frontiers | Research Topics

# Design change to fishery independent surveys: when to adjust and how to account for it

**Edited by**
Fabio Campanella, Ben Scoulding, Nathan Bacheler
and Kyle Shertzer

**Published in**
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-6106-5
DOI 10.3389/978-2-8325-6106-5

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Design change to fishery independent surveys: when to adjust and how to account for it

**Topic editors**
Fabio Campanella — Institute for Biological Resources and Marine Biotechnology, National Research Council (CNR), Italy
Ben Scoulding — Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Nathan Bacheler — Beaufort Laboratory, Southeast Fisheries Science Center (NOAA), United States
Kyle Shertzer — NOAA Southeast Fisheries Science Center, United States

**Citation**
Campanella, F., Scoulding, B., Bacheler, N., Shertzer, K., eds. (2025). *Design change to fishery independent surveys: when to adjust and how to account for it*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6106-5

# Table of contents

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 27 February 2025
DOI 10.3389/fmars.2025.1575151

# Editorial: Design change to fishery independent surveys: when to adjust and how to account for it

Kyle Shertzer[1*], Nathan M. Bacheler[1], Fabio Campanella[2] and Ben Scoulding[3]

[1]Southeast Fisheries Science Center, National Oceanic and Atmospheric Administration, Beaufort, NC, United States, [2]Institute for Biological Resources and Marine Biotechnology, National Research Council (CNR), Ancona, Italy, [3]Oceans and Atmospheric Flagship, Commonwealth Scientific and Industrial Research Organization (CSIRO), Hobart, TAS, Australia

KEYWORDS

**fishery surveys, stock assessment, fisheries, survey design, sampling method**

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Stephen J. Newman,
Western Australian Fisheries and Marine Research Laboratories, Australia

*CORRESPONDENCE
Kyle Shertzer
kyle.shertzer@noaa.gov

RECEIVED 11 February 2025
ACCEPTED 18 February 2025
PUBLISHED 27 February 2025

CITATION
Shertzer K, Bacheler NM, Campanella F and Scoulding B (2025) Editorial: Design change to fishery independent surveys: when to adjust and how to account for it.
*Front. Mar. Sci.* 12:1575151.
doi: 10.3389/fmars.2025.1575151



Editorial on the Research Topic
Design change to fishery independent surveys: when to adjust and how to account for it

Fishery independent surveys provide crucial information for monitoring and assessing marine fish stocks and ecosystems. For example, such surveys provide data on temporal fluctuations and trends in abundance, as well as age or length composition of studied populations. These data form the backbone of many stock assessments worldwide. In addition, most fishery independent surveys provide data on physical variables and on multiple species, often across a broad range of taxa, such that the information generated can help to understand and monitor communities of organisms and how they relate to their environments.

The term “fishery independent” differentiates these surveys from sampling that targets fishery operations themselves for such data as landings and discards. It implies that the survey is operated by, or in close collaboration with, scientists. A fishery independent survey generally applies standardized sampling procedures that are consistent across space and time. These procedures specify the statistical sampling design, as well as the gear type, such as handline, longline, trawl, trap, video, or acoustics. The sampling consistency across space and time allows for reasonable inference that any observed dynamics reflect those of the population or community being studied, even when the data need to be standardized to account for unequal sampling or factors outside human control (e.g., environmental variables).

Sometimes, however, changes to the sampling design or gear become desirable or necessary. They may be desirable if the benefits of modification outweigh the benefits of consistency. Examples might include improvements or innovations in sampling gear, technology, or efficiency; revised management or survey priorities; and increased funding that allows for additional gears, greater spatial coverage, or temporal resolution. In other cases, changes to the survey design may not be desirable but necessary. Examples might include reductions in funding, ship time, or human resources; modified mandates imposed by survey administrators; and when requisite supplies or equipment become unobtainable. Ironically, the longer a survey is in existence, the more valuable consistency becomes for evaluating long-term trends, and the more likely a change in survey design—whether desirable or necessary—will be forced or require consideration.

This Research Topic compiled case studies that describe and evaluate changes to fishery independent surveys from a wide range of aquatic systems. The case studies document the rationale for making changes and how those changes were accounted for in monitoring and assessment.

Evaluations of survey design change generally fall into one of two categories: those that evaluate potential or inevitable changes prior to their occurrence and those that evaluate their effects *post hoc*. The former is particularly relevant for emerging technologies or infrastructure. White et al. used a modeling approach to evaluate survey designs for using active acoustics to sample fish aggregations, and they found that a parallel line design outperformed a "star" design in most of the scenarios tested. Bolser et al. evaluated the potential for acoustic data, including data collected by uncrewed surface vehicles, to estimate biomass-at-age of Pacific hake, providing a methodology for estimation along with advice and caveats for application. Methratta et al. considered offshore wind energy development that is now underway in the northeast United States (US). These projects are expected to affect current surveys that have been in place for decades, and the authors evaluated whether project-level monitoring by wind energy developers would be sufficient to mitigate the effects on surveys. They concluded that current efforts were insufficient, and offered recommendations for how to mitigate impacts of offshore wind development on existing fishery independent surveys in their and other systems.

Several other papers evaluated effects of potential, but not yet implemented, changes to the surveys in the Bering and Chukchi Seas. Bryan and Thorson analyzed 1) the performance of spatio-temporal statistical models when estimating relative abundance in a new climate-adaptive spatial stratum and 2) whether annual sampling at reduced intensity or biennial sampling would provide the most informative data, if effort reductions were necessary. DeFilippo et al. evaluated effects of reduced sampling intensity in areas of currently high sampling rates, which could provide useful guidance whether sampling effort is reduced or redistributed. Oyafuso et al. used simulation tests to analyze three different statistical designs for the US Chukchi Sea bottom trawl survey: simple random, stratified random, and systematic. They found best performance from the stratified random design.

In not all cases is it possible to evaluate changes prior to their implementation. Several papers demonstrated the value of *post hoc* evaluations through statistical modeling, with focus on data products used in stock assessments. Along these lines, Hendon et al. evaluated a bottom longline survey and Pollack et al., a long-term groundfish trawl survey, both in the US Gulf of America (also called Gulf of Mexico). They highlighted the positive effects that design changes, including spatial expansion in sampling, had on the survey products. Vecchio et al. evaluated effects of spatial expansion in a trap survey conducted in the US Atlantic. Chang et al. considered the fluctuating sampling protocol of an ichthyoplankton survey in the Hudson River Estuary. Schrandt et al. described the evolution of estuarine surveys in the US state of Florida. They focused on the need to balance utility of long-term data with shifts in funding and management priorities, offered advice on how to do so, and highlighted the benefits of reconnaissance sampling prior to survey modifications.

When possible and funding allows, the effects of changing from one sampling procedure to another can be informed by pairing the two procedures in simultaneous data collection. This pairing of methods allows for direct comparison of data collected before and after the change, with the potential benefit of a continuous time series. Bacheler et al. examined fish counts from a video survey in the US Atlantic that upgraded the video cameras used for sampling. A paired-gear study, using both the old and new cameras, allowed for data calibration such that fish counts could be utilized across the full time series of the survey. Latour et al. described a trawl survey conducted in the Chesapeake Bay, the largest estuary in the US. The survey underwent multiple, simultaneous improvements, including a new sampling vessel, and it utilized paired-tow studies to calibrate data from before and after the change. The authors offered cogent advice that, among other topics, highlights the value of making multiple changes simultaneously when forward planning is feasible.

This Research Topic compiled 13 papers addressing design change to fishery independent surveys. The compilation provides lessons learned from real-world examples across a variety of aquatic systems. Collectively, these papers can inform those in the future faced with potential or inevitable changes to survey design.

## Author contributions

KS: Writing – original draft, Writing – review & editing. NB: Writing – review & editing. FC: Writing – review & editing. BS: Writing – review & editing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

TYPE Original Research
PUBLISHED 28 April 2023
DOI 10.3389/fmars.2023.1183955

Check for updates

OPEN ACCESS

EDITED BY
Francois Bastardie,
Technical University of Denmark, Denmark

REVIEWED BY
Euan S. Harvey,
Curtin University, Australia
Gianna Fabi,
National Research Council (CNR), Italy

*CORRESPONDENCE
Nathan M. Bacheler
nate.bacheler@noaa.gov

RECEIVED 10 March 2023
ACCEPTED 14 April 2023
PUBLISHED 28 April 2023

CITATION
Bacheler NM, Shertzer KW, Schobernd ZH and Coggins Jr. LG (2023) Calibration of fish counts in video surveys: a case study from the Southeast Reef Fish Survey. *Front. Mar. Sci.* 10:1183955. doi: 10.3389/fmars.2023.1183955



# Calibration of fish counts in video surveys: a case study from the Southeast Reef Fish Survey

Nathan M. Bacheler*, Kyle W. Shertzer, Zebulon H. Schobernd and Lewis G. Coggins Jr.

Beaufort Laboratory, Southeast Fisheries Science Center, National Marine Fisheries Service, Beaufort, NC, United States

Changes to sampling gears or vessels can influence the catchability or detectability of fish, leading to biased trends in abundance. Despite the widespread use of underwater video cameras to index fish abundance and the rapid advances in video technology, few studies have focused on calibrating data from different cameras used in underwater video surveys. We describe a side-by-side calibration study (*N* = 143 paired videos) undertaken in 2014 to account for a camera change in the Southeast Reef Fish Survey, a regional-scale, multi-species reef fish survey along the southeast United States Atlantic coast. Slope estimates from linear regression for the 16 species included in the analyses ranged from 0.21 to 0.98, with an overall mean of 0.57, suggesting that original cameras (Canon Vixia HF-S200) observed an average of 43% fewer fish than newer cameras (GoPro Hero 3+). Some reef fish species had limited calibration sample sizes, such that borrowing calibration information from related or unrelated species was justified in some cases. We also applied calibrations to 11-year video time series of relative abundance of scamp *Mycteroperca phenax* and red snapper *Lutjanus campechanus* (*N* = 13,072 videos), showing that calibrations were critical to separating changes in camera sightability from true changes in abundance. We recommend calibrating data from video cameras anytime changes occur, and pairing video cameras to the extent possible to control for the spatial and temporal variability inherent in fish populations and environmental conditions. Following these guidelines, researchers will be able to maintain the integrity of valuable long-term video datasets despite intentional or unavoidable changes to video cameras over time.

KEYWORDS
**fishery-independent survey, calibrate, reef fish, catchability, index of abundance, camera, video, survey**

## Introduction

Estimating the abundance of marine fish or invertebrates over large spatial or temporal scales is typically accomplished with data from scientific surveys, and these abundance estimates or indices are critically informative inputs to stock assessments (Dennis et al., 2015; Maunder and Piner, 2015). A wide variety of sampling gears have been used by scientific surveys to estimate fish abundance (Murphy and Jenkins, 2010; Goethel et al., 2022). Trawls are typically used on unconsolidated sediments and can often be used to estimate absolute abundance or density of fish given the known area sampled (Kimura and Somerton, 2006). On untrawlable habitats like natural or artificial reefs, numerous gears have been used but the resulting abundance estimates are often relative (i.e., indices of abundance) because estimating the area over which sampling occurs is challenging (Maunder and Punt, 2004; Bacheler et al., 2022a). It is nearly universally assumed that indices of abundance from survey data vary in proportion to the actual abundance of the population; in other words, catchability (i.e., the efficiency of a sampling gear) is nearly always assumed to be constant over space and time, even though its absolute value is generally unknown (Hangsleben et al., 2013).

There are several reasons why catchability of a survey gear may not be constant over time. Catchability is often considered to be the product of availability (i.e., proportion of the stock occurring in the survey area) and gear efficiency (i.e., the proportion of animals caught or detected that are available to the gear; Arreguín-Sánchez (1996)). Changes in the spatial footprint of a survey or seasonal or diurnal migrations of a species of interest influence that species' availability to the survey (Aguzzi and Company, 2010). Moreover, environmental variability can influence the efficiency of sampling gears (Bacheler et al., 2014; Bacheler and Shertzer, 2020), which, if left unaccounted for, will be confounded with temporal trends in abundance (Tyre et al., 2003). Another reason why catchability can vary in a survey is due to changes in gears, vessels, or sampling characteristics over time (Pelletier, 1998; Cadigan and Dowden, 2010; Thorson and Ward, 2014). These changes may be unavoidable (e.g., when a survey vessel or outdated equipment needs to be replaced), while other changes may be discretionary (e.g., due to improved performance or ease of use of new gears). Regardless, any of these changes require calibration between the old and new sampling methodologies because any change in gears or vessels can influence the relative catch rates (Pelletier, 1998; Kimura and Somerton, 2006).

Over the last few decades, underwater video has become a common tool for indexing the abundance and distribution of fish species in many places throughout the world (Mallet and Pelletier, 2014; Whitmarsh et al., 2017; Bacheler and Ballenger, 2018). Underwater video has evolved into a valuable sampling gear that can be standardized to provide indices of abundance for a wide variety of pelagic and demersal fish species (Priede and Merrett, 1996; Heagney et al., 2007; Brooks et al., 2011; Santana-Garcon et al., 2014; Bacheler and Ballenger, 2018). Some video surveys use unbaited cameras while others are baited, and different fish communities may be sampled based on bait choices (Harvey et al., 2007; Dorman et al., 2012).

While improvements have been made to most sampling gears over time, underwater video cameras have evolved particularly dramatically over the last half century (Mallet and Pelletier, 2014). The original cameras used to quantify fish species diversity and abundance were large (~ 1 m high, 0.5 m in diameter), low quality, and needed a connection to a power supply on land or ship (Kumpf and Lowenstein, 1962; Myrberg et al., 1969). Nowadays, underwater video cameras are small, cheap, reliable, fully digital, record in high definition, and have small and long-lasting batteries (Struthers et al., 2015). Just as changes to trawl nets or the vessels dragging them influence the catchability of fish, the improvements of video cameras over time likely improved the sightability of fish (i.e., ability to see fish that are present). Even within the advanced underwater video cameras available today, there is enormous variability in video size, shape, color, quality, and light sensitivity that can influence the sightability of fish. Despite the vast changes in video cameras over time and the fact that numerous fishery-independent surveys use video cameras, there are few published examples where fish counts have been calibrated between cameras (but note that calibrations commonly occur when measuring fish length; Harvey and Shortis, 1998; Balletti et al., 2014; Letessier et al., 2015; Shafait et al., 2017).

Here, we describe a calibration study that was undertaken to account for a camera change in the Southeast Reef Fish Survey (SERFS), a large-scale fishery-independent trap and video survey that provides key relative abundance data for many reef-associated fish species along the southeast United States Atlantic coast (hereafter, SEUS) between North Carolina and Florida. There were four objectives of our work. The first objective was to describe the statistical design and methodological approach of our calibration experiment with paired camera given the paucity of examples in the literature. Our second objective was to estimate species-specific calibration factors for multiple economically important reef fish species in the SEUS. Our third objective was to consider alternatives to species-specific calibrations when sample sizes were limited, for instance, by borrowing information from related or unrelated species. The fourth objective was to evaluate two approaches for applying calibration factors between video cameras: calibrating the data before inclusion in a standardization model or after standardization has occurred (i.e., calibrating the index itself). Through this case study, we describe the importance of video calibrations, detail the lessons learned from our video calibration experiment, and provide guidance to researchers around the world on how to calibrate for changes in video sampling gears.

# Materials and methods

## Objective 1: calibration experiment design

Video data for this study were provided by SERFS, a regional-scale trap and video survey occurring in the SEUS. SERFS is made up of three groups that sample reef fish species collaboratively using identical methods. The first group is the Marine Resources Monitoring, Assessment, and Prediction program, housed at the South Carolina Department of Natural Resources, which has been funded by the National Marine Fisheries Service (NMFS) to sample with chevron traps in the region since 1990. The second group is the Southeast Area Monitoring and Assessment Program – South Atlantic Region reef fish complement, which has provided additional funding to South Carolina Department of Natural Resources to conduct reef fish surveys in the region since 2009.

The third group is the Southeast Fishery-Independent Survey, which was created by NMFS in 2010 to work with their partners listed above to increase fishery-independent sample sizes in the SEUS and incorporate underwater video into the survey.

SERFS used a simple random sampling design to select stations for sampling each year. Approximately 2,000 stations were randomly selected for sampling each year out of a sampling frame of approximately 4,300 stations on known hardbottom reef habitat. A majority of stations sampled each year and included in our analyses were randomly selected for sampling, but some stations not selected for sampling were sampled opportunistically to increase sampling efficiency. A small number of new hardbottom sampling stations were discovered and sampled each year, and were included in our analyses if hardbottom was observed on video. Sampled stations were always separated by at least 200 m in a given year to provide independence between samples. Five research vessels have been used to carry out this work: the R/V *Palmetto*, R/V *Savannah*, NOAA Ship *Nancy Foster*, NOAA Ship *Pisces*, and NOAA Ship SRVx *Sand Tiger*. All video sampling occurred during daylight hours between the spring and fall each year.

For 20 years, SERFS used chevron traps alone to sample reef fish species in the SEUS. Chevron traps are large, arrowhead-shaped traps that were baited with *Brevoortia* spp. and soaked for ~90 min (Figure 1). Beginning in 2011, all chevron traps deployed by SERFS included two attached cameras – one placed over the trap mouth that looked outward and used to count fish and quantify seafloor habitat, and one placed over the trap nose that also looked outward, but this second camera was only used to quantify seafloor habitat in the opposite direction of the first camera. In 2011–2014, Canon Vixia HF-S200 video cameras in Gates HF-21 housings were attached over the mouth of the trap and used to count fish. In 2015, GoPro Hero 3+ cameras replaced the Canon cameras because GoPros are smaller, cheaper, higher resolution, and have a larger field of view, which we expected would increase fish counts relative to Canon cameras (Figure 2).

To address our first objective and account for this camera switch, we conducted a calibration experiment during the 2014 field season. We attached Canon Vixia HF-S200 and GoPro Hero 3 + cameras side-by-side on traps, looking outward over the trap mouth (Figures 1, 2). In addition to pairing video cameras in space, we also paired video reading in time (see below). A total of 143 chevron traps were deployed in 2014 that included both video cameras placed side-by-side over trap mouths, looking outward. Two video cameras malfunctioned, leaving 141 paired video samples that were available for reading. Of these, 54 paired samples were deemed to have sufficient numbers of priority fish species (e.g., red snapper) from cursory examinations to make complete reading of these calibration videos worthwhile.

## Objective 2: estimating calibrations for reef fish species

We focused our analyses on economically-important species of reef fish species across various families that had sufficient calibration sample sizes (minimum sample size threshold: $N \geq 4$). Note that two species of lionfish *Pterois* spp. (i.e., devil firefish *Pterois miles* and red lionfish *Pterois volitans*) exist in the SEUS and are difficult to distinguish visually (Hamner et al., 2007), so they were treated as a single species here.

All videos were read using the MeanCount approach, which was calculated as the mean number of individuals of a particular species that was observed in a series of snapshots within a video (Schobernd et al., 2014). Schobernd et al. (2014) showed that MeanCount was

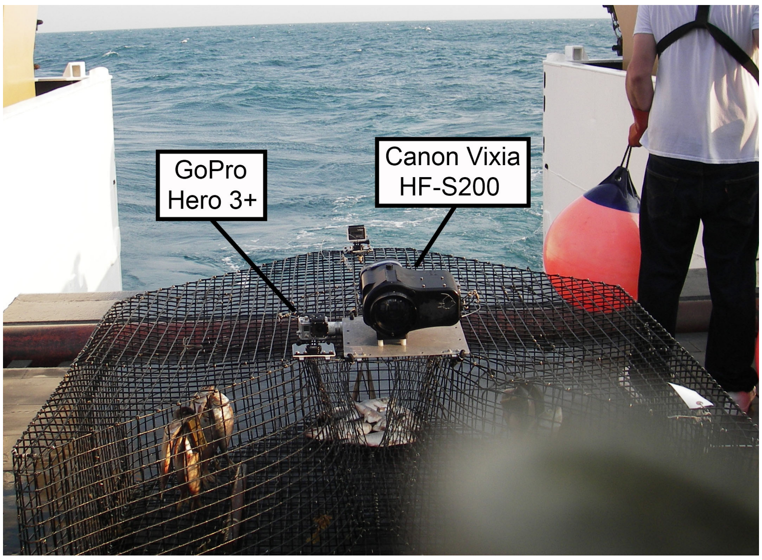



FIGURE 1
Side-by-side calibration experiment conducted by the Southeast Reef Fish Survey along the southeast United States Atlantic coast in 2014. Canon Vixia HF-S200 and GoPro Hero 3+ cameras were attached to baited traps side-by-side looking outward and read for fish at exactly the same times using the MeanCount approach (Schobernd et al., 2014).

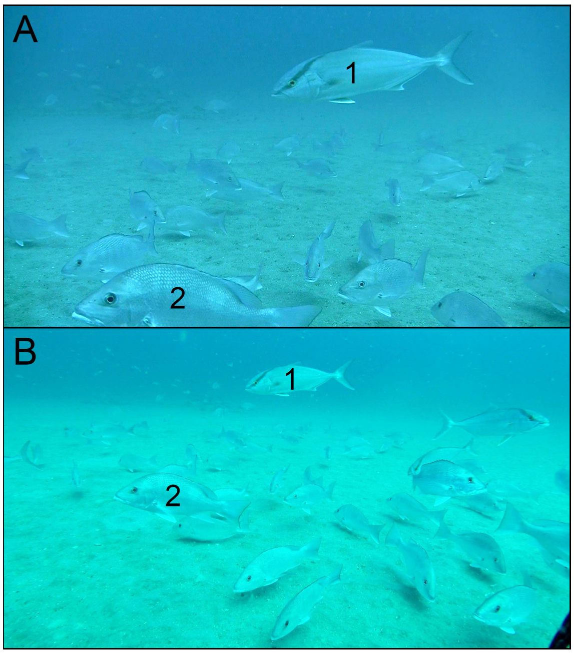



FIGURE 2
Differences in field of view between two underwater video cameras, paired in space and time, from a sample of the Southeast Reef Fish Survey off Jacksonville, Florida, taken in 2014. **(A)** Image from a Canon® Vixia HF-S200 video camera in Gates underwater housing. **(B)** Image from GoPro Hero 3+ camera in stock underwater housing. Numbers are shown so readers can identify and compare the same individual fish in each image: (1) almaco jack *Seriola rivoliana*; (2) red snapper *Lutjanus campechanus*.

proportional to actual abundance using laboratory, simulation, and field data, while other common video reading metrics like *MaxN* (i.e., MinCount; Ellis and DeMartini, 1995) were often nonlinearly related to actual abundance (but see Campbell et al., 2018). Here, we followed SERFS video reading protocol: all species were counted on a total of 41 snapshots, starting 10 min after the trap landed on the bottom and spaced 30 seconds apart for a total of 20 min (Bacheler et al., 2020). Four video readers with extensive training in fish identification read these calibration videos, and a portion of videos from each reader were read by other readers to ensure accuracy.

Species-specific calibration factors were estimated using linear models. Linear models related the MeanCounts of a particular species on Canon cameras ($\text{MeanCount}_{Canon}$) as the response variable to the counts of that species on GoPro cameras ($\text{MeanCount}_{GoPro}$) as the predictor variable as follows:

$$\text{MeanCount}_{Canon} = a + b(\text{MeanCount}_{GoPro}) \tag{1}$$

where $a$ is the model intercept and $b$ is the slope. We also provide the $R^2$ value for each model, which indicates how predictable Canon counts are from GoPro counts. Note that we are using Canon counts as the response variable and GoPro counts as the predictor variable here, implying that our slope and intercept estimates would be used to reduce GoPro counts to make them comparable to Canon counts (see *Results* below). Alternatively, we could have used GoPro counts as the response variable and Canon counts as the predictor variable, in which case slope and intercept estimates would be used to increase Canon counts to make them comparable to GoPro counts. Ultimately, we are interested in using these data to develop time series of relative abundance, which were robust to the direction of calibration. These and all following analyses were conducted in R (R Core Team, 2021).

## Objective 3: limited sample sizes

Calibrations for species with relatively low sample sizes were in some cases poorly estimated (see *Results* below), so our third objective was to evaluate whether borrowing information from related or unrelated species might be warranted in some cases. Ideally, sufficient data will be collected for all species of interest during calibration experiments, but in our case, despite collecting 143 paired calibration videos, some important but rare species had quite low sample sizes.

One potential solution for estimating calibrations for species with low sample sizes is to apply an overall calibration calculated across all species. This approach would only be justified if variability among families was low. To evaluate whether this was the case, we calculated family-level calibrations for each of the family groupings in our calibration dataset, using linear models as described in Equation 1 above. Similar calibrations among families would suggest low variability among taxa and therefore an overall calibration might be substituted for species with low sample sizes. Alternatively, significant variability among families would suggest applying an overall calibration to taxa would not be valid. Note that commonly observed species will tend to drive family-level calibrations much more so than rarer species.

Another possible solution is to apply the family-level calibrations to species with low sample sizes. In this situation, there may be behavioral or anatomical similarities among species of a particular family that could justify borrowing information from related species. For families containing more than one species with sufficient calibration sample sizes, we compared family-level calibrations to those of each species within that family. Similar calibrations for species within a family might justify the use of a family-level calibration for species with insufficient sample sizes. In contrast, if there is sufficient variability in calibrations among species within a family, applying a family-level calibration to any particular species is likely not justified.

## Objective 4: compare approaches to apply calibrations when developing indices of relative abundance

Our fourth objective was to evaluate two approaches for applying calibrations: calibrating the data before inclusion into a standardization model or after standardization has occurred. Calibrating at the data level would be preferable in various situations where standardization models are not being used, for example, in ecological studies or specific research projects. Alternatively, calibrating a final index of abundance is much faster and easier and would be preferable in most cases where video data are being standardized with a statistical model. It was unclear, however, if these two methods of applying calibrations provided standardized indices of abundance that are equivalent.

To evaluate this objective, we tested whether indices of abundance from two representative species varied based on how the calibrations were applied (i.e., at the data or index level). We used SERFS video data in 2011–2021 for scamp *Mycteroperca phenax* and red snapper *Lutjanus campechanus*. These species were selected because they have displayed opposite patterns in relative abundance during the 2011–2021 time frame, with red snapper increasing (SEDAR, 2021a) and scamp declining (SEDAR, 2021b).

Here we followed the standardization procedures used for SERFS video data in the stock assessments of these two species (SEDAR, 2021a; SEDAR, 2021b). The response variable was SumCount, defined as the total number of individuals (scamp or red snapper) observed across all frames of a unique video sampling event. For these analyses, SumCount was used instead of MeanCount because the negative binomial distribution operates on discrete variables (MeanCount is continuous), but we note that SumCount relates linearly to MeanCount because the number of frames per event is constant. The distribution of SumCount for most species contained a large proportion of zeros and had an extended tail of positive values. Therefore, the modeling approach applied a zero-inflated negative binomial (ZINB) formulation, in which a negative binomial sub-model describes the count data and a binomial sub-model describes the occurrence of positive versus zero counts (Zeileis et al., 2008; Zuur et al., 2009). Explanatory variables included year ($y$), season ($t$), depth ($d$), latitude ($lat$), temperature ($temp$), water clarity ($wc$), current direction ($cd$), biotic density ($bd$), and substrate composition ($sc$; see Bacheler et al. (2014) for details). Year was necessarily included, because for an index of abundance, the year effect is of primary interest. The other variables were included or excluded based on a step-wise backward model selection procedure, starting with the full model formulation and then removing variables that did not contribute to explaining variance in the data, based on the Akaike information criterion (AIC) and likelihood ratio tests (Zuur et al., 2009). For this procedure, the initial full model was:

$$SumCount = y + t + d + lat + temp + wc + cd + bd + sc \mid y + t + d + lat + temp + wc + cd + bd + sc \,. \tag{2}$$

In this formulation, variables to the left of the vertical bar apply to the negative binomial sub-model and variables below it apply to the binomial sub-model. The final model included only those variables that were retained after applying the model selection procedure. Model fitting used the *zeroinfl* function in the *countreg* package of R (Zeileis and Kleiber, 2017).

Uncertainty in the resulting index of abundance was computed using a bootstrap procedure with $N$ = 1000 replicates. For each replicate, a new data set of the original size was created by drawing video observations (rows) at random with replacement. The final model configuration was fitted to each replicate data set, and resulting variability (i.e., 95% confidence intervals) in the relative abundance was computed for each year.

The calibration method was applied in two different ways, either at the index level or the data level. To calibrate at the index level, the index was first computed from the original, uncalibrated data. Then, the species-specific linear model (Equation 1) was applied to the standardized index for years 2015 and onward. To account for uncertainty in the calibration itself, parameters of the linear model were included as part of the bootstrap process described above. This was accomplished by drawing at random a new intercept and slope for each bootstrap iteration, in which each draw came from a bivariate normal distribution with means equal to the parameter point estimates and the covariance matrix as estimated by the linear regression. For both species, the point estimate of the intercept was negative, which is expected given that GoPro cameras have a wider field of view and therefore more fish should be observed on GoPros than on Canons. Thus, to preserve that feature in the bootstrap procedure, we truncated the bivariate distribution to provide only negative intercept values.

To calibrate at the data level, the data themselves were adjusted prior to fitting the models. For each positive observation of SumCount in 2015-2021 in the original data, a calibrated SumCount value was drawn from a binomial distribution, $B(n_i,\ p_i)$. Here, $n_i$ is the original SumCount for observation $i$, and $p_i$ was determined by the linear regression (Equation 1) as:

$$p_i = (a + bn_i)/n_i \tag{3}$$

In the bootstrap process, uncertainty in the regression parameters ($a$ and $b$) were incorporated using the same bivariate normal distribution as in the index-level calibration. The index was fitted to each calibrated data set, but the final index did not require any adjustment.

# Results

## Objective 1: calibration experiment design

A total of 27 fish species were observed and counted across the 54 calibration videos collected in 2014. The most commonly observed species on calibration videos was gray triggerfish *Balistes capriscus* ($N$ = 41 videos), followed by vermilion snapper *Rhomboplites aurorubens* ($N$ = 40), red snapper ($N$ = 31), and black sea bass *Centropristis striata* ($N$ = 28; Table 1). The least commonly observed species among those included in our analyses were red grouper *Epinephelus morio* ($N$ = 4), mutton snapper *Lutjanus analis* ($N$ = 6), and hogfish *Lachnolaimus maximus* ($N$ = 8). Eleven fish species were counted on calibration videos but excluded from all analyses because they did not reach the minimum sample size threshold.

## Objective 2: estimating calibrations for reef fish species

Sixteen species met our minimum sample size threshold of being observed on at least 4 calibration videos, and calibrations were

TABLE 1 Calibration information for the 16 reef fish species with a sample size (*N*) of at least 4, their associated families, and a fit across all 16 species ("overall calibration") as part of the 2014 side-by-side calibration study by the Southeast Reef Fish Survey along the southeast United States Atlantic coast.

| Taxa | Scientific name | *N* | Reduce GoPro counts | | | Increase Canon counts | | |
|---|---|---|---|---|---|---|---|---|
| | | | Slope | Intercept | $R^2$ | Slope | Intercept | $R^2$ |
| Overall calibration | | 299 | 0.572 | -0.06 | 0.89 | 1.562 | 0.25 | 0.89 |
| Balistidae | | 41 | 0.446 | 0.01 | 0.93 | 2.080 | 0.07 | 0.93 |
| Gray triggerfish | *Balistes capriscus* | 41 | 0.446 | 0.01 | 0.93 | 2.080 | 0.07 | 0.93 |
| Carangidae | | 40 | 0.709 | -0.01 | 0.94 | 1.327 | 0.03 | 0.94 |
| Almaco jack | *Seriola rivoliana* | 25 | 0.619 | 0.00 | 0.87 | 1.420 | 0.01 | 0.87 |
| Greater amberjack | *Seriola dumerili* | 15 | 0.722 | -0.03 | 0.93 | 1.294 | 0.06 | 0.93 |
| Haemulidae | | 19 | 0.884 | -0.31 | 0.94 | 1.064 | 0.46 | 0.94 |
| White grunt | *Haemulon plumierii* | 19 | 0.884 | -0.31 | 0.94 | 1.064 | 0.46 | 0.94 |
| Labridae | | 8 | 0.978 | -0.02 | 0.90 | 0.934 | 0.03 | 0.90 |
| Hogfish | *Lachnolaimus maximus* | 8 | 0.978 | -0.02 | 0.90 | 0.934 | 0.03 | 0.90 |
| Lutjanidae | | 100 | 0.581 | -0.06 | 0.96 | 1.659 | 0.17 | 0.96 |
| Vermilion snapper | *Rhomboplites aurorubens* | 40 | 0.546 | -0.10 | 0.91 | 1.663 | 0.31 | 0.91 |
| Red snapper | *Lutjanus campechanus* | 31 | 0.610 | -0.10 | 1.00 | 1.635 | 0.18 | 1.00 |
| Gray snapper | *Lutjanus griseus* | 14 | 0.737 | -0.04 | 0.95 | 1.290 | 0.08 | 0.95 |
| Lane snapper | *Lutjanus synagris* | 9 | 0.906 | -0.04 | 0.86 | 0.973 | 0.05 | 0.86 |
| Mutton snapper | *Lutjanus analis* | 6 | 0.214 | 0.05 | 0.27 | 1.938 | -0.06 | 0.27 |
| Scorpaenidae | | 14 | 0.726 | -0.09 | 0.74 | 1.041 | 0.28 | 0.74 |
| Lionfish | *Pterois* spp. | 14 | 0.726 | -0.09 | 0.74 | 1.041 | 0.28 | 0.74 |
| Serranidae | | 53 | 0.293 | 0.07 | 0.79 | 2.699 | 0.09 | 0.79 |
| Black sea bass | *Centropristis striata* | 28 | 0.288 | 0.09 | 0.77 | 2.683 | 0.23 | 0.77 |
| Scamp | *Mycteroperca phenax* | 12 | 0.622 | -0.03 | 0.96 | 1.543 | 0.07 | 0.96 |
| Gag | *Mycteroperca microlepis* | 9 | 0.295 | 0.01 | 0.96 | 3.269 | -0.02 | 0.96 |
| Red grouper | *Epinephelus morio* | 4 | 0.757 | -0.18 | 0.90 | 1.238 | 0.25 | 0.90 |
| Sparidae | | 24 | 0.604 | -0.14 | 0.94 | 1.553 | 0.34 | 0.94 |
| Red porgy | *Pagrus pagrus* | 24 | 0.604 | -0.14 | 0.94 | 1.553 | 0.34 | 0.94 |

"Reduce GoPro counts" indicates a model relating counts from GoPro cameras to Canon cameras, whereas "Increase Canon counts" indicates a model relating counts from Canon cameras to GoPro cameras.

variable among these species (Table 1; Figure 3). Slope estimates ranged from 0.214 for mutton snapper to 0.978 for hogfish, but 8 of 16 species and generally those species with the largest sample sizes had slope estimates within a fairly narrow range of 0.440 to 0.740 (Table 1). The slope of the overall model that included all 16 species was 0.572, suggesting that, on average, 42.8% fewer fish were observed on Canon compared to GoPro cameras. Species-specific model intercepts ranged from -0.31 (white grunt *Haemulon plumierii*) to 0.09 (black sea bass), with 11 of 16 being negative as expected given we are calibrating a camera that observed more fish (i.e., GoPro) to a camera that observed fewer fish (i.e., Canon; Table 1). All but one species-specific $R^2$ value was greater than 0.70, and 11 of 16 were at least 0.90, suggesting GoPro counts predicted Canon counts well for most species (Table 1; Figure 3).

