Also note that spatially distributed relative abundances are not by size or age.

Relative abundance among areas might be static over time, or we can calibrate $\tilde{\beta}_{jy}$ to change across *y* (for example representing fish movement), and in this paper we examine both scenarios. The choice not to sum $\tilde{\beta}_{jy}$ to unity is one of convenience, that allows us more flexibility to parameterize our model such that fishers are making meaningful choice across space.

After biomass from the operating model is distributed spatially according to observed β_{jy}, then in each year, depending on the calibration of fisher behavior, fishing occurs across the grid resulting in some observed "sampling" pattern, and the resulting catches are used to construct an estimate of abundance in that year. Thus the fishing model simulates artificial catch data comprising *N* independent haul occasions each year given the true spatial abundance for a single year. Similar to fisheries observers that record catches on fishing vessels, we then assume data for *Y* total years that have observer coverage, in each replicate of the simulation. The population is simulated for 100 years total, allowing burn-in, but only the last *Y* years have fishing activity and observer coverage.

We take fishing mortality from the operating model as exogenous, simulating haul-level data until total catches equal what is dictated by the operating model. Therefore, the fishing mortality from the biological operating model dictates total catches, while the fishing model only simulates how fishers would make spatial decisions and potentially choose areas with greater realized CPUE. This is akin to a fishery where, for example, management sets a total allowable catch (TAC), and the TAC is always fulfilled. Note that fishery size selectivity is assumed to be independent of catch area, and we do not estimate an impact on size/age structure. Possible ramifications of this assumption are examined in the Discussion.

2.2.1 Modeling fishing behavior

For each simulated haul, we first specify the catch Y_{ij} at area *j* on the *i*th haul occasion as a linear combination of covariates X_{ij} with dimension *1xn* where *n* equals the number of vessel-specific covariates, scaled by area-specific coefficients β_{jy} with dimension *nx1*, an area-specific constant c_j, plus some stochastic variation u_{ij} (Chen et al. 2023), where $X_{ij}'\beta_{jy}$ represents vector multiplication. Vessel-specific covariates could include vessel characteristics for example. Therefore, while each location might have larger or smaller catches denoted by c_j, different vessels might also catch more or less fish on average, depending on their vessel characteristics such as size or gear. We allow part of this variation u_{ij}^f to be known by the fisher, while the remainder u_{ij}^s is unobservable: the specific form of our selection bias results from private information known to the fisher that is unobserved by the analyst.

(2) $$Y_{ij} = c_j + X_{ij}'\beta_{jy} + u_{ij}$$

$$= c_j + X_{ij}{}'\beta_{jy} + u_{ij}^f + u_{ij}^s$$

We assume the errors are non-correlated and normally distributed (the specified parameters of these errors are listed in Table 2). The joint likelihood can be parameterized for different distributional assumptions, however, an assumption must be made.

Our fishing model assumes a fisher starting at area k chooses among M discrete areas on haul i, for $i = 1 \ldots N$. Equation 3 describes this choice problem as the standard random utility model, such that the fisher moves to the j^{th} area as long as the utility U_{ijk} at that area is greater than at all other areas: by maximizing utility fishers will choose to fish where they derive the greatest net benefit, and this utility describes the objective function that fishers optimize over. In the case where all utilities are equal the fisher would randomly choose a location. We assume all areas are potentially accessible to the fisher from the current area, where M is the total number of alternatives. However, an area may not practicably be accessible depending on its distance, or a fisher's disutility for travel, et cetera.

(3) $$\text{Choose area } j \text{ on } i^{th} \text{ haul } \textit{if and only if } U_j > U_m \; \forall \, m \neq j$$

Part of the fisher's utility, denoted V_{ijk}, is allowed to depend on any number of individual- or area-specific variables Z_{ijk} known to the analyst and is therefore the deterministic portion of the fisher's utility. Notably, the fisher's utility depends on their expectations of catches at an area, $E_u[Y_{ij}]$, where α is the marginal utility from catch, their starting area k, and γ a vector of parameters that scales area- and individual-specific variables Z_{ijk}, where α and γ are estimated.

For the fisher's expectation $E_u[Y_{ij}]$, the fisher's information set includes the realization of u_{ij}^f, as well as average catches $X_{ij}{}'\beta_{jy}$. Other covariates such as lagged catches, capturing historical experience, could also be included in X_{ij} (and is generally common in the fisheries economics literature). Therefore,

$$
\begin{aligned}
U_{ijk} &= V_{ijk} + \varepsilon_{ij} \\
&= \alpha E_u[Y_{ij}] + Z_{ijk}{}'\gamma + \varepsilon_{ij}
\end{aligned} \tag{4}
$$

where ε_{ij} varies in an unobservable way and is assumed to be independently and identically distributed generalized extreme value type I (Gumbel).

2.2.2 Spatial abundance and fishing behavior scenarios

This section (2.2.2) outlines the specific approaches we model in this manuscript. Because different fishery, spatial, or fleet characteristics might result in different behavior and observed catches, we investigate when selection creates more bias in observed catches, and under which scenarios our proposed correction works more or less effectively. For example, an area's distance from the vessel can be important, depending on the relative abundance at more distant areas as well as a fisher's disutility for travel. By parameterizing these factors and testing different values across our scenarios, we can control fisher spatial coverage, which then affects the variation in vessel types and trip types we see in our data, and correspondingly the efficacy of a statistical correction for selection bias in catch data.

