ICES Journal of Marine Science, 2023, **80**, 1028–1042
DOI: 10.1093/icesjms/fsad037
Advance access publication date: 9 March 2023
Original Article

Standardizing fishery-dependent catch-rate information across gears and data collection programs for Alaska sablefish (*Anoplopoma fimbria*)

M. L. H. Cheng [1,*], C. J. Rodgveller[2], J. A. Langan[1], and C. J. Cunningham[1]

[1]Department of Fisheries at Lena Point, College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Road, Juneau, Alaska 99801, USA
[2]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, 17109 Point Lena Loop Road, Juneau, Alaska 99801, USA

[*] *Corresponding author: tel: +(603)-409-7618*; e-mail: lhcheng@alaska.edu.

Indices of abundance used to inform stock assessment models are commonly derived from fishery-dependent data sources. However, fishery catch-per-unit-effort (CPUE) are often confounded by a myriad of factors for which corrections must be made using model-based standardization methods. The Alaska sablefish (*Anoplopoma fimbria)* fishery provides a fitting case study of such issues, wherein a regulatory change in 2017 disrupted historic fishery dynamics, promoting a rapid transition in use of pot gear over demersal hook-and-line gear in the Gulf of Alaska. To address this, we combined across both observer and logbook programs (data sources) and gear types to develop an intercalibrated abundance index. We first regressed observer records against vessel logbooks to understand potential biases that may arise from combining data sources during the CPUE standardization process. Here, we found that both data sources exhibited strong agreement in reported CPUEs when compared on a set-by-set basis. Therefore, we intercalibrated both CPUE data sources and developed an index of abundance that incorporated catch records from both demersal hook-and-line and pot gear fisheries for sablefish in Alaska, to account for the recent rapid change in gear use. This standardized index of abundance compared favourably with an index generated from a fishery-independent hook-and-line survey currently used in management, suggesting it is representative of sablefish population trends. Our findings not only represent a valuable contribution to the management of sablefish in Alaska, but also provide a widely applicable framework for standardizing fishery-dependent CPUE data to support the management of multi-gear fisheries.

Keywords: catch-per-unit-effort (cpue), fishery-dependent data, indices of abundance, multi-gear fisheries.

Introduction

Contemporary integrated stock assessment models incorporate a variety of data sources into a single analysis to provide science-based advice, guiding the fishery management process (Maunder and Punt, 2013). These assessments are often fit to relative indices of abundance, developed using catch-rate data (i.e. catch-per-unit-effort; CPUE) from both fishery-dependent and fishery-independent data sources, and provide information on the scale of population biomass and trends across time (Maunder and Punt, 2004). Fundamentally, the use of CPUE as an index of abundance assumes that changes in CPUE are directly proportional to changes in abundance (N) such that $CPUE = q*N$, where the catchability coefficient q is the proportion of a stock captured with one unit of effort. Indices of abundance derived from fishery-independent sources are often preferred because they are collected through standardized survey designs (e.g. systematic or randomized), with careful consideration of the comparability of gear, effort, and survey sampling methods to ensure constant catchability (q) across time. However, collection of these data are often expensive, logistically challenging, and constrained to a short period of time within each year, thus limiting the understanding of spatial and temporal patterns for a given species that fishery-independent data can provide. As a result of these challenges, many global fisheries are assessed using indices of abundance developed from fishery-dependent catch-rate data because their lower relative cost and accessibility make them a viable complement, and in some cases a necessary alternative to standardized survey data.

Fishery-dependent data that provide information on catch and effort most often come in two forms: (1) self-reported fishery logbooks maintained by commercial fishers, and (2) records collected from fishery observer programs. Self-reported fishery logbooks may be subject to errors in reported catch or effort, potentially driven by incentive structures affecting harvesters, or insufficient training in data collection methods (Gilman *et al.*, 2019). In contrast, fishery-dependent data from observer programs (e.g. North Pacific Observer Program) in the United States are generally considered to be less error-prone because they are collected via sampling of catch from a random subset of hauls conducted in directed fisheries by scientifically trained observers governed by a pre-defined sampling strategy (AFSC, 2022).

Nevertheless, both forms of fishery-dependent data often exhibit characteristics of preferential sampling (i.e. fishing effort is spatially associated with fish abundance) and reflect a diversity of fishing gears, collection methods, and distributions in fishing location in both space and time. Such features can result in either hyperstability or hyperdepletion of catch rates, where changes in CPUE are no longer proportional to changes

Received: 19 September 2022; **Revised:** 20 February 2023; **Accepted:** 21 February 2023

in abundance (i.e. q is no longer constant) (Walters, 2003). Hyperstability occurs when CPUE increases or remains stable as abundance decreases, resulting in the "illusion of plenty" (Rose and Kulka, 1999; Erisman *et al.*, 2011). Conversely, hyperdepletion occurs when CPUE declines faster than abundance, where vulnerable portions of the stock are depleted, but the less vulnerable portion of the stock remain abundant (Hilborn and Walters, 1992).

Considering that indices of abundance derived from fishery-dependent sources can potentially be misleading, they are often "standardized" for use in stock assessment to control for factors with the potential to influence catchability. The index standardization process is commonly employed via model-based methods using generalized linear models (GLMs), generalized additive models (GAMs), or spatiotemporal models (Grüss *et al.*, 2019; Kai, 2019). These model-based approaches control for the effect of factors affecting CPUE (e.g. gear type, data source, and vessel characteristics), and the year trend extracted from the model is interpreted as a standardized index of relative abundance (Maunder and Punt, 2004).

Indices of abundance derived from self-reported logbooks and/or observer data are invaluable because they can expand the spatiotemporal coverage of fishery-independent data, and are able to provide complimentary information to fishery-independent surveys when properly standardized (Fox and Starr, 1996; Ye and Dennis, 2009). However, few studies have explored the utility of combining both logbook and observer data sources to develop indices of abundance, and the comparability of data resulting from self-reported logbooks and observers (Starr and Vignaux, 1997; Walsh *et al.*, 2002; Nakano and Clarke, 2006). Thus, there remains a need to understand factors driving consistencies and differences between observer and logbook data and inform development of unified fishery-dependent indices of abundance from both data sources that encapsulate the maximum available spatiotemporal coverage of fishing effort.

For the purpose of addressing these knowledge gaps, the sablefish (*Anoplopoma fimbria*) fishery in Alaska represents an ideal case study because of the extensive sampling and availability of both observer ($n = 46{,}590$) and logbook ($n = 99{,}244$) data sources. Furthermore, these fishery-dependent data are complemented by an annual sablefish longline survey conducted throughout Alaska by the National Oceanic Atmospheric Administration (NOAA) that is considered to provide an accurate measure of abundance trends (Sigler, 2000). Sablefish are a demersal groundfish species, typically occupying depths of about 200–1000 m (Wolotira, 1993), that exhibit high movement rates and episodic recruitment (Goethel *et al.*, 2020). Juvenile sablefish display ontogenetic movements, occupying shallow nearshore bays and inlets before moving into deeper and colder continental shelves and slopes (Hanselman *et al.*, 2019). In recent years, several high recruitment events (2014, 2016, and 2017) have resulted in small sablefish being captured at high rates throughout the directed fishery, which primarily operates using hook-and-line and pot gear (Goethel *et al.*, 2020). The federal Alaska sablefish stock assessment employs an age- and sex-structured integrated model fit to age and length compositions, catch, and indices of abundance from fishery-dependent and fishery-independent data (Goethel *et al.*, 2020). The directed fishery is treated as a single fleet assuming logistic selectivity, where data across hook-and-line and pot gears are aggregated. Currently, the assessment assumes a single homogeneous population in Alaska, with the fishery comprising six management regions: Aleutian Islands (AI), Bering Sea (BS), Western Yakutat (WY), Western Gulf of Alaska (WGOA), Central Gulf of Alaska (CGOA), and Eastern Yakutat/Southeast (EY/SE) (Figure 1).

Traditionally, the Alaska sablefish fishery has operated using hook-and-line gear, with small portions of fleets using pot gear in the BS and AI region. However, due to concerns with sperm whale depredation of hook-and-line gear, the North Pacific Fishery Management Council (NPFMC) passed a regulation to allow the use of pot gear across the Gulf of Alaska in 2017. Since then, the proportion of fleets utilizing pot gear in Alaska has increased rapidly and is expected to continue on this trend (Goethel *et al.*, 2022). However, the rapid change in gear use throughout Alaska is problematic for the fishery-dependent index used within the assessment, where data resulting from pot gear in the directed fishery are not currently used to inform the fishery-dependent index (Goethel *et al.*, 2022). Further, progressively fewer hook and line observations are available to inform this index in recent years, underscoring the need to transition towards a multi-gear fishery-dependent index of abundance. As discussed, the fishery-dependent index of abundance only utilizes data from hook-and-line gear, and the assessment treats the directed fishery as a single fleet that aggregates data across gear types (Goethel *et al.*, 2021); this inconsistency may potentially result in misleading conclusions on stock dynamics. Finally, the current assessment approach utilizes a nominal index, and does not account for factors resulting in differences in catch rates that are unrelated to year-to-year changes in abundance (e.g. vessel attributes). Thus, there exists a critical need to explore statistical methods for deriving intercalibrated and standardized indices of abundance using model-based approaches across gear types (e.g. hook and line, and pot) and data collection programs (e.g. logbook and observer).

The present study aims to: (1) understand consistencies and differences between observer and logbook data sources for the sablefish fishery in Alaska, and (2) develop a robust, standardized index of abundance by integrating catch records from hook-and-line and pot gear, derived from both observer and logbook data sources. Here, we provide a framework for managers of multi-gear fisheries seeking to understand the potential for integrating observer and logbook data to develop informative standardized indices of abundance using fishery-dependent data.