## Objective 3: limited sample sizes

The 16 species included in our analyses were represented by 8 families, and these family-level calibrations were generally similar to the species-specific calibrations of members of those families. Five of the eight families only included a single species, so in these cases, the species-level calibrations were identical to the family-level calibrations (i.e., Balistidae, Haemulidae, Labridae, Scorpaenidae, Sparidae; Table 1). The Carangidae family included two species, Lutjanidae included five species, and Serranidae included four species. Family-level slope estimates ranged from 0.293 (Serranidae) to 0.978 (Labridae), with Lutjanidae and Sparidae having slopes most closely resembling the slope of the overall model (Figure 4). Family-level intercepts ranged from -0.31

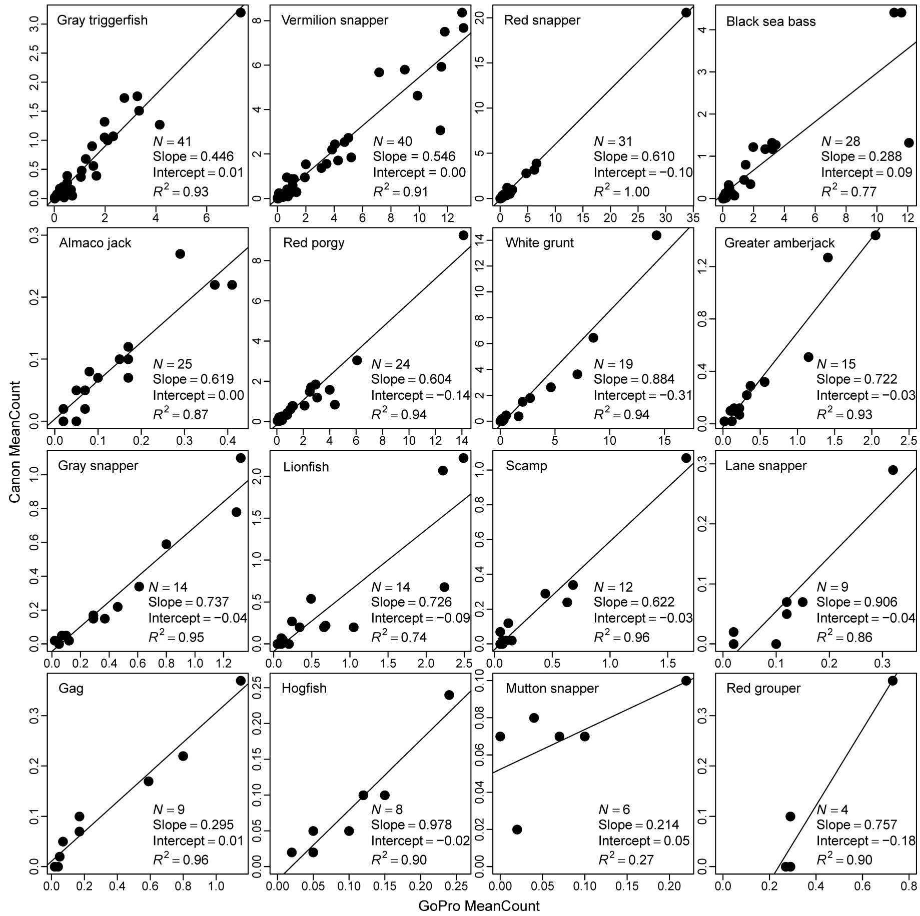



FIGURE 3
Species-specific calibrations of Canon® Vixia HF-S200 and GoPro Hero 3+ cameras for 16 reef-associated fish species along the southeast United States Atlantic coast in 2014. Only species observed on at least 4 videos were included here. Slope, intercept, and the $R^2$ value were estimated using linear models.

(Haemulidae) to 0.07 (Serranidae), and the intercept for Lutjanidae (-0.06) again most closely matched the intercept of the overall model (Table 1; Figure 4). The $R^2$ values for the family-level calibrations ranged from 0.74 to 0.96, with most being at least 0.90 (Table 1).

There were three families that contained more than one species, so calibrations from each of these families were compared to each of the species comprising these families. For Lutjanidae, the two species with the largest sample sizes (i.e., vermilion and red snapper) had calibration slope estimates that were very similar to one another (0.064 difference) and to the family-level calibration (< 0.040; Figure 5). For the remaining lutjanids, as species-specific sample sizes declined, the degree to which their slopes diverged from the family-level calibration increased, with the slope for mutton snapper ($N$ = 4) being most different from Lutjanidae (0.367 difference; Figure 5). For Serranidae, the slope for black sea bass (0.288) and gag *Mycteroperca microlepis* (0.295) appeared to drive the overall family-level slope (0.293), while the slopes for scamp (0.622) and red grouper (0.757) were very different. Species in the family Carangidae had similar slopes, with the lowest being almaco jack *Seriola rivoliana* (0.619) and the highest being greater amberjack *Seriola dumerili* (0.722), with an overall family-level slope of 0.709 (Figure 5).

## Objective 4: determining optimal approach to apply calibrations

A total of 13,072 videos were included in the scamp and red snapper analyses used to compare two approaches for applying calibrations (Table 2). Sampling was generally consistent across years except for 2020, when no sampling occurred due to the COVID-19 pandemic (Table 2). After model selection, the final scamp negative binomial sub-model included all predictor variables except season, depth, biotic density, and substrate composition, while the scamp binomial sub-model included all variables except current direction. For red snapper, all variables were included in final models except temperature in the negative binomial sub-model and water clarity in the binomial sub-model.

Using 2011–2021 video data from SERFS (Table 2), the nominal index of abundance for scamp declined between 2011 and 2013, increased in 2014 and 2015, and declined again from 2015 until

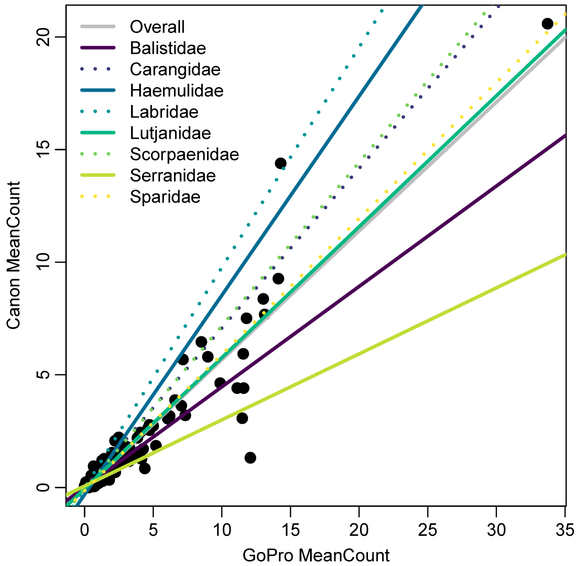



FIGURE 4
Family-specific calibrations between GoPro Hero 3+ and Canon Vixia HF-S200 cameras from data collected by the Southeast Reef Fish Survey along the southeast United States Atlantic coast in 2014. The overall relationship was estimated using data across all families. Trendlines were estimated using a linear model fit to data across all species within each of the eight families listed.

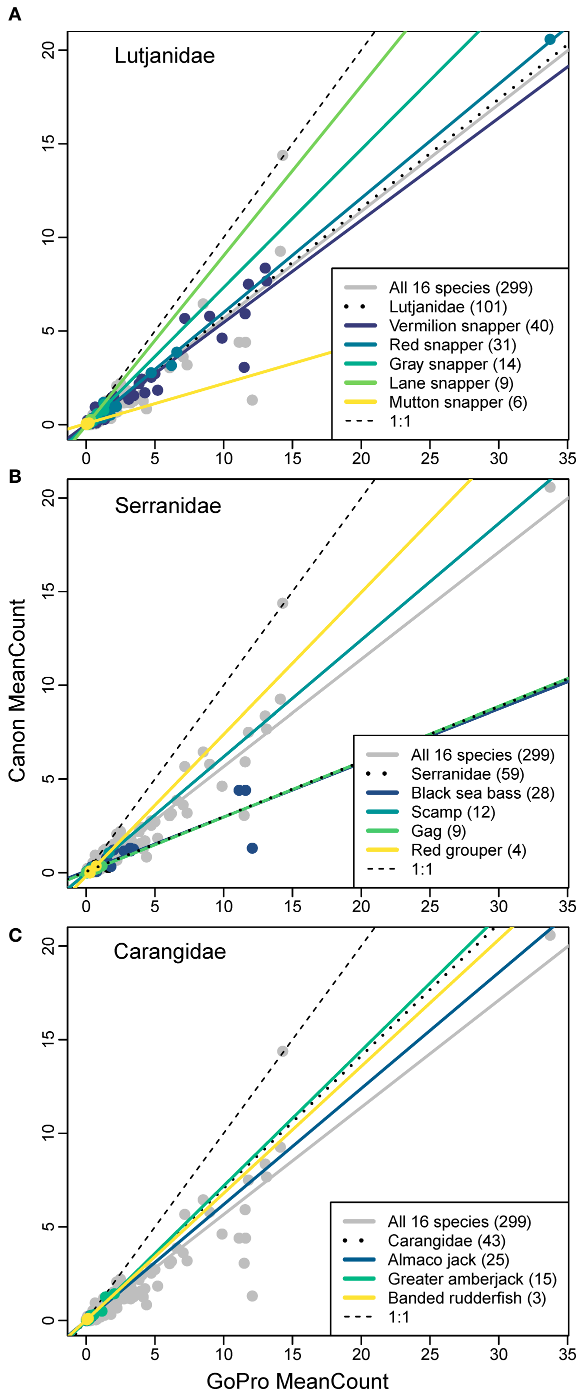



FIGURE 5
Calibrations between Canon Vixia HF-S200 and GoPro Hero 3+ cameras for various species within three families **(A)** Lutjanidae; **(B)** Serranidae; **(C)** Carangidae from data collected by the Southeast Reef Fish Survey along the southeast United States Atlantic coast in 2014. The overall calibration is shown in gray, family level calibrations are shown by black dotted lines, species' calibrations are shown by solid colored lines, and the 1:1 relationship is indicated by thin black dashed line. Number in legend indicates sample size for each taxa.

2021 (Figure 6). The standardized index of abundance for scamp completely removed the nominal increase that was evident between 2013 and 2014, instead declining across all years except between 2014 and 2015, when cameras were switched from Canon (2011–2014) to GoPro (2015–2021) in the survey. Calibrating the standardized scamp video index using both approaches removed the increase in relative abundance between 2014 and 2015 entirely, suggesting that increase was due to the increased video counts of scamp on GoPro cameras (relative to Canon cameras) and not an actual increase in abundance. Most importantly, calibrating scamp video data before standardization at the data level, or after standardization at the index level, had very little influence on the resulting scamp index of abundance (Figure 6).

In contrast to scamp, red snapper increased substantially over the course of the study. The nominal red snapper index of abundance increased nearly linearly in all years except between 2015 and 2016, when abundance slightly declined (Figure 6). The standardized index was somewhat different from the nominal index, being lower in 2011–2014 and slightly higher in 2015–2021. Calibration of the standardized index increased abundance in 2011–2014 and decreased abundance in 2015–2021. Consistent with the results for scamp, the standardized red snapper index of abundance was mostly unaffected by which calibration approach was used (Figure 6).

## Discussion

Underwater video cameras are widely used to monitor fish abundance and biodiversity (Mallet and Pelletier, 2014), and these cameras have evolved drastically over the last few decades (Struthers et al., 2015). For any video survey, the benefits of utilizing improved technology may at some point outweigh the benefits of maintaining

TABLE 2 Annual video sampling information for the Southeast Reef Fish Survey, 2011–2021, included in the analyses.

| Year | Camera | *N* | Calibration *N* | Mean latitude (°N; range) | Mean depth (m; range) |
|---|---|---|---|---|---|
| 2011 | Canon Vixia HF-S200 | 580 | 0 | 30.7 (27.2 – 34.5) | 42 (15 – 94) |
| 2012 | Canon Vixia HF-S200 | 1,083 | 0 | 31.9 (27.2 – 35.0) | 40 (15 – 105) |
| 2013 | Canon Vixia HF-S200 | 1,221 | 0 | 31.3 (27.3 – 35.0) | 38 (15 – 98) |
| 2014 | Canon Vixia HF-S200 | 1,382 | 143 | 31.9 (27.2 – 35.0) | 39 (16 – 109) |
| 2015 | GoPro Hero 3+ | 1,405 | 0 | 31.9 (27.3 – 35.0) | 39 (15 – 110) |
| 2016 | GoPro Hero 3+ | 1,404 | 0 | 32.2 (27.2 – 35.0) | 41 (16 – 115) |
| 2017 | GoPro Hero 3+ | 1,424 | 0 | 32.0 (27.2 – 35.0) | 40 (15 – 111) |
| 2018 | GoPro Hero 3+ | 1,654 | 0 | 32.0 (27.2 – 35.0) | 40 (16 – 114) |
| 2019 | GoPro Hero 3+ | 1,545 | 0 | 32.1 (27.2 – 35.0) | 41 (14 – 110) |
| 2020 | NA | 0 | 0 | – | – |
| 2021 | GoPro Hero 3+ | 1,374 | 0 | 31.9 (27.2 – 35.0) | 38 (16 – 109) |
| Overall | – | 13,072 | 143 | 31.9 (27.2 – 35.0) | 40 (14 – 115) |

*N* is the number of videos collected and analyzed each year and "Calibration *N*" is the number of paired calibration videos (i.e., Canon Vixia HF-S200 and GoPro Hero 3+) collected. No videos were collected in 2020 due to the COVID-19 pandemic.

the original gear, yet little to no attention has been given to calibrating video cameras when switches or replacements have occurred. Calibrating different video cameras that are used in fisheries surveys or ecological studies is critical because small differences in camera lenses, sensors, and other characteristics can influence the ability of readers to identify individual fish and therefore affect fish sightability. For instance, a greater field of view of a camera will increase the volume of water sampled, but will also make all objects appear somewhat smaller and thus harder to identify. Therefore, it is critical that researchers carry out empirical calibrations using their cameras, study areas, and target species; theoretical calibrations based on the respective water volume sampled for each camera are unlikely to track closely with empirical calibration results. Video counts are assumed to reflect the true abundance of fish in a way that is not confounded by differing sightabilities of cameras, and this can only be achieved using calibration experiments.

There were substantial differences in fish count calibrations among species from Canon and GoPro cameras. There are likely three main reasons why variability in calibrations was observed among species. The first reason was sample size, whereby species with higher sample sizes had more similar calibrations than species with low sample sizes, whose calibration relationships were much more variable. The second reason was likely due to behavioral variability among species. Some reef fish species are strongly attracted to baited gears, some are indifferent, and some are wary and keep their distance. For example, most species displaying attraction to baited gears (e.g., most lutjanids, carangids, red porgy *Pagrus pagrus*, gray triggerfish; Bacheler et al., 2022b) had moderate calibration slopes (i.e., 40-50% fewer fish seen on Canon cameras compared to GoPro) because they could be easily viewed by cameras and differences in fields of view between cameras was the primary determinant of the calibration relationships. Black sea bass, on the other hand, were strongly attracted to bait (Bacheler et al., 2013) but highly demersal, leading GoPro cameras with a wider field of view to see individuals close to the camera and on the bottom that Canon cameras could often not see (i.e., 71% fewer individuals observed on Canon cameras). The third reason that may explain some of the variability among species could be differences in their appearance (e.g., size, shape, and color) that might make some species more readily visible and identifiable on some cameras compared to others. For instance, some distinctly shaped fish could be identified far in the background on some cameras compared to nondescript species, which could also influence calibration relationships.

A challenge with video surveys that target multiple species is that calibration factors need to be estimated for all species, but estimating calibrations for rare species can take considerable effort. We evaluated among- and within-family variability in video calibrations to determine if calibrations for related or unrelated species could be used for rare species. Results were species- and family-specific and difficult to generalize. For instance, the behavior of almaco jack, greater amberjack, and banded rudderfish *Seriola zonata* in family Carangidae is similar (i.e., shoaling, strong curiosity about bait and sampling gears; Campbell et al., 2021), and the resulting calibrations were likewise similar, suggesting that rarely encountered *Seriola* spp. could justifiably borrow a carangid family-level calibration. That was not the case for species in the family Serranidae, whose calibration slopes were highly variable. Even two species in the same genus of Serranidae like scamp and gag had very different calibration slopes (0.327 difference), suggesting that borrowing calibration information within the family Serranidae is not prudent. These results suggest the safest approach is to collect enough calibration data so that slopes can be estimated well for even the rarest species.

We showed that calibrating video data before standardization or calibrating the index after standardization had a negligible influence

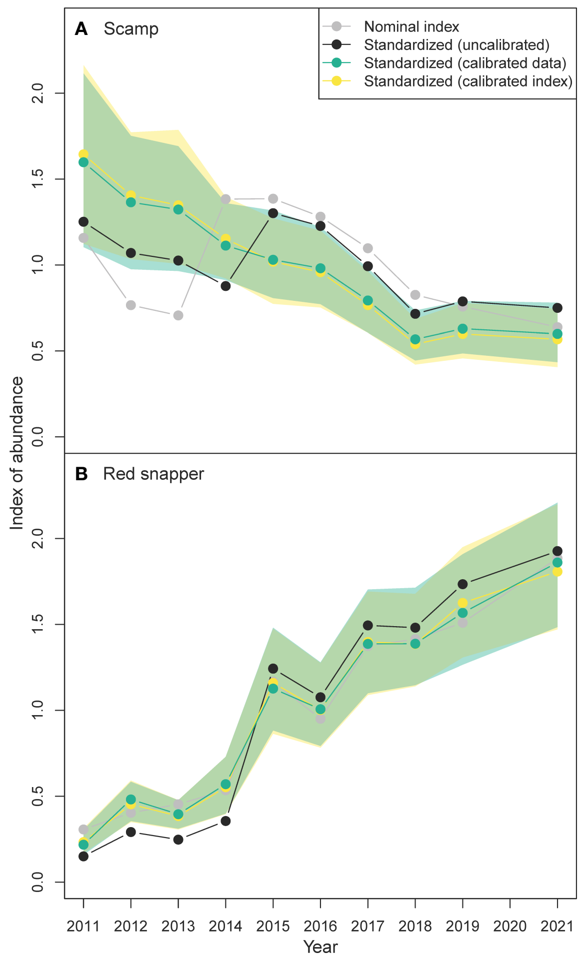



FIGURE 6
Indices of abundance for **(A)** scamp *Mycteroperca phenax* and **(B)** red snapper *Lutjanus campechanus* calculated four different ways: (1) a nominal index where no standardization or calibration occurred (gray filled points and line), (2) a standardized index that did not include calibration for a camera switch that occurred between 2014 and 2015 (black filled points and line), (3) a standardized index that included video data that were pre-calibrated before inclusion in a standardization model (blue-green filled points, line, and shaded 95% confidence interval), and (4) a standardized index that was post-calibrated at the index level (yellow filled points, line, and shaded 95% confidence interval). Overlapping 95% confidence intervals are shown in green.

on the final index of abundance. This is a particularly encouraging result because it provides researchers flexibility in how to calibrate among video cameras – in some cases it may be easier to calibrate at the data level, whereas in others it is likely much easier to calibrate after index standardization. Note that when calibrating a camera seeing fewer fish to a camera seeing more fish, it is not straightforward to increase the lower camera counts to make them consistent with the higher counts when calibrating at the data level, because it is unclear how to expand when counts are zero on the original camera. To avoid that issue here when calibrating at the data level, we reduced the more recent, higher counts to make them comparable to the earlier, lower counts. In most cases, though, it would probably be preferable to calibrate the previous gear to match the newer gear even if it does not make a difference for a relative index.

Calibration is challenging because sampling gears are often not conducive to being paired in space and time. For instance, trawls are often calibrated by comparing catches from the same general water body in the same season, being pulled near one another at the same time, or being dragged on the same line in succession (Mahon and Smith, 1989; Miller, 2013; Benoit and Cadigan, 2014). Given that fish and environmental conditions are patchily distributed in space and time, often at small scales (Ciannelli et al., 2010; Bacheler et al., 2017), a large amount of residual variation is typically introduced around the estimated calibration factor when gears are not paired in space or time (Pelletier, 1998). Cameras are much more conducive to being paired, given their small size. When calibrating two cameras, we recommend pairing their deployments so that the spatial and temporal variability inherent in fish abundance and environmental conditions can be completely controlled for, and counting fish from the paired samples in exactly the same way (i.e., choosing the same sequential images to analyze). If camera systems are contained in stand-alone metal frames or landers (e.g., Cappo et al., 2004; Merritt et al., 2011; Bacheler and Shertzer, 2015; Amin et al., 2017), it will be necessary to develop a way for these different systems to be paired (i.e., attaching two landers together side-by-side) while not changing the efficiency of each gear compared to when they are deployed independently.

Indices of abundance for scamp and red snapper were improved considerably by accounting for a camera change in 2015 and using a statistical model to standardize video counts among years. Many fishery-independent surveys have used design-based estimators where average catch rates within predetermined sampling strata in a sampling design are calculated, and then an area-weighted sum of abundance in each stratum is produced (Smith, 1990). Due to some downsides of the design-based approach, it has become more common to use statistical models to control for variability in sampling or environmental conditions during the survey (e.g., Helser et al., 2004; Maunder and Punt, 2004; Bacheler and Ballenger, 2018). In our study, nominal (i.e., design-based) indices for scamp and, to a lesser extent, red snapper were different and more variable than standardized indices of abundance (i.e., model-based), the latter of which appeared able to control for environmental variability and changes in the spatial and temporal aspects of sampling among years. Nonetheless, only when the calibration between cameras was accounted for properly did the indices of abundance show smooth declining or increasing trends as expected, highlighting the importance of video calibration studies.

There were some shortcomings of our approach. First, higher calibration sample sizes would have improved calibration relationships, particularly for less commonly observed but important species like red grouper, gag, and hogfish. For instance, red grouper has been observed on 1.4% of videos collected by SERFS in recent years (Bacheler et al., 2019), so approximately 1,429 paired calibration videos would be required to attain an $N$ = 20 for that species. If it is impossible to collect that many calibration videos, we provide a framework for potentially borrowing information from related or unrelated species. A second shortcoming is that we used linear models to relate video counts from one camera to another, when in fact relationships may be nonlinear. In preliminary analyses, we compared linear and nonlinear model fits, which were similar for most species but additional parameters were required for nonlinear models, so linear models were almost always selected by Akaike information criterion (Burnham and Anderson, 2002). Therefore, linear models were used for all comparisons. Third, we selected scamp and red snapper for application of the video calibration (Objective 4), but we cannot conclude that these examples represent all other species. We chose these species for two reasons: (1) indices of abundance were developed for these two species for recent assessments (SEDAR, 2021a; SEDAR, 2021b), and (2) they showed opposite patterns of relative abundance over time, with scamp declining and red snapper increasing, which may have affected calibrations.

Changes in sampling gears can strongly influence the catchability of fish (Arreguín-Sánchez, 1996), leading to biased spatial or temporal trends in relative abundance (Maunder and Punt, 2004). This is especially the case for underwater cameras that are now used widely to provide relative abundance information for various species of fish (Mallet and Pelletier, 2014). Video gears are rapidly evolving, getting smaller and cheaper with higher resolution (Struthers et al., 2015), yet there has been a paucity of examples where calibrations between different video sampling gears occurred. We showed that a camera switch in SERFS in 2015 resulted in much higher fish counts on the new camera compared to the old, which necessitated a calibration experiment to maintain the temporal continuity of the SERFS video survey. We recommend calibrating data from video cameras any time changes occur, and pairing video cameras to the extent possible to control for the spatial and temporal variability inherent in fish populations and environmental conditions. Following these guidelines, researchers will be able to maintain the integrity of valuable long-term video datasets despite changes in their sampling gear that occur out of necessity or by choice.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The animal study was reviewed and approved by U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training.

## Author contributions

NB and ZS obtained funding for the calibration experiment, NB, ZS, and LC conceived the research project, ZS coordinated the field data collection, LC, KS, and NB analyzed data, NB and KS drafted the manuscript, and all authors contributed ideas and editorial work on the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was funded by the National Marine Fisheries Service.

## Acknowledgments

We thank SERFS staff members at the South Carolina Department of Natural Resources and the National Marine Fisheries Service, various volunteers, and the captains and crew of the research vessels for data collection. We also thank R. Cheshire and K. Craig for providing comments on earlier versions of this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Author disclaimer

Mention of trade names or commercial companies is for identification purposes only and does not imply endorsement by the National Marine Fisheries Service, NOAA. The scientific results and conclusions, as well as any views and opinions expressed herein, are those of the authors and do not necessarily reflect those of any government agency.

# References


Aguzzi, J., and Company, J. B. (2010). Chronobiology of deep-water decapod crustaceans on continental margins. *Adv. Mar. Biol.* 58, 155–225. doi: 10.1016/B978-0-12-381015-1.00003-4

Amin, R., Richards, B. L., Misa, W. F. X. E., Taylor, J. C., Miller, D. R., Rollo, A. K., et al. (2017). The modular optical underwater survey system. *Sensors* 17, 2309. doi: 10.3390/s17102309

Arreguín-Sánchez, F. (1996). Catchability: a key parameter for fish stock assessment. *Rev. Fish Biol. Fish.* 6, 221–242. doi: 10.1007/BF00182344

Bacheler, N. M., and Ballenger, J. C. (2018). Decadal-scale decline of scamp *(Mycteroperca phenax*) abundance along the southeast united states Atlantic coast. *Fish. Res.* 204, 74–87. doi: 10.1016/j.fishres.2018.02.006

Bacheler, N. M., Berrane, D. J., Mitchell, W. A., Schobernd, C. M., Schobernd, Z. H., Teer, B. Z., et al. (2014). Environmental conditions and habitat characteristics influence trap and video detection probabilities for reef fish species. *Mar. Ecol. Prog. Ser.* 517, 1–14. doi: 10.3354/meps11094

Bacheler, N. M., Geraldi, N. R., Burton, M. L., Muñoz, R. C., and Kellison, G. T. (2017). Comparing relative abundance, lengths, and habitat of temperate reef fishes using simultaneous underwater visual census, video, and trap sampling. *Mar. Ecol. Prog. Ser.* 574, 141–155. doi: 10.3354/meps12172

Bacheler, N. M., Gillum, Z. D., Gregalis, K. C., Pickett, E. P., Schobernd, C. M., Schobernd, Z. H., et al. (2022b). Comparison of video and traps for detecting reef fishes and quantifying species richness in the continental shelf waters of the southeast USA. *Mar. Ecol. Prog. Ser.* 698, 111–123. doi: 10.3354/meps14141

Bacheler, N. M., Gillum, Z. D., Gregalis, K. C., Schobernd, C. M., Schobernd, Z. H., and Teer, B. Z. (2020). Spatial patterns in relative abundance and habitat use of adult gray snapper off the southeastern coast of the united states. *Mar. Coast. Fish.* 12, 205–209. doi: 10.1002/mcf2.10118

Bacheler, N. M., Runde, B. J., Shertzer, K. W., Buckel, J. A., and Rudershausen, P. J. (2022a). Fine-scale behavior of red snapper *(Lutjanus campechanus*) around bait: approach distances, bait plume dynamics, and effective fishing area. *Can. J. Fish. Aquat. Sci.* 79, 458–471. doi: 10.1139/cjfas-2021-0044

Bacheler, N. M., Schobernd, Z. H., Berrane, D. J., Schobernd, C. M., Mitchell, W. A., and Geraldi, N. R. (2013). When a trap is not a trap: converging entry and exit rates and their effect on trap saturation of black sea bass *(Centropristis striata*). *ICES J. Mar. Sci.* 70, 873–882. doi: 10.1093/icesjms/fst062

Bacheler, N. M., Schobernd, Z. H., Gregalis, K. C., Schobernd, C. M., Teer, B. Z., Gillum, Z., et al. (2019). Patterns in fish biodiversity associated with temperate reefs on the southeastern US continental shelf. *Mar. Biodiv.* 49, 2411–2428. doi: 10.1007/s12526-019-00981-9

Bacheler, N. M., and Shertzer, K. W. (2015). Estimating relative abundance and species richness from video surveys of reef fishes. *Fish. Bull.* 113, 15–26. doi: 10.7755/FB.113.1.2

Bacheler, N. M., and Shertzer, K. W. (2020). Catchability of reef fish species in traps is strongly affected by water temperature and substrate. *Mar. Ecol. Prog. Ser.* 642, 179–190. doi: 10.3354/meps13337

Balletti, C., Guerra, F., Tsioukas, V., and Vernier, P. (2014). Calibration of action cameras for photogrammetric purposes. *Sensors* 14, 17471–17490. doi: 10.3390/s140917471

Benoit, H. P., and Cadigan, N. (2014). Model-based estimation of commercial-sized snow crab *(Chionoecetes opilio*) abundance in the southern gulf of st. Lawrence 1980-2013, using data from two bottom trawl surveys. *DFO Can. Sci. Advis. Sec. Res. Doc.*

Brooks, E. J., Sloman, K. A., Sims, D. W., and Danylchuk, A. J. (2011). Validating the use of baited remote underwater video surveys for assessing the diversity, distribution and abundance of sharks in the Bahamas. *Endanger. Species Res.* 13, 231–243. doi: 10.3354/esr00331

Burnham, K. P., and Anderson, D. R. (2002). *Model selection and multimodal inference: a practical information-theoretic approach. 2nd edn* (New York: Springer).

Cadigan, N. G., and Dowden, J. J. (2010). Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data. *Fish. Bull.* 108, 15–30.

Campbell, M. D., Huddleston, A., Somerton, D., Clarke, M. E., Wakefield, W., Murawski, S., et al. (2021). Assessment of attraction and avoidance behaviors of fish in response to the proximity of transiting underwater vehicles. *Fish. Bull.* 119, 216–230. doi: 10.7755/FB.119.4.2

Campbell, M. D., Salisbury, J., Caillouet, R., Driggers, W. B., and Kilfoil, J. (2018). Camera field-of-view and fish abundance estimation: a comparison of individual-based model output and empirical data. *J. Exp. Mar. Biol. Ecol.* 501, 46–53. doi: 10.1016/j.jembe.2018.01.004

Cappo, M., Speare, P., and Ge'ath, G. (2004). Comparison of baited remote underwater video stations (BRUVS) and prawn (shrimp) trawls for assessments of fish biodiversity in inter-reefal areas of the great barrier reef marine park. *J. Exp. Mar. Biol. Ecol.* 302, 123–152. doi: 10.1016/j.jembe.2003.10.006

Ciannelli, L., Knutsen, H., Olsen, E. M., Espeland, S. H., Asplin, L., Jelmert, A., et al. (2010). Small-scale genetic structure in a marine population in relation to water circulation and egg characteristics. *Ecol.* 91, 2918–2930. doi: 10.1890/09-1548.1

Dennis, D., Plagányi, É., Van Putten, I., Hutton, T., and Pascoe, S. (2015). Cost benefit of fishery-independent surveys: are they worth the money? *Mar. Pol.* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

Dorman, S. R., Harvey, E. S., and Newman, S. J. (2012). Bait effects in sampling coral reef fish assemblages with stereo-BRUVs. *PloS One* 7 (7), e41538. doi: 10.1371/journal.pone.0041538

Ellis, D. M., and DeMartini, E. E. (1995). Evaluation of a video camera technique for indexing the abundances of juvenile pink snapper, *Pristipomoides filamentosus*, and other Hawaiian insular shelf fishes. *Fish. Bull.* 93, 67–77.

Goethel, D. R., Omori, K. L., Punt, A. E., Lynch, P. D., Berger, A. M., de Moor, C. L., et al. (2022). Oceans of plenty? challenges, advancements, and future directions for the provision of evidence-based fisheries management advice. *Rev. Fish Biol. Fish.* doi: 10.1007/s11160-022-09726-7

Hamner, R. M., Freshwater, D. W., and Whitfield, P. E. (2007). Mitochondrial cytochrome b analysis reveals two invasive lionfish species with strong founder effects in the western Atlantic. *J. Fish Biol.* 71, 214–222. doi: 10.1111/j.1095-8649.2007.01575.x

Hangsleben, M. A., Allen, M. S., and Gwinn, D. C. (2013). Evaluation of electrofishing catch per unit effort for indexing fish abundance in Florida lakes. *Trans. Am. Fish. Soc* 142, 247–256. doi: 10.1080/00028487.2012.730106

Harvey, E. S., Cappo, M., Butler, J. J., Hall, N., and Kendrick, G. A. (2007). Bait attraction affects the performance of remote underwater video stations in assessment of demersal fish community structure. *Mar. Ecol. Prog. Ser.* 350, 245–254. doi: 10.3354/meps07192

Harvey, E. S., and Shortis, M. (1998). Calibration stability of an underwater stereo-video system: implications for measurement accuracy and precision. *Mar. Technol. Soc J.* 329, 3–17.

Heagney, E. C., Lynch, T. P., Babcock, R. C., and Suthers, I. M. (2007). Pelagic fish assemblages assessed using mid-water baited video: standardizing fish counts using bait plume size. *Mar. Ecol. Prog. Ser.* 350, 255–266. doi: 10.3354/meps07193

Helser, T. E., Punt, A. E., and Methot, R. D. (2004). A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fish. Res.* 70, 251–264. doi: 10.1016/j.fishres.2004.08.007

Kimura, D. K., and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Rev. Fish. Sci.* 14, 245–283. doi: 10.1080/10641260600621761

Kumpf, H. E., and Lowenstein, J. M. (1962). Undersea observation station. *Sea Front.* 8, 198–206.