We investigate what happens if there is good fisher coverage (Scenario 1), if some areas are not sampled by fishers (Scenarios 2 and 3), and when the abundance at areas follows different temporal patterns (Scenario 4). Scenario 1 represents a base case that is likely farthest from reality, but provides the most the most data for our methods. Scenario 2 adds more nuance and detail to the model of fisher behavior and better approximates the fact that for most exploited fisheries, the spatial extent of fishing grounds do not overlap the entire spatial domain of the species. Fishers may not visit areas where fishing is poor or areas of high abundance that are too distant. Scenarios 3 and 4 then extend the biological model to include untrawlable habitat (for example, certain rockfish species), or temporal movement, that might occur due to changing marine temperatures, for example. Common steps between all four scenarios in how we model fisher behavior and produce different realized catches are outlined below.

1. The catch data Y_{ij} are generated for each area on a *3x3* or *2x4* grid from the data-generating process in equation (2), depending on the scenario, where the elements of u_{ij} are stochastic, c_j and X_{ij} are fixed. We choose these grid sizes as a practical matter, to simplify the model with the goal of comparing asymptotic behavior. Catches at areas are scaled over years to match the exogenous abundance trend from the operating model; i.e. if abundance decreases, the deterministic part of catch (β_{jy}) must decrease to match. We ignore size and age structure because selectivity is already implicitly accounted for in the distribution of exploitable biomass A_y^t from the operating model. The variable X_{ij} can be thought of as a vessel-specific covariate that affects catchability (e.g. vessel power). Note that any number of covariates affecting catch could be included, depending on the data the analyst has on hand. We only generate one covariate that affects catch. We assume

haul duration here is constant for all hauls, normalizing catches to catch per unit effort. This simplification is not necessary for the statistical correction, but allows us to forgo modeling haul duration, which is used to standardize CPUE.

2. Fishers start at some given area and make their decision - representing a single choice occasion - according to the information available to them. The starting area may represent a port, the area of the last haul, or may potentially abstract from how the fisher arrived at that area altogether, depending on the scenario specification. We allow a fisher to choose a single area according to the selection criteria in equation (3), given stochastic ε_{ij}, fixed γ and Z_{ijk}, where the fisher knows X_{ij}, β_{jy}, and the realization of u_{ij}^{f}, but not u_{ij}^{s}. Notably, this implies the fisher's expectation of catch is a part of the selection criteria, allowing them to choose sites with greater relative abundance, if they desire. The fisher then moves to the site with the largest of its expected utility values. The vessel fishes at this new site and the realized catch (including both u_{ij}^{f} & u_{ij}^{s}) at the chosen area is included in the data available in the assessment.

3. Steps 1 and 2 are repeated for *N* independent haul occasions until the sum of catches from all hauls equals the total catches from the operating model. In practice, the sum of catches from N independent hauls will exceed total catch in the OM slightly; fishing ceases when catches from the last haul simulated meets or exceeds the total catch. This comprises the catch data for a single year *y*, and this process is repeated for all *Y* years. This provides a simulated yearly time series of hauls and haul-level data for all years in a single replicate, based on economic factors driving fishery behavior and the spatial distribution of

abundance. Note that we simulate hauls, not individual fishers. For each haul, realizations of X_{ij} and Z_{ijk} are drawn. To the extent X_{ij} and Z_{ijk} represent vessel-specific covariates, there are a minimum number of vessels equal to the number of unique permutations of X_{ij} and Z_{ijk}. However, different vessels could also have similar vessel characteristics (two vessels may both be 100 meters long). We simulate individual hauls until catches bind, but the number of vessels is not directly modeled.

We compare the four scenarios describing different spatial fishing patterns and spatial abundances in Table 2, which represent conditions where the statistical correction is hypothesized to be more or less successful. Column I lists the name of each scenario. Column II broadly describes the scenario and what we expect will result from that specific fisher behavior and abundance pattern. Column III describes fisher behavior and spatial coverage resulting from that sampling. Finally Column IV describes the spatial abundance pattern of the fishery.

Our four scenarios characterize different spatial fishing patterns and spatial abundances, which enables us to examine when a statistical correction for selection bias may perform better or worse. Figure 3 demonstrates an example where fishing occurred in a sample replicate-year for Baseline Scenario 1 (left panel) and Partial Fisher Coverage Scenario 2 (right panel). In both of these scenarios, the spatial distribution of abundance is static across all years, shaded in blue. In Scenario 1 however, each “trip” (an example trip is illustrated in red) consists of only a single haul, and a vessel’s starting area is randomized.

This scenario might correspond to a world where fishers have preferred starting areas, based on their past experience, and the fisher steams to that area to begin their fishing trip, but this data-generating process abstracts from how fishers might make multiple hauls in a trip. In this scenario there are a large number of observations in each area (labeled in green). These are catch observations as if there were a hypothetical observer on board; note length observations are not included in the fishery model but rather passed directly from the operating model because the population length structure does not vary in space.