Methods

Data sources

We obtained fishery-dependent observer data for both hook-and-line and pot fleets from the NOAA, Alaska Fisheries Science Center (AFSC), North Pacific Observer Program, and vessel logbooks from the joint National Marine Fisheries Service (NMFS) International Pacific Halibut Commission (IPHC) logbook program. Vessels greater than 60 feet are required to provide logbooks; however, they are not required for vessels under 60 feet. Additionally, catch-rate data collected from observers are indiscriminate of vessel length. During 1995, the Alaska sablefish fishery shifted from an open-access fishery to an Individual Fishing Quota (IFQ) system, resulting in substantial increases in fishery efficiency and commensurately large differences in catch rates (~1.8 times higher under IFQs) between these two periods (Sigler and Lunsford, 2001). These

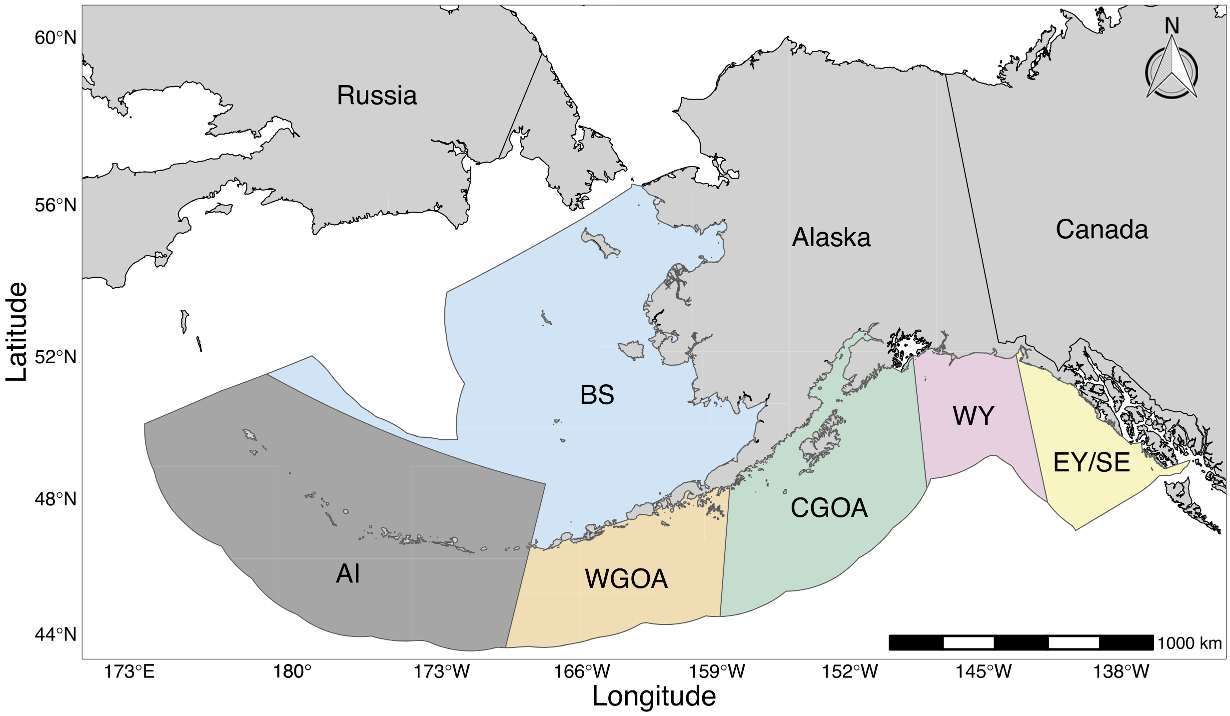

Figure 1. Map of sablefish fishery management boundaries in the U.S. Exclusive Economic Zone surrounding Alaska. Canada and Russia are shown on the map for reference. AI, BS, WGOA, CGOA, WY, and EY/SE represent the Aleutian Islands, Bering Sea, Western Gulf of Alaska, Central Gulf of Alaska, Western Yakutat, and Eastern Yakutat/Southeast respectively.

large differences in catch rates driven by the 1995 shift in fishery management have the potential to obscure trends derived from the CPUE standardization process if data prior to 1995 are included. Consequently, only data from 1995 to present (2020) were included in these analyses.

Logbook data are available in a standardized format starting from 2002 to 2020. These data are self-reported by vessel captains with information on catch, effort, variables associated with the fished set (e.g. depth and geospatial coordinates), and a dockside landing weight associated with each fishing trip. However, only sablefish catch records are avaliable, hence, catch composition information are not available from logbook data. It is important to note that logbook reported set weights are approximate because weights reported by vessel captains are estimated rather than values measured directly. Conversely, fishery data from the observer program (1995–2020) are collected by at-sea observers, where all fishing sets, or more often a randomly selected subset of fishing sets, are sampled to represent catch by sector under regular fishing activities (AFSC, 2022). In 2020, the total weight of observed sablefish catch was 9% of the total catch (Goethel *et al.*, 2021). Other commonly captured species in the sablefish fishery include Pacific halibut (*Hippoglossus stenolepis*), several rockfish species (*Sebastes spp.*), Greenland turbot (*Reinhardtius hippoglossoides*), and Pacific cod (*Gadus macrocephalus*) (Goethel *et al.*, 2020). For observer records, catch composition data are reported for the aforementioned species. Additionally, catch, effort, and auxiliary information (e.g. depth and geospatial coordinates) are also reported for each fishing set. Across both observer and logbook data sets, the units of effort for hook-and-line and pot gear are defined as number of hooks and number of pots, respectively.

We also obtained a fishery-independent index of relative population weights (RPW) for sablefish derived from the NOAA AFSC longline survey to serve as a basis for comparison, to evaluate the plausibility of standardized year trends estimated from fishery-dependent data. RPWs are depth-stratified relative indices of abundance (weight per unit effort), which are widely used across stock assessments in Alaska. The survey has been conducted annually by the AFSC since 1987 (see Rutecki *et al.*, 2016 for a detailed overview on survey methods) and abundance indices derived from the longline survey are considered representative of sablefish dynamics in Alaska (Sigler, 2000).

Data processing

For data obtained from the observer program, the target of a set is not explicitly recorded. As such, we conducted a hierarchical clustering procedure following similar methods to He *et al.* (1997) using catch composition information, geospatial coordinates, vessel length (ft), bottom depth (m), day of year, and effort information (number of hooks and number of pots) to identify different targeting strategies (see Supplementary Material 1). Catch composition information was available for species commonly observed in the sablefish fishery, including Greenland turbot, Pacific cod, Pacific halibut, Shortspine thornyhead, and the cumulative weight of several species of the *Sebastes* genus. Through clustering analysis (see Supplementary Material 1), we identified three optimal clustering groups: (1) sablefish targets (n = 35893), (2) turbot targets (n = 4633), and (3) mixed Halibut and sablefish targets (n = 6777). Finally, records from the observer data were removed if there were reported gear performance issues (e.g. lost gear or depredation), and if there were obvious data errors (e.g. unrealistically large fishing depth values).

Considering differences in sampling programs and auxiliary information reported between the two data sources, data processing for logbook data were treated differently. In logbooks, vessel captains declare the target of the set. Thus, records were selected if they were reported as: (1) sablefish targets, (2) mixed halibut and sablefish targets, or (3) targets of sablefish and other species. Trip records were removed if CPUE was unable to be computed (i.e. no appropriate catch or effort data), missing required data fields, and other obvious errors (e.g. geospatial coordinates are on land). Additionally, adjustments to reported catch were made to account for variation in reporting of catch in the logbooks. Specifically, fish can be processed at-sea in multiple ways (e.g. eastern cut and western cut) or delivered unprocessed (round weight), and may potentially impact estimates of weight reported. Although all weights within the logbook data should be reported in units of round weight, there are some instances where weights are reported for different processed types. Thus, considering that observers report sablefish catch in round weight, corrections for reported catch within the logbook data were made to reflect round weight, allowing catch from both data sources to be comparable. To this end, catch records for fishing sets reported as eastern cut (pectoral girdle is removed along with the head) were divided by 0.63 and those reported as western cuts were (pectoral girdle is intact but head is removed) divided by 0.68, if these adjustments resulted in greater coherence between the sum of all trip records and the reported IFQ dockside landing trip weight, to calculate approximate round weights. Values used for correction factors are consistent with those used to adhere to NOAA recordkeeping and reporting requirements (National Oceanic and Atmospheric Administration, 2001). In addition to adjustments made for processed cut types, catch per set was also adjusted by multiplying the IFQ dockside landing trip weight by the ratio of the logbook reported set weight and logbook reported trip weight. This was a necessary step because catch reported in logbooks are estimates of weight made without an accurate scale measurement (Sigler and Lunsford, 2001). As noted, these corrections for processed sablefish product types were necessary because we wanted to maintain the interpretability of our results when comparing observer and logbook data sources on a set-by-set basis prior to our CPUE standardization analysis, to understand biases that may arise from combining data sources. However, for our subsequent CPUE standardization analysis, we controlled for differences among processed product types by treating product type as a covariate (instead of using adjusted weight values) to better reflect the total number of parameters estimated. By treating the “round weight” category as the reference level, estimated coefficients for each product type describe the average difference in weight for each product form relative to round weights. Finally, records from trips with severe over and underestimates of catch were removed (i.e. reported trip weight was more or less than the measured dockside delivery weight by 10000 lbs.).

Statistical comparison of logbook and observer CPUE

Both logbook and observer data can be identified for corresponding sets (i.e. observer data are a subset of the logbook data), allowing the opportunity to compare both data sources and to understand potential factors resulting in consistencies and differences between the two data sources, on a set-by-set basis. However, given that a unique identification code did not exist to identify corresponding records on a set-by-set basis for the two data sources, a series of stringent decision rules were developed to ensure that records were paired as accurately as possible. Specifically, fishing sets from observer and logbook data were considered matching records if: (1) reported fishing dates from both data sources occurred on the same day, (2) sets originated from the same gear type, vessel length (ft), and fishery management area, (3) the difference of fishing depths for a set was within 5 fathoms (9.1 m), and (4) the reported geospatial locations of fishing were within 1 km of each other (orthodromic distance). Note that vessel length in the logbook data are reported as factors (vessels under 60 ft, and over 60 ft, equivalent to under and over 18.3 m), whereas vessel length in the observer data are reported in continuous values (ft). Thus, for the purpose of identifying matched sets in a consistent manner, we converted vessel lengths from the observer data to factors.