Letessier, T. B., Juhel, J. B., Vigliola, L., and Meeuwig, J. J. (2015). Low-cost small action cameras in stereo generates accurate underwater measurements of fish. *J. Exp. Mar. Biol. Ecol.* 466, 120–126. doi: 10.1016/j.jembe.2015.02.013

Mahon, R., and Smith, R. W. (1989). Comparison of species composition in a bottom trawl calibration experiment. *J. Northw. Atl. Fish. Sci.* 9, 73–79. doi: 10.2960/J.v9.a6

Mallet, D., and Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications, (1952–2012). *Fish. Res.* 154, 44–62. doi: 10.1016/j.fishres.2014.01.019

Maunder, M. N., and Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES J. Mar. Sci.* 72, 7–18. doi: 10.1093/icesjms/fsu015

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* 70, 141–159. doi: 10.1016/j.fishres.2004.08.002

Merritt, D., Donovan, M. K., Kelley, C., Waterhouse, L., Parke, M., Wong, K., et al. (2011). BotCam: a baited camera system for nonextractive monitoring of bottomfish species. *Fish. Bull.* 109, 56–67.

Miller, T. J. (2013). A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. *Can. J. Fish. Aquat. Sci.* 70, 1306–1316. doi: 10.1139/cjfas-2013-0136

Murphy, H. M., and Jenkins, G. P. (2010). Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Mar. Freshw. Res.* 61, 236–252. doi: 10.1071/MF09068

Myrberg, A. A., Banner, A., and Richard, J. D. (1969). Shark attraction using a video-acoustic system. *Mar. Biol.* 2, 264–276. doi: 10.1007/BF00351149

Pelletier, D. (1998). Intercalibration of research survey vessels in fisheries: a review and an application. *Can. J. Fish. Aquat. Sci.* 55, 2672–2690. doi: 10.1139/f98-151

Priede, I. G., and Merrett, N. R. (1996). Estimating of abundance of abyssal demersal fishes: a comparison of data from trawls and baited cameras. *J. Fish Biol.* 49 (Suppl A), 207–216. doi: 10.1111/j.1095-8649.1996.tb06077.x

R Core Team (2021). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Santana-Garcon, J., Braccini, M., Langlois, T. J., Newman, S. J., McAuley, R. B., and Harvey, E. S. (2014). Calibration of pelagic stereo-BRUVs and scientific longline surveys for sampling sharks. *Meth. Ecol. Evol.* 5, 824–833. doi: 10.1111/2041-210X.12216

Schobernd, Z. H., Bacheler, N. M., and Conn, P. B. (2014). Examining the utility of alternative video monitoring metrics for indexing reef fish abundance. *Can. J. Fish. Aquat. Sci.* 71, 464–471. doi: 10.1139/cjfas-2013-0086

SEDAR (2021a) *SEDAR 73 stock assessment report: south Atlantic red snapper*. Available at: https://sedarweb.org/documents/sedar-73-stock-assessment-report-south-atlantic-red-snapper/.

SEDAR (2021b) *SEDAR 68 stock assessment report: Atlantic scamp grouper*. Available at: https://sedarweb.org/documents/sedar-68-atlantic-scamp-final-stock-assessment-report/.

Shafait, F., Harvey, E. S., Shortis, M. R., Mian, A., Ravanbakhsh, M., Seager, J. W., et al. (2017). Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo-video measurements. *ICES J. Mar. Sci.* 74, 1690–1701. doi: 10.1093/icesjms/fsx007

Smith, S. J. (1990). Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Can. J. Fish. Aquat. Sci.* 47, 894–903. doi: 10.1139/f90-103

Struthers, D. P., Danylchuk, A. J., Wilson, A. D. M., and Cooke, S. J. (2015). Action cameras: bringing aquatic and fisheries research into view. *Fisheries* 40, 502–512. doi: 10.1080/03632415.2015.1082472

Thorson, J. T., and Ward, E. J. (2014). Accounting for vessel effects when standardizing catch rates from cooperative surveys. *Fish. Res.* 155, 168–176. doi: 10.1016/j.fishres.2014.02.036

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecol. Appl.* 13, 1790–1801. doi: 10.1890/02-5078

Whitmarsh, S. K., Fairweather, P. G., and Huveneers, C. (2017). What is big BRUVver up to? methods and uses of baited underwater video. *Rev. Fish Biol. Fish.* 27, 53–73. doi: 10.1007/s11160-016-9450-1

Zeileis, A., and Kleiber, C. (2017). "Countreg: count data regression," in *R package version0.0/r34.*

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in r. *J. Stat. Software* 27, 1–25. doi: 10.18637/jss.v027.i08

Zuur, A. F., Ieno, E. N., Walkder, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed effects models and extensions in ecology with r* (New York: Spring Science and Business Media, LLC).

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 06 June 2023
DOI 10.3389/fmars.2023.1214526



OPEN ACCESS

EDITED BY
Kyle Shertzer,
NOAA Southeast Fisheries Science Center,
United States

REVIEWED BY
James Gartland,
College of William & Mary, United States
Kai Wieland,
Technical University of Denmark, Denmark

*CORRESPONDENCE
Zack S. Oyafuso
zack.oyafuso@noaa.gov

†These authors contributed equally to this work

‡Senior authorship

§First authorship

RECEIVED 29 April 2023
ACCEPTED 22 May 2023
PUBLISHED 06 June 2023

CITATION
Oyafuso ZS, Barnett LAK, Siple MC, Cooper DW and Kotwicki S (2023) Evaluating potential changes to the US Chukchi Sea bottom trawl survey design via simulation testing.
*Front. Mar. Sci.* 10:1214526.
doi: 10.3389/fmars.2023.1214526



# Evaluating potential changes to the US Chukchi Sea bottom trawl survey design via simulation testing

Zack S. Oyafuso*†§, Lewis A. K. Barnett†, Margaret C. Siple†, Daniel W. Cooper† and Stan Kotwicki†‡

National Oceanic and Atmospheric Administration Alaska Fisheries Science Center, Groundfish Assessment Program, National Oceanic and Atmospheric Administration National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, WA, United States

The US Chukchi Sea consists of the waters off the northwest of Alaska and is a naturally dynamic ice-driven ecosystem. The impacts from climate change are affecting the Arctic marine ecosystem as well as the coastal communities that rely on healthy marine ecosystems. In anticipation of increased ecosystem monitoring in the area, there is an opportunity to evaluate improved sampling designs for future ecological monitoring of the Chukchi Sea, an area that is sampled less comprehensively compared to other regions in Alaska. This analysis focused on standardized NOAA-NMFS-AFSC bottom trawl surveys (otter and beam trawls) and three types of survey designs: simple random, stratified random, and systematic. First, spatiotemporal distributions for 18 representative demersal fish and invertebrate taxa were fitted using standardized catch and effort data. We then simulated spatiotemporal taxon densities to replicate the three survey design types to evaluate design-based estimates of abundance and precision across a range of sampling effort. Modest increases in precision were gained from stratifying the design when compared to a simple random design with either similar or lower uncertainty and bias of the precision estimates. There were often strong tradeoffs between the precision and bias of the systematic estimates of abundance (and associated variance) across species and gear type. The stratified random design provided the most consistent, reliable, and precise estimates of abundance indices and is likely to be the most robust to changes in the survey design. This analysis provides a template for changing bottom trawl survey designs in the Chukchi Sea and potentially other survey regions in Alaska going forward and will be important when integrating new survey objectives that are more ecosystem-focused.

KEYWORDS

**bottom trawl surveys, sampling design, groundfish, Chukchi Sea, simulation testing**

## Introduction

The recent environmental and ecological changes occurring in the Pacific Arctic Ocean are unprecedented (Huntington et al., 2020). The diminishing extent of the sea ice observed in the past century is perhaps the most apparent of the changes occurring in the Arctic ocean (Polyak et al., 2010). The Arctic ice pack reached its lowest point in 2012 relative to 1979-2000 (Parkinson and Comiso, 2013). Sea ice and the cold conditions associated with it are important to atmospheric and oceanographic regulation (Budikova, 2009). The edges of the sea ice are active in primary and secondary production, creating important foraging habitats for fish and marine mammals (Post et al., 2013). Seals haul out on the surface of the ice to rest and nurse their pups, and polar bears and walruses depend on the ice to hunt. Many Arctic communities hunt these mammals for subsistence. Warmer waters can expand the habitat ranges of more temperate species. For example, the discovery of large populations of mature walleye pollock (a common and commercial Bering Sea species) in the Russian western portion of the Chukchi Sea (e.g., Emelin et al., 2022; Maznikova et al., 2023b) led to the development of a fishery in the region in 2021.

The US portion of the Pacific Arctic Ocean includes the eastern Chukchi Sea which is connected to the Bering Sea *via* the Bering Strait and extends to the Beaufort Sea to the northeast. Bottom trawl surveying of groundfish and benthic invertebrates has been conducted by the National Marine Fisheries Service (NMFS) and its predecessor, the Bureau of Commercial Fisheries, sporadically since the 1950s. Increased monitoring of the Chukchi Sea is likely, given the poleward expansion of many Bering Sea species like walleye pollock, Pacific cod, and various flatfishes into the northern Bering Sea (Stevenson and Lauth, 2019; Spies et al., 2020) and further into the Chukchi Sea (Datsky et al., 2022; Cooper et al., 2023; Levine et al., 2023; Maznikova et al., 2023b) in recent anomalously warm years.

In the past ten years, there have been increased efforts to conduct integrated ecosystem-wide monitoring across the entire Chukchi Sea (Baker et al., 2023). To increase the monitoring of groundfish and benthic invertebrates in the Chukchi Sea, it has been proposed to extend the current Bering Sea NMFS bottom trawl survey (BTS) conducted by the Alaska Fisheries Science Center (AFSC), similar to the extension of the Bering Sea survey into the northern Bering Sea since 2010. Thus, the naive assumption for future Chukchi Sea NMFS survey designs is to extend the fixed NMFS Bering Sea 20-nmi systematic grid onto the Chukchi Sea shelf as done in 2012 (Goddard et al., 2014). However, until funding is available for a groundfish survey in the Chukchi Sea, there is an opportunity to evaluate survey designs that could provide reliable abundance estimates while allowing for more flexibility in survey extent and total survey effort than a systematic survey would. Systematic sampling has its advantages, especially in survey logistics (e.g., stations are equally spaced) and variance reduction for homogeneously distributed populations. Randomized designs, especially with stratification, can allow for higher flexibility to different levels of total survey effort while providing robust and unbiased survey estimates of abundance and variance. Stratum boundaries and station allocations among strata can also be optimized to weight species of importance (Oyafuso et al., 2021).

We evaluated the bias and precision of survey estimates of abundance using a systematic fixed-grid survey design along with two types of randomized designs in the US Chukchi Sea BTS. Spatiotemporal distributions for 18 representative demersal fish and invertebrate taxa were fitted based on historical bottom trawl catch and effort data. The models used to fit these spatiotemporal relationships were then used to simulate taxon densities on which surveys under different designs could be conducted. Three conventional survey designs were evaluated: simple random sampling (SRS), stratified random sampling (STRS), and a fixed-grid systematic (SYS) grid similar to what is employed in the NMFS Bering Sea BTS. Design-based estimates of abundance and precision from the three survey designs across a range of sampling effort were calculated, from which the performance of each design was evaluated. We evaluated the advantages and tradeoffs of using a systematic grid as previously done in the NMFS Chukchi Sea BTS and then highlighted potential improvements to the survey by using randomized designs. This analysis is intended to provide a template for a modified Chukchi Sea groundfish survey design going forward and will be important when transitioning to ecosystem-focused survey objectives.

## Methods

### Survey area and historical datasets

The US Chukchi Sea sampling frame was defines as a 2-nmi resolution grid (N = 15,736 cells or sampling units) that extends north of the Bering Strait and is bounded by the Barrow Canyon 100-m isobath to the north, US-Russia Maritime Boundary to the west, and the 10-m isobath along the Alaska coastline to the east.

Readers are referred to Stauffer (2004) and Deary et al. (2021) for a detailed specification of the gears used in this study. We will briefly introduce and identify the major differences between the two gears used.

83-112 Eastern otter trawl ("otter trawl" hereafter): Surveys from two years, 1990 and 2012, were included in this analysis due to the consistencies in the sampling protocol. In 1990, 48 stations were sampled along 11 transect lines perpendicular to shore near Point Hope, Alaska (Barber et al., 1997). In 2012, a systematic sampling design was employed based on a 30-nmi square grid with the planned trawl stations located at the approximate center of each grid cell, resulting in a total of 73 sampling locations, 71 of which were successful and included in the analysis. The wings and throat sections of the trawl net have a 10.2 cm mesh size. The codend has a 8.9 cm mesh size and a smaller-meshed 32-mm liner for retaining smaller organisms. Otter trawl tows were trawled at a target speed of 3 knots for 15 minutes. Acoustic net mensuration sensors were used to assess trawl performance and to provide net width for calculating effort (total area swept, the product of net width and distance trawled with bottom contact).

Plumb staff beam trawl ("beam trawl" hereafter): Surveys from three years, 2012, 2017, and 2019 were included in this analysis and used the same systematic grid as the 2012 otter trawl survey. In 2012, a tickler chain preceded the trawl footrope (Gunderson and Ellis, 1986; Kotwicki et al., 2017). Beam trawl tows from 2017 and 2019 were conducted as part of the Arctic Integrated Ecosystem Survey component of the Arctic Integrated Ecosystem Research Program. The body of the trawl has 7-mm mesh with a 4-mm mesh at the cod end. In 2017 and 2019, the tickler chain was removed, and the trawl was modified with a footrope of 10.2-cm rubber discs over a steel chain as in Abookire and Rose (2005). In all beam trawl survey years, effort was calculated similar to the otter trawl, with a bottom contact sensor to determine distance fished by the trawl. Effective trawl width of the beam trawl was assumed to be 2.26 m in 2012 (Gunderson and Ellis, 1986; Kotwicki et al., 2017), and 2.1 m in 2017 and 2019 (Abookire and Rose, 2005). Beam trawl tows were trawled at a target speed of 1.5 knots for 2.9-7.5 minutes. Catch samples from the beam and otter trawls were identified and sorted to the lowest possible taxonomic group, weighed, and counted. Field identifications of a subset of age-0 gadids in 2017 and 2019 were confirmed with genetic techniques (see Wildes et al., 2022).

## Species list

The set of taxa we chose to include in this analysis was influenced by cultural importance to Bering Strait and Chukchi Sea communities, commercial and ecological importance, availability in the dataset, adequate catchability to the two bottom trawl gears, and the ability to fit informative spatiotemporal distribution models to survey catch data. Taxonomic groupings were defined from a prior northern Bering Sea analysis of bottom trawl surveys conducted from 2010-2021 (Markowitz et al., 2022). These taxonomic groupings were important representatives of the demersal marine community as identified by Bering Sea native communities (Markowitz et al., 2022). We do not have similar distinctions for those communities living within the Chukchi Sea, however these taxonomic groupings represent a diverse range of fish and invertebrate taxa in an area proximal to the Chukchi Sea *via* the Bering Strait. Taxa were further filtered to those with reasonably high catchability for each of the two gears (Lauth et al., in review) and models were fit separately for each taxon and gear type to reflect those differences in catchability.

## Conditioning and operating models

We conditioned univariate spatiotemporal distribution models on historical catch and effort survey data for a particular gear and taxon using the VAST (vector autoregressive spatio-temporal) R Package [v. 4.0.2; Thorson and Barnett (2017); Thorson (2019)]. The VAST model applied here is a spatiotemporal generalized linear mixed-effects model where Gaussian Markov random effects describe spatial and/or spatiotemporal variation (spatial variation that is constant or time-varying, respectively) in density and temporal variation in the mean density is modeled as a fixed effect of survey year. Continuous spatial and/or spatiotemporal random fields were approximated using the INLA R package [www.r-inla.org; Rue et al. (2009)] using a mesh with 200 spatial "knots" where the values of spatial variables between knot locations are calculated *via* bilinear interpolation. Spatiotemporal fields were modeled as independent and identically distributed among years. If a model with spatiotemporal variation included resulted in a decreased (i.e. ≥2-unit decrease) AIC value relative to the model estimated with only spatial variation, it was chosen as the operating model for a given taxon/gear combination. Otherwise, a model with only estimated spatial variation was chosen. The "Poisson-link" reformulation of a conventional delta model was used (Thorson, 2018), and a gamma distribution was specified for modeling biomass density.

The density ($kgkm^{-2}$) of each taxon was predicted onto the Chukchi spatial domain based on the maximum likelihood estimates of the parameters of the chosen model for each gear type. The total abundance index ($I_{st}$) of taxon $s$ in year $t$ was calculated using an epsilon bias-correction technique (Thorson and Kristensen, 2016) and represents the "true" abundance from which to evaluate the design-based abundance indices of the different surveys tested. Using the fitted spatiotemporal model as an operating model, population densities were simulated for each taxon with observation error to represent samples obtained by simulating surveys under different sampling designs as in the "Survey Simulation" section below.

## Survey designs

Three survey designs were tested: SRS, STRS, and a fixed-station systematic grid (SYS) under a range of total sampling effort from roughly 50 - 175 total stations. Distance from shore and latitude were used as stratum variables for the STRS designs and the SamplingStrata R package Barcaroli (2014) was used to optimize the placement of stratum boundaries and allocation of effort across strata subject to user-defined pre-specified precision targets for each taxon. A full explanation of the optimization methods can be found in Barcaroli (2014) and an application of the STRS survey design optimization in the Gulf of Alaska is described in Oyafuso et al. (2021; 2022). Appendix A provides more detail into how the STRS optimization was parameterized for the Chukchi BTS. For each gear type we optimized stratum boundaries for three- and four-stratum solutions, as this range of strata created the most reasonable solutions given the range of sample sizes analyzed.

## Survey simulation

The estimated abundance index $\hat{I}_{st}$ for taxon $s$ in year $t$ and associated variance for the three designs were calculated following Wakabayashi et al. (1985):

$$\hat{I}_{st} = \sum_{l=1}^{L} A_l \overline{CPUE_{lst}}$$

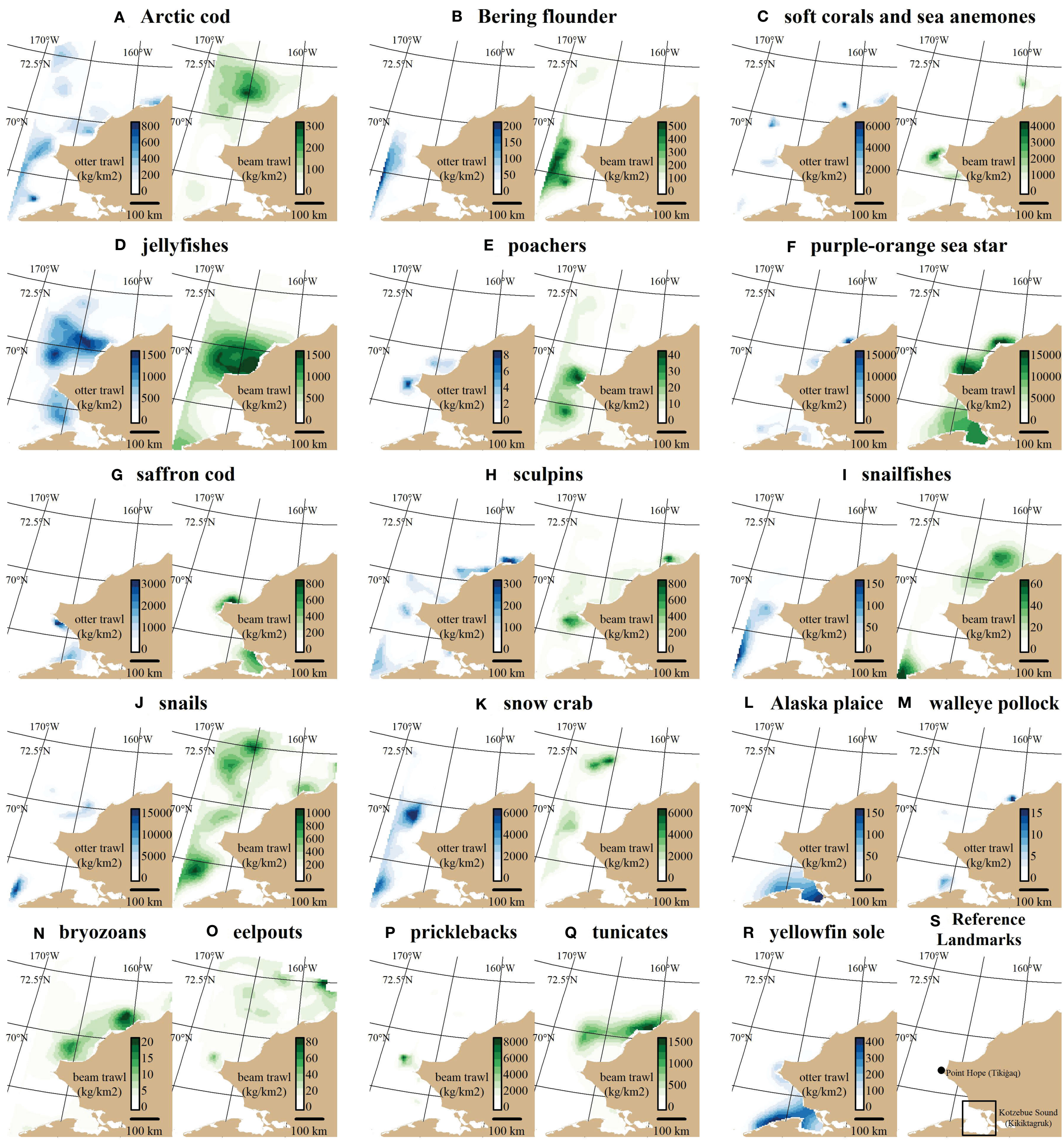



FIGURE 1
Predicted densities ($kgkm^{-2}$) for each taxon **(A–R)** under each gear type shown for the most recent survey year for a given gear type (2012 for the otter trawl (blue gradient) and 2019 for the beam trawl (green gradient)). **(S)** shows locations of major landmarks described in the text. Some taxa under a particular bottom trawl gear did not have an adequately fitting spatiotemporal distribution model (see Table 1 for specifications).

$$Var(\widehat{I_{st}}) = \sum_{l=1}^{L} A_l^2 Var(\overline{CPUE_{lst}})$$

where $\overline{CPUE_{lst}}$ is the mean CPUE (units $kgkm^{-2}$) in stratum $l$ ($L$ total strata), taxon $s$, and year $t$, and $A_l$ is the total area (units $km^2$) of stratum $l$.

The above equations can be used for calculating total abundance and variance under SRS and SYS by assuming one stratum, $L = 1$. While some studies indicate more appropriate variance estimators for systematic designs (e.g., Aune-Lundberg and Strand, 2014), the naive approach of assuming SRS estimators was used to calculate the abundance index and variance for the SYS simulations.

Each survey was replicated for $M = 1,000$ iterations. It was assumed that all sampling units were available for trawling, however in practice, variation in bottom rugosity and currents may render some sampling units untrawlable (i.e., unavailable to the sampling

TABLE 1 List of the fish and invertebrate taxa and associated gears included in the analysis.

| Scientific Name | Common Name | Gear |
|---|---|---|
| *Pleuronectes quadrituberculatus* | Alaska plaice | otter trawl |
| *Boreogadus saida* | Arctic cod | beam and otter trawl |
| *Hippoglossoides robustus* | Bering flounder | beam and otter trawl |
| Family: Zoarchidae | eelpouts | beam trawl |
| Family: Agonidae | poachers | beam and otter trawl |
| Family: Stichaeidae | pricklebacks | beam trawl |
| *Eleginus gracilis* | saffron cod | beam and otter trawl |
| Family: Cottidae | sculpins | beam and otter trawl |
| Family: Liparidae | snailfishes | beam and otter trawl |
| *Gadus chalcogrammus* | walleye pollock | otter trawl |
| *Limanda aspera* | yellowfin sole | otter trawl |
| Phylum: Bryozoa | bryozoans | beam trawl |
| Class: Scyphozoa | jellyfishes | beam and otter trawl |
| *Asterias amurensis* | purple-orange sea star | beam and otter trawl |
| Class: Gastropoda | snails | beam and otter trawl |
| *Chionoecetes opilio* | snow crab | beam and otter trawl |
| Class: Anthozoa | soft corals and sea anemones | beam and otter trawl |
| Subphylum: Tunicata | tunicates | beam trawl |

frame). Due to the limited data used to condition the operating model, high positive outliers in density masked the trends in the performance metrics. Thus, prior to calculating the performance metrics, positive outliers greater than three standard deviations above the mean among survey replicates were removed.

## Performance metrics

Three performance metrics were used to evaluate survey designs. The True CV ($TrueCV_{st}$) is the variability of the estimated abundance index across the survey replicates and is defined as the standard deviation of the estimated indices of abundance normalized by the true value, $\frac{\sqrt{Var(\hat{I_{st.}})}}{I_{st}}$, where $\hat{I_{st.}}$ refers to the vector of estimated indices for taxon $s$ and year $t$ across the $M$ replicates. The True CV provides two pieces of information about the precision of the survey design: 1) if the True CV is low for simulated densities generated from one type of survey (e.g., $< 0.2$), that is an indication that the survey is appropriate for a species with that type of distribution (i.e., the data quality is high); and 2) a very low True CV (e.g., $< 0.05$) can indicate that any survey will have a hard time estimating the variability in the density of the target species, in which case the relative root-mean-square error (RRMSE) of the CV is a useful diagnostic for determining whether a proposed survey can provide a reliable estimate of CV. The RRMSE of CV is defined as $\frac{\sqrt{\sum_{m=1}^{M}(CV_{stm} - TrueCV_{st})^2/M}}{\hat{CV_{st.}}}$ where $\hat{CV_{st.}}$ refers to the vector of estimated sample CVs for taxon $s$ and year $t$ across the $M$ replicates. Lastly, bias is the residual of a quantity relative to its assumed "true" value. Bias of the estimated index of abundance from a sample is relative to the assumed true index conditioned by the data. Bias of the estimated sample CVs associated with the index of abundance is relative to the True CV.

## Code repository

The code used to perform this analysis and format this manuscript is currently stored in a code repository in Z. Oyafuso's NOAA GitHub account and can be accessed at https://github.com/zoyafuso-NOAA/chukchi_survey_evaluation.

# Results

## Species distributions

The species included in this analysis exhibited a diversity of spatiotemporal distributions (Figure 1; see Appendix B for full spatiotemporal distributions and diagnostic plots). Alaska plaice (Figure 1L), saffron cod (Figure 1G), and yellowfin sole (Figure 1R) were restricted to the southeastern portion of the domain which includes Kotzebue Sound. Bryozoans (Figure 1N), tunicates (Figure 1Q), sculpins (Figure 1H), poachers (Figure 1E) and jellyfishes (Figure 1D) were more commonly observed in the middle of the domain around Point Hope. Purple-orange sea stars (Figure 1F) had a broad nearshore distribution along much of the coastline of the domain, whereas eelpouts (Figure 1O), snailfishes (Figure 1I), and Bering flounder (Figure 1B) had more offshore distributions along the western edge of the domain. Snails were commonly observed across the spatial domain across both gears (Appendix B14). Arctic cod were commonly observed with broad distributions across the domain (Figure 1A), although with higher densities at beam trawl stations in the northern part of the domain in 2019 compared to beam trawl stations in 2012 and 2017 (Appendix B2). Soft corals and sea anemones (primarily the sea raspberry *Gersemia rubiformis* and miscellaneous anemones; Figure 1C) and walleye pollock (uncommonly observed; Figure 1M) had patchier distributions. Snow crab had higher offshore densities near the western boundary of the domain (Figure 1K) but were present in high densities in the northern part of the domain as well (Appendix B15).

 10.3389/fmars.2023.1214526

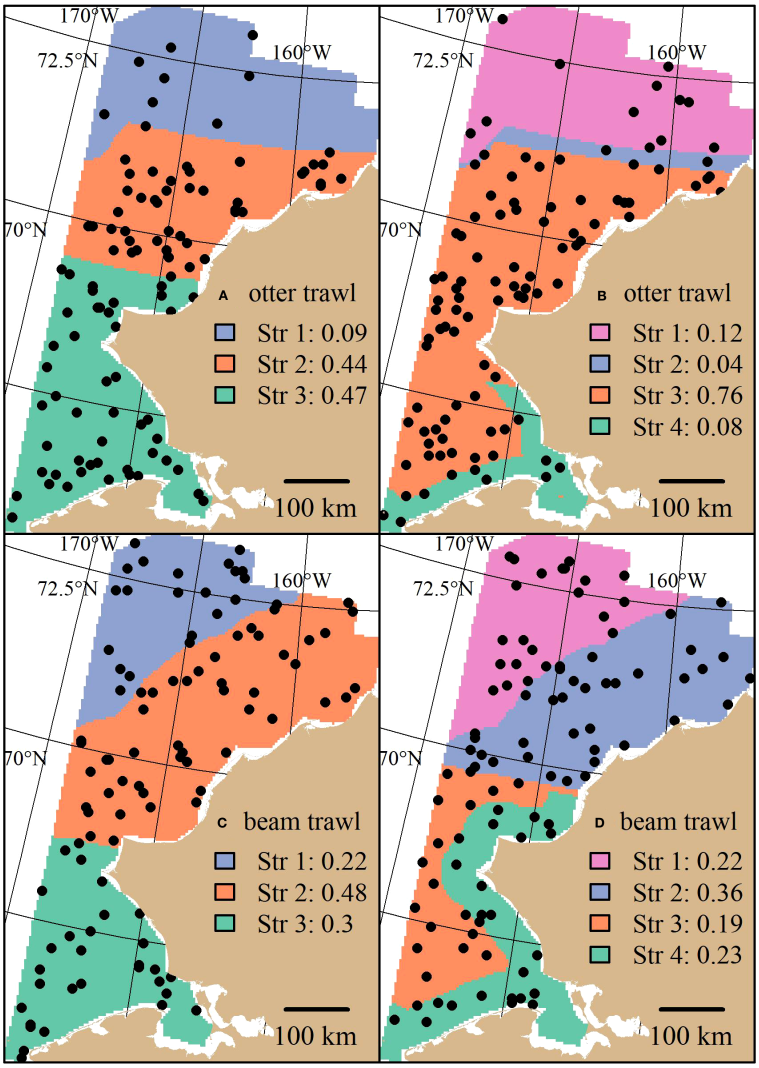



FIGURE 2
Stratified random designs resulting from the stratified random design optimization algorithm using three and four strata for the otter **(A, B)** and beam **(C, D)** trawl gears. Distance to shore and latitude characterize the different strata. An example of 100 stations randomly drawn from the optimal allocation are superimposed as points. The proportion of stations allocated across strata are shown in the legend.

## Multispecies STRS design optimization

Stratum boundaries of both otter and beam trawl survey optimizations generally separated the domain of the Chukchi Sea into two latitudinal sections split at roughly 69 and 70 degrees N latitude (Figure 2). The three-stratum otter trawl solution (Figure 2A) consists of two latitudinal boundaries at roughly 70 and 71 degrees N latitude. The four-stratum otter trawl solution (Figure 2B) shares the northern latitudinal boundary at 71 degrees N latitude but also adds a nearshore stratum in the southern part of the domain. The three-stratum beam trawl solution (Figure 2C) has a southern stratum with a northern boundary at roughly 69 degrees N latitude and two inshore/offshore strata in the northern section of the domain. The four-stratum beam trawl solution (Figure 2D) is similar to the three-stratum beam trawl solution but two inshore/offshore strata in the southern section of the domain.

Sampling densities for the otter trawl STRS designs were generally higher in the southern and central strata and less so in the northern strata. Sampling densities for the beam trawl solutions were proportional to stratum area. For the subsequent survey simulation section, the four-stratum solution for the beam trawl and the three-stratum solution for the otter trawl were used as the representatives of the STRS design in the survey simulations.

## Survey performance

The random designs (SRS and STRS) monotonically decreased in True CV with increased sample size for both gears. Since CV and precision are conversely related (lower True CV is interpreted as higher precision and vice versa), we will describe survey performance using both terms. The STRS designs often provided lower True CVs

10.3389/fmars.2023.1214526

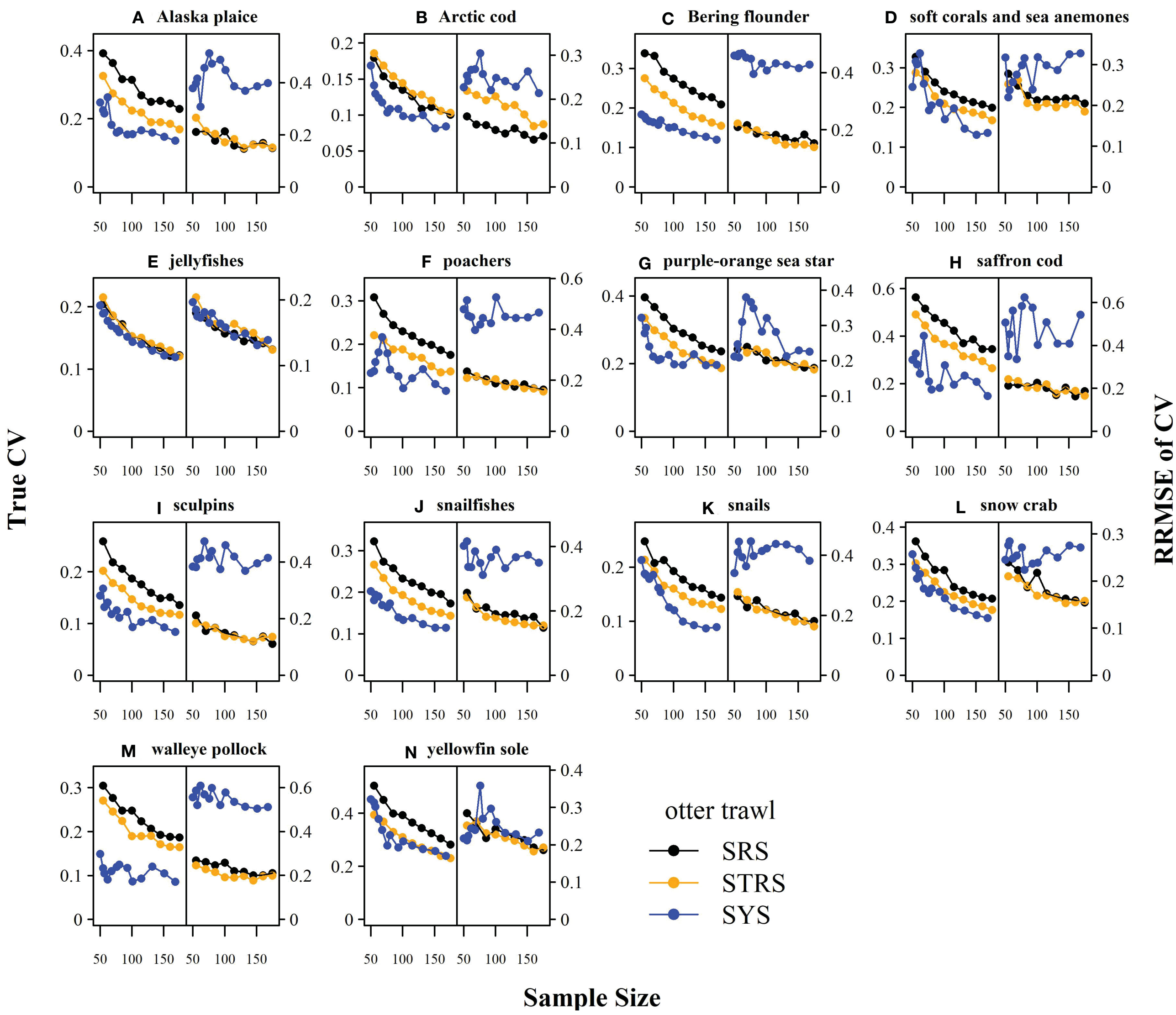



FIGURE 3
True CV (left-side of panel) and relative root mean square error (RRMSE) of CV (right-side of panel) across a range of total sampling effort for each taxon **(A–N)** and survey design for the otter trawl gear. SRS, simple random sampling; STRS, stratified random sampling optimized over the species set; SYS, fixed-grid systematic sampling.

than the SRS designs at equivalent sample sizes, especially for taxa collected *via* the otter trawl (Figure 3). The increase in precision from a random to a stratified design was less for the taxa sampled with the beam trawl, with many taxa performing similarly to the SRS design (Figure 4). Given the limited data used to condition the operating model, the inconsistent +/- 5% bias observed with the estimated index is fairly low (Figures 5, 6).