As an alternative, in Scenario 2, vessels always start at Area 1. They then take multiple hauls in a trip, where the red arrows illustrate five choice occasions in a single hypothetical trip. When all vessels start in Area 1, they tend to migrate towards the areas with greater abundance. Then the probability of moving to an area with low abundance is relatively smaller, although fishers may still select areas with lower abundance as travelling larger distances is costly, and expected catches at a location vary stochastically, so fishers may not stay in an area. Note that areas with low abundance in the southwest have fewer observations. For catches to be identifiable there needs to be at least as many observations as the number of location-specific parameters which equals the number of covariates in X_{ij} and the number of parameters in the polynomial approximation. However, in practice, a larger sample is required for good performance because the correction relies on a good amount of variation in the probabilities used in the correction functions.

In Scenario 3, there is a buffer in the form of habitat (e.g. reefs or rocks) that prevents fishing (Figure 4). We investigate the effect of not being able to estimate abundance in some areas, including what happens if the westward area contains most of the abundance in a year. Fishers start in Area 1, and make multiple hauls in a trip. Abundance across space is not static across years, but random across space as described in Column IV of Scenario 3, where each panel is one realization of the random distribution (in one year). In some years the unfished area might have high abundance, while in other years there might be low abundance in the unfished area, but generally there are not enough observations to estimate catches at the most westward areas.

Finally in Scenario 4 there is an eastward trend in spatial abundance. The panels in Figure 5 correspond to years 1 (upper left), 5 (upper right), 9 (lower left), and 13 (lower right) respectively, for an example replicate. Fishers start in Area 1, and make multiple hauls in a trip. As each year passes, the spatial distribution of the population in terms of abundance shifts eastward, while fishing vessels still do not frequent the sites furthest away.

For each scenario the sum of catches from all hauls equals the total catches from the operating model, which are then passed to an econometric model.

2.3 Econometric model of fisher behavior and corrected abundance indices

The role of the econometric model is to recover the true spatial abundance from the observed catches (i.e., decisions) and limited knowledge about fisher behavior. We investigate three different constructions of an abundance index: an uncorrected index, a corrected index, and a randomly sampled index. While the first two differ as to whether a statistical correction is applied, the randomly sampled index does not use the selection criteria in equation (3), but rather randomly chooses a site independent of its characteristics for each haul, including areas inaccessible to the fishery in scenarios 3 and 4. This is meant to mimic the effect of including an additional (more frequent and ongoing) fishery-independent survey which provides an abundance index rather than an absolute abundance estimate.

The assumption that the error distribution is extreme value type I allows a convenient expression of the probability of choosing area *j*, where *m* denotes all other areas besides area *j*.

$$Prob\left(U_{ijk} > U_{ijm}, \forall\, m \neq j\right) = \frac{\exp\left(\frac{\alpha}{\sigma_{scale}}\left(c_j + X_{ij}{}'\beta_{jy} + u_{ij}^{f}\right) + \frac{1}{\sigma_{scale}}\left(Z_{ijk}{}'\gamma\right)\right)}{\sum_{m=1}^{m=M} \exp\left(\frac{\alpha}{\sigma_{scale}}\left(c_m + X_{im}'\beta_{my} + u_{im}^{f}\right) + \frac{1}{\sigma_{scale}}\left(Z_{imk}{}'\gamma\right)\right)} \quad (5)$$

This is the common conditional logit model (McFadden 1974). Note that parameters α and γ are only identified up to a scale parameter σ_{scale}; that is, we can identify the extent fishers value catch relative to distance α/γ, for example, but not their absolute values (McFadden 1974). Then, when we substitute in the fisher's expectation of catch, the probability of choosing the *j*th area is a function of the private information available to the fisher. For instance, when $\alpha > 0$ the probability of choosing a site increases with catch, the analyst who is creating an abundance index is more likely to observe an area when the fisher knows catches are larger at an area, and would more likely observe larger catches. Note α is a free parameter and agnostic to sign.

It follows that catch data, conditional on observation, will also incorporate this private information, which is correlated with the probability of choosing an area. This is because the fisher observes the realization of u_{ij}^{f} in equation 2 (but not u_{ij}^{s}). As a result, when fishers choose areas (and catches) that result in the greatest expected utility, and visit areas when they have private information fishing is good at the area. Specifically, note that u_{ij}^{f} enters directly into the probability of choosing an area in equation 5.

(6) $$E[Y_{ij}|observe\ Y_{ij}] = c_j + X_{ij}'\beta_{jy} + E\left[u_{ij}^{f}|observe\ Y_{ij}\right]$$

While the mean of the unconditional expectation of u_{ij}^{f} is zero, the conditional mean is not. If we interpret $X_{ij}'\beta_{jy}$ as an area-specific average, observed catches will deviate from the average at an area equal to a selection bias term $E\left[u_{ij}^{f}|observe\ Y_{ij}\right]$. The resulting fishery-dependent data are more likely to be characterized by observations with larger catches, relative to a randomized site sampling (i.e., a scientific survey or uninformed fisher). Scientific surveys usually follow some semi-deterministic sampling plan, but with a design to avoid statistical bias.