Observations from year 2003 were removed because there were insufficient data to obtain reliable estimates ($n = 13$). This process resulted in the identification of 10976 matched records. CPUE for hook-and-line gear was computed as reported catch (kg) divided by the number of hooks. For pot gear, CPUE was computed as reported catch (kg) divided by the number of pots retrieved. Following the identification of matched records, two separate models were developed to compare CPUE between observer and logbook data. Given that several zeros existed, we added a small constant (0.001) to both observer and logbook CPUE values to accommodate log transformation. For the first model, log-transformed CPUE values from the observer dataset were regressed against log-transformed CPUE values from the logbook dataset

$$\log\left(\text{Observer CPUE}_i + 0.001\right) = \alpha + \beta_c \log\left(\text{Logbook CPUE}_i + 0.001\right) + \varepsilon_i, \quad (1)$$

where α is the intercept, β_c is the estimated coefficient for the log-transformed logbook CPUE, subscript i is the ith matched record, and $\varepsilon_i \sim t(\mu, \sigma^2, \nu)$, where ε_i are observation errors following a scaled student’s t-error distribution. Residuals following a normal distribution were initially assumed. However, residuals appeared overdispersed, suggesting that a scaled student’s t-error distribution would be more appropriate, as it better accommodates frequent large positive and negative residuals (i.e. heavy tails). Note that the formulation in Equation (1) treats logbook CPUE as exact, and errors-in-variables or non-parametric regression may be more appropriate. Nonetheless, we also conducted a median-based regression to ensure that the interpretation of results remained similar among regression model variants.

In addition to comparing observer and logbook CPUE values on a set-by-set basis, it is also imperative develop an understanding of factors that may be associated with differences among data sets. This is a necessary step because it helps to inform whether further considerations when combining data sources are required (e.g. differentially weighting data sources across areas, gears, or years if such differences are substantial). To understand potential factors associated with differences between the two data sources, a linear mixed-effects model

Table 1. Descriptions and *a priori* justification for variables included in the full CPUE standardization model [Equation (3)] for the US Alaska sablefish fishery from 1995 to 2020. AI, BS, WGOA, CGOA, WY, and EY/SE represent the Aleutian Islands, Bering Sea, Western Gulf of Alaska, Central Gulf of Alaska, Western Yakutat, and Eastern Yakutat/Southeast, respectively.

Variables	Justification	Description
γ_{Year_i}	Describes changes in abundance over time (Maunder and Punt, 2004)	Categorical (25 levels); 1995–2020
$\delta_{\text{Gear Type}_i}$	Controls for differences in observed catch among gear types and the conversion factors for CPUE (i.e. hook-and-line; catch/hooks deployed, pot gear; catch/pots retrieved)	Categorical (2 levels); hook-and-line gear, pot gear
$\zeta_{\text{Fishery Management Area}_i}$	Accounts for potential differences in sablefish abundance among fishery management areas	Categorical (6 levels); AI, BS, WGOA, CGOA, WY, and EY/SE
$\beta_{\text{Vessel Length}_i}$	Controls for catchability differences in catch-rates because of vessel characteristics	Categorical (2 levels); under 60 feet, Over 60 feet
$\vartheta_{\text{Data Source}_i}$	Accounts for potential differences of how catch may be reported and observed between data sources	Categorical (2 levels); logbook, observer
$\varphi_{\text{Target Strategy}_i}$	Controls for different targeting strategies when operating in a multi-species fishery (i.e. targeting both halibut and sablefish)	Categorical (4 levels); sablefish targets, mixed sablefish and halibut, mixed sablefish and other species, turbot targets
$\omega_{\text{Processed Type}_i}$	Fish can be processed at-sea in multiple ways (e.g. eastern cut and western cut) or delivered unprocessed (round weight), which may potentially impact estimates of weight reported. Furthermore, logbook data contain multiple processed types, whereas observer data are always reported as round weight.	Categorical (3 levels); round weight (RWS), western cut (WCS), and eastern cut (ECS)
$f(\textbf{Day of Year}_i)$	Sablefish abundances have been shown to vary seasonally, which may influence catch-rates.	Continuous; day of year fishing occurred, represented with a cyclic regression spline.
$f(\textbf{Bottom Depth}_i)$	Sablefish undergo ontogenetic shifts and typically occupy depths of about 200 −1000 m, thus, differences in fishing depths may impact catch-rates (Wolotira, 1993; Goethel *et al.*, 2020)	Continuous; depth where fishing occurred, represented with a thin-plate regression spline.
$f(\textbf{lat}_i, \textbf{lon}_i)$	Sablefish catchability may differ based upon where fishing took place, thus influencing catch-rates. Controls for interannual variation in spatial distribution of fishing effort (Walters, 2003).	Continuous; interaction term representing an average spatial field for CPUE, represented as a tensor product smooth with thin-plate regression splines for both latitude and longitude.
$f_{\text{Gear Type}}(\textbf{lat}_i, \textbf{lon}_i)$	Catchability for sablefish may differ spatially based upon gear type (hook-and-line and pot gear).	Continuous; interaction term representing an average spatial field for CPUE for each gear type, represented as a tensor product smooth with thin-plate regression splines for both latitude and longitude. Each gear type has an independent estimated spatial field.

was developed:

$$\log\left(\text{Observer CPUE}_i + 0.001\right) - \log\ \left(\text{Logbook CPUE}_i + 0.001\right)$$
$$= \alpha + \gamma_{\text{Year}_i}\ + \delta_{\text{Gear Type}_i} + \zeta_{\text{Fishery Management Areas}_i}$$
$$+ \beta_{\text{Vessel Length}_i}\ + \theta_{\text{Vessel}_i} + \varepsilon_i, \tag{2}$$

where the difference between the log-transformed observer CPUE and logbook CPUE is equivalent to the log ratio of the observer CPUE and logbook CPUE, and α is the model estimated intercept. Subscript i corresponds to the ith set. γ_{Year_i} represents a year factor (2004–2020), $\delta_{\text{Gear Type}_i}$ is a factor representing the average gear effect for hook-and-line and pot gear, $\zeta_{\text{Fishery Management Area}_i}$ are the six different fishery management areas (BS, AI, CGOA, WGOA, WY, and EY/SE), $\beta_{\text{Vessel}}\text{Length}_i$ is the estimated linear slope coefficient for vessel length (ft), and $\theta_{\text{Vessel}_i} \sim Normal(0, \sigma_\theta^2)$ represents a random intercept for individual vessels that accounts for unexplained differences in catch rates among vessels. We used all-subsets of models following Equation (2), and model selection was based upon Bayesian information criterion (BIC) values.

CPUE standardization

To standardize CPUE among observer ($n = 46{,}590$) and logbook ($n = 99{,}244$) data sources and gear types (hook-and-line vs. pot), we developed a standardized and intercalibrated index of abundance using GAMs with the package "mgcv" in the R environment (version 4.1.2) (Wood, 2017; R Core Team, 2021). For the purpose of combining logbook and observer data sources in the CPUE standardization analysis, we removed matched records from the logbook data ($n = 10{,}976$), in favour of those from the observer data to avoid pseudoreplication from duplicated observations (subset of logbook records were observed by at-sea observers). Matched records were identified from our previous analysis of comparing observer and logbook records. Given the stringent decision rules developed for the purpose of matching the two datasets, it is possible that some of the unmatched records are also duplicated. However, we do not believe that this constitutes a large proportion of the total dataset and thus, likely do not alter the conclusions of this study. Furthermore, we converted vessel lengths to factors (two levels; vessels under 60 ft, and over 60 ft), because they were reported inconsistently between the two data sources. This process resulted in $n = 145{,}834$ records. To develop our model for the purpose of standardizing CPUE, we selected a range of explanatory covariates that we hypothesized may influence catch-rate (see Table 1). These explanatory covariates control for observed differences in catch rates unrelated to changes in sablefish abundance (e.g. vessel attributes, gear, and observer effects). The resulting index of abundance is extracted from the estimated year effect, representing year-to-year changes in abundance (Maunder and Punt, 2004). Year, vessel length, target strategy, processed type, and gear were retained in every model to capture temporal trends and changes in fishery dynamics (see Table 1). We grouped together target strategies identified via clustering for observer records that were most similar to those specified in logbooks, which included: (1) sablefish targets, and (2) mixed halibut and sablefish targets. However, the third cluster identified in observer records (primarily turbot targets) was not reflected as a category within the logbook data and thus, was included as an additional level.

Model selection was then done using permutations of all other variables via a 5-fold cross validation and comparison of BIC values. Iteratively, models were defined with different combinations of predictor variables that were fit to training subsets of the available data, and used to predict the withheld testing data. Metrics to assess out-of-sample predictive performance included root mean square errors (RMSE), mean absolute error (MAE), and R^2 of predicted and observed values for the testing dataset. The full model for CPUE standardization was

$$\begin{aligned} E(\text{weight}_i) = {} & \text{offset}\left[\log\left(\text{effort}_i\right)\right] + \gamma_{\text{Year}_i} + \delta_{\text{Gear Type}_i} \\ & + \zeta_{\text{Fishery Management Area}_i} + \beta_{\text{Vessel Length}_i} \\ & + \vartheta_{\text{Data Source}_i} + \varphi_{\text{Target Strategy}_i} + \omega_{\text{Processed Type}_i} \\ & + f\left(\text{Day of Year}_i\right) + f\left(\text{Bottom Depth}_i\right) \\ & + f\left(\text{lat}_i, \text{lon}_i\right) + f_{\text{Gear Type}}\left(\text{lat}_i, \text{lon}_i\right) + \varepsilon_i, \quad (3) \end{aligned}$$

where subscript i represents the ith set. $E(\text{weight}_i)$ is the expected mean weight with a log-link function following Tweedie distribution, which are well-suited to accommodate both positive continuous and zero values. The Tweedie distribution has previously been shown to perform well for the purpose of CPUE standardization when compared to other distributions (e.g. delta-lognormal and negative binomial) and thus, other distributions were not futher explored in this study (Shono, 2008; Thorson *et al.*, 2021). Here, $\log(\text{effort}_i)$ is the natural logarithm of effort associated with set i and is treated as an offset within the model. We treated effort as an offset rather than transforming it as a catch rate, because it is considered good practice to retain the original scale of the response variable for CPUE standardization when possible (Thorson, 2019). For hook-and-line gear, $\log(\text{effort}_i)$ represents the total number of hooks, and for pot gear, it represents the total number of pots deployed for observation i. Individual vessel IDs were not included in these models as random effects due to inconsistent reporting of vessel IDs for both data sources (i.e. vessel codes are reported between the two data sources, such that one unique vessel could be modelled as 2). Given data confidentiality reasons, we are unable to distribute data utilized in the present study. However, we have developed an associated R package that implements the CPUE standardization process outlined here, and examples of how Equation (3) was implemented (https://github.com/chengmatt/idxstd4R).