The SYS design often provided the lowest True CVs compared to the two random designs; however, this design displayed inconsistent behavior, as the True CV did not always decrease with sample size. Furthermore, there was a tradeoff observed for many taxa under both gears, where lower True CVs were associated with much higher RRMSE of CV (Figures 3, 4). The higher RRMSE of CV of the fixed systematic grid was attributed to a high positive bias of the simulated sample CVs relative to the True CV (Figures 5, 6). The average bias of the abundance indices for the SYS designs across taxa were not consistent across total sample size, with as much as a 25% fluctuation in average bias (Figures 5, 6).

## Discussion

When considering changes to ecological surveys, one must weigh the advantages of consistency with historical designs in the same or adjacent regions against potential gains in efficiency and flexibility of a new design. A SYS design, as currently implemented in the Bering Sea BTS, may be a logical choice for a Chukchi BTS as a natural extension to the established Bering Sea SYS design. Surveys conducted under a SYS design provide good spatial

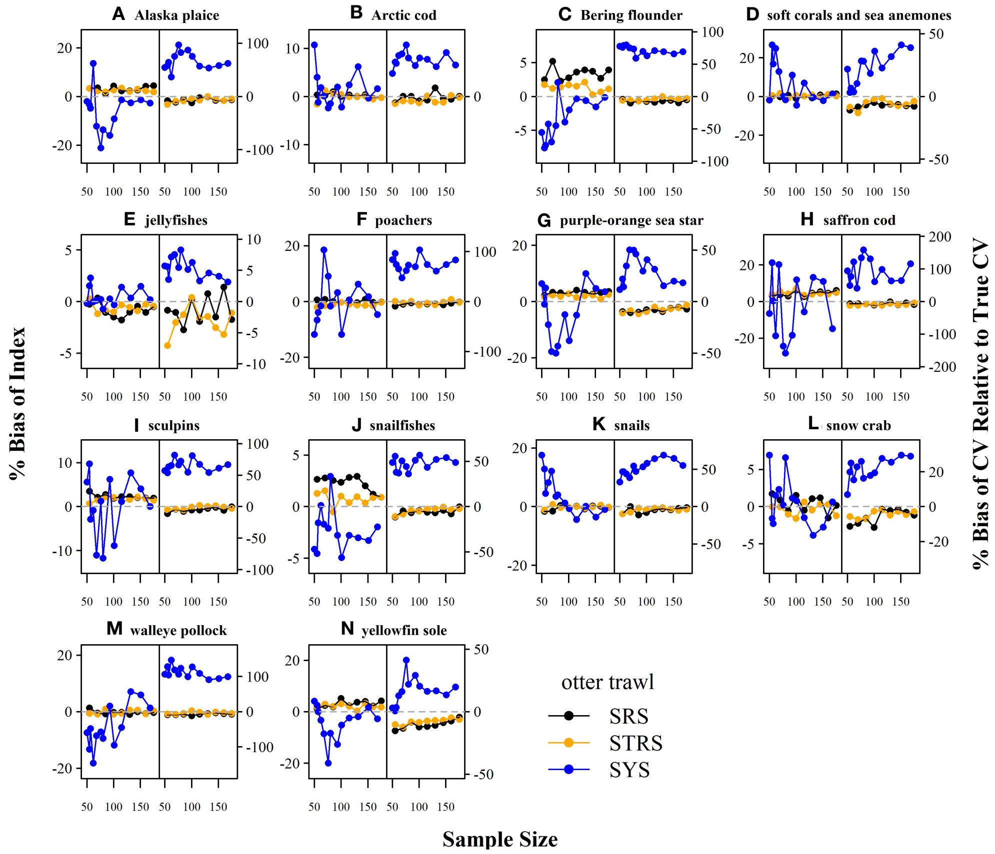



FIGURE 4
True CV (left-side of panel) and relative root mean square error (RRMSE) of CV (right-side of panel) across a range of total sampling effort for each taxon **(A–N)** and survey design for the beam trawl gear. SRS, simple random sampling; STRS, stratified random sampling optimized over the species set; SYS, fixed-grid systematic sampling.

coverage of the sampling domain and can thus be advantageous in the early, data-limited stages of a survey time series. Having evenly spaced sampling stations is also logistically advantageous, in that the completion rate of stations per day is more consistent than with stations chosen under randomized designs. When minimizing the survey CV is the top priority, systematic survey designs should ideally be created with random starting locations to slightly vary the locations of stations within the sampling frame. However, the SYS design as currently implemented in all Bering Sea BTS is the most practical survey design due to those aforementioned logistical survey planning advantages.

The main tradeoff of the logistical advantages of the SYS design was the reduced quality of the statistical data products that might result from such a design, as observed in our simulation testing. We found that randomized designs provided more reliable estimates of abundance and precision than SYS designs for the US Chukchi Sea. While the True CVs for many taxa were lower under SYS, the estimates of the variance were less reliable (i.e., RRMSE of CV) when compared to both randomized designs. The tradeoff between the RRMSE of CV and True CV has been shown previously in the Gulf of Alaska when comparing proposed optimized STRS designs with historical STRS designs using similar simulation testing (Oyafuso et al., 2022). Variance is a critical measure of the quality of a survey and can be used as a data weight in stock assessment models, however the estimation of variance can be unreliable depending on the design of the survey, along with other considerations like variation in catchability (Kotwicki and Ono, 2019). The stratified random designs created in our analysis provided an advantageous combination of increased precision relative to SRS and increased reliability of the estimated CVs relative to the True CVs.

A challenge of designing STRS surveys in a region like the Chukchi Sea with highly dynamic oceanographic conditions is that historical data to inform the design (i.e., stratification and effort allocation across strata) may not represent the current ecosystem state, similar to the challenge of forecasting species distributions to novel environmental conditions due to climate change (Brodie et al., 2022). While the last NMFS beam trawl survey in the Chukchi Sea occurred in 2019, the most recent Chukchi Sea NMFS otter trawl survey occurred in 2012. Within the same range of time (i.e., the last ten years), there have been significant poleward shifts in the distributions of many subarctic taxa common to the Bering Sea (Kotwicki and Lauth, 2013; Stevenson and Lauth,

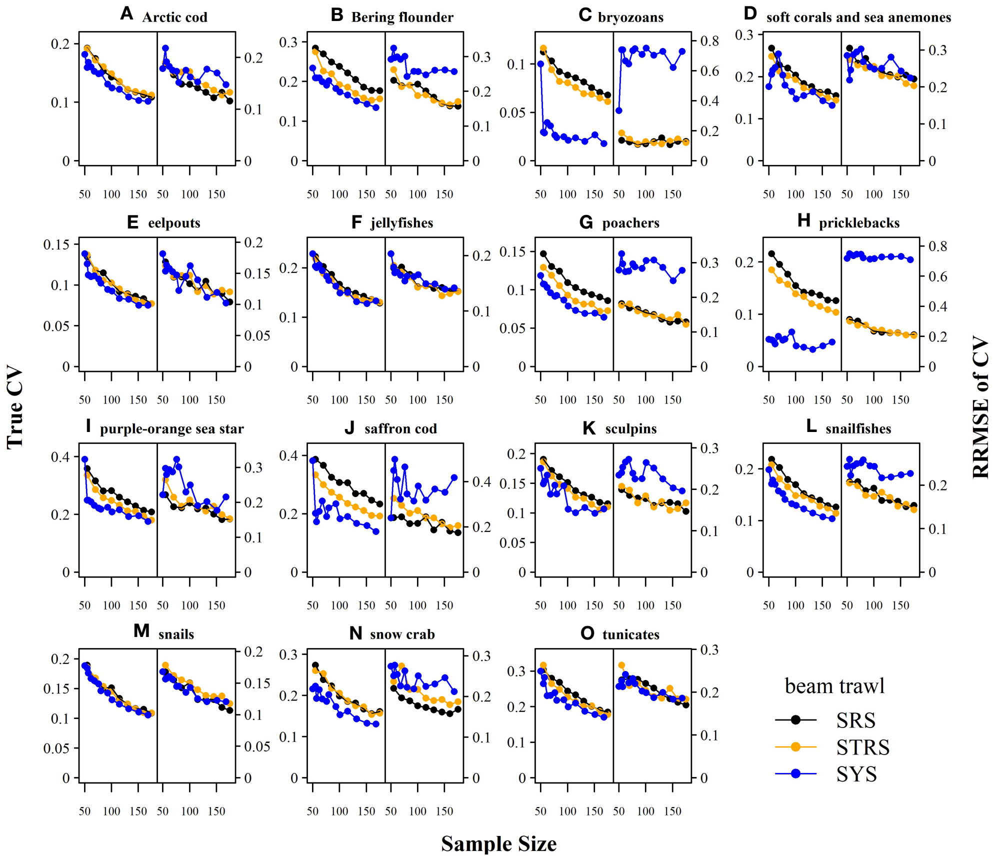



FIGURE 5
Average percent bias of the 1) estimated abundance relative to the true abundance (left-side of panel) and 2) estimated sample coefficient of variation (CV) relative to the True CV (right-side of panel) across sample size for each taxon **(A–O)** and survey design for the otter trawl gear. Dashed grey line at zero included for reference. SRS, simple random sampling; STRS, stratified random sampling optimized over the species set; SYS, fixed-grid systematic sampling.

2019; Maznikova et al., 2023a) but previously seldom observed in the Chukchi Sea, including many Bering Sea gadids like walleye pollock (Datsky et al., 2022; Wildes et al., 2022). With continued sampling of the region, the design of a STRS survey could be easily modified to reflect the species distributions observed in more recent years. The discussion of the range of years to include when planning surveys is outside the scope of this paper, however our approach to updating STRS designs is amenable to testing and planning STRS designs that incorporate varying ranges of years to provide more weight to contemporary data.

We investigated survey designs implemented with both otter and beam trawl gears in order to anticipate survey designs consistent with the standardized bottom trawl gears used for NMFS-AFSC BTS. The patterns among survey designs previously discussed were present in both the beam and otter trawl gears. However, there were some differences in the optimized STRS designs calculated for each gear type. The STRS designs for both gears had similar stratifications that split the Chukchi spatial domain by two or three latitudinal regions and inshore/offshore strata. However, the sampling densities for the otter trawl solutions were higher in the southern and central strata compared to the northern strata, whereas the beam trawl sampling densities were nearly proportional to stratum area. As a result, the performance of the STRS beam trawl survey abundance estimates were similar to the SRS design with some improvement in True CV for a handful of taxa (e.g., Bering flounder, pricklebacks, saffron cod). We presume that the expected gains in precision that come from stratification were diminished because of the strong tradeoffs that exist when optimizing over a wide set of taxa with non-overlapping spatiotemporal distributions. Lastly, additional examination of optimal number of strata along with the choice of other relevant stratum variables (e.g., sediment type, depth, temperature, etc.) similar to (Oyafuso et al., 2021; Oyafuso et al., 2022) could further improve the statistical efficiency of a Chukchi Sea STRS BTS design.

The list of taxa to include in survey planning is an important decision process and should be a part of broader discussions about survey objectives. We curated our taxa list by first considering taxa that can be appropriately sampled by either the otter and/or beam trawl gears (Lauth at al., in review). We then considered commercial

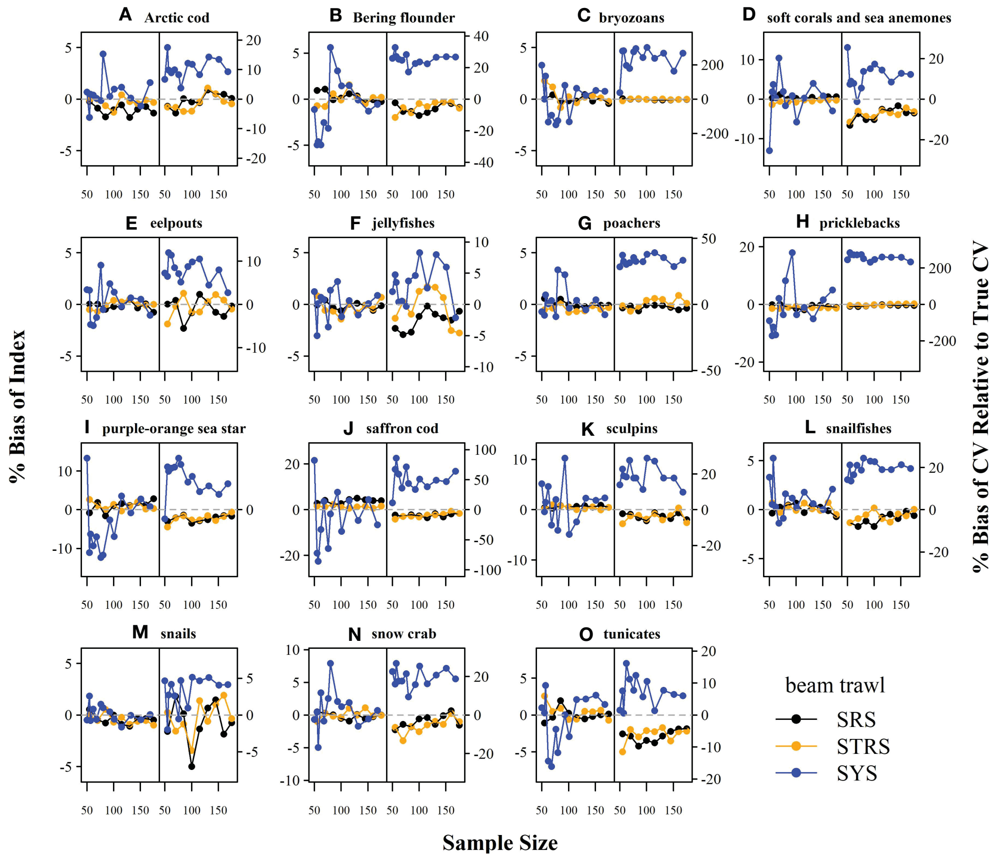



FIGURE 6
Average percent bias of the 1) estimated abundance relative to the true abundance (left-side of panel) and 2) estimated sample coefficient of variation (CV) relative to the True CV (right-side of panel) across sample size for each taxon **(A–O)** and survey design for the beam trawl gear. Dashed grey line at zero included for reference. SRS, simple random sampling; STRS, stratified random sampling optimized over the species set; SYS, fixed-grid systematic sampling.

importance given the distribution shifts of commercially important Bering Sea species into the Chukchi Sea as well as species like Arctic cod that have been observed to be trophically important in the Chukchi Sea for various seabirds and marine mammals (Kokubun et al., 2015; Quakenbush et al., 2015; Florko et al., 2021). Lastly, it is critical to engage with stakeholders to consider their values and understand how to monitor species of direct and indirect (e.g., dependent prey) importance to the resources they use. In the US Chukchi Sea, the primary stakeholders are coastal Alaska Native communities. Marine mammals are important to Alaska Native communities for subsistence and cultural value and while trawl surveys cannot monitor marine mammals, they can be used to monitor prey species on which these marine mammals depend. We have used information learned from Alaska Native communities representing the northern Bering Sea (Markowitz et al., 2022) to identify species used for subsistence or other purposes. Furthermore, we have begun more extensive efforts to consult with Alaska Native communities in the US Arctic to further tailor potential monitoring efforts to align with their values. In summary, we recommend that ecosystem monitoring surveys be designed with thorough consideration of the values and objectives of all major components of the socio-ecological system and how these relate to the limitations of what can be effectively monitored with the observational methods available.

## Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: https://github.com/zoyafuso-NOAA/chukchi_survey_evaluation.

## Author contributions

ZO, LB, and SK conceived the idea of the research project. DC provided an additional source of data. All authors contributed to the article and approved the submitted version.

## Acknowledgments

We would like to acknowledge the efforts of the NOAA AFSC Groundfish Assessment Program and Ecosystems and Fisheries-Oceanography Coordinated Investigations for successfully conducting the Chukchi Sea bottom trawl surveys. We would also like to thank NOAA-AFSC reviewers Elizabeth Logerwell, Thomas Hurst, Lukas DeFilippo, and two journal reviewers for comments that greatly improved the clarity of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1214526/full#supplementary-material

## References


Abookire, A. A., and Rose, C. S. (2005). Modifications to a plumb staff beam trawl for sampling uneven, complex habitats. *Fish. Res.* 71, 247–254. doi: 10.1016/j.fishres.2004.06.006

Aune-Lundberg, L., and Strand, G.-H. (2014). Comparison of variance estimation methods for use with two-dimensional systematic sampling of land use/land cover data. *Environ. Model. Soft.* 61, 87–97. doi: 10.1016/j.envsoft.2014.07.001

Baker, M. R., Farley, E. V., Danielson, S. L., Mordy, C., Stafford, K. M., and Dickson, D. M. S. (2023). Integrated research in the arctic – ecosystem linkages and shifts in the northern Bering Sea and eastern and western Chukchi Sea. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 208, 105251. doi: 10.1016/j.dsr2.2023.105251

Barber, W., Smith, R., Vallarino, M., and Meyer, R. (1997). Demersal fish assemblages of the northeastern Chukchi Sea, Alaska. *Fish. Bull.* 95, 195–209. Available at: https://spo.nmfs.noaa.gov/sites/default/files/pdf-content/1997/952/barber.pdf

Barcaroli, G. (2014). SamplingStrata: an r package for the optimization of stratified sampling. *J. Stat. Soft.* 61, 1–24. doi: 10.18637/jss.v061.i04

Brodie, S., Smith, J. A., Muhling, B. A., Barnett, L. A. K., Carroll, G., Fiedler, P., et al. (2022). Recommendations for quantifying and reducing uncertainty in climate projections of species distributions. *Global Change Biol.* 28, 6586–6601. doi: 10.1111/gcb.16371

Budikova, D. (2009). Role of arctic sea ice in global atmospheric circulation: a review. *Global Planet. Change* 68, 149–163. doi: 10.1016/j.gloplacha.2009.04.001

Cooper, D. W., Cieciel, K., Copeman, L., Emelin, P. O., Logerwell, E., Ferm, N., et al. (2023). Pacific cod or tikhookeanskaya treska (*Gadus macrocephalus*) in the Chukchi Sea during recent warm years: distribution by life stage and age-0 diet and condition. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 208, 105241. doi: 10.1016/j.dsr2.2022.105241

Datsky, A. V., Vedishcheva, E. V., and Trofimova, A. O. (2022). Features of the biology of mass fish species in russian waters of the Chukchi Sea. 1. commercial fish biomass. family gadidae. *J. Ichthyol.* 62, 560–585. doi: 10.1134/S0032945222040051

Deary, A. L., Vestfals, C. D., Mueter, F. J., Logerwell, E. A., Goldstein, E. D., Stabeno, P. J., et al. (2021). Seasonal abundance, distribution, and growth of the early life stages of polar cod (*Boreogadus saida*) and saffron cod (*Eleginus gracilis*) in the US arctic. *Polar Biol.* 44, 2055–2076. doi: 10.1007/s00300-021-02940-2

Emelin, P. O., Maznikova, O. A., Benzik, A. N., Sheibak, A. Y., Trofimova, A. O., and Orlov, A. M. (2022). Invader's portrait: biological characteristics of walleye pollock *Gadus chalcogrammus* in the western Chukchi Sea. *Deep-Sea Res. Part II: Top. Stud. Oceanog.* 206:1–13. doi: 10.1016/j.dsr2.2022.105211

Florko, K. R. N., Tai, T. C., Cheung, W. W. L., Ferguson, S. H., Sumaila, U. R., Yurkowski, D. J., et al. (2021). Predicting how climate change threatens the prey base of arctic marine predators. *Ecol. Lett.* 24, 2563–2575. doi: 10.1111/ele.13866

Goddard, P., Lauth, R., and Armistead, C. (2014). Results of the 2012 Chukchi Sea bottom trawl survey of bottomfishes, crabs, and other demersal macrofauna (U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-278), 110. Available at: https://repository.library.noaa.gov/view/noaa/4777.

Gunderson, D. R., and Ellis, I. E. (1986). Development of a plumb staff beam trawl for sampling demersal fauna. *Fish. Res.* 4, 35–41. doi: 10.1016/0165-7836(86)90026-3

Huntington, H. P., Danielson, S. L., Wiese, F. K., Baker, M., Boveng, P., Citta, J. J., et al. (2020). Evidence suggests potential transformation of the pacific arctic ecosystem is underway. *Nat. Climate Change* 10, 342–348. doi: 10.1038/s41558-020-0695-2

Kokubun, N., Yamamoto, T., Sato, N., Watanuki, Y., Will, A., Kitaysky, A. S., et al. (2015). Foraging segregation of two congeneric diving seabird species (common and thick-billed murres) breeding on St. George island, Bering Sea. *Biogeosci. Discuss.* 12, 18151–18183. doi: 10.5194/bgd-12-18151-2015

Kotwicki, S., and Lauth, R. R. (2013). Detecting temporal trends and environmentally-driven changes in the spatial distribution of bottom fishes and crabs on the eastern Bering Sea shelf. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 94, 231–243. doi: 10.1016/j.dsr2.2013.03.017

Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. (2017). Selectivity ratio: a useful tool for comparing size selectivity of multiple survey gears. *Fish. Res.* 191, 76–86. doi: 10.1016/j.fishres.2017.02.012

Kotwicki, S., and Ono, K. (2019). The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys. *Fish Fish.* 20, 760–774. doi: 10.1111/faf.12375

Levine, R. M., De Robertis, A., Grünbaum, D., Wildes, S., Farley, E. V., Stabeno, P. J., et al. (2023). Climate-driven shifts in pelagic fish distributions in a rapidly changing pacific arctic. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 208, 105244. doi: 10.1016/j.dsr2.2022.105244

Markowitz, E. H., Dawson, E. J., Charriere, N. E., Prohaska, B. K., Rohan, S. K., Stevenson, D. E., et al. (2022). Results of the 2021 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna (U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-452), 227. doi: 10.25923/g1ny-y360

Maznikova, O. A., Emelin, P. O., Baitalyuk, A. A., Vedishcheva, E. V., Trofimova, A. O., and Orlov, A. M. (2023a). Polar cod (*Boreogadus saida*) of the siberian arctic: distribution and biology. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 208, 105242. doi: 10.1016/j.dsr2.2022.105242

Maznikova, O. A., Emelin, P. O., Sheibak, A. Y., Nosov, M. A., and Orlov, A. M. (2023b). Can an invader support commercial fishing? a case study of walleye pollock *Gadus chalcogrammus* in the western Chukchi Sea. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 207, 105222. doi: 10.1016/j.dsr2.2022.105222

Oyafuso, Z. S., Barnett, L. A. K., and Kotwicki, S. (2021). Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. *ICES J. Mar. Sci.* 78, 1288–1300. doi: 10.1093/ICESJMS/FSAB038

Oyafuso, Z., Barnett, L., Siple, M., and Kotwicki, S. (2022). A flexible approach to optimizing the gulf of alaska groundfish bottom trawl survey design for abundance estimation (U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-434), 142 doi: 10.25923/g5zd-be29.

Parkinson, C. L., and Comiso, J. C. (2013). On the 2012 record low arctic sea ice cover: combined impact of preconditioning and an august storm. *Geophys. Res. Lett.* 40, 1356–1361. doi: 10.1002/grl.50349

Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., et al. (2010). History of sea ice in the arctic. *Quater. Sci. Rev.* 29, 1757–1778. doi: 10.1016/j.quascirev.2010.02.010

Post, E., Bhatt, U. S., Bitz, C. M., Brodie, J. F., Fulton, T. L., Hebblewhite, M., et al. (2013). Ecological consequences of sea-ice decline. *Science* 341, 519–524. doi: 10.1126/science.1235225

Quakenbush, L. T., Suydam, R. S., Bryan, A. L., Lowry, L. F., Frost, K. J., and Mahoney, B. A. (2015). Diet of beluga whales (*Delphinapterus leucas*) in Alaska from stomach contents, march–november. *Mar. Fish Rev.* 77, 70–84. doi: 10.7755/MFR.77.1.7

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J. R. Stat. soc.: Ser. b (statistical methodology)* 71, 319–392. doi: 10.1111/j.1467-9868.2008.00700.x

Spies, I., Gruenthal, K. M., Drinan, D. P., Hollowed, A. B., Stevenson, D. E., Tarpey, C. M., et al. (2020). Genetic evidence of a northward range expansion in the eastern Bering Sea stock of Pacific cod. *Evolution. Appl.* 13, 362–375. doi: 10.1111/eva.12874

Stauffer, G. D. (2004) NOAA Protocols for groundfish bottom trawl surveys of the nation's fishery resources, march 16, 2003. Available at: https://spo.nmfs.noaa.gov/content/tech-memo/noaa-protocols-groundfish-bottom-trawl-surveys-nations-fishery-resources-march-16.

Stevenson, D. E., and Lauth, R. R. (2019). Bottom trawl surveys in the northern Bering Sea indicate recent shifts in the distribution of marine species. *Polar Biol.* 42, 407–421. doi: 10.1007/s00300-018-2431-1

Thorson, J. T. (2018). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Can. J. Fish. Aquat. Sci.* 75, 1369–1382. doi: 10.1139/cjfas-2017-0266

Thorson, J. T. (2019). Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Thorson, J. T., and Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single-and multispecies models of fishes and biogenic habitat. *ICES J. Mar. Sci.* 74, 1311–1321. doi: 10.1093/icesjms/fsw193

Thorson, J. T., and Kristensen, K. (2016). Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish. Res.* 175, 66–74. doi: 10.1016/j.fishres.2015.11.016

Wakabayashi, K. R., Bakkala, G., and Alton, M. S. (1985). Results of cooperative U.S.-japan groundfish investigations in the Bering Sea during may-august 1979. Eds. R. G. Bakkala and K. Wakabayashi (U.S. Dep. Commer., NOAA Tech. Memo. NMFS F/NWC-87), 7–29.

Wildes, S., Whittle, J., Nguyen, H., Marsh, M., Karpan, K., D'Amelio, C., et al. (2022). Walleye pollock breach the Bering Strait: a change of the cods in the arctic. *Deep Sea Res. Part II: Top. Stud. Oceanog.* 204, 105165. doi: 10.1016/j.dsr2.2022.105165

TYPE Original Research
PUBLISHED 15 June 2023
DOI 10.3389/fmars.2023.1192739

Check for updates

OPEN ACCESS

EDITED BY
Ben Scoulding,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Chongliang Zhang,
Ocean University of China, China
Jorge Paramo,
University of Magdalena, Colombia
J. Marcus Drymon,
Mississippi State University, United States

*CORRESPONDENCE
Julie L. Vecchio
vecchioj@dnr.sc.gov

†These authors have contributed equally to this work and share first authorship

RECEIVED 23 March 2023
ACCEPTED 29 May 2023
PUBLISHED 15 June 2023

CITATION
Vecchio JL, Bubley WJ and Smart TI (2023) Increased fishery-independent sampling effort results in improved population estimates for multiple target species. *Front. Mar. Sci.* 10:1192739.
doi: 10.3389/fmars.2023.1192739



# Increased fishery-independent sampling effort results in improved population estimates for multiple target species

Julie L. Vecchio*†, Walter J. Bubley† and Tracey I. Smart†

Marine Resources Research Institute, South Carolina Department of Natural Resources, Charleston, SC, United States

The power of fishery-independent surveys for stock assessments and management decisions is in their consistency over time and space. Although the preference is to limit change to survey execution, such changes may be necessary. In multi-species surveys, changes that improve metrics for one species may be a detriment to survey performance for others. In 2010, the Southeast Reef Fish Survey (SERFS) was formed to better address sampling needs off the U.S. Southeast Atlantic coast by intensifying a historical chevron trap survey (MARMAP), especially at the northern and southern extent of the sampling range. We used several performance metrics (encounter rate, annual coefficient of variability, standard error, and relative abundance index values) to determine the impact of this change in survey coverage on trend estimates for three commonly encountered species with varying centers of distribution in the survey region. Gray Triggerfish (*Balistes capriscus*) is found throughout the range of both surveys (i.e. centrally-distributed), while White Grunt (*Haemulon plumierii*) and Red Snapper (*Lutjanus campechanus*) are centered near the northern and southern extent of the sampling range, respectively. For Gray Triggerfish, the survey intensification had no effect on encounter rate, but reduced the coefficient of variation and indicated that the historical index of relative abundance may have been overestimated. For White Grunt, the survey intensification slightly improved CV but did not affect the index of relative abundance value or encounter rate. For Red Snapper, SERFS increased encounter rates, reduced CV overall, and detected a population increase 5 years earlier than MARMAP. Overall, the intensification of the survey improved at least one performance metric for each species and showed few deleterious effects on performance, suggesting that intensification of the survey was a net-positive for the accurate estimation of population trends in several species of interest.

KEYWORDS

**fishery-independent index, fish population estimate, Red Snapper, fishery-independent sample changes, White Grunt, Gray Triggerfish**

# 1 Introduction

Well-regulated fisheries rely on information that can be used to monitor the status of the stock, interpret fisheries landings data, perform stock assessments, and develop regulations for managing fisheries resources (Apostolaki and Hillary, 2009; Hilborn et al., 2020; Gebremedhin et al., 2021). A key data input for these purposes is an index of relative abundance that reflects trends in the population and is proportional to the size of the stock (Hilborn and Walters, 1992). The utility of any fishery data in detecting trends, understanding variability, providing a baseline, and evaluating impacts of natural events or human-related activities, is dependent on consistency of the survey over space and time. Indices of relative abundance can be developed using data obtained by fishery-dependent or fishery-independent sources, but when fisheries are highly regulated, fishery-independent surveys are often the only method available to adequately characterize relative population size due to the limitations of economics, regulation, and recall on fishery-dependent data sources (Potts and Manooch III, 2002; Hamilton et al., 2016).

Marine fishery-independent surveys are often designed to simultaneously capture trends in abundance and life-history characteristics for several managed species concurrently due to cost and overlap of habitat utilization by numerous fishery targets (Dennis et al., 2015; Christiansen et al., 2020). While multi-species surveys are an efficient use of resources, indices of abundance for each target species may be more sensitive, or less sensitive, to population changes based on the relative sampling density in relation to species distribution. In addition, the geographic ranges of economically-important species may not follow previous survey boundaries, and their centers of distribution may change over time, necessitating the re-examination of the assumption of representative sampling for each target (Smart et al., 2020; O'Leary et al., 2021; Damiano, 2023). Surveys may be expanded or intensified due to a recognized need to increase available data for a particular species (Williams and Carmichael, 2009; Schrandt et al., 2021; Thompson et al., 2022), or they may be scaled back due to funding reductions (Zimney and Smart, 2022).

The South Carolina Department of Natural Resources has operated the Marine Resources Monitoring, Assessment, and Prediction Program (MARMAP) fishery-independent chevron trap survey in Atlantic waters off the southeastern United States since 1990 (Collins, 1990; Bubley et al., 2023). Anecdotal evidence suggested an increase in the Red Snapper, *Lutjanus campechanus*, population in the early 2000's, during a period when the existing MARMAP survey showed low and stable population numbers (SEDAR, 2009; Williams and Carmichael, 2009). The center of distribution for Red Snapper occurs near the southern terminus of the survey range, where MARMAP sample coverage was relatively poor at the time. To enhance ongoing survey efforts, especially in relation to species with centers of distribution offset from the core of MARMAP sampling, additional resources were incorporated to form the Southeast Reef Fish Survey (SERFS) in 2010. The formation of SERFS increased the density of stations available for sampling through mapping efforts using modern tools such as multibeam bathymetry, submission of potential sampling sites by the fishing industry, and deployments of cameras to confirm potential bottom for reef fish, particularly in areas at the southern and northern extent of the range (Bacheler et al., 2017). The creation of SERFS also allowed the survey to approximately triple the number of chevron traps deployed each year, effectively increasing the breadth of habitat types sampled (Glasgow et al., 2021). The effects of chevron trap survey intensification on the Red Snapper index of relative abundance were explored during a subsequent regional stock assessment (Ballenger et al., 2014). Although similar trends in relative abundance were observed between the full time series (1990-2014) and a time series beginning in 2010 with the inception of the full SERFS effort (Ballenger et al., 2014), the decision was made to use the short, five-year time series (2010-2014) due to decreased uncertainty surrounding the estimates (SEDAR, 2017).

Now that more than a decade has passed since the inception of SERFS, sufficient data are available to explore the effect of the sampling intensification on species-specific indices of relative abundance for several commonly encountered reef fish species. While increasing survey effort may be predicted to result in equivalent or improved estimates of relative abundance over surveys with lower effort, this assumption should be examined on a case-by-case basis to determine appropriate actions. Retrospective analysis provides a means to assess potential biases and ascertain the most appropriate method of employing each single-species index of relative abundance for management decision-making, with options including statistical correction to ensure continuity of the data or breaking the time series, effectively excluding historical data (Brodie et al., 2022; Zimney and Smart, 2022). The decision for appropriate incorporation of an index of relative abundance may vary by species, depending on distribution, habitat use, or behavior (Brodie et al., 2022; Zimney and Smart, 2022). Because the length of the time series is essential for understanding current population levels in relation to long-term trends, care must be taken and a variety of factors must be weighed, before a decision is made.