If an analyst could derive unbiased estimates of $\widehat{\beta_{Jy}}$, they could predict catches and create an abundance index using some weighted sum over areas, for example. However, in this case, if analysts regressed observed covariates X_{ij} on observed catches, they risk obtaining biased estimates of $\widehat{\beta_{Jy}}$, as the sample of observed catches is biased (equation 6). This would bias estimates of $\widehat{\beta_{Jy}}$ as well. A correction function approach (Dahl 2002) allows us to approximate the conditional error as a polynomial function. This allows consistent estimation of the remaining parameters in the catch equation. Specifically, the polynomial is a function of the probability of visiting an area and observing the catch (p_{ijk}), where $\boldsymbol{\beta}_{prob}$ is a vector of coefficients to be estimated that corresponds to each term of the polynomial. For example, a second-order polynomial for a fisher who moved to area *j* from area *k* might be written as $\eta\left(p_{ijk}, \boldsymbol{\beta}_{prob}\right) = c_j + \beta_{prob1_1} p_{ijk} + \beta_{prob2} p_{ijk}^2$, which is then included in the regression as additional covariates. Then, we estimate the catch function simultaneously with the fisher's choice problem, where we can simultaneously obtain the probabilities p_{ijk}.

$$(7) \quad l_{ij} = \left(\frac{2\pi^{-\frac{1}{2}}}{\sigma_{catch}} \exp\left[\frac{-\left(Y_{ij} - K\left(p_{ijk}\right)c_j - X_{ij}'\beta_{jy} - \left(1 - K\left(p_{ijk}\right)\right)\eta\left(p_{ijk}, \boldsymbol{\beta}_{prob}\right)\right)^2}{2\sigma_{catch}^2} \right] \right) * \left(\frac{\exp\left(\frac{\alpha}{\sigma_{scale}}\left(c_j + X_{ij}'\beta_{jy}\right) + \frac{1}{\sigma_{scale}}\left(Z_{ijk}'\gamma\right)\right)}{\sum_{m=1}^{m=M} \exp\left(\frac{\alpha}{\sigma_{scale}}\left(c_m + X_{im}'\beta_{my}\right) + \frac{1}{\sigma_{scale}}\left(Z_{imk}'\gamma\right)\right)} \right)$$

Note that the probabilities p_{ijk} are equal to equation (5) and are equal to the second term of the full likelihood (equation 7). Then, the first term of equation 7 is the likelihood of observing the catch, given the assumption of normality (σ_{catch}), the second term the probability of selection (equation 5), and the full likelihood is the probability of observing the catch data, conditional on the observed area being chosen. We can jointly test the significance of a correction function at the median values of the data (probabilities), where statistical significance implies the existence of selection bias.

The term $K\left(p_{ijk}\right)$ is a kernel function that shrinks as the probability of observation decreases. A kernel function allows us to estimate both levels in the catch equation as well as a constant in the polynomial correction, which would otherwise be perfectly collinear (the constant c_j appears in both the catch equation and the polynomial correction). The specific kernel function we use is equal to $1 - \exp\left(-\frac{p_{itjk}}{bw - p_{itjk}}\right)$, where *bw* is a bandwidth parameter chosen by the analyst that determines the shape of the function, but alternative forms are feasible under specific assumptions (derived in Andrews & Schafgans 1998). Intuitively, we expect the selection bias term to be non-zero when the probability of choosing that area given other observable factors is low, yet the area was chosen anyway, implying the existence of private information the analyst has not accounted for.

For each year *y*, the abundance index b_y is calculated as the area-weighted sum of abundance of areas with catch data, where $Area_j$ scales proportionally to the size of area *j*. As all our areas are the same size, $Area_j = \frac{1}{J}$. In addition, we estimate $\widehat{Y_j} = \widehat{\beta_{Jy}}$, which is the predicted catch estimated for a vessel characteristic of unity (recall we have a single vessel characteristic distributed $Uniform(1,5)$).

(8)
$$b_y = \sum_{j=1}^{J} \left(Area_j \hat{Y}_j\right)$$

Constructed estimates of abundance in all years are then used as an index in the estimating model (assessment).

2.4 Investigate performance of assessment with estimating model

Finally, we quantify the assessment impacts when using a corrected index in an estimating model that matches the structure and parameterization of the operating model (Table 1). This indicates when the use of a properly corrected index could improve stock assessment estimates, while an uncorrected index would bias results. We use simulations of the common assessment software SS3 through the R package ss3sim, to examine whether and when fishery data corrected for economic endogeneity can be valuable.

We assume quadrennial fishery-independent surveys with biomass index coefficients of variation (CVs) of 0.2 and 500 length composition samples. Length composition samples are collected every ten years starting in year 26, and every four years starting year 60. The fishery indices are annual from the start of the fishery in year 88. While fishing occurs starting in year 62, our "observer coverage" does not start until year 88. For each replicate, the uncorrected, corrected and randomly sampled (survey-like) indices are passed to the estimation model and spawning stock biomass (SSB) are compared to the truth from the operating model. Steps 2.1-2.4 are repeated for 200 replicates of independent process errors and stochastic data simulation.