The validity of indices of abundance derived from our model outputs were assessed by regressing the standardized indices of abundance against RPWs obtained from the NOAA AFSC longline survey. Furthermore, we derived a nominal CPUE index by taking the arithmetic mean of CPUE within each year and repeated the same regression procedure to illustrate the potential benefits of standardizing CPUE using model-based approaches. Considering known differences in selectivity and availability between the survey and the fishery (the survey tends to select for smaller-sized fish; Goethel *et al.*, 2022), we regressed the two fishery indices against the survey index of RPWs with a time-lag. To determine time-lags that would provide appropriate comparisons among indices, we evaluated the cross-correlation function (evaluates lags across a range) between the standardized and the nominal indices against survey RPWs by sequentially removing years from the respective indices. Sequentially removing years from the respective indices and re-evaluating the cross-correlation function aids in identifying consistent time-lags for which the indices should be regressed against, and ensures that correlations are not driven by any particular year.

Finally, to further understand and visualize the effects of covariates on the CPUE standardization process, we sequentially added in covariates identified from the candidate model, generating "step" plots (Bentley *et al.*, 2012). Covariates that appear to have a large influence on the CPUE trend across time will exhibit large differences from trends generated from its previous model sequence. Note that "step" plots allow for a qualitative assessment of the influence of particular covariates as they are sequentially included in the model, and only provide indications of the relative influence of a covariate. Nonetheless, they can provide invaluable qualitative insight as to how the inclusion of certain covariates may impact the underlying year trend estimated by the model (Bentley *et al.*, 2012).

Results

Correspondence between observer and logbook data

Comparing the natural logarithm of observer and logbook CPUE on a set-by-set basis using the standard linear regression model (Equation (1); Figure 2) revealed significant correlation between the two data sources ($R^2 = 0.912, p < 0.001$). It is important to note that log-transformations down-weight the influence of extreme CPUE values on our results. A back-transformed regression of observer and logbook CPUE can be found in Supplementary Figure S1. The fitted regression between the logbook and observer CPUE was

$$\begin{aligned} & \log\left(Observer\ CPUE_i + 0.001\right) \\ & \quad = 0.085 + 0.988 * \log\left(Logbook\ CPUE_i + 0.001\right), \quad (4) \end{aligned}$$

where the estimated intercept was 0.085 (95% *CI*: 0.076–0.093) and the estimated slope was 0.988 (95% *CI*: 0.984–0.992). Perfect correspondence between log observer and logbook CPUE values would result in an estimated intercept of 0 and a slope of 1. Therefore, across the range of log (CPUE) values observed (~ -7–4), observers tended to report slightly larger catch-rates on average relative to those reported in vessel logbooks, where the predicted regression line approaches the 1:1 line as log (CPUE) values increase (Figure 2). However, discrepancies from the predicted regression line and the 1:1 line appear minimal across the range of values observed. Median-based regression demonstrated similar results when compared to the standard linear regression model, where the estimated intercept was 0.074 (median absolute deviation:

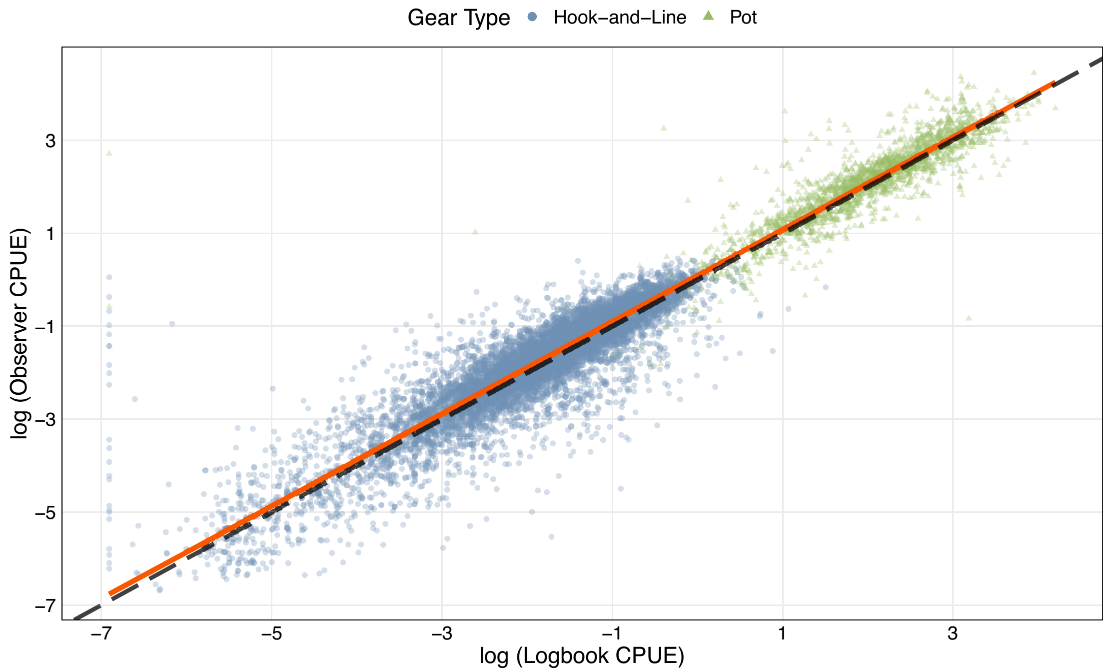

Figure 2. Comparison of observer and logbook CPUE on a set-by-set basis (kg per hook for hook-and-line gear, kg per pot for pot gear) (intercept = 0.085, slope = 0.988, R2 = 0.912). Blue circular points indicate paired observations for hook-and-line gear. Green triangular points indicate paired observations for pot gear. The black dashed line represents a 1:1 relationship. Points that lie above the dashed line represent observations where observer CPUE are larger than logbook CPUE, vice versa. The solid red line represents the model predicted trendline. Confidence intervals are not shown here because they were too narrow to be visible.

0.167), and the estimated slope was 0.983 (median absolute deviation: 0.100).

Factors affecting correspondence between observer and logbook data

Across all model subsets that were explored, BIC identified the final model as Equation (2), which contained all explanatory variables considered. The deviance explained for the final model was low (4.95%), suggesting that these differences are likely attributed to random noise and measurement error that might not be reflected with an additional covariate (Supplementary Figure S2). Predictions for the average log difference of observer CPUE and logbook CPUE demonstrated that throughout the time series for which paired records were available (2004–2020), observers generally reported higher CPUE values relative to those reported in logbooks for all gear types and regions, with the largest differences observed in 2004 and the smallest differences in 2020 (Figure 3a). Across the two different gear types, larger discrepancies in CPUE values occurred when pot gear was employed, which were statistically significant ($p < 0.01$) despite some degree of overlapping between the 95% *CI*, likely due to large sample sizes (Figure 3b). Differences were also found among the six different management areas, where model predicted average differences between the observer and logbook CPUE were highest in the AI and lowest in the BS (Figure 3c). Finally, significant negative correlations were found between the average difference of observer and logbook CPUE and vessel length (ft), albeit these differences were minimal (Figure 3d) ($p < 0.001$). Although differences between data sources were statistically significant across factors, the overall correspondence between observations was strong (Figure 2), suggesting that equally weighting observer and logbook data sources for CPUE standardization is appropriate.

CPUE standardization—model selection

To quantify interannual variation in sablefish abundance using fishery-dependent data, a range of nested candidate CPUE index standardization models (63 model variants) were explored that included different candidate predictor variables. Model selection showed that model fit and out-of-sample predictive performance metrics were extremely similar across the top-performing candidate models (Table 2). We only present the top 5 models for brevity. In general, all candidate models appeared to explain a large proportion of deviance (~85.5%). BIC was the lowest for model 2 (Equation (3) without the average spatial field). For metrics assessing out-of-sample predictive performance, RMSE, MAE, and R^2 values also indicated that model 2 (excluding the average spatial field) was considered the optimal candidate model for out-of-sample prediction. Considering the consistent performance of model 2 across all metrics that were evaluated (Table 2), the following results discussed below will be based upon the formulation for model 2 (excluding the average spatial field). Partial effect plots and model diagnostics for model 2 are presented as supplementary material (Supplementary Figures S3 and S4).