To examine the effects of survey intensification on indices of relative abundance, we selected three model species. Species whose center of distributions were in the extremes of the survey range to the north (White Grunt, *Haemulon plumierii*) and south (Red Snapper) and one with a relatively even latitudinal distribution throughout the region (Gray Triggerfish, *Balistes capriscus*) were chosen to explore a variety of distribution patterns. The goal of this work is to examine an input used in the SEDAR (Southeast Data Assessment and Review) stock assessment process, to provide the best available data for upcoming U.S. federal stock assessments in the South Atlantic region. The results from the current study will be instrumental in understanding the utility of the increase in sampling density on the indices of relative abundance for target species with varying distribution patterns and population trends, while providing guidance on the incorporation of the chevron trap survey time series for upcoming assessments and management actions.

## 2 Materials and methods

### 2.1 Sample collections

Throughout the survey period (1990-2022), chevron traps were deployed on randomly selected monitoring stations from a universe of known low- to moderate-relief hard-bottom areas from April through October each year (Collins, 1990). The sampling area included waters of the continental shelf and shelf edge between Cape Hatteras, NC, and St. Lucie Inlet, FL (Figure 1). Although the potential geographic sampling range has not changed since 1990, sampling effort and concomitant sampling density have been a function of available funding. Prior to 2010, realized sampling was concentrated between 31°N and 34°N with substantially fewer deployments in the extreme northern and southern extent of the range. From 2010 to 2022, due to the formation of SERFS, the sampling universe was intensified to ensure that the overall sampling density was approximately even (Figure 1). In the most recent years, the R/V *Palmetto*, R/V *Savannah*, and NOAA Ship *Pisces* have primarily served as the research platforms. Station depths range from 14 to 110 m. Criteria for annual station selection included sampling without replacement and that all sampled stations were farther than 200 m away from all other sampled stations that year.

Prior to deployment, chevron traps were baited with a combination of whole or cut clupeids (*Brevoortia* or *Alosa* spp., family Clupeidae; Collins, 1990). While traps soaked, bottom temperature (within 5 m of the bottom) was recorded using either a SEABIRD Conductivity Temperature Depth recorder or a VEMCO temperature logger. Traps were retrieved after approximately 90 minutes. After collection, all fishes were identified to species, and each species was counted.

### 2.2 Data analysis

Data were analyzed for the three federally managed species, Gray Triggerfish, White Grunt, and Red Snapper (Figure 2; Bubley et al., 2023). We used two datasets to explore the effect of the increased sampling that began in 2010. One was created using the current version of the sampling universe and encompasses all trap deployments meeting the above criteria over the 32-year time series (no sampling occurred in 2020 due to the COVID-19 global pandemic). This will be referred to as the “SERFS” dataset. The other dataset, referred here to as “MARMAP” was a subset of the SERFS dataset. This subset included stations that were known and sampled prior to 2010 and sampled again in 2010 or beyond, thus mimicking the situation if the sampling universe had not intensified, and the survey continued as it had prior to 2010. For each dataset, we calculated the total numbers of traps deployed each year and the proportion of traps that encountered each of the three selected species. Because the divergence in deployment rate began in 2010, we used t-tests to compare the annual encounter rates for each species from 2010 to 2022 between the SERFS and MARMAP datasets to test for impact of sampling intensification. Additionally, we used t-tests within the MARMAP dataset to test if the annual encounter rate differed between recent years (2010-2022) and older years (1990-2009), as a check for changes in population size rather than survey design (Sokal and Rohlf, 1994).

For each species and dataset (MARMAP and SERFS), we examined the Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial (ZINB) error distributions for modeled indices of relative abundance using the FishyR package in R statistical software (Ballenger, 2022). Final error distribution (ZINB in all cases) and included covariates were selected by Akaike’s Information Criterion (AIC; Akaike, 1973). ZINB models are frequently used in stock assessments to standardize

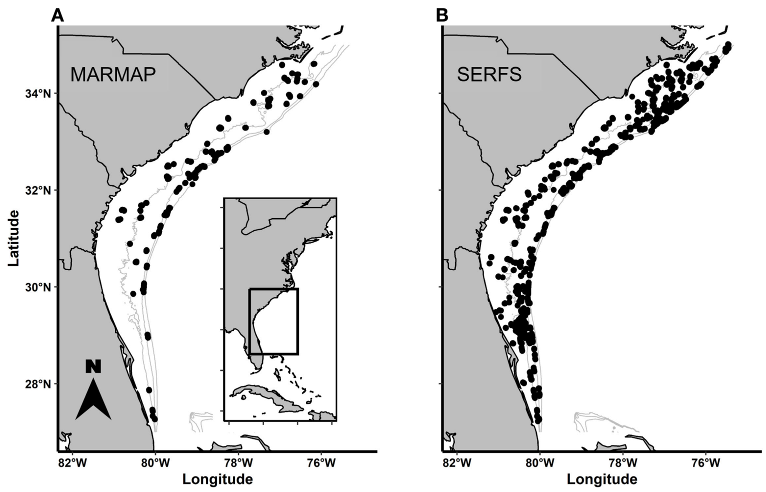



FIGURE 1
Distributions of: **(A)** MARMAP and **(B)** SERFS sampling effort throughout the range. Each dot indicates a single trap deployment during an example year of each survey.. The farthest north and farthest south sampling sites are at approximately the same latitude in each example, but density of sampling differs between the sampling designs, particularly in the northern and southern ends of the range.

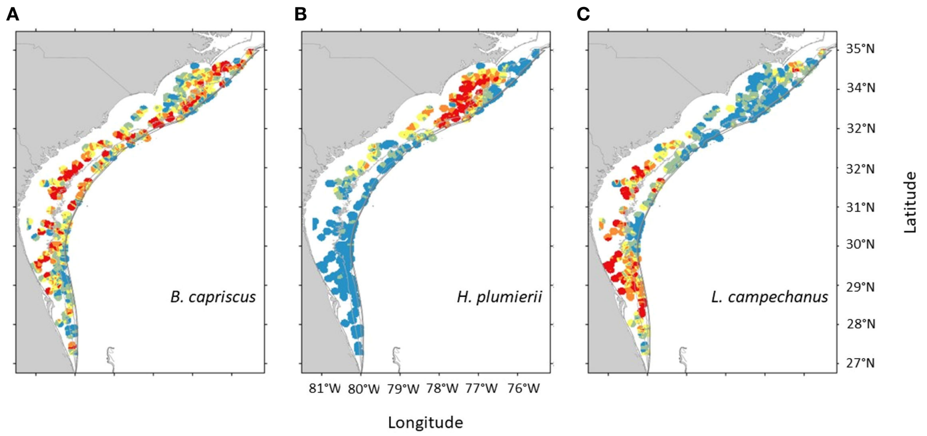



FIGURE 2
Inverse distance weighted interpolation of chevron trap catches for **(A)** Gray Triggerfish, **(B)** White Grunt, and **(C)** Red Snapper, across the survey range in the most recent five years of sampling (Bubley et al., 2023). The total range of values is species specific. Blue indicates the lowest 20% of catches, red indicates the highest 20% of catches. Green, yellow, and orange are intermediate catch rates.

abundance for the chevron trap survey due to the nature of the survey targeting a suite of species (i.e. not every trap has the same likelihood of encountering the species of interest). Both the SERFS and MARMAP models relied on the independent covariates of latitude (°N), sampling day of year, sample depth (m), and bottom temperature (°C) to estimate changes in total abundance of each species over time (Figures S1 and S2). The covariates for the MARMAP model were taken only from those stations included in the MARMAP dataset, while the covariates for the SERFS model were taken from all stations. The covariate effects in the models were fit with polynomials informed by preliminary generalized additive models (GAMs), allowing continuous covariates to be related to abundance (catch per trap) of each species in a non-linear fashion (Figures S3–S8). Year was fitted as a fixed variable and soak time included as an offset. Maximum order of polynomial allowed were limited to 4 to maintain biological relevance and reduce processing time (Wood, 2011).

Coefficients of variation (CV) and standard error per year for each species and dataset (SERFS or MARMAP) were determined by bootstrapping with 5,000 iterations. Linear regression was used to compare the CV from the full data set against the total numbers of traps deployed each year. T-tests were used to compare the slopes of the regression lines for each species, and significance was determined using a Bonferroni correction (Sokal and Rohlf, 1994). All statistical analyses were conducting in R version 4.2.2 (R Core Team, 2022).

**TABLE 1** Total numbers of traps deployed under two fishery-independent sampling scenarios.

| Period | Min Traps | Max Traps | Mean Traps (± SE) |
|---|---|---|---|
| MARMAP 1990-2009 | 224 | 426 | 318 ± 14 |
| SERFS 2010-2022 | 731 | 1883 | 1406 ± 105 |
| MARMAP 2010-2022 | 319 | 489 | 410 ± 14 |

MARMAP 1990-2009 includes all traps deployed from 1990 to 2009. SERFS 2010-2022 includes all traps deployed from 2010 to 2022. MARMAP 2010-2022 includes all traps deployed from 2010 to 2022 that would have been deployed if sampling intensification had not occured.

# 3 Results

Numbers of MARMAP stations sampled between 1990 and 2009 ranged from 224 to 404 annually with a median of 303 (Table 1). The highest concentration of traps during this period was deployed near the center of the geographic range, with less coverage near the northern and southern extent of the range (Figure 1A). Beginning in 2010, total potential sampling stations were increased through extensive habitat exploration, resulting in an approximately evenly distributed sampling effort from north to south (Figure 1B). Total numbers of SERFS stations sampled between 2010 and 2022 ranged from 731 to 1,883 annually with a median of 1,479 (Table 1). Within the MARMAP dataset, the numbers of traps remained below 500 per year throughout the time series, ranging from 319 to 489 with a median of 428 between 2010 and 2022 (Table 1).

Encounter rate between the MARMAP and SERFS sampling since 2010 varied among species. While the difference in encounter rate between MARMAP and SERFS sampling was not significant for Gray Triggerfish ($t = -1.37$, $p = 0.18$) or White Grunt ($t = 0.49$, $p = 0.62$), the difference was significant for Red Snapper ($t = -4.49$, $p < 0.001$; Table 1; Figure 3). Red Snapper encounter rate nearly doubled with the intensification of sampling (0.07 versus 0.13 in the MARMAP vs SERFS sampling). However, this comparison is confounded by potential increases in Red Snapper relative abundance during the period. For the second encounter rate comparison (MARMAP 1990-2009 versus MARMAP 2010-2022; Figure 4), encounter rates varied significantly for both Gray Triggerfish ($t = 4.2$, $p = 0.001$) and Red Snapper ($t = -4.0$, $p = 0.002$), but did not differ for White Grunt ($t = 1.6$, $p = 0.1$).

All species demonstrated a significant, decreasing linear relationship between CV for ZINB-modeled annual index of

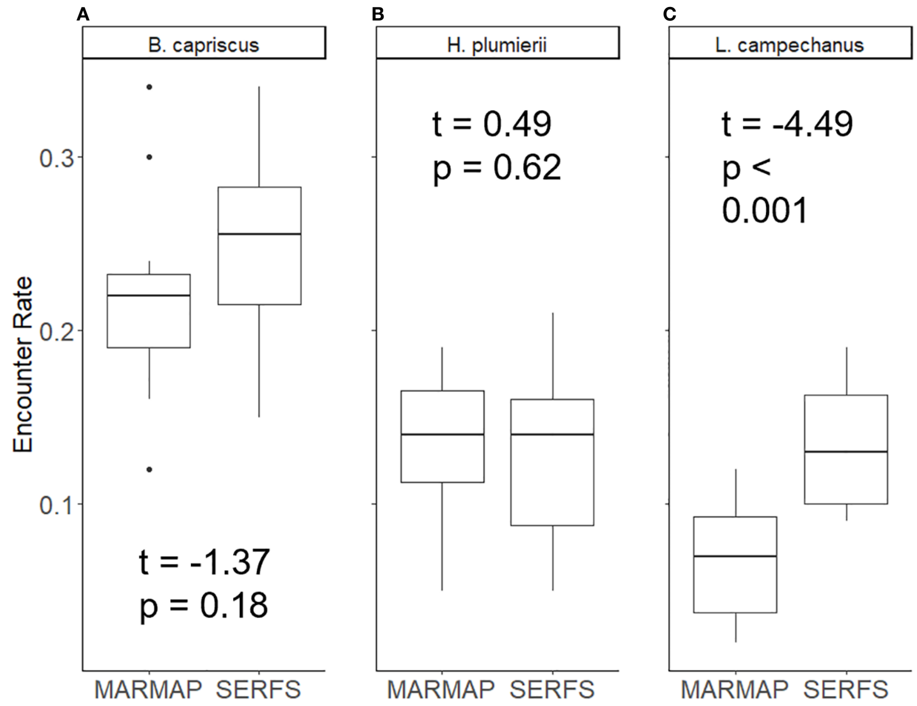



FIGURE 3
Proportion of traps encountering three species: **(A)** Gray Triggerfish, **(B)** White Grunt, and **(C)** Red Snapper under two conditions: MARMAP (data from traps deployed from 2010 to 2022 that would have been deployed even if there was no increase in sampling effort), and SERFS (full complement of traps 2010-2022). Median is the line inside the box, the edges of the box are 25th and 75th percentiles. Whiskers represent the 95% confidence interval. Dots represent samples outside of 95% confidence. T-test results are shown.

relative abundance and sample size using all available trap deployments (SERFS dataset; Figure 5). Coefficient of Variation in each species showed the largest amount of variation at low numbers of annual trap deployments (below 500 traps per year). In addition, CV varied considerably among species at low sample size with Red Snapper having the highest annual CV and Gray Triggerfish having the lowest. As number of samples increased, total variability was reduced in all three species, and CV became more similar among species. All linear models were significant, and each accounted for between 35 and 68% of variability for the species (Table 2; Figure 5). T-tests with Bonferroni correction ($\alpha$ = 0.017) indicated that the slope for Red Snapper was different from the slopes for Gray Triggerfish ($t$ = -8.65, $p < 0.001$) and White Grunt ($t$ = -11.39, $p < 0.001$), but slopes for White Grunt and Gray Triggerfish did not differ significantly ($t$ = -2.38, $p$ = 0.021).

The change in the sampling frame for the trap survey impacted indices of relative abundance in different ways among the species examined. The overall shape of the index of relative abundance was similar between models for Gray Triggerfish (Figure 6A), but the annual mean estimate was higher using the MARMAP dataset than the SERFS dataset for the early part of the time series. Once the sampling frame was intensified (after 2010), the SERFS estimate overlapped with, or tended to be higher than the MARMAP estimate, with standard errors overlapping throughout this most recent period. The White Grunt estimated index of relative abundance did not differ in shape or value between the MARMAP and SERFS models, and the standard error overlapped across the entire time series (Figure 6B). Red Snapper indices of relative abundance overlapped throughout the period 1990 to 2009 (Figure 6C). However, once sampling increased after 2010, the estimates diverged markedly. The SERFS model reflected the increase in Red Snapper abundance as early as 2010 and continued through the end of the time series. The MARMAP model, on the other hand, did not reflect an increase in Red Snapper abundance until 2014 and standard error between the data sets did not overlap until 2021. Although the SERFS model detected the increase in Red Snapper reported elsewhere better than the MARMAP model, the standard errors between 2010 and 2019 were higher in the SERFS model than the MARMAP model (Figure 6C).

When summarizing the effect of increased sampling on the performance metrics for each species, we found 89% beneficial or neutral results. For Gray Triggerfish, CV improved with increased sampling, but encounter rate and standard error for the indices of relative abundance were neutral. For White Grunt, the same pattern was observed. For Red Snapper, encounter rate increased between the surveys using the same years of data and CV declined with increasing sample size while the standard error for the index increased. Both relative abundance index value (SERFS vs. MARMAP) and encounter rate between early and late MARMAP time periods decreased for Gray Triggerfish. Both metrics were neutral for White Grunt, and both increased for Red Snapper (Table 3).

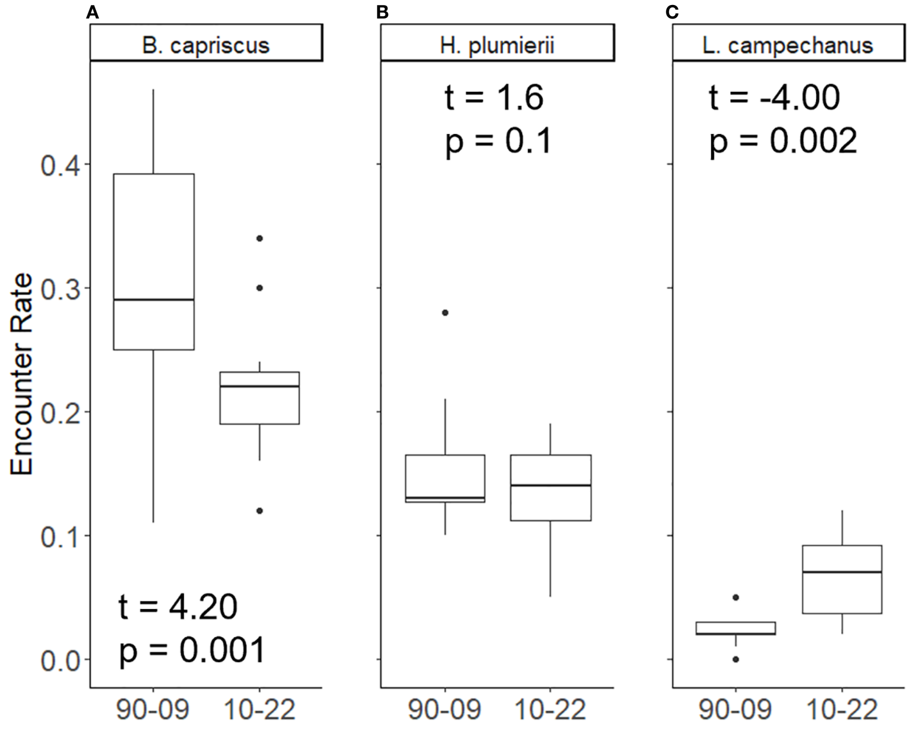



FIGURE 4
Proportion of traps encountering three species: **(A)** Gray Triggerfish, **(B)** White Grunt, and **(C)** Red Snapper under two conditions: full complement of trap deployments between 1990 and 2009, traps deployed from 2010 to 2022 that would have been deployed even if there was no increase in sampling effort. Median is the line inside the box, the edges of the box are 25th and 75th percentiles. Whiskers represent the 95% confidence interval. Dots represent samples outside of 95% confidence. T-test results are shown.

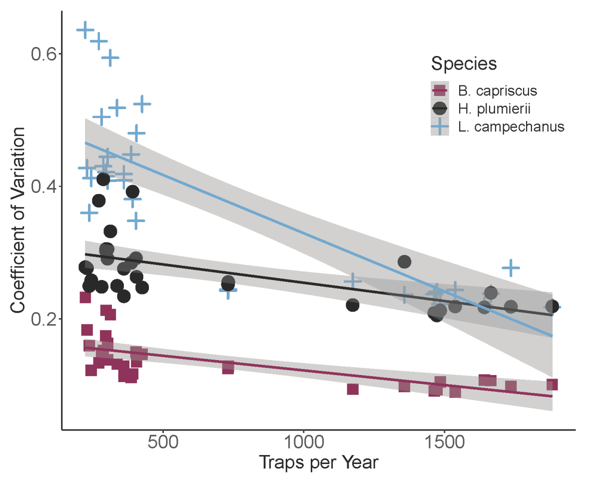



FIGURE 5
Coefficient of variation for three species as a function of total numbers of trap deployments per year. Species are Gray Triggerfish (*B. capriscus*), White Grunt (*H. plumierii)* and Red Snapper (*L. campechanus*). Each line represents a linear regression of coefficient of variation as a function of traps per year. Gray shaded areas are standard error for each regression.

## 4 Discussion

The current work suggests that increased sampling intensity of an existing multi-species survey can improve detection of a species occupying under-represented geographic locations while showing no negative impacts on the detection of other species. We showed improvements to both probability of capture and annual CV for Red Snapper with increased sampling intensity in areas where Red Snapper are known to be abundant. Meanwhile, we showed no negative impact on probability of capture or CV for White Grunt or Gray Triggerfish. The only metric examined with a detrimental effect on performance for Gray Triggerfish (encounter rate assuming the MARMAP survey design) was most likely due to population decline and not survey design. The Red Snapper standard errors were larger with the intensified design than historically once the population showed a marked increase in abundance. In summary, over 80% of metrics examined were positive or neutral, suggesting that gains outweigh losses in the survey intensification.

Despite the broad geographic range of Gray Triggerfish, annual CV values indicate that an increase in total chevron trap effort (i.e. increased sample size) improved our uncertainty estimates. Encounter rates with SERFS were higher than MARMAP during the same period (although not significantly), while the comparison between encounter rates for MARMAP before and after 2010

TABLE 2 Regression parameters for bootstrapped coefficient of variation (CV) as a function of numbers of traps per year for each of three species: Gray Triggerfish (*B. capriscus*), White Grunt (*H. plumierii*), and Red Snapper (*L. campechanus*).

| Species | Intercept (± SE) | Slope (± SE) | F | p | Adjusted $R^2$ |
|---|---|---|---|---|---|
| *B. capriscus* | 0.17 ± 0.008 | $-4.42 \times 10^{-3} \pm 8.34 \times 10^{-4}$ | 28.04 | $1.01 \times 10^{-5}$ | 0.466 |
| *H. plumierii* | 0.31 ± 0.012 | $-5.519 \times 10^{-3} \pm 1.30 \times 10^{-3}$ | 18.12 | $1.88 \times 10^{-4}$ | 0.356 |
| *L. campechanus* | 0.505 ± 0.022 | $-1.756 \times 10^{-2} \pm 2.353 \times 10^{-3}$ | 55.68 | $2.57 \times 10^{-8}$ | 0.638 |

Equation: CV = slope x (100's of traps deployed) + intercept.

indicated a decrease in encounter probability likely due to the overall population decline. Because the MARMAP sampling frame did overlap with the center of abundance for Gray Triggerfish, there was no difference in the timing of observed abundance changes. If we had maintained the MARMAP sampling design, we would have seen a decrease in encounter rate and a decreasing trend in the index of relative abundance. The differences in the overall scale of the index of relative abundance before the survey intensification could be due to the increased specificity of the independent covariates (latitude, sampling day of year, sampling depth, and bottom temperature) informing the model in the SERFS design improving estimate accuracy. The MARMAP design could have been impacted by hyperstability in that it was primarily indexing fish in their core habitat. Including sites in the SERFS design less likely to hold Gray Triggerfish, the index seemed to be more sensitive to a population decline beginning near the margins of suitability for the species (Crecco and Overholtz, 1990; Hilborn and Walters, 1992; Sarah et al., 2015). Overall, the full SERFS time series is likely appropriate to use for Gray Triggerfish with the caveat that the potential overestimate in the historical period should be examined more thoroughly.

White Grunt was included in the study because of the known center of abundance in the northern fringe of the sampling range, which had low sampling intensity prior to 2010. However, differences between the SERFS and MARMAP datasets were minimal for this species, suggesting that despite the mismatch in spatial extent, the MARMAP survey may have been adequately capturing trends in White Grunt abundance over time. Although sampling density in the northern portion of the survey was lower prior to 2010 than after, these results suggest that sampling density was sufficient for White Grunt, possibly due to the consistency of spatial use by this species in that area. The increase in sampling

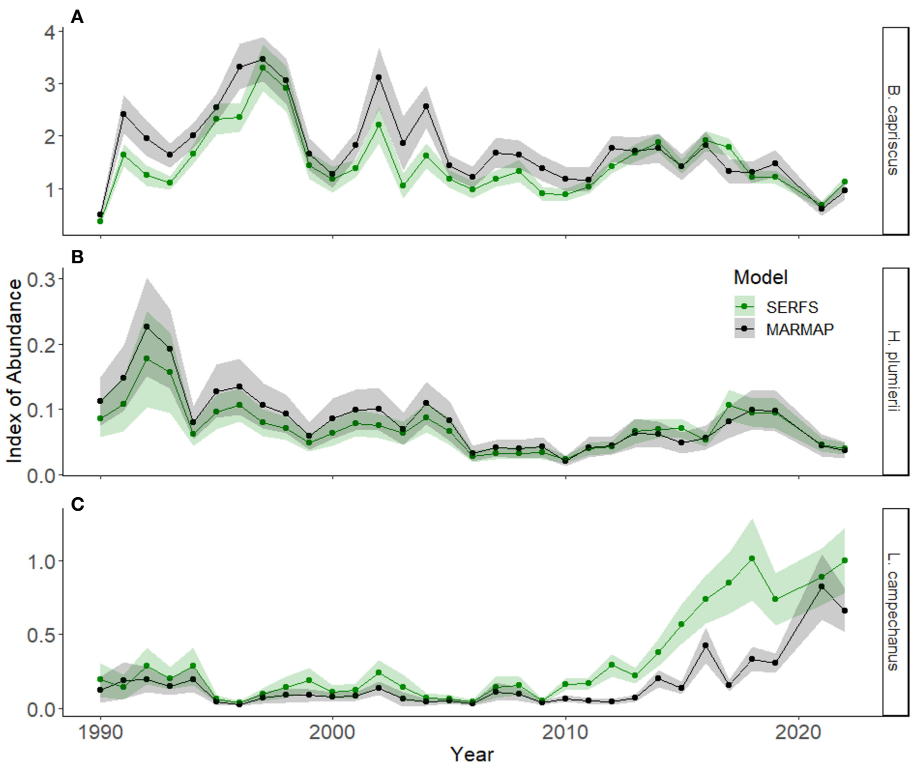



FIGURE 6
Indices of relative abundance from 1990 to 2022 derived from individuals captured in chevron traps for three species, **(A)** Gray Triggerfish (*B. capriscus*), **(B)** White Grunt (*H. plumierii*) and **(C)** Red Snapper (*L. campechanus*), using two zero-inflated negative binomial models. The SERFS model includes environmental covariates and catch information for all chevron traps deployed throughout the survey timespan. The MARMAP model includes only information from chevron traps deployed at sites that had initially been included in the sampling frame prior to 2010. Mean values each year are represented by dots. The shaded areas are standard error for each model each year.

TABLE 3 Summary of performance metrics for the representation of population trends in three species using two fishery-independent survey designs.

| | *B. capriscus* | *H. plumierii* | *L. campechanus* |
|---|---|---|---|
| Encounter Rate: Survey (MARMAP vs. SERFS) | | | |
| CV: total numbers of trap deployments | | | |
| SE: Index of Relative Abundance (MARMAP vs. SERFS) | | | |
| Encounter Rate: Time Period (MARMAP 1990-2009 vs 2010-2022) | | | |
| Index Value: MARMAP vs. SERFS | | | |

**green:** metric improved (or increased), **yellow**: no difference, **red:** metric worsened (or decreased).

effort to the south where White Grunt do not typically occur could have been detrimental to estimates for this species because of the high likelihood of extra zeroes increasing zero inflation and not being informative (Zuur et al., 2009). However, we did not observe any metric to be negatively impacted. In fact, White Grunt annual CV benefitted from increased sample size. The results here support that an index of relative abundance for White Grunt can and should use the full time series and SERFS sampling domain.

Red Snapper have been undergoing a population recovery over the last 10 to 15 years (SEDAR, 2021); however, the detection of this recovery in the chevron trap survey varied across the survey area. The SERFS dataset reflected an increase in Red Snapper abundance as early as 2010, while the increase did not become apparent in the MARMAP dataset until 2014, and this time lag appears to continue through the end of the time series. The discrepancy in detection is most likely due to the process through which Red Snapper have been undergoing population recovery, consistent with ideas on the interaction between abundance and spatial extent, in particular the basin model (Mccall, 1990), in which both abundance and distribution change in concert. Mostly likely, Red Snapper initially began recovering in the core area of their distribution in the southern region of the survey and, as density increased in that area, fish spilled over into additional areas throughout the region (Alcala et al., 2005). While the original MARMAP survey design might be able to detect early signs of population decline by means of local depletions (Lluch-Belda et al., 1989; Atkinson et al., 1997; Warren, 1997; Mcfarlane et al., 2002), the MARMAP design did not detect the Red Snapper recovery until sufficient range expansion had occurred in the species to reach the region of most intense sampling. The primary recommendation from SEDAR 41 to truncate the index of relative abundance using only SERFS sampling (or split the index of relative abundance into two time periods) is likely appropriate in future assessments based on the results presented here (SEDAR, 2017).

The analyses shown here suggest that the increase in sampling density was an effective tool for monitoring the increase in population density for a recovering species and did not negatively impact the ability of the survey to index relative abundances of other species of interest. In fact, uncertainty metrics in a species assumed to be adequately sampled by the MARMAP survey were improved by the inclusion of additional stations. Many fishery-independent surveys have not been updated due to a concern that the change will negatively impact the quality of the data. However, this study demonstrates that improvements in detection and model fit are both strengthened by the increase in station density, even if the total range of the potential survey stations changes little. For the three species examined here, we observed net positive or neutral impacts to indices of relative abundance performance, though we did not examine length or age compositions which typically accompany an index for stock assessment purposes. The method used here: comparing the estimates produced by the data that "would have been" to the estimates produced by the data that currently is, can be a powerful tool to understand whether a survey change made a major difference in the accuracy and appropriateness of the data being provided to produce management advice.

## Data availability statement

The datasets presented in this study can be found in online repository at www.seamap.org. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Ethics statement

Ethical review and approval was not required for the animal study because SCDNR is the governing body for this type of permission. Self-regulation is assumed.

## Author contributions

JV, WB, and TS contributed to the conception and design of the study. JV and WB performed the statistical analyses. JV, WB, and

TS wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Funding was provided by NOAA through the MARMAP (NA21NMF4540294) and SEAMAP-SA (NA21NMF4350372) grants to SCDNR.

## Acknowledgments

The authors thank the numerous field biologists with the MARMAP, SEAMAP-SA, and SEFIS programs and vessel crews that collected the data included in this manuscript. The authors would also like to thank reviewers for their comments and constructive criticism to improve this manuscript. This is contribution number 863 for the Marine Resources Research Institute at South Carolina Department of Natural Resources.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1192739/full#supplementary-material

## References

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle," in *Second international smyposium on information theory*. Eds. B. Petrov and F. Caski (Budapest: Akademiai Kaido), 267–281.

Alcala, A. C., Russ, G. R., Maypa, A. E., and Calumpong, H. P. (2005). A long-term, spatially replicated experimental test of the effects of marine reserves on local fish yields. *Can. J. Fisheries Aquat. Sci.* 62, 98–108. doi: 10.1139/f04-176

Apostolaki, P., and Hillary, R. (2009). Harvest control rules in the context of fishery-independent management of fish stocks. *Aqautic Living Resour.* 22, 217–224. doi: 10.1051/alr/2009022

Atkinson, D. B., Rose, G. A., Murphy, E. F., and Bishop, C. A. (1997). Distribution changes of northern cod (Gadus morhua), 1981–1993. *Can. J. Fisheries Aquat. Sci.* 54, 132–138. doi: 10.1139/f96-158

Bacheler, N. M., Geraldi, N. R., Burton, M. L., Munoz, R. C., and Kellison, G. T. (2017). Comparing relative abundance, lengths, and habitat of temperate reef fishes using simultaneous underwater visual census, video, and trap sampling. *Mar. Ecol. Prog. Ser.* 574, 141–155. doi: 10.3354/meps12172

Ballenger, J. C. (2022). FishyR: functions to facilitate standard analyses conducted in fisheries science. *R package version 0.0.0.9001*.

Ballenger, J. C., Bubley, W. J., Smart, T. I., and Reichert, M. J. M. (2014). *Red snapper fishery-independent indices of abundance in US south Atlantic waters based on a chevron trap survey* (North Charleston, SC: SEDAR), 73.

Brodie, R., Solomon, J. J., Paperno, R., Switzer, T. S., and Swanson, C. E. (2022). Development of survey methods to effectively sample juvenile red snapper along the Atlantic coast of Florida. *Trans. Am. Fisheries Society* 151 (5), 571–586. doi: 10.1002/tafs.10375

Bubley, W. J., Vecchio, J. L, and Smart, T. I. (2023). *Trends in relative abundance of reef fishes in fishery-independent surveys in waters off the southeastern United States.* MARMAP/SEAMAP-SA Reef Fish Survey Technical Report 2023-002 (Charleston, SC: South Carolina Department of Natural Resources)

Christiansen, H. M., Switzer, T. S., Keenan, S. F., Tyler-Jedlund, A. J., and Winner, B. L. (2020). Assessing the relative selectivity of multipe sampling gears for managed reef fishes in the eastern gulf of Mexico. *Mar. Coast. Fisheries* 12, 322–338. doi: 10.1002/mcf2.10129

Collins, M. R. (1990). A comparison of three fish trap designs. *Fisheries Res.* 9, 325–332. doi: 10.1016/0165-7836(90)90051-V

Crecco, V., and Overholtz, W. J. (1990). Causes of density-dependent catchability for georges bank haddock melanogrammus aeglefinus. *Can. J. Fisheries Aquat. Sci.* 47, 385–394. doi: 10.1139/f90-040

Damiano, M. D. (2023). *In search of novel solutions to emerging challenges in quantitative fisheries management* (Raleigh, NC: North Carolina State University). Avaialble at: https://www.lib.ncsu.edu/resolver/1840.20/40706

Dennis, D., Plaganyi, E. E., Van Putten, I., Hutton, T., and Pascoe, S. (2015). Cost benefit of fishery-independentsurveys: are they worth the money? *Mar. Policy* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

Gebremedhin, S., Brunell, S., Getahun, A., Anteneh, W., and Goethals, P. (2021). Scientific methods to understand fish population dynamics and support sustainable fisheries management. *Water* 13, 574. doi: 10.3390/w13040574

Glasgow, D. M., Reichert, M. J. M., and Quattro, J. (2021). Effects of environmental factors on reef fish assemblage structure in the southeastern US Atlantic. *Mar. Ecol. Prog. Ser.* 671, 147–163. doi: 10.3354/meps13764

Hamilton, R. J., Almany, G. R., Stevens, D., Bode, M., Pita, J., Peterson, N. A., et al. (2016). Hyperstability masks declines in bumphead parrotfish (*Bolbometopon muricatum*) populations. *Coral Reefs* 35, 751–763. doi: 10.1007/s00338-016-1441-0

Hilborn, R., Amoroso, R. O., Anderson, C. M., Baum, J. K., Branch, T. A., Costello, C., et al. (2020). Effective fisheries management instrumental in improving fish stock status. *Proc. Natl. Acad. Sci. United States America* 117, 2218–2224. doi: 10.1073/pnas.1909726116

Hilborn, R., and Walters, C. (1992). *Quantitative fisheries stock assessment: choice, dynamics, and uncertainty* (New York, NY: Springer-Science & Business Media, B.V). doi: 10.1007/978-1-4615-3598-0

Lluch-Belda, D., Crawford, R. J. M., Kawasaki, T., Maccall, A. D., Schwartzlose, R. A., and Smith, P. E. (1989). Worldwide fluctuations of sardine and anchovy stocks: the regime problem. *South Afr. J. Mar. Sci.* 8, 195–205. doi: 10.2989/02577618909504561

Mccall, A. D. (1990). *Dynamic geography of marine fish population* (Seattle, WA: University of Washington Press).