3. Results

3.1 Correction model performance

Figure 6 below illustrates mean abundance indices from our stochastic data simulation from scenarios 1 and 2, plotted as differences from the true abundance. Unsurprisingly, when we randomly sample one of the *M* areas on each haul occasion, instead of following the selection criteria in equation (3), the resulting abundance index closely mirrors the true abundance over time. A randomly-sampled index can be thought of mimicking an additional fishery-independent survey under a simple random sampling design.

In contrast, an uncorrected index tends to overestimate abundance when mortality is increasing. This is consistent with observed hyperstable patterns seen in many fisheries, and their associated fishery-dependent abundance indices (Rose & Kulka 1999). Fishers can perform well even when the population is less abundant, due to targeting and private information. Meanwhile, an index corrected for site selection of fishers mirrors the true abundance pattern well in baseline scenarios, and gives similar results to a randomly sampled population.

In our Baseline Scenario 1, the corrected index performs well both on average and in terms of uncertainty, when compared to a randomly-sampled index meant to mimic an additional fishery-independent survey. However, even when fisher coverage is not evenly dispersed in Partial Fisher Coverage Scenario 2, the correction still performs relatively well, however, estimates are more uncertain across replicates.

When the spatial abundance distribution is random in Random Spatial Abundance Scenario 3, again the correction performs relatively well, especially in comparison to the uncorrected index. This is because even though the fisher does not sample all areas due to cost of travel and habitat, missing observations are not systematically biased in any way – fishers are just as likely to miss high-abundance areas as low-abundance areas, when travel is costly enough. However the uncertainty around the estimates is again larger compared to our baseline randomly-sampled index. The uncorrected index still exhibits bias, as fishers seek out patches with larger than average catches, though less bias than in Scenarios 1 and 2, as in some years the highest abundance areas are not available, mitigating the impact of selection. This shows both fishery behavior and spatial abundance can drive error in the index (recall parameterizations are in Table 2).

Finally, when there is a trend, both the uncorrected and corrected indices perform poorly (Trending Spatial Abundance Scenario 4). Fishers are no longer missing high-abundance areas randomly, but rather systematically fishing on more and more abundant areas due to the eastward spatial trend. Without the ability to sample or extrapolate to areas not visited by the fisher, for example with fishery-independent surveys, the analyst is unable to correct for data they do not observe. However, Figure 6 suggests that as the majority of the biomass moves into the sampled area, the corrected estimates still exhibit less error than the uncorrected. Compared to Scenario 3, this implies non-static abundance over years is not necessarily a problem for the corrected estimates, rather it is the existence of a spatial trend causes more of an issue in estimation.

3.2 Assessment performance with corrected indices

In Figure 7 relative error in spawning biomass is plotted under partial fisher coverage (Scenario 2). For all four panels a fishery-independent survey is conducted every four years, ending 12 years prior to the end of the scenario, with an increasing fishing mortality pattern, and the base case with no additional data is plotted in the upper left panel.

The lower left panel also includes a yearly nominal uncorrected abundance index: this results in biased estimates of spawning biomass. If, instead, an additional yearly randomly sampled index is included (upper right), estimates are unbiased and also more accurate, especially at the end of the time series. This is because an additional randomly sampled index is akin to including another fishery-independent survey. However, the lower right panel shows that a similar increase in efficiency can be achieved by including a yearly corrected fishery-dependent abundance index instead.

Figure 7 illustrates that when there are fishery-independent surveys (here, four surveys in years 76-88 providing absolute estimates of abundance), but these surveys end before the end of our time series, the corrected index can add value to the stock assessment. In addition, using an uncorrected index will introduce bias and worsen performance. Meanwhile the corrected index tightens up uncertainty around the estimates, especially near the end of the time series. An assessment that relies only on a fishery-independent survey that is limited in years sampled exhibits considerably more uncertainty – unless our randomly sampled index (bottom right), mimicking another fishery-independent survey, is also included. These results are largely intuitive given the abundance indices in Figure 6, but validate that there were no unforeseen interactions or issues in the estimating model.

4. Discussion

We investigated using a structural economic model to correct for selection bias in fishery-dependent data. Our specific form of statistical selection bias occurs when fishers have some private information about catches that allows them to systematically select where to fish in a way the analyst cannot observe, including self-selecting based on their economic and vessel characteristics (Chen et al. 2023). By explicitly modeling their decision-making behavior, we can estimate the probability a fisher visits an area and use that knowledge to produce unbiased abundance indices from fishery-dependent data for use in stock assessments. We find that the inclusion of corrected indices can improve estimates of spawning biomass in common expected cases, depending on the spatial distribution of the fishery population over time.

While our proposed methods can perform relatively well, especially with sufficient richness of sample variation and number of observations, they do not represent a panacea for all index standardization issues. We outline one scenario where the correction performs particularly poorly in this paper – a spatial trend in abundance and poor fisher coverage. Intuitively, this is because even if areas are poorly sampled due to habitat or travel costs, when abundance is random, the analyst is just as likely to miss large positive shocks, as negative ones, but, in contrast, cannot capture a spatial trend.