CPUE standardization—comparison of year trends

The standardized CPUE trend across time obtained from model 2 showed a relatively stable population at the begin-

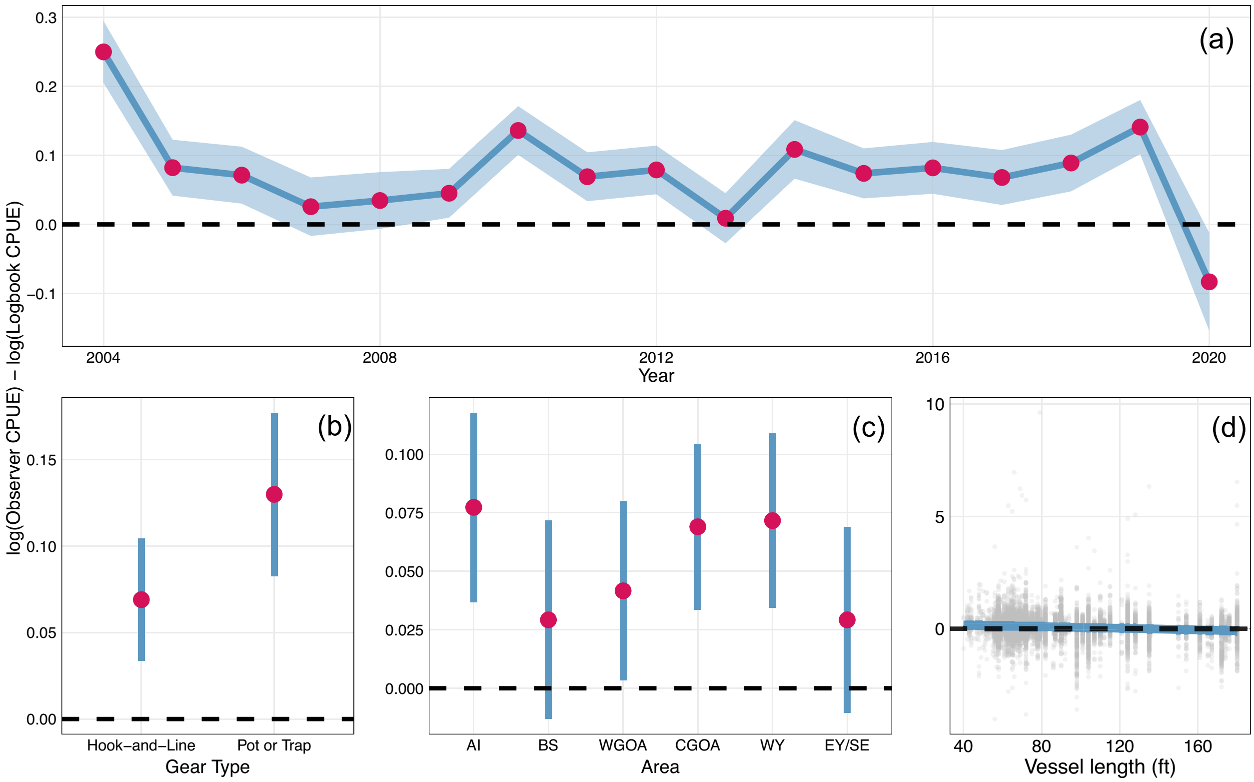

Figure 3. Model predictions of marginal effects of variables predicting the average log difference (log (Observer CPUE) – log (logbook CPUE)), conditional on all other variables being held at their respective averages. A) The average predicted log difference in CPUE across years. Points that lie above the black dashed horizontal line indicate periods where observer reported CPUE are on average larger than logbook reported CPUE. Shading represents approximate mean 95% confidence intervals. Panels B) and C) show the average predicted log difference in CPUE across gear types (B) and fishery management regions (C), respectively. AI, BS, WGOA, CGOA, WY, and EY/SE represent the Aleutian Islands, Bering Sea, Western Yakutat, Western Gulf of Alaska, Central Gulf of Alaska, and Eastern Yakutat/Southeast respectively (Figure 1). Panel D shows the predicted relationship between vessel length (ft) and the average predicted log difference in CPUE (blue line). Grey points indicate individual observations.

ning of the time series (1995–2004), followed by a steady decline (2005–2016), and a recent stabilization (2017–2020), which coincides with high recruitment events observed by the NOAA AFSC longline survey (Figure 4a; Goethel *et al.*, 2021). Similarly, the survey RPW demonstrated a fairly stable population index from 1995 to 2006, followed by similar declines from 2007 to 2015, but with diverging trends towards the end of the time-series. The RPW trend increased rapidly beginning in 2018 and the standardized CPUE index increases at a much slower rate. The nominal CPUE trend (arithmetic mean of CPUE within each year) does not show the aforementioned declines observed in both the standardized and survey indices, where it indicates a stable population from 1995 to 2016. However, the nominal fishery CPUE and survey RPW abundance indices demonstrate agreement from 2019 to 2020 (increasing trend), unlike the standardized index. Nonetheless, such agreement likely misrepresents the magnitude of recruitment events observed in the fishery, and predominately results from failing to consider differences in units of effort between hook-and-line gear and pot gear (Figure 4a). Specifically, the pot gear regulatory shift occurred in 2017, which coincides with when the nominal index began increasing. These increases observed in the nominal index are mostly driven by averaging observations across different units of effort, and not necessarily high recruitment events, considering that such increases are first observed in 2019 for the survey index (Figure 4a).

Comparison of nominal and standardized CPUE with survey estimates

Overall, the standardized CPUE index exhibited a much stronger correlation with the survey RPWs compared with the nominal CPUE index, which was largely uncorrelated with the survey index, when evaluated with a 1-year time lag given differences in fishery and survey selectivity. Evaluating the cross-correlation function for the nominal index against the survey RPWs indicated that the largest significant correlation occurred in the absence of time-lags ($p = 0.04$; Supplementary Figure S5, first row) when the entire time series was included. However, upon sequential removal of years from the time series, such correlations were no longer detected, suggesting spurious correlations driven by a large leverage point during the year 2020 (Figure 4a). In contrast, large significant correlations of 1-year time-lags were identified consistently when evaluating the standardized index against the survey RPW (Supplementary Figure S5). Significant 1-year lags are expected given differences in estimated selectivity between the survey and the fishery (Goethel *et al.*, 2021). In addition, the survey is designed to representatively index the age structure of the population, while fishers tend to target larger and older age-classes, which are more economically desirable (Goethel *et al.*, 2021).

Considering the lack of a consistent time-lag identified for the nominal index, but a consistent and significant 1-year time-lag identified between the standardized CPUE and the

Table 2. Summary of model selection information for the top 5 candidate models identified. Metrics used to assess model performance include deviance explained, BIC, RMSE ± SD, MAE ± SD, and R^2 ± SD. Bolded values represent model(s) with the best performance for a given metric.

Model	Factor variables	Continuous variables	Spatial variables	Deviance explained (%)	BIC	RMSE ± SD	MAE ± SD	R^2 ± SD
Model 1	Year, fishery management area, vessel length, gear, dataset, target strategy, and processed type	Day of year, bottom depth	Average spatial field for CPUE, average spatial field for CPUE by gear	85.55	2 288 827	1141.246 ± 14.767	744.022 ± 3.747	0.403 ± 0.008
Model 2	Year, fishery management area, vessel length, gear, dataset, target strategy, and processed type	Day of year, bottom depth	Average spatial field for CPUE by gear	85.55	**2 288 827**	**1141.246 ± 14.767**	**744.022 ± 3.747**	**0.403 ± 0.008**
Model 3	Year, vessel length, gear, dataset, target strategy, and processed type	Day of year, bottom depth	Average spatial field for CPUE, average spatial field for CPUE by gear	85.55	2 288 889	1141.550 ± 14.909	744.260 ± 3.66	0.403 ± 0.008
Model 4	Year, vessel length, gear, dataset, target strategy, and processed type	Day of year, bottom depth	Average spatial field for CPUE by gear	85.55	2 288 888	1141.548 ± 14.895	744.262 ± 3.654	0.403 ± 0.008
Model 5	Year, fishery management area, vessel length, gear, dataset, target strategy, and processed type	Day of year, bottom depth	Average spatial field for CPUE by gear	85.54	2 289 958	1148.282 ± 14.039	747.539 ± 3.527	0.395 ± 0.008

survey RPWs, we regressed both indices with a 1-year lag for comparison purposes. For the nominal CPUE index (1996–2020), we found minimal correlation nor a significant relationship ($R^2 < 0.001$, $p = 0.99$; Figure 4b). Conversely, we found a positive and significant correlation ($R^2 = 0.64$, $p < 0.001$) when standardized CPUE (1996–2020) and survey RPWs (1995–2019) were regressed (Figure 4c). A major outlier in the otherwise strong and positive relationship between the standardized CPUE and survey RPWs occurred during the years 2019–2020, which is a likely result of recent large recruitment events and differential selectivity between the survey and the fishery (Goethel *et al.*, 2021).

Step plots—year trends

"Step" plots generated following the formulation of model 2 shows that when gear type (i.e. hook-and-line vs. pot gear) and dataset (i.e. logbook and observer) are included into the model, the year trend changes drastically, suggesting that differences in catch rates among gear types and data sources have a large influence on the apparent year index. If not accounted for, this has the potential to obscure underlying trends in stock biomass (Figure 5). Further, incorporating a smooth spatial field for each gear type as a covariate appeared to dampen the recent increases in CPUE observed from 2018 to 2020, thus suggesting the importance of accounting for spatial variation in catchability across gear types (Figure 5).

Discussion

The results of this study provide a framework for developing fishery-dependent indices of abundance in cases where catch records come from multiple gear types and are available from different sources. Here, we illustrate the utility of our proposed framework and the validity of indices of abundance derived using this framework. Additionally, we demonstrated the consequences of naively treating CPUE as a nominal index, and differential units of catchability and effort across space when developing intercalibrated indices of abundance across multiple fishing gears.