Mcfarlane, G. A., Smith, P. E., Baumgartner, T. R., and Hunter, J. R. (2002). "Climate variability and pacific sardine populations and fisheries," in *Fisheries in a changing climate*. Ed. N. A. Mcginn (Bethesda, MD: American Fisheries Society), 195–214.

O'Leary, C. A., Kotwicki, S., Hoff, G. R., Thornson, J. T., Kulik, V. V., Ianelli, J. N., et al. (2021). Estimating spatiotemporal availability of transboundary fishes to fishery-independent surveys. *J. Appl. Ecol.* 58, 2146–2157. doi: 10.1111/1365-2664.13914

Potts, J., and Manooch III, C. S. (2002). Estimated ages of red porgy (Pagrus pagrus) from fishery-dependent and fishery-independent data and a comparison of growth parameters. *Fishery Bull.* 100, 81–89.

R Core Team (2022). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Sarah, S. L., Ralf, D., and Axel, T. (2015). Combining area closures with catch regulations in fisheries with spatio-temporal variation: bio-economic implications for the north Sea saithe fishery. *Mar. Policy* 51, 281–292. doi: 10.1016/j.marpol.2014.08.017

Schrandt, M. N., Shea, C. P., Kurth, B. N., and Switzer, T. S. (2021). Amending survey design to improve statistical inferences: monitoring recruitment of juvenile reef fish in the eastern gulf of Mexico. *Fisheries Res.* 241, 106015. doi: 10.1016/j.fishres.2021.106015

SEDAR (2009). *SEDAR 15: stock assessment report 1 south Atlantic red snapper* (North Charleston, SC: South Atlantic Fishery Management Council).

SEDAR (2017). *SEDAR 41- stock assessment report - revision 1 south Atlantic red snapper* (North Charleston, SC: SEDAR).

SEDAR (2021). *SEDAR 73 south Atlantic red snapper stock assessment report* (North Charleston, SC: SEDAR).

Smart, T. I., Bubley, W. J., Glasgow, D. M., and Reichert, M. J. M. (2020). Spatial distribution changes and habitat use in red porgy in waters off the southeast U.S. Atlantic coast. *Mar. Coast. Fisheries* 12, 381–394. doi: 10.1002/mcf2.10135

Sokal, R. R., and Rohlf, F. J. (1994). *Biometry: the principles and practices of statistics in biological research third (3rd) edition* (New York, NY: W.H. Freeman and Co.).

Thompson, K. A., Switzer, T. S., Chirstman, M. C., Keenan, S. F., Gardner, C., Overly, K. E., et al. (2022). A novel habitat-based approach for combining indices of abundance from multiple fishery-independent video surveys. *Fisheries Res.* 247, 106178. doi: 10.1016/j.fishres.2021.106178

Warren, W. G. (1997). Changes in the within-survey spatio-temporal structure of the northern cod (Gadus morhua) population 1985–1992. *Can. J. Fisheries Aquat. Sci.* 54, 139–148. doi: 10.1139/f96-147

Williams, E. H., and Carmichael, J. (2009). *Final report: south Atlantic fishery independent monitoring program workshp* (Beaufort, NC: NOAA-NMFS).

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Society: Ser. B (Methodological)* 73, 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Zimney, A., and Smart, T. I. (2022). Effects of incomplete sampling and standardization on indices of abundance from a fishery-independent trawl survey off the southeastern united states. *Fishery Bull.* 120, 252–267. doi: 10.7755/FB.120.3-4.6

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed effects models and extensions in ecology with r* (New York, NY: Springer Science +Business Media, LLC).

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 05 July 2023
DOI 10.3389/fmars.2023.1198260

Check for updates

OPEN ACCESS

EDITED BY
Nathan Bacheler,
Beaufort Laboratory, Southeast Fisheries Science Center (NOAA), United States

REVIEWED BY
James Morley,
East Carolina University, United States
Janet Nye,
University of North Carolina at Chapel Hill, United States

*CORRESPONDENCE
Meaghan D. Bryan
meaghan.bryan@noaa.gov

RECEIVED 31 March 2023
ACCEPTED 15 June 2023
PUBLISHED 05 July 2023

CITATION
Bryan MD and Thorson JT (2023) The performance of model-based indices given alternative sampling strategies in a climate-adaptive survey design.
*Front. Mar. Sci.* 10:1198260.
doi: 10.3389/fmars.2023.1198260



# The performance of model-based indices given alternative sampling strategies in a climate-adaptive survey design

Meaghan D. Bryan* and James T. Thorson

Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, United States

Species-distribution shifts are becoming commonplace due to climate-driven change. Difficult decisions to modify survey extent and frequency are often made due to this change and constraining survey budgets. This often leads to spatially and temporally unbalanced survey coverage. Spatio-temporal models are increasingly used to account for spatially unbalanced sampling data when estimating abundance indices used for stock assessment, but their performance in these contexts has received little research attention. We therefore seek to answer two questions: (1) how well can a spatio-temporal model estimate the proportion of abundance in a new "climate-adaptive" spatial stratum? and (2) when sampling must be reduced, does annual sampling at reduced density or biennial sampling result in better model-based abundance indices? We develop a spatially varying coefficient model in the R package VAST using the eastern Bering Sea (EBS) bottom trawl survey and its northern Bering Sea (NBS) extension to address these questions. We first reduce the spatial extent of survey data for 30 out of 38 years of a real survey in the EBS and fit a spatio-temporal model to four commercially important species using these "data-reduction" scenarios. This shows that a spatio-temporal model generally produces similar trends and density estimates over time when large portions of the sampling domain are not sampled. However, when the central distribution of a population is not sampled the estimates are inaccurate and have higher uncertainty. We also conducted a simulation experiment conditioned upon estimates for walleye pollock (*Gadus chalcogrammus*) in the EBS and NBS. Many species in this region are experiencing distributional shifts attributable to climate change with species historically centered in the southeastern portion of the survey being increasingly encountered in the NBS. The NBS was occasionally surveyed in the past, but has been surveyed more regularly in recent years to document distributional shifts. Expanding the survey to the NBS is costly and given limited resources the utility of reducing survey frequency versus reducing sampling density to increase survey spatial extent is under debate. To address this question, we simulate survey data from alternative sampling designs that involve (1) annual full sampling, (2) reduced sampling in the NBS every year, or (3) biennial and full sampling in the NBS. Our results show that annual sampling, even with reduced sampling density, provides less biased abundance information than biennial sampling. We therefore conclude that ideally fishery-independent surveys should be conducted annually and spatio-temporal models can help to provide reliable estimates.

KEYWORDS

spatio-temporal models, fishery-independent sampling, abundance indices, climate change, Bering Sea

# 1 Introduction

Marine species worldwide are responding to climate-driven shifts in ocean conditions by shifting their spatial distribution (Pinsky et al., 2013; Poloczanska et al., 2013; Pecl et al., 2017). For example, decreased springtime sea-ice production in the eastern Bering Sea (EBS) is causing a decline in the spatial area of near-freezing seafloor water temperatures during summer (termed "cold pool extent"). These cold seafloor waters previously inhibited the northward extent of summertime movement for commercially important Pacific cod (*Gadus macrocephelus*) and walleye pollock (*Gadus chalcogrammus*), and is hypothesized to have provided a refuge from predation for snow (*Chionoecetes opilio*) and tanner crabs (*Chionoecetes bairdi*). Interannual changes in cold-pool extent have therefore been linked to changes in the spatial distribution, diet, and productivity of these and other species (Thorson et al., 2021). In particular, the decline in cold-pool extent has led walleye pollock, Alaska plaice (*Pleuronectes quadrituberculatus*), and other species to occupy new habitat in the northern Bering Sea, causing a substantial fraction of the managed stock to move outside of the area conventionally monitored for use in stock assessment and fisheries management (O'Leary et al., 2020; O'Leary et al., 2022).

As stocks migrate beyond the boundaries of conventional resource surveys, it complicates traditional approaches to stock assessment. Assessment scientists can respond by:

A. Ignoring the portion of the stock beyond conventional boundaries, in some cases, despite genetic or tagging evidence that stocks are well-mixed and likely subject to the same fishery;
B. Combining abundance estimates in an *ad-hoc* manner, whereby years with more extensive surveys are likely to capture a larger portion of stock abundance, and potentially correcting for this effect via modifications to the stock-assessment model;
C. Creating a spatially stratified assessment model and seek to include survey data from different regions in only those years where it is available, while also estimating annual movement rates to provide a mechanistic model for shifting availability in different areas.

There are substantial limitations with each of these potential responses. For example, response A will not properly measure the total stock that is subject to fishing (presumably resulting in overly conservative catch quotas), while response B will either confound survey extent and abundance index trends or require estimating a complicated process for time-varying catchability. Finally, response C will require estimating many additional movement parameters, which is likely difficult without extensive tagging information (Thompson and Thorson, 2019).

The development of model-based abundance indices has become more common in the fisheries and ecosystem literature. Approaches including delta-generalized linear and mixed models (delta-GLMs and delta-GLMMs), as well as spatio-temporal models have been frequently used. The intention of using these approaches is to provide accurate and improved estimates of precision while incorporating information about factors that cannot be accounted for in the statistical design of surveys. A key aspect shared among the approaches is separating the survey catch process into two components: the probability of encountering a species and the probability of a positive catch rate when encountered (Lo et al., 1992; Stefansson, 1996; Maunder and Punt, 2004). In doing so, covariates that are hypothesized to change stock range and abundance can be accounted for in each component of the catch process outside the assessment model. For example, catchability covariates, such as changes in survey gear and fishing power differences among survey vessels, have been included in delta-GLMs and delta-GLMMs to account for their impacts on the survey abundance (Thorson and Ward, 2013). Accounting for spatial and spatio-temporal variation has also been shown to be important given inter-annual and spatial variation in abundance due to changes in fishing pressure and movement patterns (Shelton et al., 2014; Thorson and Barnett, 2017; Grüss and Thorson, 2019; Perretti and Thorson, 2019). Spatio-temporal models evolved from delta-GLMM approaches to explicitly model spatial and spatio-temporal variation. Programs like the vector-autoregressive spatial temporal (VAST) R package can provide estimates for multiple locations over time by assuming that observation and process errors are more similar to its nearest neighbor (Thorson and Barnett, 2017; Thorson, 2019b). VAST also models catchability covariates and habitat covariates (e.g., bottom temperature or cold pool extent in the eastern Bering Sea) separately. Catchability covariates are those expected to impact catch rates such as vessel and gear characteristics and fishing method. Including them in the model reduces bias in the spatio-temporal variation and increases precision in the density estimates (Thorson, 2019b). Habitat covariates are applied to the expected density and catch rates and are extremely useful to include when survey observations are spatially coarse or missing entirely

(Thorson, 2019a; Thorson, 2019b). Habitat covariates are associated with each location and can aid in extrapolating population density in areas with limited or no data.

Accurate and precise estimates of abundance from fishery-independent surveys are important to effectively manage our fishery resources. Modifications to survey strategies are often needed due to logistical (e.g., staffing issues, gear and vessel failures, and inclement weather) and budgetary constraints (ICES, 2020). Population distribution shifts further complicate the decision-making process in light of logistic and budgetary considerations. Outright cancellation of a survey in a given year, reduced spatial coverage, and reduced sampling intensity are potential survey modifications. The biggest concern is that modifications to survey designs can lead to spatially and/or temporally unbalanced time series that result in biased or imprecise (or both) abundance estimates (ICES, 2020). Directional bias can lead to unintended over- or under-exploitation, while imprecision will increase the uncertainty in our stock assessments and management advice. Therefore, understanding the minimum sampling needs to produce reliable abundance estimates is important to the entire management system from data collection, stock assessment, and management decision making.

Spatial distribution shifts of several commercially important species in Alaska, as the cold-pool extent weakens, underscores the need to design fishery-independent surveys that can capture these shifts and adequately estimate abundance/biomass (Mueter and Litzow, 2008; Stevenson and Lauth, 2019; Spies et al., 2020). The EBS bottom trawl survey (BTS) represents a long-term (i.e., over 30 years) annual time-series that not only collects population information but also important environmental data for the region. This survey has extended northward infrequently over time (e.g., 1982, 1985, 1988, 1991, 2010, and 2017-2019) to ascertain the prevalence of abundance outside the standard EBS BTS area. In the majority of years this extension was exploratory; however, the survey was officially expanded in 2017. The observed fish populations in the standard EBS trawl survey and the northern survey extension, or northern Bering Sea (NBS), is the same; therefore, it would be worthwhile to combine these data to derive a single, spatially and temporally comprehensive index. Spatio-temporal modeling using habitat covariates is a promising approach to fill in these spatial and temporal data gaps (Thorson, 2019b) and has been used to derive abundance indices for walleye pollock and Pacific cod (Thompson and Thorson, 2019; O'Leary et al., 2020). Therefore, evaluating the spatio-temporal model's ability to effectively estimate abundance in infrequently sampled survey areas is of utmost importance. Identifying appropriate levels of sampling frequency and intensity in this "newer" stratum is also needed in the face of budget limitations and the need to survey the NBS more consistently to better capture shifts in abundance with advancing climate change. Therefore, we aim to answer two questions with this project: (1) how well can spatio-temporal index standardization estimate the proportion of abundance in a new "climate-adaptive" spatial stratum? and (2) does annual sampling at reduced density or biennial sampling result in better model-based abundance indices? We address the first question empirically, where we first drop survey data from large areas of the EBS BTS in years when the NBS was not surveyed. We then fit a spatio-temporal model using a habitat covariate to extrapolate abundance in the missing areas and compare the estimates to the estimates from a full model to determine whether (A) estimates using reduced data are accurate and (B) uncertainty estimates when reducing data still include the estimates arising from fitting to all data. The second question is addressed with a simulation experiment conditioned upon estimated densities when fitting to all available survey data for walleye pollock in the EBS and the NBS. We simulate survey data from alternative sampling designs that involve full sampling every year, reduced sampling in the NBS every year, or full sampling in the NBS every other year. Similar to the empirical experiment, we fit a spatio-temporal model using a habitat covariate to the simulated data. We then measure bias in the NBS abundance estimate.

# 2 Methods

## 2.1 Survey area and data

A fishery-independent EBS BTS has been conducted annually since 1982 (Bakkala, 1993) (Figure 1). The one exception was in 2020 due to the global pandemic. The EBS BTS is conducted from June to August of each year and follows a standardized methodology using the same standard trawl in all years (Stauffer, 2004). The standard survey includes 376 stations covering depths from 20m to 200m. The EBS BTS has been extended beyond its core area to the NBS in 1982, 1985, 1988, 2010, and 2017-2019 and used the same standardized methods as the EBS BTS (Markowitz et al., 2022). The number of stations surveyed in the NBS extension varied in the early years, but included an additional 143 stations with depths ranging from 10m - 80m in 2017 and 2019. The number of stations in the NBS was reduced to 41 stations in 2018. Station level catch rates are derived and represent numbers and weight per area-swept. The station level catch rates are then used to develop spatially-aggregated biomass/abundance indices to provide information about stock trends over time in Alaska Region stock assessment models.

The EBS BTS also collects important oceanographic data used to develop environmental indices that help to explain species distributions (Stevenson and Lauth, 2019). One such index is the cold pool index (CPI). The CPI is an annual index derived from bottom temperature measurements taken at each survey station and measures the spatial extent ($km^2$) of the cold pool. The cold pool is defined by bottom temperatures in the Bering Sea that are < 2°C (Wyllie-Echeverria and Wooster, 1998). This dynamic feature of the EBS is largely determined by annual sea ice coverage that regulates bottom temperature. Sea ice retreat in late winter/early spring leads to a small cold pool with bottom temperatures > 2°C. Conversely, later sea ice retreat maintains bottom temperature below 2°C and results in a larger cold pool. Species distributional changes with northward movement linked to the cold pool extent for some species in the Bering Sea have been predicted and documented (Stabeno et al., 2012; Stevenson and Lauth, 2019).

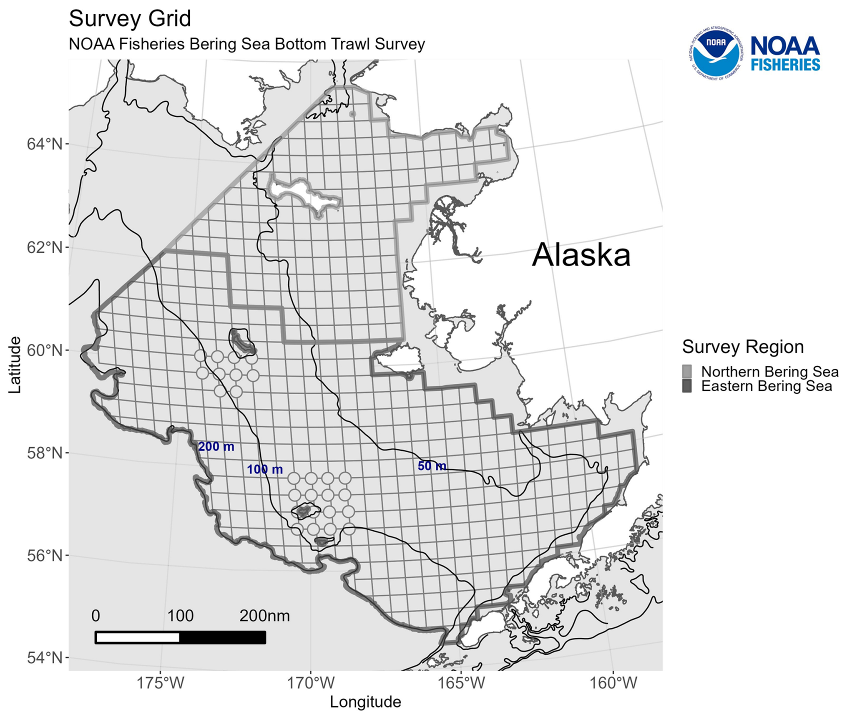



FIGURE 1
The Bering Sea bottom trawl survey grid, including the eastern Bering Sea standard survey area and the extension in the northern Bering Sea. Each square and circle (corner station) represents a survey station grid cell. The 50m, 100m, and 200m bathymetry lines are included for reference. https://www.fisheries.noaa.gov/alaska/science-data/near-real-time-temperatures-bering-sea-bottom-trawl-survey-2023

Model-based indices using the vector autoregressive spatio-temporal (VAST) model have been developed for walleye pollock and Pacific cod to combine the data from the standard EBS survey and the NBS extension (Thompson and Thorson, 2019; O'Leary et al., 2020). The CPI was included as a habitat covariate in the model to account for its impact on the distribution of these species over time. It is therefore imperative that we evaluate how well a spatio-temporal model can estimate abundance in a climate adaptive survey area, like the NBS extension, in years when not surveyed. Additionally, given the shifts in species distributions the expansion of the EBS BTS to the NBS needs to be conducted more frequently. An understanding of the required frequency and intensity of sampling is of utmost importance, so that survey resources can be allocated efficiently and effectively.

## 2.2 Empirical analysis

We conducted an empirical analysis to evaluate how well a spatio-temporal model can estimate biomass in a new or infrequently surveyed spatial stratum. We obtained and used EBS BTS catch rate data for four species of interest; walleye pollock, Pacific cod, yellowfin sole (*Limanda aspera*), and snow crab. All are among the most commercially important species in this region. They also exhibit different spatial distributions, where walleye pollock and Pacific cod have more widespread distributions, yellowfin sole is generally concentrated in the eastern portion of the EBS, and snow crab are predominately in the northwest. The full dataset (14,089 samples at approximately 376 unique stations) was reduced by dropping survey stations from one of four large areas in the EBS that were designed to mimic a circumstance where survey data were periodically unavailable in the eastern, northern, southern, or western portion of the full survey extent (Figure 2). The number of sampled stations retained was 11,399, 12,322, 10,367, and 11,357 when the eastern, western, northern, and southern stations were removed, respectively. We adopted the NBS sampling frequency, when the stations were dropped in all years of the time series except for those when the NBS was surveyed (i.e., keeping data across the full survey area only in 1982/1985/1988/1991/2010 and 2017-2019). This was done to mimic the unbalanced survey design of the NBS extension. The reduced dataset was then fitted to a spatio-temporal model developed in VAST (Thorson, 2019b) within R to estimate biomass indices for each species (Thorson and Barnett, 2017). The biomass indices from the full and reduced data sets were then compared.

The spatio-temporal model used for this analysis followed the guidelines in (Thorson, 2019b) and accounted for cold-pool effects. Biomass per unit area observations, $b_i$, from all EBS BTS grid cells for 1982-2019 were fit using a Poisson-link delta-gamma distribution. We wanted to estimate biomass in large unsampled areas in some years; therefore, the extrapolation to these areas was informed by estimating a zero-centered spatially varying coefficient (SVC) that measures the local response to an annual index of cold-pool extent index (Thorson, 2019b; Thorson et al., 2023). The SVC was estimated for both the linear predictors of the delta model. The variance of the SVC to cold-pool extent was estimated at zero for

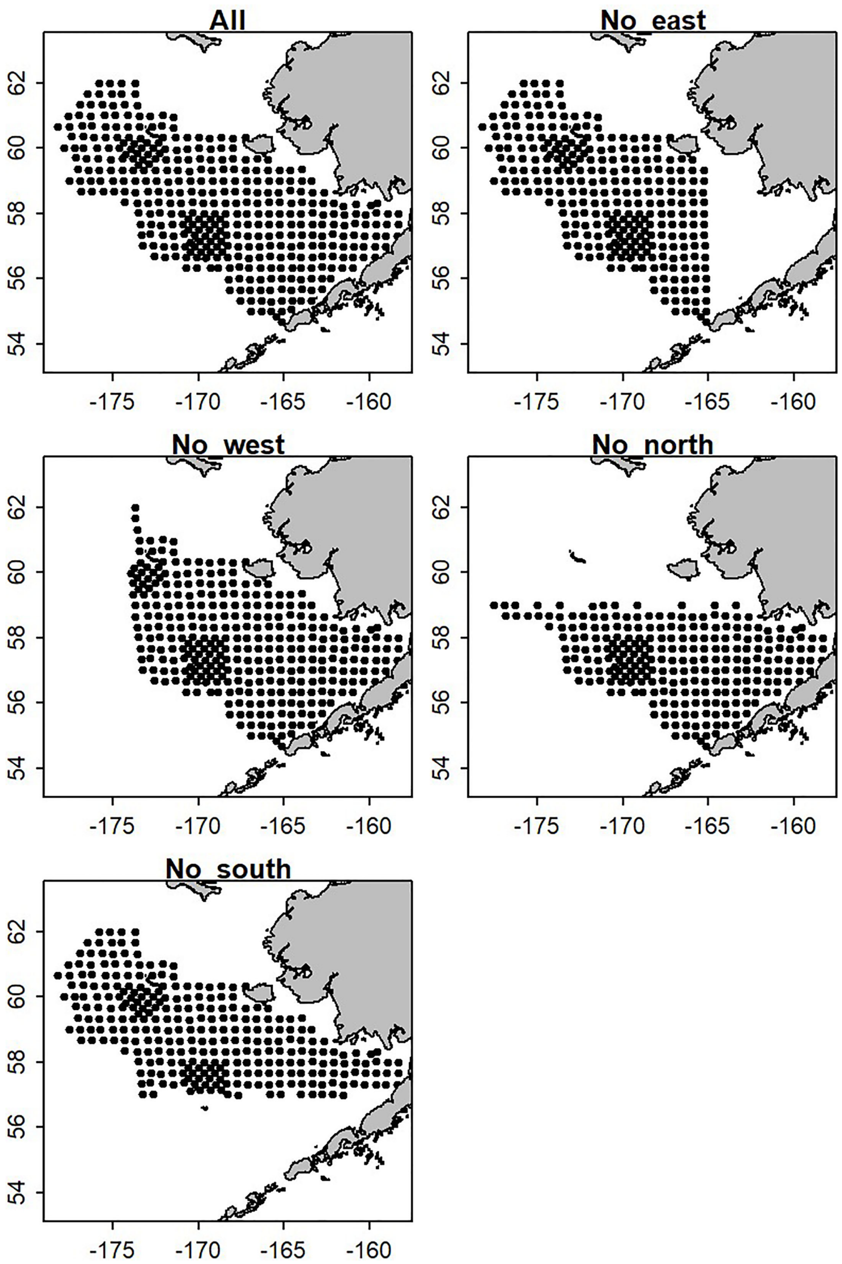



FIGURE 2
The survey footprint of the eastern Bering Sea bottom trawl survey standard area and four scenarios where data were removed from large areas.

the second linear predictor of the yellowfin sole model. Therefore, it was removed from the yellowfin sole model for all scenarios. The two linear predictors of the delta model represent the encounter probability, $p_i$, and positive biomass per unit area, $r_i$.

The probability distribution of biomass sample $b_i$ was specified as:

$$\Pr(b_i > 0) = 1 - p_i$$

and

$$b_i | B > 0 \sim \text{Gamma}(B;\ \theta_b^{-2}, r_i\theta_b^2),$$

where we specified a gamma distribution for positive catch rates where $r_i$ is the mean and $\theta_b$ is the coefficient of variation and we use the shape-scale parameterization. We estimated geometric anisotropy (i.e., the tendency for correlations to decline more rapidly in certain cardinal directions) (Thorson et al., 2015), and a spatial and spatio-temporal term for both linear predictors was included in the model. We also used epsilon bias-correlation to correct for retransformation bias (Thorson and Kristensen, 2016).

The linear predictors for numbers density, $n_i$, and biomass per individual, $w_i$, were

$$\log(n_i) = \beta_n(t) + \omega_n(s) + \epsilon_n(s,t) + \gamma_n(s)T(t)$$

and

$$\log(w_i) = \beta_w(t) + \omega_w(s) + \epsilon_w(s,t) + \gamma_w(s)T(t),$$

where $\boldsymbol{\omega}_n$ and $\boldsymbol{\omega}_w$ represent the time-invariant spatial variation, $\boldsymbol{\epsilon}_n(t)$ and $\boldsymbol{\epsilon}_w(t)$ represent the time-varying spatial variation, $\beta_n(t)$ and

$\beta_w(t)$ represent the annual intercepts which are treated as fixed effects, $T(t)$ is the cold-pool index (CPI) in each year, and $\boldsymbol{\gamma}_n$ and $\boldsymbol{\gamma}_w$represents the log-linear impact of the CPI which vary spatially. We treat spatial terms ($\boldsymbol{\omega}_n$ and $\boldsymbol{\omega}_w$) and SVCs ($\boldsymbol{\gamma}_n$ and $\boldsymbol{\gamma}_w$) as Gaussian Markov random fields (GMRFs) and estimate them as random effects while estimating their variance as fixed effects. Similarly, we estimate spatio-temporal terms ($\boldsymbol{\epsilon}_n(t)$ and $\boldsymbol{\epsilon}_w(t)$) as GMRFs that follow a first-order autoregressive process, and estimate their variance and temporal autocorrelation as fixed effects. The spatial domain included 100 knots and 2000 extrapolation grid cells that define the value of Gaussian Markov random fields (GMRFs) at the location of those knots, and the value of GMRFs elsewhere is calculated via bilinear interpolation.

Linear predictors were then transformed to calculate encounter probability $p_i$ and positive catch rate $r_i$ following the Poisson-linked delta model:

$$p_i = 1 - e^{a_i n_i}$$

and

$$r_i = \frac{a_i n_i}{p_i} w_i ,$$

where $a_i$ is the area swept for each sample (Thorson, 2018). The spatio-temporal terms were estimated following a first-order autoregressive process across years to better estimate density hotspots. The temporal intercepts were treated as fixed effects for each linear predictor and year. Treating the temporal intercepts as fixed effects reduces the correlation structure among years so that the estimates can be used for assessment purposes. Lastly, the SVC $\gamma_p(s)$ and $\gamma_r(s)$ were estimated and assumed independent for $p_i$ and $r_i$.

## 2.3 Simulation experiment

We also conducted a simulation experiment to evaluate the impact of sampling intensity and frequency on our biomass estimates in a new climate adaptive area. The operating model made the same structural assumptions as the estimation model used in the empirical analysis. The main difference between the two is that the spatial extent of the survey-sampling grid included the EBS and NBS bottom trawl survey stations (Figure 3). The OM was conditioned on the EBS BTS standard survey (1982-2018) and the NBS 2017 data through an initial model fit. Given the inconsistent frequency and unbalanced design of the NBS extension, we used the location of bottom trawl survey data in the NBS in 2017 to define the location of NBS sampling in 1982-2016 and 2018 to enable the OM to simulate data in the NBS in all years. The CPI was used in the model as an annual habitat covariate while estimating a zero-centered SVC for both predictors. In each simulation replicate, we then simulated new values for all fixed and random effects from the joint precision matrix estimated when fitting to real world data. We then simulated new survey observations conditional upon these simulated values for fixed and random effects. Therefore, each simulation replicate for a given sampling scenario differs in terms of true underlying densities, as well as resulting simulated samples of those densities.

The estimation model made the same structural assumptions as the OM; however, we reduced the number of knots to 50 from 250. Additionally, we subset the data to include the years 2000-2018. Using a subset of the data and reducing the number of knots reduced runtime. A total of 100 replicates were simulated for three survey sampling scenarios. The sampling scenarios included 1) annual and full sampling, 2) annual and full sampling in the EBS

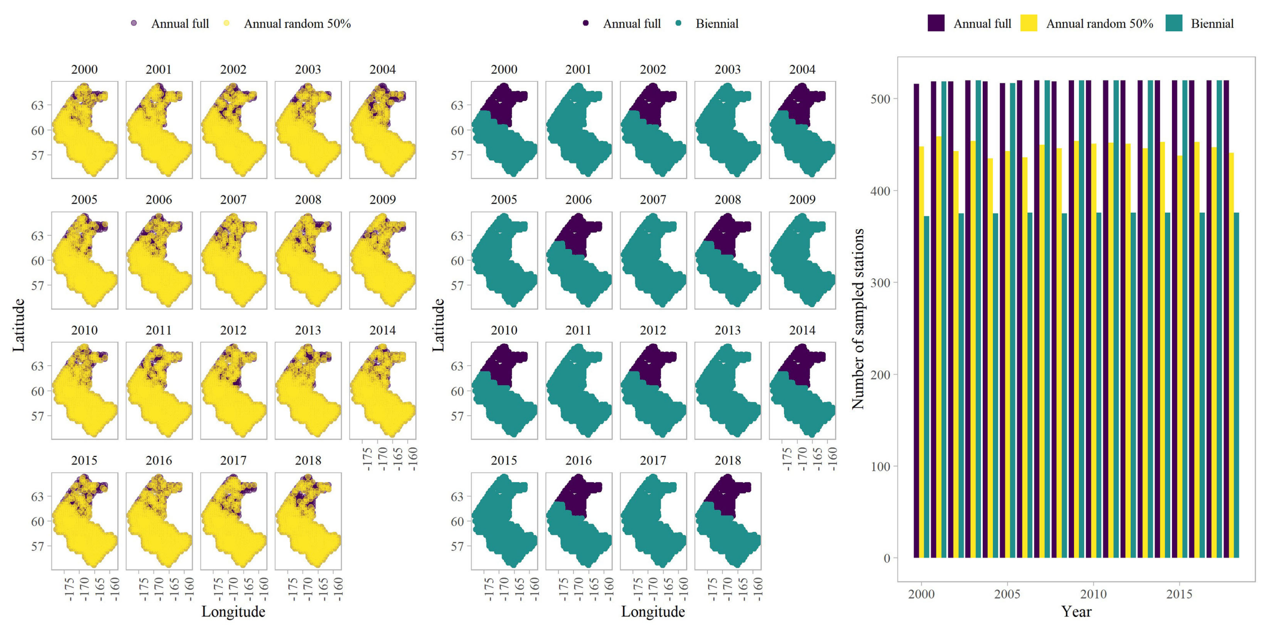



FIGURE 3
An example of the simulated sampling design scenarios: annual scenarios (left panel) and biennial scenario (middle panel), as well as the number of sampled stations per year (right panel).

and a 50% reduction in survey stations while maintaining annual sampling in the NBS, and 3) annual and full sampling in the EBS and biennial sampling in the NBS, where odd years were sampled at all NBS stations. We randomly selected ~50% of NBS stations in scenario 2. We then evaluated performance by comparing the true proportion of biomass in the NBS in a given simulation replicate with the estimated proportion from each of three sampling scenarios. We specifically calculated bias on a log scale and the median absolute error in the proportion of biomass in the NBS to determine differences in the model estimation capabilities among the scenarios.

# 3 Results

## 3.1 Empirical analysis

Eliminating data from the eastern, western, northern, or southern portions for the majority of years and comparing results with those when using all data shows that the scale and trends in estimated biomass are generally similar for all four species (Figure 4). Removing data from large areas, thereby reducing the survey footprint, leads to greater uncertainty in the density estimates for all species and inaccuracy for some species. An interaction between species and the area removed was apparent, where greater uncertainty in the density estimates arose for particular areas for each species (e.g., comparing Figures 4, 5). For example, standard errors were larger when the west and north data were removed for walleye pollock, where they have tended to have increased density in years with low CPI (Figure 4, top row). Standard errors and inaccuracy were highest when the eastern data were removed for yellowfin sole (Figure 4, third row), which corresponds to core habitat for this species (Figure 5, third row; Table 1). Finally, removing the northern data resulted in the highest standard errors and inaccuracy for snow crab, which again corresponds to core habitat for snow crab (Figures 4, 5, bottom row; Table 1). The increased standard errors were similar across the removed areas for Pacific cod since this species has a more even distribution in the EBS than the other focal species (Figure 4, second row and where average density is similar across all scenarios in Table 1).