Notably, our results illustrate that the models cannot correct for data they do not observe. This can occur e.g. for trans-boundary stocks, or where there are Marine Protected Areas (MPAs) in which densities increase or stay steady, while fishing impacts result in a different relative trend outside the MPAs. In these cases, the proposed methods should be used in conjunction with existing index standardization best practices, as well as techniques to extrapolate over less-sampled areas (Maunder & Punt 2004), such as coupling these methods with spatiotemporal smoothers (Thorson et al. 2020).

This study represents a step forward to applying these methods to real data, but there are some important caveats and next steps to highlight. Building toward real applications of our method is an important next step, and we suggest areas of future research. Our results are based on a single simulated fish stock (cod like) and exploitation pattern (one-way trip). The four fishing model scenarios (Table 2) tested do not capture the full richness of fisher behavior nor spatial distribution of fish, such as competition across fishers, or depletion over the season or localized areas (Walter et al. 2007). We do not anticipate the biology to greatly influence the method, but alternative fishing histories, such as a severely depleted stock, could interact with the fishing and econometric models. Such a scenario would require including some time-varying variable in the fishing model to account for changes within-season of the desirability of sites.

Our choice of a 3x3 or 2x4 grid is unlikely to represent the true spatial extent of many fisheries, but may adequately represent the variability in expected catch and other factors, and illustrates a proof of concept for our methods, allowing us to report asymptotic results (which require thousands of model runs). Computationally, a model with a larger number of areas might take a few days to converge. This is manageable for practical use, where an analyst may only need to update their fishery-dependent abundance index once a year, perhaps by running a few different candidate models (investigating the inclusion of various covariates or distributional assumptions). However, because we are interested in asymptotic behavior, our workflow requires running a model for each year fishing occurs, for 200 replicates, which requires thousands of model runs. A smaller grid size allows us to report asymptotic results.

Fishers may also have an incentive to misreport harvest, or there could exist measurement error in catch observations or the location of harvests. Broadly, this will vary across fisheries: many exploited fisheries have observer coverage or automatic identification system (AIS) tracking such that misreporting is difficult on the part of the fisher. To the extent measurement error does occur, we only expect a systematic bias in catches or location to affect our results. For example, Scenario 3 suggests that when missing observations are not systematically biased in any way, this increases uncertainty but does not bias estimates of catch. Conversely, if fishers are systematically underreporting catches, another form of correction may be required (for example the existence of another covariate that is correlated with the magnitude of misreporting), though consistent fractional underreporting would not necessarily impact the validity of a relative index.

Targeting is also often a multispecies process, but here we examined a simplified scenario with a single stock. While our stylized model only includes one species in the fisher's utility function, it is relatively straightforward to include multiple species, as long as the analyst has catch data for those species. Any number of covariates can be included in the fisher's utility function, including but not limited to bycatch species, management actions, time in season (to account for depletion for example), herding of vessels, or state dependence. These can be statistically tested by nesting models and performing likelihood ratio tests, for example, or by comparing information criterion. The marginal utility from catch is an estimated parameter and is agnostic as to sign; bycatch for example might have a negative coefficient. Assumptions as to the joint distribution of catch for the two species may need to be made. If the analyst believes the catch of the other species might also suffer from selection bias, additional correction functions could also be added for that species, although the number of covariates would quickly increase.

Fishers are also likely to target based on expected revenue, which includes prices that may vary based on the size of fish (Chen & Haynie 2024, Thébaud et al. 2023, Zimmermann & Heino

2013). We have included a sensitivity analysis in Appendix A that examines fisher behavior that is dependent on size-targeting. In general, if the estimating model can incorporate price data, for example from fish tickets that report revenue, probabilities of selection are still estimated accurately and the correction still performs well. This is also consistent with other research that suggests that price data, which is widely available, can be used to capture interannual variation and improve stock assessment methods (Lancker et al. 2023). A deeper investigation of the impacts of size-targeting would be prudent, especially as this effect could be influential for many important stocks.

In addition, if model of site selection is misspecified, the probabilities that enter our polynomial correction may be misspecified. Our specific method uses a conditional logit estimation, requiring a parametric assumption of the distribution of the error term. Broadly, the polynomial only requires consistent estimates of the probabilities of selection, and conditional logit provides consistent estimates under specific conditions (McFadden 1974). Some of these conditions (such as independence from irrelevant alternatives) may be strong. However, there are many alternative methods, such as mixed logit, to estimate the probabilities that relax these assumptions, as well (Train 2009). Another potential method is to estimate “cell probabilities” (Dahl 2002). These would be constructed by, for example, grouping vessels with similar characteristics together (e.g. vessel tons or horsepower), creating cells, and calculating the proportion of individuals in each cell that move from one location to another. These are distribution-free, do not assume a parametric framework, and are computationally straightforward to include in a linear regression.