Comparison of observer and logbook CPUE

The direct comparison of observer and logbook CPUE demonstrated minor discrepancies between the two data sources when compared on the log scale (Figure 2). Furthermore, the model estimated intercept and slope both revealed that across the range of CPUE values observed, observers tended to report slightly higher catch rates on average, although the difference was negligible. Considering that catch rates derived from logbooks are estimates (Sigler and Lunsford, 2001), and that observer catch rates are collected using a standardized methodology, discrepancies between the two data sources are not surprising. Similar to past studies that have compared catch rates derived from both logbook and observer programs (Starr and Vignaux, 1997; Walsh *et al.*, 2002; Baum and Blanchard, 2010), we also found close correspondence between the two data sources, supporting the validity of using both logbook and observer data to develop intercalibrated standardized indices of abundance. Nevertheless, understanding the nature of logbook data (e.g. data collection patterns, fishery tactics), and cooperation with management bodies are important considerations in utilizing fishery logbook programs for the development of fishery-dependent indices, which are likely to be

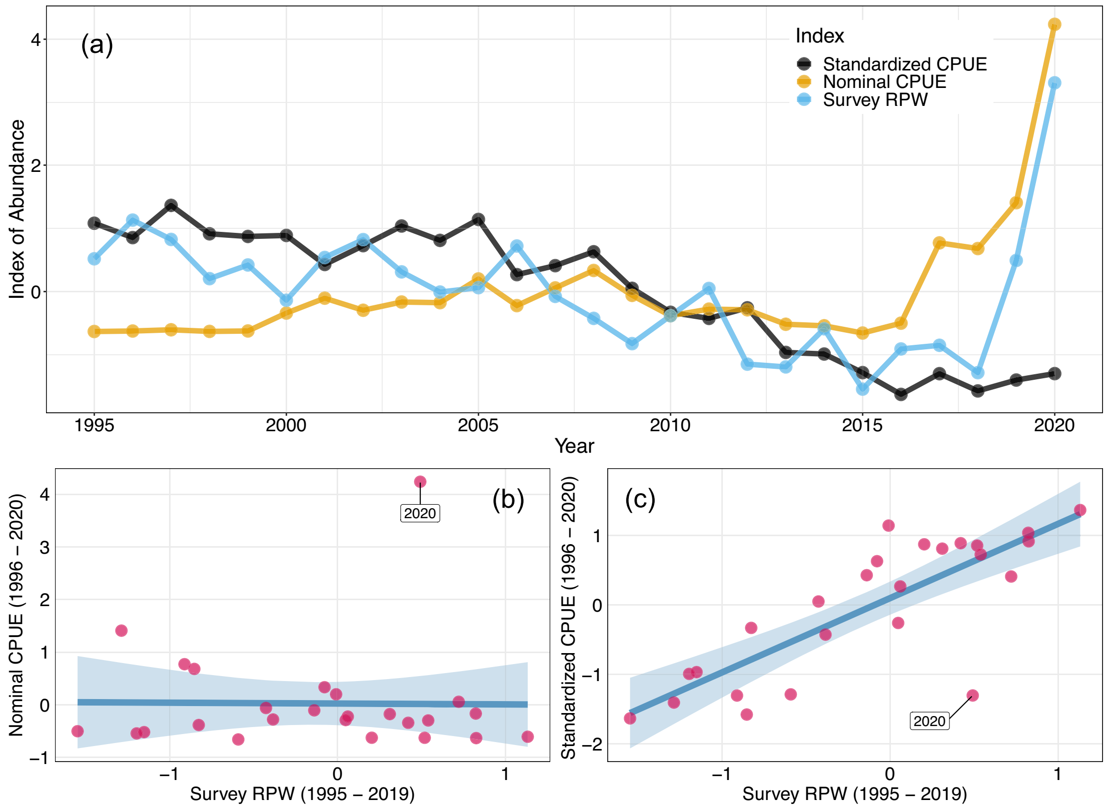

Figure 4. Estimated relative indices of abundance for sablefish in Alaska and comparisons of nominal and standardized indices with survey derived estimates for relative population weight (RPW). A) Z-scaled relative indices of abundance from the survey (blue) and fishery (nominal index: orange, standardized index: black). Panels B and C show survey derived RPWs regressed against the standardized fishery CPUE and nominal fishery CPUE, respectively, with a 1-year lag (survey RPW: 1995 – 2019, nominal and standardized CPUE: 1996 – 2020). The solid blue line represents the trendline predicted via linear regression (Nominal CPUE: R2 <0.001, p =0.95; Standardized CPUE: R2 = 0.64, p < 0.01), and the blue shading indicates corresponding 95% confidence intervals. The year 2020 references the fishery indices (2019 data point for the survey) and is labelled to highlight its large deviation from the estimated trendline.

fishery and region specific. Thus, we urge analysts to consider comparing catch rates derived from logbooks and observer programs as a first step before developing an integrated index of abundance, in order to better understand the potential for biases stemming from differences in observation methods.

Across all factors that were explored, we found that year, gear type, area, and vessel size resulted in small differences on average between observer and logbook CPUE, where observers reported slightly larger CPUE values on average in comparison to paired logbook records (Figure 3). However, the % deviance explained for the model elucidating differences between the average log difference of observer and logbook CPUE was low (4.95%). Although differences appeared negligible in the present study, developing an understanding of factors driving any differences is valuable when intercalibrating data sources for CPUE standardization, which can facilitate fine-scale considerations of data weighting schemes and overall comparability. Nonetheless, our model suggests that observers reported slightly larger CPUE values relative to logbook records consistently across years (Figure 3a), likely attributed to reported catch from logbooks being coarse estimates (i.e. fisher estimation error) (Sigler and Lunsford, 2001). However, it is important to acknowledge that observer catch rates are also estimates (i.e. subsamples of catch weight are extrapolated to a given set) and thus could also contribute to said differences. The average log difference of observer and logbook CPUE reached a time series low in the year 2020, presumably due to a small, unrepresentative sample of observations; 2020 had fewer matched observations ($n = 163$) relative to other years (average number of matched observations per year excluding year 2020 = 675). Conversely, the time series high in year 2004 coincides with a sharp increase in logbook participation (Goethel *et al.*, 2020) and may suggest initial unfamiliarity with data collection methods associated with the logbook program.

Pot gear revealed slightly larger differences than hook-and-line gear, which may stem from how catch is sorted, observed, and processed when employing different gear types. Area-specific differences were highest in the AI and lowest in the BS on average (Figure 3c), but these differences were minimal. Nonetheless, these patterns could be driven by regional differences in fisher behavior, but this remains unclear from the available data. While we found negative and significant correlations between the average log difference of observer and logbook CPUE and vessel length (ft) (Figure 3d), this relationship appears to be weak and uncertain, particularly at larger values of vessel length, which were sparsely observed (Figure 3d). Thus, we do not attempt to draw strong conclusions upon this apparent relationship. Previous research has highlighted that at-sea observers on smaller vessels may be disruptive to typical fishing operations and may affect the performance of the reporting of catch from observers and vessel logbooks (Benoît

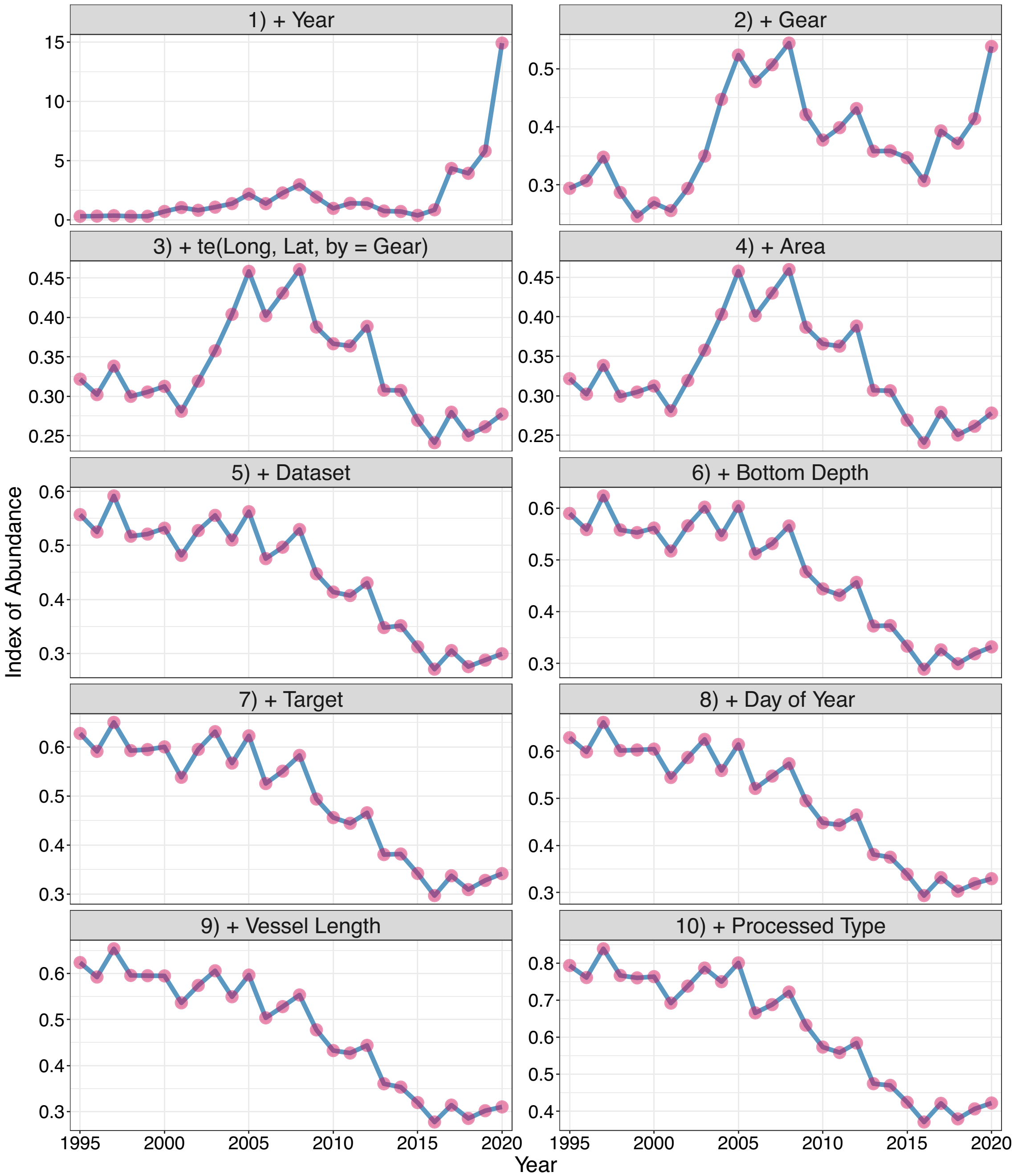

Figure 5. Step plots displaying changes in the standardized CPUE index trend for Alaska sablefish with the addition of explanatory variables. Numbers denoted across panels represents the order in which explanatory variables were sequentially added into the model. As such, the bottom right panel represents the trend described by the formulation of model 2 (see Table 2).

and Allard, 2009; Gilman *et al.*, 2019). Consequently, the utility of electronic-monitoring (EM) systems (video monitoring) have resulted in increasing EM deployments in some fisheries. (Emery *et al.*, 2018; Gilman *et al.*, 2019). In Alaska, EM systems are currently being utilized on a portion of observed fixed gear vessels (AFSC and AKRO, 2021). EM systems have been shown to be representative of fishing activities as reported by observers in other fisheries, and could supplement and/or complement human-observer effort on smaller vessels given the potential impacts on fishing operations (Ruiz *et al.*, 2015; Bartholomew *et al.*, 2018). Thus, these data could also potentially be utilized to develop an intercalibrated index of abundance; for Alaska sablefish, this will require greater consistency and collection of necessary auxiliary information (e.g. depth of fishing, vessel length, and appropriate units of effort) to be usable.