## 3.2 Simulation experiment

We evaluated model convergence prior to processing the results and a total of 96, 87, and 96 model runs out of 100 converged for the annual, annual reduced, and biennial sampling scenarios, respectively. Of the model runs that converged, 82 simulations were in common among the scenarios. Greater bias was associated with biennial sampling than the annual sampling strategies (Figure 6). This was driven by the bias in the estimated proportion of biomass in years when the NBS was not surveyed (i.e., odd years). Bias in the biennial sampling was greater than the annual sampling strategies in the non-surveyed years, whereas the estimates were more similar among the sampling strategies when there was a temporal overlap in sampling. The biennial sampling strategy also exhibited greater uncertainty than the annual sampling strategies (Figure 6). The results are not unexpected given the total loss of information in the NBS in the even years.

# 4 Discussion

Our study demonstrated that spatio-temporal models can successfully fill in data gaps in many circumstances when estimating abundance. Our empirical approach adopted the northern Bering Sea bottom trawl survey's sampling frequency and removed observations from four large areas of the eastern Bering Sea standard survey area. We specifically showed that a spatio-temporal model using an environmental covariate (1) results in accurate biomass indices when the core of the stock's range is not excluded from sampling, and (2) when the core of the stock's range is excluded from data, the confidence intervals increase in width to still capture the abundance index that would arise using full data. We also used a simulation experiment to explore likely performance under alternative sampling strategies involving an infrequently surveyed area or a newer climate-adaptive survey area. The simulation experiment shows that the model has minimal bias and is precise when full sampling coverage is available in every year. However, if a reduction in sample sizes is necessary, reducing sampling density and maintaining annual sampling is more advantageous than maintaining sampling density at biennial sampling intervals. The overlap in the sampling and species spatial distributions, as well as survey frequency and intensity influenced the uncertainty estimates in our predictions. The benefits of adequate spatial coverage and annual sampling (i.e., reduced uncertainty in our biomass estimates) were obvious from our empirical and simulation experiments. For example, the better performance of the annual, reduced scenario in the simulation exercise indicates that having some information every year will improve annual estimates as opposed to have full information every other year. Spatio-temporal models, like VAST, rely on information from nearby locations and among years for extrapolation. It is therefore intuitive that excluding data from a large subarea that contains the center of a species distribution results in increased uncertainty. Similar results were shown for walleye pollock, where uncertainty estimates from a model similar to the one used in this study and using similar data was greater in years when the survey did not sample in the northern Bering Sea (O'Leary et al., 2020). Grüss and Thorson (2019) produced similar results to this study when simulating two scenarios: 1) the removal of the northwestern Gulf of Mexico (GOM) survey sample for red snapper (*Lutjanus campechanus*) and 2) not surveying the GOM over a number of early years. Uncertainty in their abundance estimates increased when either large spatial areas were not surveyed over time or a number of consecutive years were not sampled.

The uncertainty associated with survey biomass is an important stock assessment input, where estimates are calculated outside of the assessment and then inputted as the coefficient of variation (CV) into an assessment model. The CV value provides the model with a relative data weight associated with each observation and in

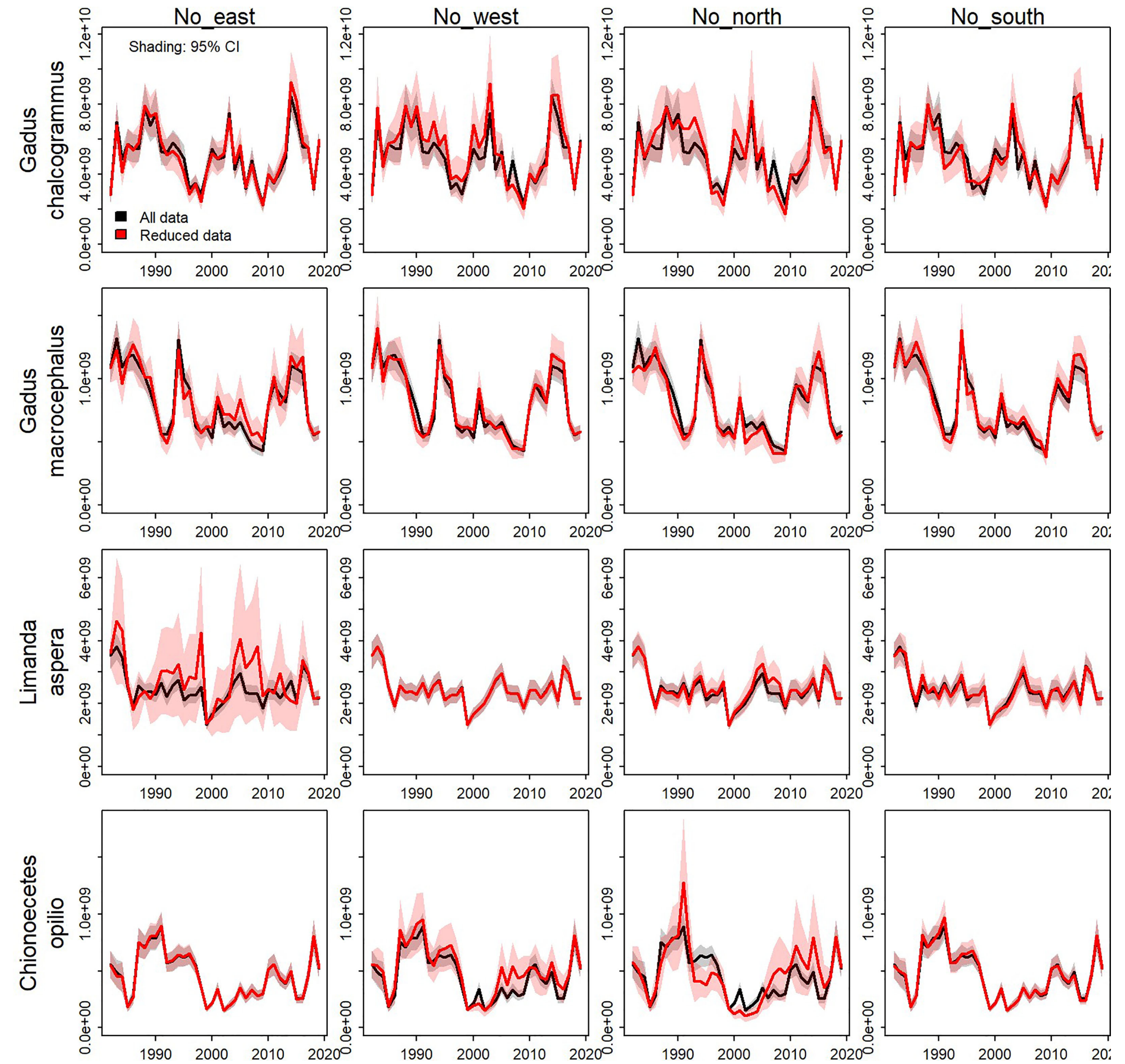



FIGURE 4
Biomass index estimates when the model was fit to all data (black line) and reduced data (red line) by species (rows) and areas removed from the data set (columns).Shaded regions represent the 95% confidence interval.

relation to other sources of information. The data weights are used in the likelihood component of the assessment model and effectively determines how well the model will fit individual data points, as well as the entire time series (Francis, 2011). Stock assessment model outcomes can be highly sensitive to input data weights leading to greater uncertainty in estimates of current stock size and stock status and in the estimation of management reference points (Francis, 2011; Maunder et al., 2017; Punt, 2017). Hence, studies like the one presented here are important to conduct and ascertain how a change in survey sampling will affect the estimate of biomass and uncertainty.

Fishery-independent surveys collect a wide variety of data that go beyond biomass/abundance and include length and age composition data, as well as environmental data. Composition data provides information about changes in size and age structure, recruitment, growth, natural mortality, and in some cases sex ratio. Composition data are often included as proportions within a size or age class and the input sample size provides a measure of uncertainty. Biological (e.g., ontogenetic habitat, depth, and food preferences) and environmental drivers (e.g., bottom temperature) can lead to strong distribution patterns among lengths/ages within a species. Spatio-temporal models have been shown to effectively estimate compositional data and improve estimates of multinomial sample size (Thorson and Haltuch, 2019; O'Leary et al., 2020). This study focused on the impact of changing survey sampling on abundance estimates, but it is equally important to conduct a similar evaluation for composition data. A similar study should be conducted to determine the potential bias and

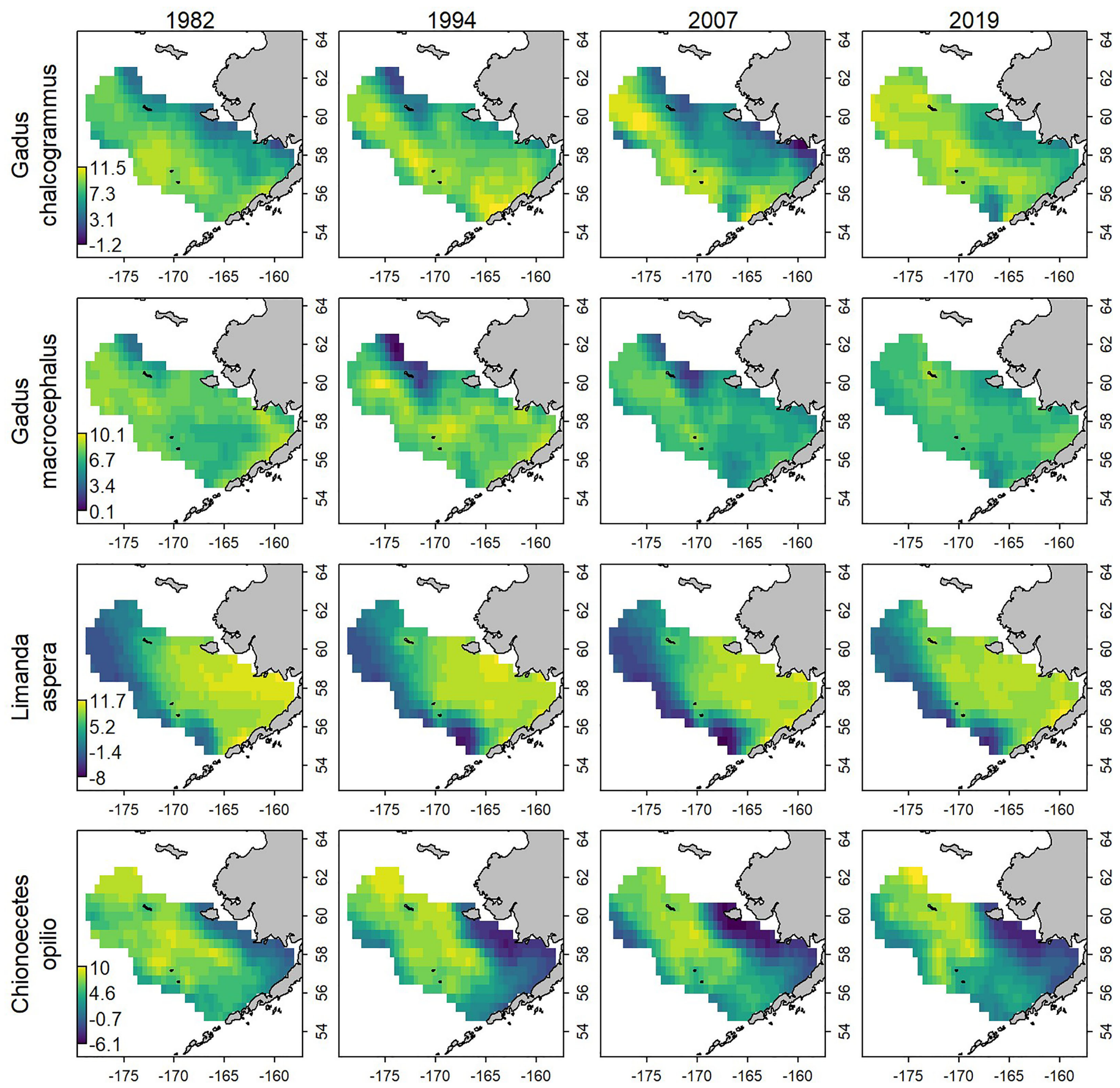



FIGURE 5
Estimate of log-biomass density, $\log_e(\text{kg/km}^2)$, for evenly spaced years (columns) for each species (row) analyzed in the empirical analysis, estimated using the "full data set". See Figure 2 for how these ranges overlap with the data-removal experiments.

uncertainty in the proportions at size/age and input sample size with changes in survey sampling. A loss of critical environmental data will be expected with changes in survey sampling. In the eastern Bering Sea the cold pool is an important oceanographic feature that is known to control species distribution in this region. Using an environmental index as a habitat covariate in a spatio-temporal model has been shown to be an effective way to model changes in species distributions and reduce uncertainty in abundance estimates (Thorson, 2019b; O'Leary et al., 2020). In our empirical analysis and simulations, we assumed that environmental data were available thereby we assumed we had perfect information about CPI. In reality, removing survey stations from a large area of the overall survey grid as was done for the empirical analysis would lead to a loss of information and affect our ability to calculate the CPI. In these cases, CPI could instead be calculated from other information, e.g., the Bering-10K Regional Ocean Modelling System which has been validated previously for this use (Kearney et al., 2021). Having an incomplete or alternative environmental index would lead to greater uncertainty in the model-based index. Therefore, our estimates may be optimistic and the loss of environmental information resulting from decreased sampling should be evaluated in the future.

The need to restructure sampling strategies will always be in the forefront for survey programs due to funding uncertainties, changes in species distributions and stock status, and the frequency of other unanticipated events like the COVID pandemic, or reduced or cancelled surveys due to inclement weather or a lack of funding. One certainty is that including biased inputs into a stock assessment

TABLE 1 The average (top rows) or coefficient of variation (bottom rows) for estimated density across years when fitting to all BT data in the eastern Bering Sea, computed at the set of extrapolation-grid cells that were retained across all years for a given sampling design.

| | All | No_east | No_west | No_north | No_south |
|---|---|---|---|---|---|
| Average density | | | | | |
| *Gadus chalcogrammus* | 7.987 | 8.155 | 7.826 | 8.121 | 7.786 |
| *Gadus macrocephalus* | 6.821 | 6.791 | 6.812 | 7.011 | 6.794 |
| *Limanda aspera* | 4.79 | 3.285 | 6.091 | 5.629 | 5.112 |
| *Chionoecetes opilio* | 3.756 | 5.012 | 3.309 | 3.106 | 3.84 |
| Average CV | | | | | |
| *Gadus chalcogrammus* | 0.158 | 0.156 | 0.165 | 0.141 | 0.17 |
| *Gadus macrocephalus* | 0.129 | 0.131 | 0.131 | 0.098 | 0.144 |
| *Limanda aspera* | 0.186 | 0.232 | 0.135 | 0.083 | 0.236 |
| *Chionoecetes opilio* | 0.184 | 0.249 | 0.149 | 0.173 | 0.129 |

These retained extrapolation-grid cells either included the entire eastern Bering Sea extent ("All" in 2nd column), or dropped extrapolation-grid cells in the eastern, western, northern, or southern areas (3rd-6th columns) for each species (rows). For example, the design that dropped data in the eastern Bering Sea (3rd column) for *Limanda aspera* has lower average density and higher average CV (compared with the values calculated for the entire eastern Bering Sea), indicating that the survey design dropping eastern stations is excluding the core habitat area for that species.

will lead to biased results and management advice. Survey abundance is an assumed absolute measure or a relative measure and proportional to population biomass. The results from an assessment model will be impacted by the bias in the survey estimate and in turn lead to biased population estimates and management reference points. Therefore, obtaining unbiased survey estimates is integral to any assessment. The spatio-temporal model that we presented can provide unbiased estimates; however, it cannot ameliorate problems with non-representative sampling, as was demonstrated for yellowfin sole and snow crab when we excluded

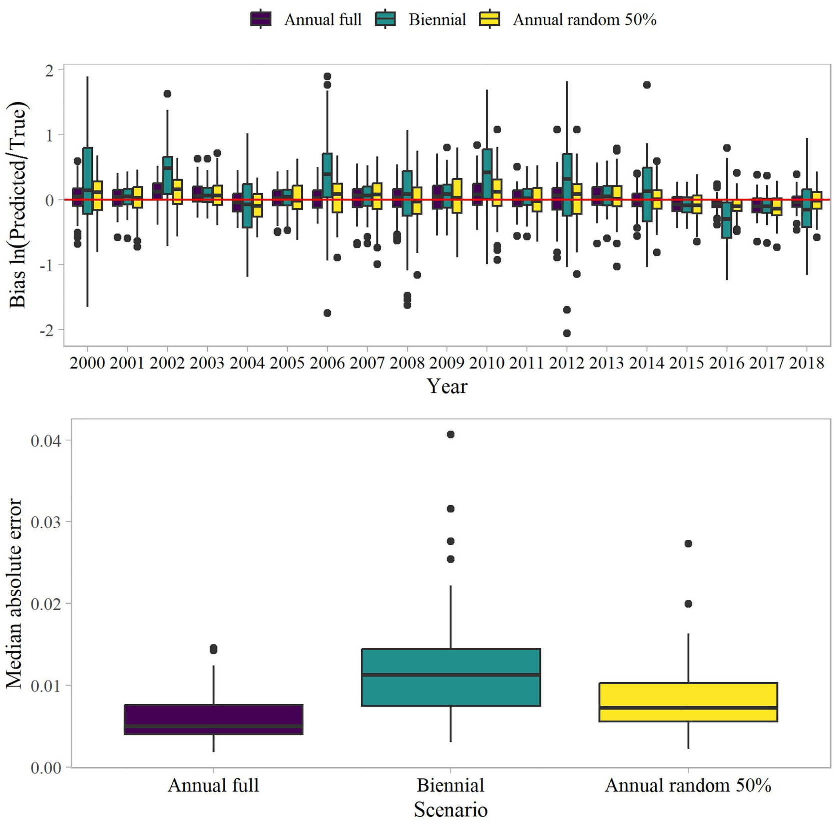



FIGURE 6
Bias in the estimated proportion of biomass in the northern Bering Sea for different sampling design scenarios (top panel) and the median absolute error among sampling scenarios (bottom panel). The boxplots are defined by the median (line), interquartile range (IQR, box), the furthest points from the 1.5 the IQR (whiskers), and outliers (points).

data from their core distributions. The empirical results also showed a non-uniform response among species and a trade-off among species and the removal of regional data. This has important implications on future surveys. As species distributions shift due to a changing climate, our surveys must adapt to effectively monitor variability and changing centers of distribution. The consequence of not doing so will likely be biased estimates; however, adaptability is no small task. Sampling optimization for all species within a multispecies fishery-independent survey is incredibly difficult. Spatio-temporal models and optimization methods should be used to identify trade-offs in bias/inaccuracy and uncertainty among species and strive to achieve representative sampling for as many species possible (Oyafuso et al., 2021).

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

All authors conceived the ideas, designed methodology, and analyzed the data and were equally involved in manuscript writing. All authors contributed to the article and approved the submitted version.

## Acknowledgments

We would like to thank the NOAA Alaska Fisheries Science Center (AFSC) Groundfish Assessment Program and their tremendous efforts during the eastern and northern Bering Sea shelf bottom trawl survey to collect the data and make it available for this analysis. We would also like to thank NOAA-AFSC reviewers Sandra Lowe and Drs. Steve Barbeaux and Zack Oyafuso for their valuable comments, as well as the journal reviewers.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Bakkala, R. G. (1993). *Structure and historical changes in the groundfish complex of the eastern Bering Sea* (United States Department of Commerce: NOAA Technical Memorandum NMFS 114), 86.

Francis, R. I. C. C. (2011). Data weighting in statistical fisheries stock assessment models. *Can. J. Fish. Aquat. Sci.* 68, 1124–1138. doi: 10.1139/f2011-025

Grüss, A., and Thorson, J. T. (2019). Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. *ICES J. Mar. Sci.* 76, 1748–1761. doi: 10.1093/icesjms/fsz075

ICES (2020). *Workshop on unavoidable survey effort reduction (WKUSER) (No. 2:72), ICES scientific reports* (Copenhagen: ICES).

Kearney, K. A., Alexander, M., Aydin, K., Cheng, W., Hermann, A. J., Hervieux, G., et al. (2021). Seasonal predictability of sea ice and bottom temperature across the eastern Bering Sea shelf. *J. Geophys. Res.: Ocean.* 126, e2021JC017545. doi: 10.1029/2021JC017545

Lo, N. C., Jacobson, L. D., and Squire, J. L. (1992). Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish. Aquat. Sci.* 49, 2515–2526. doi: 10.1139/f92-278

Markowitz, E. H., Dawson, E. J., Charriere, N. E., Prohaska, B. K., Rohan, S. K., Stevenson, D. E., et al. (2022). *Results of the 2021 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna* (United States Department of Commerce: NOAA Technical Memorandum NMFS-AFSC-452), 215.

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* 70, 141–159. doi: 10.1016/j.fishres.2004.08.002

Maunder, M. N., Punt, A. E., Valero, J. L., and Semmens, B. X. (2017). Data conflict and weighting, likelihood functions and process error. *Fisheries Research* 192, 1–4. doi: 10.1016/j.fishres.2017.03.006

Mueter, F. J., and Litzow, M. A. (2008). Sea Ice retreat alters the biogeography of the Bering Sea continental shelf. *Ecol. Appl.* 18, 309–320. doi: 10.1890/07-0564.1

O'Leary, C. A., DeFilippo, L. B., Thorson, J. T., Kotwicki, S., Hoff, G. R., Kulik, V. V., et al. (2022). Understanding transboundary stocks' availability by combining multiple fisheries-independent surveys and oceanographic conditions in spatiotemporal models. *ICES J. Mar. Sci.* 79, 1063–1074. doi: 10.1093/icesjms/fsac046

O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. (2020). Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (Gadus chalcogrammus) stock assessment. *Fish. Oceanog.* 29, 541–557. doi: 10.1111/fog.12494

Oyafuso, Z. S., Barnett, L. A. K., and Kotwicki, S. (2021). Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. *ICES J. Mar. Sci.* 78, 1288–1300. doi: 10.1093/icesjms/fsab038

Pecl, G. T., Araujo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., et al. (2017). Biodiversity redistribution under climate change: impacts on ecosystems and human well being. *Science* 355, eaai9214. doi: 10.1126/science.aai9214

Perretti, C. T., and Thorson, J. T. (2019). Spatio-temporal dynamics of summer flounder (Paralichthys dentatus) on the northeast US shelf. *Fish. Res.* 215, 62–68. doi: 10.1016/j.fishres.2019.03.006

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarimiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science* 341, 1239–1242. doi: 10.1126/science.1239352

Poloczanska, E., Brown, C., Sydeman, W., Kiessling, W., Schoeman, D. S., and Moore, P. J. (2013). Global imprint of climate change on marine life. *Nature Climate Change* 3, 919–915. doi: 10.1038/nclimate1958

Punt, A. E. (2017). Some insights into data weighting in integrated stock assessments. *Fish. Res.* 192, 52–65. doi: 10.1016/j.fishres.2015.12.006

Shelton, A. O., Thorson, J. T., Ward, E. J., and Feist, B. E. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Can. J. Fish. Aquat. Sci.* 71, 1655–1666. doi: 10.1139/cjfas-2013-0508

Spies, I., Gruenthal, K. M., Drinan, D. P., Hollowed, A. B., Stevenson, D. E., Tarpey, C. M., et al. (2020). Genetic evidence of a northward range expansion in the eastern Bering Sea stock of pacific cod. *Evolution. Appl.* 13, 362–375. doi: 10.1111/eva.12874

Stabeno, P. J., Farley, E. V.Jr., Kachel, N. B., Moore, S., Mordy, C. W., Napp, J. M., et al. (2012). A comparison of the physics of the northern and southern shelves of the eastern Bering Sea and some implications for the ecosystem. *Deep-Sea Res. II* 65–70, 14–30. doi: 10.1016/j.dsr2.2012.02.019

Stauffer, G. (2004). *NOAA Protocols for groundfish bottom trawl surveys of the nation's fishery resources* (United States Department of Commerce: NOAA Technical Memorandum NMFS-F/SPO-65), 205.

Stefansson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.* 53, 577–588. doi: 10.1006/jmsc.1996.0079

Stevenson, D. E., and Lauth, R. R. (2019). Bottom trawl surveys in the northern Bering Sea indicate recent shifts in the distribution of marine species. *Polar. Biol.* 42, 407–421. doi: 10.1007/s00300-018-2431-1

Thompson, G. G., and Thorson, J. T. (2019). "Assessment of the pacific cod stock in the Eastern Bering Sea," in *Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea and Aleutian islands* (605 W. 4th Avenue Suite 306, Anchorage: North Pacific Fishery Management Council).

Thorson, J. T. (2018). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Can. J. Fish. Aquat. Sci.* 75, 1369–1382. doi: 10.1139/cjfas-2017-0266

Thorson, J. T. (2019a). Measuring the impact of oceanographic indices on species distribution shifts: the spatially varying effect of cold-pool extent in the eastern Bering Sea. *Limnol. Oceanog.* 64, 2632–2645. doi: 10.1002/lno.11238

Thorson, J. T. (2019b). Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.*, 143–161. doi: 10.1016/j.fishres.2018.10.013

Thorson, J. T., Arimitsu, M. L., Barnett, L. A. K., Cheng, W., Eisner, L. B., Haynie, A. C., et al. (2021). Forecasting community reassembly using climate-linked spatio-temporal ecosystem models. *Ecography* 44, 612–625. doi: 10.1111/ecog.05471

Thorson, J. T., Barnes, C. L., Friedman, S. T., Morano, J. L., and Siple, M. C. (2023). Spatially varying coefficients can improve parsimony and descriptive power for species distribution models. *Ecography* 2023, e06510. doi: 10.1111/ecog.06510

Thorson, J. T., and Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES J. Mar. Sci.* 74, 1311–1321. doi: 10.1093/icesjms/fsw193

Thorson, J. T., and Haltuch, M. A. (2019). Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. *Can. J. Fish. Aquat. Sci.* 76, 401–414. doi: 10.1139/cjfas-2018-0015

Thorson, J. T., and Kristensen, K. (2016). Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fish. Res.* 175, 66–74. doi: 10.1016/j.fishres.2015.11.016

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West coast groundfishes. *ICES J. Mar. Sci.* 72, 1297–1310. doi: 10.1093/icesjms/fsu243

Thorson, J. T., and Ward, E. J. (2013). Accounting for space–time interactions in index standardization models. *Fish. Res.* 147, 426–433. doi: 10.1016/j.fishres.2013.03.012

Wyllie-Echeverria, T., and Wooster, W. S. (1998). Year-to-year variations in Bering Sea ice cover and some consequences for fish distributions. *Fish. Oceanog.* 7, 159–170. doi: 10.1046/j.1365-2419.1998.00058.x

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 06 July 2023
DOI 10.3389/fmars.2023.1214949

Check for updates

OPEN ACCESS

EDITED BY
Nathan Bacheler,
Beaufort Laboratory, Southeast Fisheries Science Center (NOAA), United States

REVIEWED BY
Josep Lloret,
University of Girona, Spain
Brendan J. Runde,
The Nature Conservancy, United States

*CORRESPONDENCE
Elizabeth T. Methratta
Elizabeth.Methratta@noaa.gov

RECEIVED 30 April 2023
ACCEPTED 15 June 2023
PUBLISHED 06 July 2023

CITATION
Methratta ET, Lipsky A and Boucher JM (2023) Offshore wind project-level monitoring in the Northeast U.S. continental shelf ecosystem: evaluating the potential to mitigate impacts to long-term scientific surveys.
*Front. Mar. Sci.* 10:1214949.
doi: 10.3389/fmars.2023.1214949



# Offshore wind project-level monitoring in the Northeast U.S. continental shelf ecosystem: evaluating the potential to mitigate impacts to long-term scientific surveys

Elizabeth T. Methratta[1]*, Andrew Lipsky[1] and Jason M. Boucher[2]

[1]Northeast Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service, Narragansett, RI, United States, [2]Northeast Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service, Woods Hole, MA, United States

Offshore wind will generate much needed renewable energy in the U.S. and worldwide, but this industry will also affect other ocean uses. In the Northeast U.S. continental shelf (NES) ecosystem, these effects include the impact that wind development will have on the design and execution of long running scientific surveys conducted by National Marine Fisheries Service of the National Oceanic and Atmospheric Administration (NOAA Fisheries) which play a critical role in the provision of scientific information for stock assessment and advice for fisheries management. Recognizing these impacts, the federal government has established a Survey Mitigation Strategy that identifies a need to evaluate whether the information yielded from project-level monitoring studies conducted by wind developers might be suitable for integration with data from NOAA Fisheries surveys, thereby ameliorating the impacts to the surveys. To address this need, we compiled and tabulated information from all currently available project-level monitoring studies and compared elements of the design and methodology of each study with that of the comparable NOAA Fisheries survey. Based on this information, we evaluated their suitability for filling expected gaps in long term surveys, for addressing impacts at the population level, and for understanding interactions between fish stocks and habitat alterations. We found that project-level monitoring studies as currently designed for the NES ecosystem will not yield information that can mitigate impacts to NOAA Fisheries scientific survey time series from offshore wind development. We provide recommendations on how to enhance the ability of project-level monitoring studies to mitigate impacts to long term scientific surveys.

KEYWORDS

**renewable energy, survey mitigation, impact assessment, stock assessment, fisheries independent surveys**

## Introduction

The U.S. plans to develop 30 GW of offshore wind energy by the year 2030 as part of a multi-faceted effort to combat climate change. The nation's first utility scale wind developments are slated to be constructed in the Northeast U.S. continental shelf (NES) ecosystem (Figure 1), which is also home to one of the world's most productive fishing grounds, protected and endangered species, and sensitive habitats. As the nation's steward of natural marine resources, the National Marine Fisheries Service of the National Oceanic and Atmospheric Administration (NOAA Fisheries) conducts 14 scientific surveys in the NES ecosystem (Table 1) that will be impacted by offshore wind development, some with time series exceeding 60 years. As the footprint of offshore wind energy development grows, additional surveys may be impacted and the impacts to existing surveys will likely increase. These surveys support management of more than 40 fisheries, more than 30 marine mammal species, and 14 threatened and endangered species through stock assessment and the provision of management advice. Moreover, these scientific surveys support numerous other NOAA Fisheries' science products, including ecosystem and climate assessments. NOAA Fisheries surveys have occurred in the region since the early 1960s, leading the NES ecosystem to be one of the best studied marine ecosystems in the world.

Because of the substantial spatial overlap between offshore wind development and NOAA Fisheries scientific surveys (e.g., Bottom Trawl Survey, Figure 1), wind development will impact the surveys through the following four means (Hare et al., 2022): 1) Preclusion - displacement of survey by wind infrastructure; 2) Impacts to Statistical Survey Design - current statistical survey methods will no longer be able to be executed due to reduced spatial sampling frame; 3) Change in Habitat and Concomitant Effects on

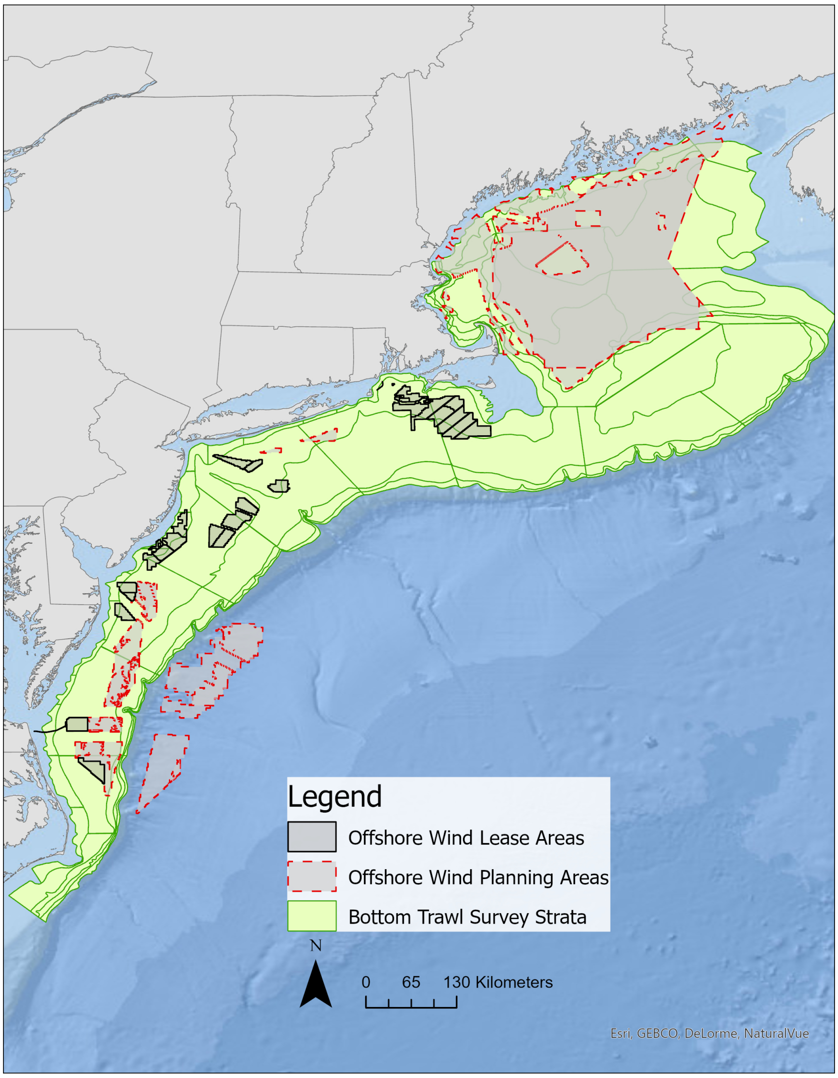



FIGURE 1
NOAA Fisheries fall and spring Bottom Trawl Survey strata overlaid by current offshore wind leases and planning areas in the NES ecosystem. Planning areas are initial areas identified by BOEM as potentially suitable for wind development. These areas are winnowed down through analysis and input from stakeholders to wind energy areas (WEAs). Lease areas are specific areas of the WEAs that are leased to developers for wind project development. Map provided by Angela Silva (NOAA Fisheries).