The assumed distribution of the error terms might also be incorrect in our method, although robustness checks examining an assumption of normality suggests good performance with alternative distributions (Chen et al. 2023). We also assume that the error terms in the fisher's catch and utility are independent, but they could be correlated. To the extent they are positively correlated, we would expect this to exacerbate the selection bias described, and attenuate the bias if they are not correlated.

We also examine only one specific stock assessment framework – Stock Synthesis. Other assessment methods may not use fishery-dependent data to construct relative abundance indices, may instead rely predominantly on catch-at-age data, or use fishery-independent indices in conjunction to tune time series. We are unaware of any fisheries papers that demonstrate a new method on multiple platforms, however, we note that the bias we investigate occurs because the sample observed is conditioned on factors such as the preferences and choices of fishers, and this results in a sample that is potentially different from the population. While the specific data and conditioning may depend on individual fisheries and modeling situations, we expect different assessment models would give similar results. However, subsequent studies replicating our analysis in alternative platforms (e.g., Nielsen & Berg 2014, Stock & Miller 2021) tailored to region- or stock-specific challenges would be worthwhile to corroborate the findings here.

All our various caveats above emphasize that the process of selecting a final model from competing choices can be time-consuming and subjective, however, due to the variety of modeling assumptions and specifications this can be a necessary and common undertaking. The fishers may choose not to fish if profits are too low; we examine the possibility of an opt-out option in Appendix B. The researcher must check the robustness of different catch and utility functions, as well varying assumptions of the catch and fisher error terms, and decide both the fineness and size of the spatial extent, etc.

While we find that our correction improves estimates of spawning biomass, the magnitude of this improvement is lower when an existing fishery-independent survey, that is well designed for the stock, can be used. Thus, we expect our methods to be more useful for situations where fishery-dependent data play a more prominent or exclusive role; an example might be tuna fishery management. Another avenue for future research is to understand how survey frequency impacts the value of our method. Instead of occurring once every few years in a limited season as is typical for fishery-independent surveys, fishing may generate data every year across many months, thereby covering more months of the year for more years. In addition, if fishery-independent surveys become less frequent due to decreases in financial resources, these methods provide a potential supplement to that sparse data (ICES, 2020). These methods may also provide an avenue to explain the results of stock assessments to fishers, when stock assessment predictions diverge from fishing conditions and catches observed “on the ground”.

The proposed methods therefore are useful for fisheries where fishery-independent surveys do not occur, or to take better advantage of fishery-dependent data that is collected for regulatory purposes. In addition, we note that because a polynomial approximation is relatively straightforward to implement and is statistically testable for whether or not selection bias occurs, we emphasize these methods do not represent an either-or proposition. Rather, they are an additional tool allowing stock assessors to potentially better use the wealth of data generated by observer coverage and the fishing process itself to improve understanding of trends in abundance of stocks.

Acknowledgements

The authors thank Kiva Oken, Leif Anderson, Ian Taylor, Lisa Pfeiffer, and two anonymous reviewers for valuable comments on previous versions of the manuscript. In addition, we thank participants at the North American Association of Fisheries Economists Forum. Financial assistance was provided by FishSET, the Pacific States Marine Fisheries Commission, and NOAA Fisheries Office of Science and Technology.

Competing interests: The authors declare there are no competing interests.

This research did not receive any additional specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

A repository containing replication code for this paper can be found at: https://github.com/allen-chen-noaa-gov/ss3simfork.

Author contribution statement: Conceptualization YAC, ACH; Data curation YAC; Formal analysis YAC; Funding acquisition ACH; Investigation YAC, CCM, OSH, ACH; Methodology YAC, CCM; Project administration YAC, ACH; Software YAC; Resources ACH; Supervision YAC, ACH; Validation YAC; Visualization YAC, CCM; Writing – original draft YAC; Writing – review & editing YAC, CCM, OSH, ACH.

References

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., & Valero, J.L. (2014). ss3sim: An R package for fisheries stock assessment simulation with Stock Synthesis. *PLOS ONE*. 9(4): e92725. http://doi.org/10.1371/journal.pone.0092725.

Andrews, D., & M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. The Review of Economic Studies, 65(3), 497-517.

Bishop, J., Venables, W. N., & Wang, Y.-G. (2004). Analysing commercial catch and effort data from a Penaeid trawl fishery: A comparison of linear models, mixed models, and generalised estimating equations approaches. *Fisheries Research*, *70*(2-3 SPEC. ISS.), 179–193.

Chen, Y.A., A.C. Haynie, & C.M. Anderson. (2023). Full information selection bias correction for discrete choice models with observation-conditional regressors. *Journal of the Association of Environmental and Resource Economists*, volume 10, number 1, January.

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, *70*(6), 2367–2420.

Eales, J. & Wilen, J.E. (1986). An Examination of Fishing Location Choice in the Pink Shrimp Fishery. *Marine Resource Economics*, 2(4), 331–351.

Lancker, K., Voss, R., Zimmermann, F., & Quaas, M. F. (2023). Using the best of two worlds: A bio-economic stock assessment (BESA) method using catch and price data. Fish and Fisheries, 24, 744–758. https://doi.org/10.1111/faf.12759

Pfeiffer L., Carvalho P., Abelman A., Lee M., McManus B., Chen Y.A., Harsch M., & Haynie A.. (2024). FishSET: Spatial Economics Toolbox for Fisheries. R package version 1.1.0.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161.