CPUE standardization

Across the top candidate models, we found that model performance was similar, irrespective of different model comparison metrics. The candidate models presented in Table 2 accounted for a large amount of deviance explained (approximately 85%) with similar values for both in-sample (deviance explained and BIC) and out-out-sample performance (RMSE, MAE, and R^2) across all models. Our study highlights model 2 as our final preferred model for index standardization, however, other models presented in Table 2 would have likely sufficed for informing interannual trends in biomass as a final model considering the broad similarities in model selection metrics across all models. Interestingly, the average spatial field covariate was excluded from three of the top performing models, whereas the average spatial field by gear covariate was supported in all top performing models (Table 2). This suggests that the average spatial field covariate may not have provided substantial improvements in predictive power and accounting for spatially varying catchability for each gear type likely accounted for much of the variation resulting from spatial differences in gear-specific fishing effort. Nonetheless, conducting larger folds of cross-validation could have resulted in more precise estimates of RMSE, MAE, and R^2, and a clearer choice for selection of the final model but was not done due to the computational expense involved.

The year trend extracted from the standardization process indicated an initial stable population, followed by a gradual decrease and recent stabilization (Figure 4a). This pattern is similar to the survey index and generally conforms to our understanding of sablefish dynamics in Alaska, where the population was fairly stable in the early 2000s, followed by population declines. However, the standardized index does not capture increases in the survey RPW index starting in 2019, stemming from high recruitment events in 2014, 2016, and 2017. Compared to the survey and standardized indices, the arithmetic mean nominal CPUE trend indicates a steady population followed by unprecedented increases from 2017 to 2020, and fails to capture population declines that are apparent from the standardized and survey indices, from 2005 to 2015. Although the nominal and survey indices demonstrate similar increasing trends from 2019 to 2020, such marked increases in the nominal index begin in 2017, which coincide with the pot gear regulatory shift in the Gulf of Alaska (Figure 4a). Considering that the fishery tends to target larger fish, and that the survey provides a representative index of the age structure within the population (Goethel *et al.*, 2021), the nominal index likely mischaracterizes the timing and magnitude of recruits observed in the fishery. This suggests that unreliable and overly optimistic conclusions may be drawn if this index is considered alone (Figure 4a). The unrealistic increase exhibited by the nominal CPUE trend is likely a result of naively treating CPUE as a ratio-estimator and failing to consider differences in the units of effort between the two gear types (i.e. many sablefish may be caught per pot, while only one sablefish per hook is possible). These conclusions are further supported by a lack of consistent and significant correlation between the nominal index and survey RPWs and a 1-year time-lag relationship identified for the standardized index (Figure 4b and c). Consequently, our findings further demonstrate the importance of the CPUE standardization process and the need to control for the effects of factors unrelated to year-to-year changes (e.g. gear type; Maunder and Punt, 2004), especially when developing a multi-gear fishery-dependent index of abundance.

Similarly, visualizations of "step" plots indicated that the inclusion of gear type greatly altered the index of abundance and suggests that it may be the most influential covariate in our CPUE standardization process, further highlighting the need to consider differences in the units of effort between hook-and-line and pot gear. Interestingly, the addition of the data source factor (logbook and observer) also resulted in a large change in the index of abundance, contrary to what our previous analyses indicated (Figure 2). This potentially suggests that despite logbook and observer data reporting similar catch rates, the presence of an observer (i.e. observer effect) might impact how and where fishing operations occur (Babcock *et al.*, 2003) and, in turn, associated catch rates. Thus, we further recommend analysts to consider controlling for potential effects resulting from having an observer on board, especially when developing fishery-dependent indices that utilize both observer and logbook data. In addition, the inclusion of a spatial field by gear type appeared to dampen the recent estimated upward trend (2017–2020), highlighting that spatial processes resulting in variation in catch rates among gear types are an important consideration, which is in agreement with other studies (Walters, 2003; Walter *et al.*, 2014; Grüss *et al.*, 2019). The "step plots" presented in this study provide a qualitative measure of the incremental contribution of covariates to the estimated year trend rather than their overall influence on the final model and thus should be interpreted with caution. Nonetheless, we recommend analysts explore the use of "step plots" when conducting CPUE standardization, which can provide additional insight into the standardization effects of explanatory variables (Bentley *et al.*, 2012). We also recommend that analysts evaluate standardized models for each data source and gear type as independent indices, prior to combining data sources and gear types. This process may provide helpful context in understanding the relative impact of different data sources and gear types, when developing intercalibrated indices of abundance (see Supplementary Material 2).

There were several factors that could have influenced the year trend extracted from the standardization process but could not be included in this analysis due to data limitations. Specifically, we did not consider fishing duration information (e.g. soak times) as an additional measure of fishing effort because it was not available. Although, hook-spacing has been suggested to affect catch rates of targeted species for hook-and-line gear (e.g. through hook-competition; Monnahan and Stewart, 2018), the standardized index of abundance does not account for this source of variation because it was reported irregularly for both logbook and observer data. Nevertheless, we do not expect these unaccounted factors to have greatly influenced the interpretation of these findings.

The NOAA AFSC longline survey is considered representative of sablefish dynamics in Alaska, and thus provides a novel basis for comparison and model validation of standardized CPUE indices (Sigler, 2000). Strong correlations between model estimated year trends (1996–2020) and the survey estimated RPWs (1995–2019) (Figure 4c) indicate that model the fishery-dependent model estimates of relative abundance are representative of sablefish dynamics in Alaska when regressed with a 1-year lag. Furthermore, evaluating the cross-correlation function for the standardized index against survey RPWs demonstrated consistent and significant 1-year lags. However, survey estimates for 2020 were not in agreement with model estimates in 2019, likely due to a large influx of juvenile sablefish observed by the survey that were not vulnerable and/or primarily targeted by the fishery at the time (i.e. differences in gear selectivity and/or fish availability). In addition, no consistent correlations were identified by a cross-correlation function between the arithmetic mean nominal CPUE and the survey estimates, demonstrating that naively treating CPUE as a nominal ratio-estimator would result in unreliable and spurious conclusions on stock status that must be corrected through model-based standardization approaches.

CPUE standardization is integral to ensuring that fishery-dependent indices of abundance are representative and can be used with confidence to inform stock assessment (Maunder and Punt, 2004). The present study illustrates that both logbook and observer data exhibit close correspondence in reported catch rates for the sablefish fishery in Alaska, and the efficacy of developing a multi-gear, fishery-dependent index of abundance utilizing both observer and logbook data sources. Our results provide an applicable framework for other fisheries that seek to leverage multiple fishery-dependent data sources and highlights important considerations when developing indices of abundance to support multi-gear fisheries. Within integrated stock assessments, abundance indices are often associated with a fleet-specific selectivity pattern. Although there has been growing interest in developing intercalibrated abundance indices that leverage multiple gear types (Perretti and Thorson, 2019; Thorson *et al*., 2020), it is unclear how these indices should be treated within integrated stock assessments. Given the potential for differential selectivity patterns among gear types to obscure fishery selection patterns, there remains a need to determine best practices for utilizing multi-gear indices, which will likely depend on the treatment of fleet structure, fishery dynamics, and the degree to which selectivity patterns differ (i.e. contrasting selectivity patterns; dome-shaped and logistic). Nonetheless, when *a priori* knowledge of selectivity patterns for multi-gear indices do not exist, assuming flexible forms of selectivity (e.g. splines, double normal) could be robust (Privitera-Johnson *et al*., 2022). In the context of the present study, hook-and-line and pot selectivity has been shown to demonstrate comparable patterns in contact selectivity, although further investigations are required to draw more robust conclusions in this respect (Sullivan *et al*., 2022). Further, it remains unclear how selectivity may be impacted by availability processes due to spatial differences in fleet operations.

Future research could expand upon this research through simulation analysis of: (1) contrasting patterns of observer and logbook spatiotemporal coverage, (2) various levels of observer bias, (3) different levels of preferential sampling (e.g. Ducharme-Barth *et al*., 2022), (4) evaluating model performance of different CPUE standardization methods (e.g. vector autoregressive spatiotemporal models, boosted regression trees) (Mateo and Hanselman, 2014; Thorson, 2019), and (5) best practices for selectivity treatments when using multi-gear indices. While the methods here may be improved upon with future research, our approach represents a large step forward in improving the incorporation of fishery-dependent data sources into the stock assessment and fishery management process.

Acknowledgements

We would like to thank the Alaska sablefish industry for their great efforts in providing vessel logbooks, the International Pacific Halibut Commission for collecting logbooks dockside, providing quality control, and maintaining these data, and the NOAA North Pacific Observer Program for their long-term data collection. This project would not have been possible without their commitment. We also thank three anonymous reviewers for their insightful comments, which greatly improved this manuscript. This publication is funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CIOCES) under NOAA cooperative agreement NA20OAR4320271, contribution number 2022–1232. Findings and conclusions of this study are those of the authors, and do not represent the views of the National Marine Fisheries Service, NOAA.

Supplementary data

Supplementary material is available at the *ICESJMS Journal* online version of the manuscript.

Authors' contributions

MC and CC designed the study. CR provided data for this manuscript. MC, CR, JL, and CC contributed to the development of the methodology. MC wrote the initial draft of the manuscript and all authors contributed to writing and editing of the manuscript and approved the final draft.

Conflict of Interest statement

The authors have no conflict of interest to declare.

Data availability

Fishery data may be shared with the public in an aggregated or summarized format that maintains confidentiality requirements under Public Law 94-265. Confidential data shall only be disclosed to the public if required by the Freedom of Information Act (FOIA), 5 U.S.C. 552, the Privacy Act, 5 U.S.C. 552a, or by court order. These data may be available to individuals under data sharing agreements with the NOAA. To request NOAA fishery data or determine eligibility for a data sharing agreement, please contact the Director of the Fishery Monitoring and Analysis Division, from the NOAA Alaska Fisheries Science Center (https://www.fisheries.noaa.gov/about/fisheries-monitoring-and-analysis). Additionally, the IPHC data use and confidentiality policy requires that data be shared in an aggregated or summarized format or may be available at a finer level of stratification with written authorization from each data source (https://iphc.int/the-commission). Data requests for research purposes should be made through the IPHC Executive Director (https://iphc.int/staff//).

References

AFSC (Alaska Fisheries Science Center)and AKRO (Alaska Regional Office). 2021. North Pacific Observer Program 2020 Annual Report. AFSC Processed Report 2021-03. Alaska Fisheries Science Center, NOAA National Marine Fisheries Service, Seattle, WA. 143pp.

AFSC. 2022. Observer Sampling Manual. Fisheries Monitoring and Analysis Division, North Pacific Groundfish Observer Program. AFSC, Seattle, WA. 584pp.

Bartholomew, D. C., Mangel, J. C., Alfaro-Shigueto, J., Pingo, S., Jimenez, A., and Godley, B. J. 2018. Remote electronic monitoring as a potential alternative to on-board observers in small-scale fisheries. Biological Conservation, 219: 35–45.

Baum, J. K., and Blanchard, W. 2010. Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. Fisheries Research, 102: 229–239.

Benoît, H. P., and Allard, J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards?. Canadian Journal of Fisheries and Aquatic Sciences, 66: 2025–2039.

Bentley, N., Kendrick, T. H., Starr, P. J., and Breen, P. A. 2012. Influence plots and metrics: tools for better understanding fisheries catch-per-unit-effort standardizations. ICES Journal of Marine Science, 69: 84–88.

Ducharme-Barth, N. D., Grüss, A., Vincent, M. T., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., *et al.* 2022. Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: a simulation study and fishery application. Fisheries Research, 246: 106169.

Emery, T. J., Noriega, R., Williams, A. J., Larcombe, J., Nicol, S., Williams, P., Smith, N., *et al.* 2018. The use of electronic monitoring within tuna longline fisheries: implications for international data collection, analysis and reporting. Reviews in Fish Biology and Fisheries, 28: 887–907.

Erisman, B. E., Allen, L. G., Claisse, J. T., Pondella, D. J., Miller, E. F., and Murray, J. H. 2011. The illusion of plenty: hyperstability masks collapses in two recreational fisheries that target fish spawning aggregations. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1705–1716.

Fox, D. S., and Starr, R. M. 1996. Comparison of commercial fishery and research catch data. Canadian Journal of Fisheries and Aquatic Sciences, 53: 2681–2694.

Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., and Barkai, A. 2019. Increasing the functionalities and accuracy of fisheries electronic monitoring systems. Aquatic Conservation: Marine and Freshwater Ecosystems, 29: 901–926.

Goethel, D. R., Rodgveller, C. J., Echave, K. B., Shotwell, S. K., Siwicke, K. A., Malecha, P. W., Cheng, M., *et al.* 2022. Assessment of the Sablefish Stock in Alaska. North Pacific Fishery Management Council, Anchorage, AK. 182pp.

Goethel, D., Hanselman, D. H., Rodgveller, C. J., Fenske, K. H., Shotwell, S. K., Echave, K. B., Malecha, P. W., *et al.* 2020. 3. Assessment of the Sablefish Stock in Alaska. North Pacific Fishery Management Council, Anchorage, AK. 257pp.

Goethel, D., Hanselman, D., Rodgveller, C., Echave, K. B., Williams, B., Shotwell, S. K., Sullivan, J., *et al.* 2021. 3. Assessment of the Sablefish Stock in Alaska. North Pacific Fishery Management Council, Anchorage, AK. 347pp.

Grüss, A., Walter, J. F., Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., and Schirripa, M. J. 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fisheries Research, 213: 75–93.

Hanselman, D., Rodgveller, C., Fenske, K., Shotwell, S. K., Echave, K., Malecha, P., and Lunsford, C. 2019. Assessment of the Sablefish Stock in Alaska. North Pacific Fishery Management Council, Anchorage, AK. 263pp.

He, X., Bigelow, K. A., and Boggs, C. H. 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Research, 31: 147–158.

Hilborn, R., and Walters, C. J. 1992. Quantitative Fisheries Stock Assessment. Springer, Boston, MA. http://link.springer.com/10.1007/978-1-4615-3598-0 (last accessed 20 December 2021).

Kai, M. 2019. Spatio-temporal changes in catch rates of pelagic sharks caught by Japanese T research and training vessels in the western and central North Pacific. Fisheries Research, 19: 177–195.

Mateo, I., and Hanselman, D. H. 2014. A comparison of statistical methods to standardize catch-per-unit-effort of the Alaska longline sablefish fishery. NOAA Technical Memorandum NMFS-AFSC-269. National Oceanic and Atmospheric Administration, Springfield, VA. 81pp.

Maunder, M. N., and Punt, A. E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research, 70: 141–159.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Monnahan, C. C., and Stewart, I. J. 2018. The effect of hook spacing on longline catch rates: implications for catch rate standardization. Fisheries Research, 198: 150–158.

Nakano, H., and Clarke, S. 2006. Filtering method for obtaining stock indices by shark species from species-combined logbook data in tuna longline fisheries. Fisheries Science, 72: 322–332.

National Oceanic and Atmospheric Administration. 2001. Fisheries of the Exclusive Economic Zone Off Alaska; Revisions to Recordkeeping and Reporting Requirements; Proposed Rule. Department of Commerce, 50 CFR Part 679. National Oceanic and Atmospheric Administration, Washington, DC.

Perretti, C. T., and Thorson, J. T. 2019. Spatio-temporal dynamics of summer flounder (*Paralichthys dentatus*) on the Northeast US shelf. Fisheries Research, 215: 62–68.

Privitera-Johnson, K. M., Methot, R. D., and Punt, A. E. 2022. Towards best practice for specifying selectivity in age-structured integrated stock assessments. Fisheries Research, 249: 106247.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/ (last accessed 21 February 2023).

Rose, G. A., and Kulka, D. W. 1999. Hyperaggregation of fish and fisheries: how catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined, 56: 10.

Ruiz, J., Batty, A., Chavance, P., McElderry, H., Restrepo, V., Sharples, P., Santos, J., *et al.* 2015. Electronic monitoring trials on in the tropical tuna purse-seine fishery. ICES Journal of Marine Science, 72: 1201–1213.

Rutecki, T. L., Rodgveller, C. J., and Lunsford, C. R. 2016. National Marine Fisheries Service longline survey data report and survey history, 1990-2014. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC: 329. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle WA.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fisheries Research, 93: 154–162.

Sigler, M. F. 2000. Abundance estimation and capture of sablefish (*Anoplopoma fimbria*) by longline gear, 57: 14.

Sigler, M. F., and Lunsford, C. R. 2001. Effects of individual quotas on catching efficiency and spawning potential in the Alaska sablefish fishery. Canadian Journal of Fisheries and Aquatic Sciences, 58: 1300–1312.

Starr, P. J., and Vignaux, M. 1997. Comparison of data from voluntary logbook and research catch-sampling programmes in the New Zealand lobster fishery. Marine and Freshwater Research, 48: 1075.

Sullivan, J., Dimond, J. A., and Malecha, P. 2022. Slinky pot and hook-and-line comparison project during the experimental leg of the 2021 AFSC sablefish longline survey. Alaska Fisheries Science Center, NOAA, National Marine Fisheries Service, Auke Bay Laboratories, Juneau, AK. 18pp.

Thorson, J. T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210: 143–161.

Thorson, J. T., Cunningham, C. J., Jorgensen, E., Havron, A., Hulson, P.-J. F., Monnahan, C. C., and von Szalay, P. 2021. The surprising sensitivity of index scale to delta-model assumptions: recommendations for model-based index standardization. Fisheries Research, 233: 105745.

Thorson, J. T., Maunder, M. N., and Punt, E. 2020. The development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance. Fisheries Research, 230: 105611.

Walsh, W. A., Kleiber, P., and McCracken, M. 2002. Comparison of logbook reports of incidental blue shark catch rates by Hawaii-based longline vessels to ®shery observer data by application of a generalized additive model. Fisheries Research, 16: 79–94.

Walter, J. F., Hoenig, J. M., and Christman, M. C. 2014. Reducing bias and filling in spatial gaps in fishery-dependent catch-per-unit-effort data by geostatistical prediction, I. Methodology and simulation. North American Journal of Fisheries Management, 34: 1095–1107.

Walters, C. 2003. Folly and fantasy in the analysis of spatial catch rate data. Canadian Journal of Fisheries and Aquatic Sciences, 60: 1433–1436.

Wolotira, R. J. 1993. Geographic and bathymetric distributions for many commercially important fishes and shellfishes off the West Coast of North America, based on research survey and commercial catch data,1912-84. NOAA Technical Memorandum, NMFS-AFSC-6: 191. National Oceanic and Atmospheric Administration, Washington, DC.

Wood, S. N. 2017. Generalized additive models: an introduction with R. *In* Chapman & Hall/CRC Texts in Statistical Science. 2nd edn. CRC Press/Taylor & Francis Group, Boca Raton, FL. 476pp.

Ye, Y., and Dennis, D. 2009. How reliable are the abundance indices derived from commercial catch–effort standardization?. Canadian Journal of Fisheries and Aquatic Sciences, 66: 1169–1178.

Handling Editor: Shijie Zhou