Hilborn, R. & Walters, C. J. (1992). Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, New York. 570 p.

Hoyle, S. D., & Okamoto, H. (2013). Target changes in the tropical WCPO Japanese longline fishery, and their effects on species composition. *Western and Central Pacific Fisheries Commission Scientific Committee Ninth Regular Session*, *August*.

ICES. (2020). ‘Workshop on unavoidable survey effort reduction (WKUSER)’. ICES Scientific Reports. doi: 10.17895/ices.pub.7453.

Johnson, K.F., Anderson, S.C., Doering, K., Monnahan, C.C., Stawitz, C.C., & Taylor, I.G. (2019). ss3sim: Fisheries Stock Assessment Simulation Testing with Stock Synthesis. R package version 1.0.3.

Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-pre, K.A., Anderson, S.C., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R.R., Muradian, M.L., Ono, K., Szuwalski, C.S., Valero, J.L., Whitten, A.R., & Punt, A.E. (2015). Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. *ICES Journal of Marine Science* 72, 137–150. https://doi.org/10.1093/icesjms/fsu055.

Maunder, M. & Punt, A.E. (2004). Standardizing catch and effort data: A review of recent approaches. *Fisheries Research*. 70. 141-159. 10.1016/j.fishres.2004.08.002.

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. Academic Press.

Methot, R.D., & Wetzel, C.R. (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142, 86-99.

Nielsen, A., & Berg, C.W. (2014). Estimation of time-varying selectivity in stock assessments using state-space models. *Fisheries Research.* Volume 158, Pages 96-101, https://doi.org/10.1016/j.fishres.2014.01.014.

Ono, K., Licandeo, R., Muradian, M.L., Cunningham, C.J., Anderson, S.C., Hurtado-Ferro, F., Johnson, K.F., McGilliard, C.R., Monnahan, C.C., Szuwalski, C.S., Valero, J.L., Vert-Pre, K.A., Whitten, A.R., & Punt, A.E. (2015). The importance of length and age composition data in statistical age-structured models for marine species. *ICES Journal of Marine Science* 72, 31–43. https://doi.org/10.1093/icesjms/fsu007.

Privitera-Johnson, K.M., Methot, R.D., & Punt A.E. (2022). Towards best practice for specifying selectivity in age-structured integrated stock assessments. *Fisheries Research*, Volume 249, https://doi.org/10.1016/j.fishres.2022.106247.

Rose, G.A., & D.W. Kulka. (1999). Hyperaggregation of fish and fisheries: how catch-perunit-effort increased as the northern cod (Gadus morhua) declined. Canadian Journal of Fisheries and Aquatic Sciences 56 (Supplement 1):118–127

Rufener, M.C., Kristensen, K., Nielsen J.R., and Bastardie, F. (2021). Bridging the gap between commercial fisheries and survey data to model the spatio-temporal dynamics of marine species. Ecological Applications 31(8): http://doi.org/10.1002/eap.2453

Rufener, M.C., Nielsen, J.R., Kristensen, K., and Bastardie, F. (2023). Closing certain essential fish habitats to fishing could be a win-win for fish stocks and their fisheries- insights from the western Baltic cod fishery. Fish. Res. 268 https://doi.org/10.1016/j.fishres.2023.106853

Stafford T.M.. (2018). Accounting for outside options in discrete choice models: An application to commercial fishing effort, Journal of Environmental Economics and Management, Volume 88, Pages 159-179, https://doi.org/10.1016/j.jeem.2017.10.006.

Stock, B.C., & Miller, T.J. (2021). The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. *Fisheries Research*, Volume 240, https://doi.org/10.1016/j.fishres.2021.105967.

Thorson, J.T., Maunder, M.N., & Punt, A.E. (2020). The development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance. *Fisheries Research*, Volume 230, 105611, https://doi.org/10.1016/j.fishres.2020.105611.

Thorson, J. T., Shelton, A. O., Ward, E.J., & Skaug, H.J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, Volume 72, Issue 5, May/June, 1297–1310, https://doi.org/10.1093/icesjms/fsu243.

Train, Kenneth E. (2009). Discrete choice methods with simulation. Cambridge University Press, New York, (2nd), 1148. https://doi.org/10.1016/S0898-1221(04)90100-9.

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch data. *Canadian Journal of Fisheries and Aquatic Sciences* 60. 1433-1436. 10.1139/f03-152.

Walter, J. F. III, Hoenig, J. M., and Gedamke, T. (2007). Correcting for effective area fished in fishery-dependent depletion estimates of abundance and capture efficiency. ICES Journal of Marine Science Volume 64, Issue 9, December 2007, Pages 1760–1771, https://doi.org/10.1093/icesjms/fsm147

Zimmermann F,, Heino M., (2013). Is size-dependent pricing prevalent in fisheries? The case of Norwegian demersal and pelagic fisheries, ICES Journal of Marine Science, Volume 70, Issue 7, November 2013, Pages 1389–1395, https://doi.org/10.1093/icesjms/fst121