TABLE 1 NOAA Fisheries scientific surveys.

| Name of NOAA Scientific Survey | Year Started | Survey Design (Sampling Gear) | Major Applications |
|---|---|---|---|
| Continuous Plankton Recorder | 1961 | Towed Continuous Plankton Recorder | Abundance, distribution, biomass |
| Autumn Bottom Trawl Survey | 1963 | Random Stratified (Bottom Trawl) | Abundance, distribution, length, age, sex, weight, diet, and maturity samples, components of Ecosystem Monitoring survey |
| Spring Bottom Trawl Survey | 1968 | Random Stratified (Bottom Trawl) | Abundance, distribution, length, age, sex, weight, diet, and maturity samples, components of Ecosystem Monitoring survey |
| Sea Scallop Dredge Survey/Integrated Benthic Habitat Survey | 1979 | Random Stratified (Dredge)<br>Line Transect (HabCam) | Abundance, distribution, biomass, size, and sex of sea scallops (*Placopecten magellanicus*) and other benthic fauna |
| Atlantic Surfclam and Ocean Quahog Surveys | 1980 | Random Stratified (Hydraulic Dredge) | Abundance, distribution, biomass, size, and sex of Atlantic surfclam (*Spisula solidissima*) and ocean quahog (*Arctica islandica*) |
| Northern Shrimp Survey | 1983 | Random Stratified (Commercial Shrimp Trawl) | Abundance, distribution, biomass, and size |
| Gulf of Maine Cooperative Bottom Longline Survey | 2014 | Randomly Stratified (Bottom Longline) | Abundance, distribution, length, age, sex, weight, diet, and maturity samples, components of Ecosystem Monitoring survey |
| Ecosystem Monitoring Survey (6 times per year) | 1977 | Random Stratified [linked to Bottom Trawl Survey Design] and fixed Stations (Plankton, Oceanographic, and Visual Sampling) | Phytoplankton, zooplankton, ichthyoplankton, carbonate chemistry, nutrients, marine mammals, sea birds |
| North Atlantic Right Whale Aerial Surveys | 1998 | Line Transects (Visual) | Right Whale (*Eubalaena glacialis*) population estimates; dynamic area management |
| Marine Mammal and Sea Turtle Aerial Surveys | 1993 | Line Transects (Visual) | Abundance and spatial distribution of marine mammals and sea turtles for stock assessments |
| Marine Mammal, Sea Turtle, and Seabird Ship-based Surveys | 1991 | Line Transects (Visual along with Plankton and Oceanographic Sampling) | Abundance and spatial distribution of marine mammals, sea turtles, and sea birds for stock assessments |
| Seal Aerial Abundance Surveys | 1990 | Surveys over Haul-out Sites and Pupping Colonies (photographic) | Abundance, distribution, migration (tagging) for assessments of harbor and gray seals |
| Coastal Shark Bottom Longline Survey | 1986 | Fixed station (bottom longline) | Abundance, distribution, life history, migrations (tagging) |
| Cooperative Atlantic States Shark Pupping and Nursery Longline/Gillnet Survey | 1998 | Random stratified and fixed station (longline and gillnet) | Abundance, distribution, life history, migrations (tagging) |

Population Structure - changes in habitat will have affect species distribution, abundance, and vital population rates both inside and outside of wind project areas and lead to changes in population structure after construction and for the lifetime of the project; and 4) Practical Sampling - navigating survey vessels around wind energy areas will increase transit time and sampling time. In the immediate term, these impacts will cause gaps in long term time series of the surveys. The long term implications of survey impacts and the knock-on effect of reduced data quality to support stock assessment and management advice, will ultimately lead to greater uncertainty in fisheries management. Given the severity of the impacts that wind development will have on NOAA Fisheries scientific surveys, NOAA Fisheries and the Bureau of Ocean Energy Management (BOEM) have joined together to establish the Federal Survey Mitigation Strategy for Northeast U.S. Region (Hare et al., 2022). The Mitigation Strategy is intended to guide the Mitigation Program which will include survey-specific mitigation plans for each impacted survey, including both vessel and aerial surveys (Survey-Specific Mitigation Plans). Although specific to the Northeast U.S. Region (Maine to North Carolina), the strategy is generally applicable to other regions of the country. These issues are also being faced in many other counties and they are currently being addressed by the International Council for the Exploration of the Seas (ICES) Working Group for Offshore Wind and Fisheries (ICES, 2022) and have been incorporated into the objectives of the ICES roadmap for Offshore and Marine Renewable Energy (ICES, 2023).

A potential source of information that could mitigate impacts to NOAA Fisheries scientific surveys, evaluate population-level impacts, and inform our understanding of how stocks respond to habitat alteration is the project-level monitoring undertaken by wind developers. Project-level monitoring could be designed to address questions about changes in habitat, the underlying mechanisms, and how measured responses do or do not confer population level effects. For each wind project, some individual states require wind developers to conceive and execute project-level monitoring that evaluates the impacts derived from the construction, installation, and operation of wind structures (e.g., Vineyard Wind, 2023). BOEM also provides a set of guidelines for project monitoring and incorporates any of the state-level requirements or voluntary activities proposed by a developer into the project's regulatory approval (BOEM, 2023).

The potential impacts on habitat and fisheries resources are substantial and include effects from a wide array of impact producing factors (IPFs) including electromagnetic fields (EMF), noise, benthic habitat alteration, artificial reef and FAD effects, pelagic habitat alteration, and hydrodynamic changes (NOAA, 2023). Each of these IPFs cause habitat alterations that have the potential to affect vital population rates and thus populations. Despite the requirements to conduct project-level monitoring, there currently exist no requirements for what a monitoring plan should contain, although various sets of guidelines have been provided by BOEM, state agencies, and regional working groups (MADMF, 2018; ROSA, 2021; BOEM, 2023). The collection of monitoring data in and around individual wind projects represents a potential opportunity to mitigate impacts to NOAA Fisheries scientific surveys. These data could inform our understanding of impacts to populations and provide data that could supplement information lost from the survey time series due to wind development. This opportunity was recognized by the Federal Survey Mitigation Strategy which called for an evaluation and integration, where feasible, of wind energy development monitoring studies with NOAA Fisheries surveys (Hare et al., 2022, Action 2.2.1).

The purpose of this paper is to evaluate whether project-level monitoring studies as currently designed by wind developers will mitigate impacts to the long-term scientific resource surveys conducted by NOAA Fisheries with a focus on fisheries resource surveys. Specifically, we evaluated whether they can yield information that can fill expected gaps in long term surveys, examine impacts at the population level, and/or inform our understanding of how populations will respond to wind-derived habitat alterations. To that end, we 1) Collated all existing offshore wind development fisheries and benthic monitoring plans and tabulated several key aspects of each plan; 2) Determined whether each survey proposed by project-level monitoring collects information that is functionally equivalent to the comparable NOAA Fisheries scientific survey; and 3) Provide recommendations on how project-level monitoring plans could be adapted to enhance their suitability for providing information that can support existing scientific survey time series and thus the mitigation of wind-derived impacts on fisheries management. For clarity, we use the term "monitoring plan" throughout this paper to refer to the combined benthic and fisheries monitoring plan for an individual wind project. The term "study" is used to refer to a specific experiment within a monitoring plan. The term "survey" is used to refer to NOAA Fisheries scientific surveys.

# Methods

All accessible fisheries and benthic monitoring plans were gathered and collated. The plans are publicly available and can be accessed on websites maintained by BOEM, other agencies, or obtained directly from the developers. From each monitoring plan, we extracted and compiled the following information into a table:

## Characteristics of the monitoring studies

These included: 1) Name of the wind project; 2) Stated research question or hypothesis; 3) Impact producing factors studied; 4) Target taxa, species, or habitat; 5) Gear type; 6) Location; 7) Data types to be collected; 8) Where the study will be conducted (e.g., wind project, controls, cable route); 9) Month and/or Season of sampling; 10) Temporal duration during each phase of wind development; 11) Experimental design; 12) Statistical method for station selection (e.g., random, systematic, etc.); 13) Regional survey the plan states that it is comparable to or is modeled after, if any; and 14) Comparable NOAA Fisheries scientific survey, i.e., survey that is currently sampling the same species or habitat as the proposed monitoring; and 15) Description of how QA/QC'd data will be accessible and readily available.

## Utility of the monitoring plan in mitigating wind-derived impacts to long-term surveys

We further examined whether the monitoring studies as proposed: 16) Include supplementing the comparable NOAA Fisheries survey as stated an objective; 17) Are calibrated to an existing NOAA Fisheries survey; 18) Address Preclusion; 19) Address impacts to statistical survey design; 20) Address habitat change and responses to habitat change after construction and for the lifetime of the wind project; 21) Address practical sampling issues; and 22) Provide a functionally equivalent sample to the comparable NOAA Fisheries survey.

# Results

We identified monitoring plans from 9 different offshore wind projects that were available for review including: Atlantic Shores (2021a; 2021b); Empire Wind (2022); New England Wind (2021a; 2021b); Ocean Wind (2021; 2022); Revolution Wind (2021); South Fork Wind (2020); Sunrise Wind (2021); Vineyard Wind (2023), and Coastal Virginia Offshore Wind (CVOW, 2022a; CVOW, 2022b). Thus 27% of the N=33 total leased areas in the NES ecosystem (BOEM, 2023) have proposed fisheries and benthic monitoring plans that were accessible. Among these, there were 67 unique monitoring studies proposed across a range of taxonomic groups and habitat types (Supplemental Tables S1-S5).

## Characteristics of the monitoring studies

### Research question, objective, or hypotheses and IPFs evaluated

All of the proposed studies provided research questions, objectives, and/or hypotheses that were to be addressed (Figure 2;

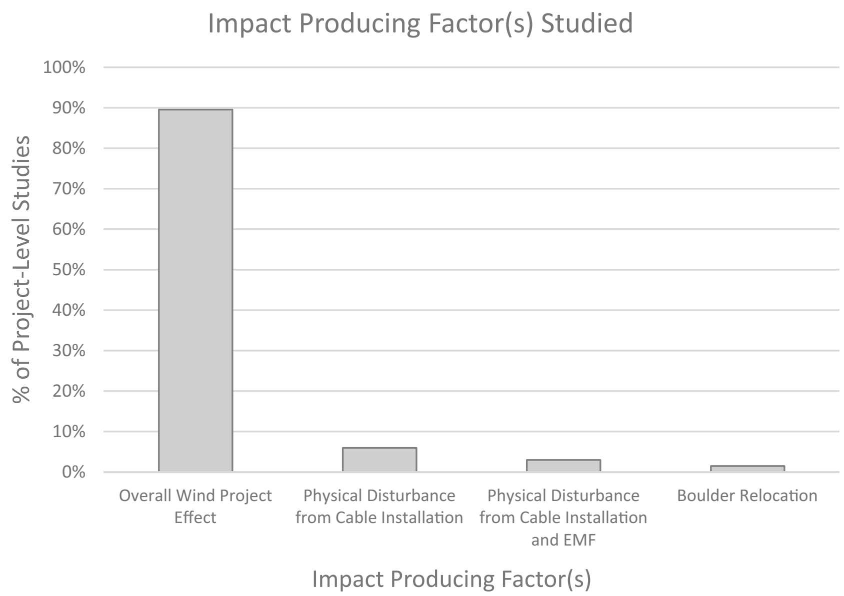



FIGURE 2
The proportion of project level studies (N=67) that studied particular effects or impact producing factors (IPFs).

Supplemental Table S1). A small proportion of studies proposed to explore a specific IPF. All of these were local scale studies that did not explore potential impacts to populations-level impacts at the spatial scale of the stock. Among the 67 studies were 6 (9%) studies aimed at studying how physical disturbance during the cable installation process affected benthic habitat and/or benthic habitat function. Also, two (3%) studies specifically planned to test hypotheses related to electro-magnetic frequency (EMF) impacts along cables although one of these had a hypothesis that combined the effects of both physical disturbance and EMF. One other study planned to look at the effect of boulder relocation on the epibenthic community. The vast majority of studies (n=60; 90%) focused on the overall impact of wind project structures (presence vs. absence) on target species that are expected to occur in the area.

## Study target and sampling gear

Eight (89%) of the N=9 wind project monitoring projects included a trawl study (e.g., otter trawl, beam trawl) to investigate demersal fish and invertebrates, representing 12% of the N=67 overall proposed studies (Figures 3A, B; Supplemental Table S2). There were 2 (22%) wind projects (Ocean Wind 1 and Atlantic Shores) that proposed dredge studies to examine surf clams and other shellfish, representing 3% of all studies proposed. A total of 11 (16%) studies across 7 (78%) wind projects proposed to utilize fish pots or traps to study structure associated species including black sea bass (*Centropristis striata*), tautog, scup, American lobster (*Homarus americanus*), Jonah crab (*Cancer borealis*), and whelk (*Busycon spp.*). A total of 17 (25%) studies plan optical sampling modalities. This included 9 (13%) studies across 6 (67%) wind projects that plan to use remotely operated vehicle (ROV)/video studies of hard bottom habitat; 2 (3%) studies across 2 (22%) wind projects that plan to use ROV/video to study benthic megafauna or epibenthos; 2 (3%) studies across 2 (22%) wind projects that plan to use drop camera methods to examine benthic macrofauna; 1 (2%) study at a single wind project that plans to use drop camera methods to study submerged aquatic vegetation (SAV) coverage; 3 (4%) studies across 2 wind projects (22%) that plan to use baited remote underwater video (BRUV) methods to study structure-oriented or large pelagic fish; and 1 (2%) plan view camera study of scallops at a single wind project. Eight (12%) acoustic telemetry studies across 6 (67%) wind projects were proposed that targeted highly migratory species, lobsters, elasmobranchs, and other demersal finfish such as summer flounder and black sea bass. All acoustic telemetry studies involved fixed receiver arrays but two also included data collected from an autonomous glider. Sediment profile imaging/Plan View camera methods (SPI/PV) were the most common method for studying soft bottom habitat function (11 (67%) studies across 5 (56%) wind projects) including within the array and along the cable route. Only 2 (3%) of these studies planned to also collect physical sediment samples with a grab. Neuston nets will be used to sample lobster larvae and other planktonic organisms in 2 (22%) wind projects. The non-extractive method of eDNA sampling is planned for 2 (22%) wind projects. Gillnet, the least common method proposed for studying the demersal fish community, was planned for 1 wind project.

## Location

Of the N=67 studies, there were 7 (10%) studies proposed along wind project export cables and 1 (2%) study proposed along the inter-array cable (Supplemental Table S2). The remaining 59 (88%) studies focused on structures within the footprint of the

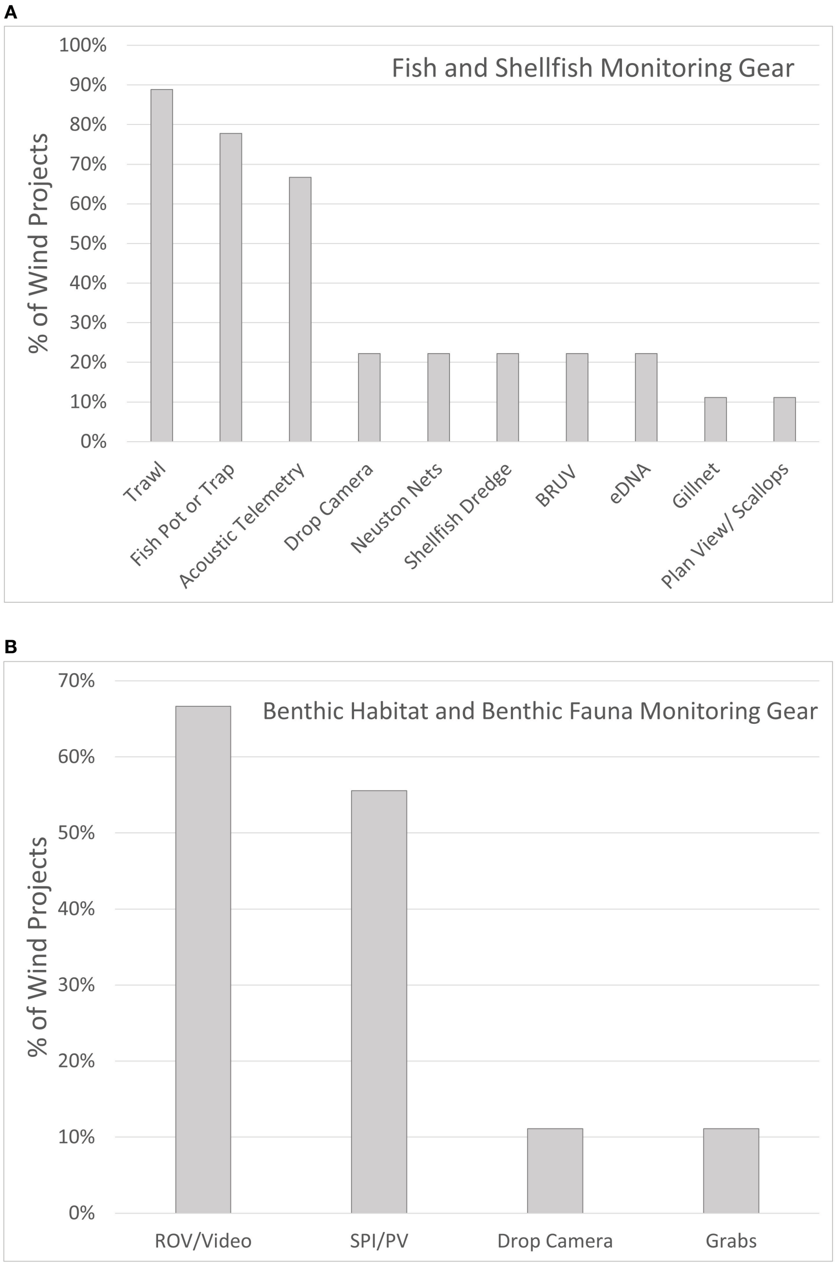



FIGURE 3
Gear types used. The proportion of wind projects (N=9) that used each of the gear types proposed to examine **(A)** Fish and shellfish and **(B)** Benthic habitat and/or benthic fauna.

wind project; 30 (45%) included comparisons with controls located outside of the wind energy area. Some telemetry projects noted that existing arrays located outside of the wind project array would supplement their study (e.g., Bangley et al., 2020).

## Data types to be collected

Abundance, community composition, length, weight, and reproductive status were the most common types of data planned for collection for finfish while less common were diversity, species richness, and diet composition of specific species (Supplemental Table S3). Studies of lobsters frequently planned collect data on carapace length, sex, egg status, v-notch status, cull status, and shell disease incidence. For Jonah crab collecting data on carapace width, ovigery, sex, shell disease incidence, cull status, and mortality data were common objectives. Hard bottom habitat monitoring planned to focus on measures of %cover; relative abundance of microbiota; estimated biomass/biovolume; invasive species. Studies employing acoustic telemetry aimed to collect data to characterize presence, residency, movement, and in some instances connectivity among lease areas. Dredge studies of shellfish focused on biomass, volume, size, and age measurement.

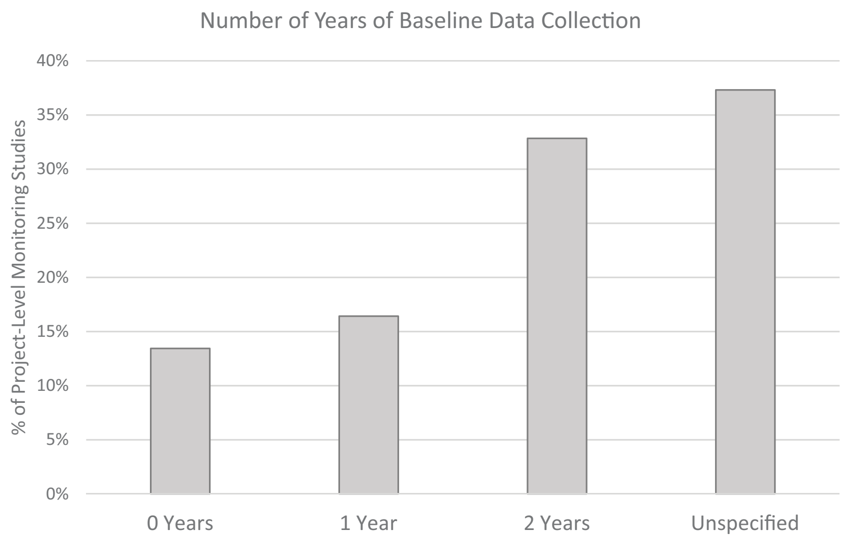



FIGURE 4
Baseline data collection. The proportion of project monitoring studies (N=67) proposing 0, 1, 2, or an unspecified number of years of baseline data collection.

## Month or season of sampling

The month or season of sampling varied among monitoring plan goals and target species (Supplemental Table S3). Most studies of fisheries resource species aimed to collect data when species were expected to be present and/or when commercial fisheries for the target species were most active. Several plans noted that particular months would be avoided to minimize interactions with protected species.

*Temporal Duration.* Of the N=67 studies proposed, 11 (16%) planned only a single year of baseline data collection (Figure 4; Supplemental Table S3). Twenty-two (33%) planned 2 years of baseline data collection, 25 (37%) planned baseline data collection for an unspecified number of years. The remaining 9 studies were only planned for the post construction time period.

## Experimental design

Twenty-six (39%) of N=67 proposed studies proposed a before-after-control-impact (BACI) design, 21 (31%) planned a before-after-gradient (BAG) design, and 4 (6%) planned a hybrid BACI/BAG approach (Supplemental Table S3) (Methratta, 2020). Six (9%) proposed a Before-After study in the impact area with no control. Seven (10%) proposed a post-construction study only in the impact area with no control. There were 3 (4%) proposed studies that did not have a clear experimental design stated or implied.

## Statistical method for station selection

Random selection was the most common method for station assignment (Supplemental Table S3). Several studies sought to achieve a spatial balance by spatially gridding the study area and randomly selecting a station within each grid cell. Systematic random sampling was another method used to avoid areas with set fishing gear. Stratification by habitat type was applied in some instances to address spatial variability.

## Comparability with regional assessments

Twenty-three (34%) of N=67 studies noted that they were using methods consistent with another regional or state survey (Supplemental Table S3). For example, trawl studies consistently noted that they would follow the protocols of Northeast Area Monitoring and Assessment Program (NEAMAP) while optical studies of sea scallops and other shellfish noted that their study was similar to other regional drop camera or HABCam studies (Bonzek et al., 2017; Bethoney and Stokesbury, 2018; Coonamessett Farm, 2023). Another 11 (16%) noted that they were using similar methods to those applied in another lease area. The one study that planned to use a gillnet specifically noted that their results would not be comparable to other regional datasets due differences in catchability.

## Comparable NOAA Fisheries scientific survey

Of the 44 studies targeting fisheries species, 36 (82%) have spatial and temporal overlap with a comparable existing NOAA Fisheries scientific surveys (Supplemental Table S4). Because the focus of the studies is on comparing impacts within project areas, the majority do not employ methods that provide comparable measurements to the NOAA Fisheries surveys (Supplemental Table S4).

## Provides for QA/QC'd data to be accessible and readily available

The majority of studies presented no plan to share or make accessible the QA/QC'd raw data collected in the study (Supplemental Table S3). The one exception were acoustic telemetry studies, most of which planned to share data with other telemetry researchers on the ACT and/or MATOS networks (Atlantic Cooperative Telemetry network, Mid-Atlantic Acoustic Telemetry Observation System).

## Utility of the monitoring plan in mitigating wind-derived impacts to long-term surveys

### Includes supplementing the NOAA Fisheries survey as a stated objective

None of the proposed studies stated that supplementing NOAA Fisheries scientific surveys was one of their objectives (Supplemental Table S4).

### Will be calibrated to an existing NOAA Fisheries survey

None of the proposed studies indicated that the study would be calibrated with an existing NOAA Fisheries scientific survey (Supplemental Table S4).

### Addresses preclusion

None of the proposed studies indicated that the study would address the issue of sampling preclusion of the comparable NOAA Fisheries scientific survey within the wind project (Supplemental Table S5). Studies that describe the use of gear and methods consistent with regional protocols, such as trawl studies that state that Northeast Ecosystem Assessment and Monitoring Program (NEAMAP) methods are employed, lack specificity of how the study design employs these standards, e.g. gear and vessel configuration; fish diet and condition methods; and relative to NOAA Fisheries multi-species groundfish survey, none of the proposed studies include sampling at night which contrasts with federal trawl survey protocols.

### Addresses impacts to statistical survey design

None of the proposed studies indicated that the study would address issues with the statistical survey design of the comparable NOAA Fisheries scientific survey within the wind project (Supplemental Table S5).

### Addresses habitat change and responses to habitat change

All (100%) of the 67 proposed studies intended to address either habitat change or a biological response to habitat change caused by wind development (Supplemental Tables S1-S3; S5). However, study design issues reduce the likelihood that these studies will be able to address this question. These issues include insufficient baseline study duration (most proposed 0, 1, or 2 years), experimental designs that utilize a control that is likely within the zone of impacts, and unknown statistical power for most of the variables mentioned (although some studies conducted power analyses for measures of abundance). In addition, no studies indicated how the local-scale investigations proposed might inform impacts of wind development for the lifetime of the wind project or at the population level of the species studied.

### Addresses practical sampling issues

None of the proposed studies indicated that they would address practical sampling issues (e.g., increased transit time) due to wind development (Supplemental Table S5).

### Provides a functionally equivalent sample to the comparable NOAA Fisheries survey

The 2 (3%) drop camera studies proposed to study benthic habitat and macroinvertebrate abundance and distribution have the potential to provide a sample that is functionally equivalent to the comparable NOAA Fisheries survey (Supplemental Table S5). The remaining 65 (97%) studies will not be able to do so.

## Discussion/conclusions

Project-level monitoring for offshore wind projects as currently designed for the NES ecosystem will not yield information that can be integrated into NOAA Fisheries scientific survey time series, nor are they designed with that intention. Therefore, they cannot help to mitigate scientific survey impacts from offshore wind development. In order for data yielded by project-level studies to be used in stock assessments, samples would need to be functionally equivalent to those collected by the NOAA Fisheries survey of the same population. Achieving functional equivalency for any of the surveys would be challenging and time intensive, often taking 10+ years of comparison of the datasets (Miller et al., 2010; ASMFC, 2020). For example, bottom trawl sampling was proposed for nearly every offshore wind project. In every instance, the plans noted that some or all of the protocols of the Northeast Area Monitoring and Assessment Program (NEAMAP) would be followed. Typically the studies emphasized matching such elements as gear type (demersal otter trawl), net mesh size, gear deployment and haul protocol, and sample handling with those of NEAMAP. Although this is laudable from the perspective of maintaining regional consistency among wind projects, unfortunately it does not ensure functional equivalency with the NOAA Fisheries bottom trawl survey, which means that those data cannot be automatically incorporated into stock assessment models. The same challenge is faced by nearly all of the project-level monitoring studies that are targeting species that are also sampled by NOAA Fisheries surveys.

An exception may be the drop camera studies used to evaluate abundance and other indices for sea scallops (Bethoney and Stokesbury, 2018). This study methodology employs a systematic grid design with samples collected at regular spatial intervals that are at equal distances from each other. Originally designed by the University of Massachusetts School of Marine Science and Technology (SMAST) in collaboration with commercial scallop fishermen, information collected by the SMAST survey has been incorporated into scallop stock assessment through the Stock Assessment Workshop process) (e.g., NEFSC, 2018).

The provision of scientific advice to inform fisheries management decisions is underpinned by population assessments developed with acceptable levels of accuracy and precision. The fisheries independent data collections undertaken by NOAA Fisheries each year provide unbiased data that are used in the calculation of biological indices for stock assessment purposes. The tremendous strength of NOAA Fisheries scientific surveys is rooted in their long-term spatial and temporal consistency in method and execution which includes rigorous statistical survey designs that are individualized for the taxa, species, or habitats that they are designed to sample. The development of offshore wind will disrupt the collection of data for every NOAA Fisheries survey and will thus create spatial and temporal gaps in every data set it collects. Fishery independent surveys are designed to provide essential, unbiased data for stock assessments, including indices of abundance, size and age composition, growth rates, and additional life history parameters (Lynch et al., 2018). Loss of accessible areas will have the knock-on effect of introducing bias into the assessments, with the potential for an index to deviate from the true trend if a species does not have a uniform distribution across the entire area. For example, structure-oriented species attracted to offshore developments may become unavailable to the survey. This would cause the index of abundance to artificially decrease, with the stock appearing to decline in the assessment. Some species also exhibit age-specific habitat preference (Macpherson and Duarte, 1991; Swain, 1993; Methratta and Link, 2007; Pappal et al., 2012), with individuals moving shallower or becoming less structure-oriented as they grow. The potential exists for an assessment to interpret the absence of certain ages as age-specific mortality, which could be attributed to recruitment failure or age truncation. This will also have a direct impact on many forms of Biological Reference Point (BRP), which is a measure of stock status that reflects the combination of several components of stock dynamics (growth, recruitment and mortality, fishing mortality) into a single index. The severity of these impacts on the accurate, timely, and precise assessments of stock condition will vary by stock due to specified stock assessment methods and their sensitivity to changes in sampling methods and potential interactions of offshore wind impact producing factors on stock attributes. However, identifying approaches that could potentially fill these gaps is essential to being able to provide fisheries managers with valid scientific advice. So what potential solutions might be considered?

For some population assessments, it may still be necessary to physically sample inside of wind projects to collect the necessary data, including for stocks where there is a likelihood that variance structure inside wind energy development areas may differ from outside of wind energy areas (e.g., Reubens et al., 2013; Roach et al., 2022). Given the amount of time to evaluate functional equivalency, it would be prudent to initiate a process by which samples that are functionally equivalent to long term surveys could be collected (e.g., Miller et al., 2010; ASMFC, 2020). In addition to deriving estimates of abundance, NOAA Fisheries surveys collect biological samples to provide essential data on population growth, reproduction, and age structure, among other information (e.g., bottom trawl survey, Azarorvitz et al., 1981). To provide relevant data, study design within a project should aim to supplement the existing NOAA Fisheries surveys for all data types, not only provide a replacement abundance value for stations precluded by the development of the project. It will be essential to monitor growth and reproductive rates of structure-oriented species, such as black sea bass (Bacheler and Ballenger, 2016), for the life of the project as these vital population rates may differ significantly between habitats.

For other populations, estimation modeling methods may provide an alternative approach to developing biological indices in areas that cannot be sampled by existing long term surveys (e.g., Thorson et al., 2019). Applying such methods with a high level of statistical validity and reliability would require making a set of assumptions about how populations respond to habitat change in the zone of impact of wind development compared to areas outside the zone of impact, how these changes vary over space, and how these patterns change through time. Project-level monitoring aimed at validating these assumptions, could provide a path toward understanding regional changes in population status.

The vast majority of project-level monitoring studies aim to address one fundamental question: what effect will offshore wind development have on response metrics of a target species? Successfully addressing this question is an essential part of being able to validate the assumptions regarding how populations respond to wind-derived habitat change. Whether or not this question can be answered by the studies proposed will depend on whether sufficient baseline data are collected, whether a suitable experimental design and methodology are planned, and whether the study possesses sufficient statistical power to detect a change. The NES ecosystem is a dynamic system under multiple pressures including climate change and fishing (Nye et al., 2009; Hare et al., 2016). Given underlying trends, the ability to distinguish offshore wind impacts from existing patterns of variability will require a robust understanding of baseline conditions. In order to evaluate inter-annual variability in the baseline state, at least 3-5 years of data would be needed which is greater than the 0-2 years proposed by most studies. Three to five years may still not be sufficient to adequately assess baseline temporal variability (e.g., Willsteed et al., 2018a; Willsteed et al., 2018b), but would provide a minimal view of the temporal sampling distribution. Post construction study duration is another key element of study design given that the longest running studies from European wind energy areas have reported new and significant effects more than 9 years since operations began (e.g., Buyse et al., 2022; Degraer et al., 2021). In the U.S., Wilber et al. (2022) reported an increase in structure oriented species such as Atlantic cod (*Gadus morhua*) and black sea bass following installation, but understanding how these effects change over time and whether these impacts convey population-level effects will require studies that span the lifetime for the wind project. Studies employing control sites are also vexed by the challenge of finding control sites that have sufficiently similar habitat conditions as the impact site but are outside the potential zone impacts of wind development which may extend 10s-100s of km (Popper and Hawkins, 2019; Christiansen et al., 2022; Daewel et al., 2022). Power analysis was a common tool employed in the study design to assess the number of samples needed to detect a change; typically power is only analyzed for measures of abundance so it remains uncertain whether the sample size chosen will be sufficient to detect changes in other metrics important in describing fisheries resource populations. Many of the proposed studies would require some modification to reduce the uncertainty in their ability to detect changes caused by wind development.

Toward mitigating the impacts of offshore wind development on long term scientific surveys, we make the following recommendations: 1) Advance the actions outlined in the BOEM and NOAA Survey Mitigation Strategy which includes evaluating how project-level monitoring may contribute to mitigating impacts to long term scientific surveys and developing regional monitoring standards (Hare et al., 2022); 2) Coordinate with scientists at NOAA Fisheries early on when developing monitoring plans to discuss methodologies and best practices; 3) For each study, clarify whether the study is designed to address project-level changes (local variation) and whether they are designed to collect functionally equivalent samples (abundance and biological) to ensure proper management of fishery resources at the population level; 4) Align sampling methodologies with studies within the region studying the same target taxa, species, or habitat (e.g., sampling gear; sampling protocols; handling of samples; measurements made, etc.); 5) Extend the temporal sampling frame to at least 3-5 years of baseline data; after construction, conduct sampling for the lifetime of the project; 6) Design experiments that can provide information that informs our understanding of how populations respond to habitat change caused by wind development and that can validate the assumptions of estimation models for biological indices for areas that will no longer be sampled by NOAA Fisheries surveys; and 7) Provide ready access to raw QA/QC'd data files. These recommendations are applicable in the U.S. and are also relevant in other countries where similar issues are occurring.

## Author contributions

AL and ETM contributed to the concept and layout of the paper. ETM wrote the first draft of the paper. AL, JMB, and ETM wrote text and reviewed for accuracy. All authors contributed to the article and approved the submitted version.

## Acknowledgments

We are grateful to our colleagues at the Northeast Fisheries Science Center (NEFSC) for many discussions on offshore wind topics. We thank Peter Chase and Jon Hare at NEFSC for providing constructive comments that improved this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Author disclaimer

The views expressed herein are those of the authors and do not necessarily reflect the views of the Department of Commerce or its sub-agencies.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1214949/full#supplementary-material

## References

ASMFC (2020) *2020 American shad benchmark stock assessment and peer review report* (Arlington, VA). Available at: https://asmfc.org/uploads/file/63d8437dAmShadBenchmarkStockAssessment_PeerReviewReport_2020_web.pdf (Accessed April 24, 2023).

Atlantic Shores (2021a). *Atlantic Shores benthic monitoring plan. Atlantic Shores Offshore Wind Construction and Operations Plan for Commercial Lease OCS (OCS-A 0499), Appendix II-H*, (Atlantic Shores Offshore Wind, Atlantic City, NJ). 29. Available at: https://www.boem.gov/renewable-energy/state-activities/atlantic-shores-offshore-wind-construction-and-operations-plan. (Accessed June 21, 2023).

Atlantic Shores (2021b). *Atlantic Shores fisheries monitoring plan. Atlantic Shores Offshore Wind Construction and Operations Plan for Commercial Lease OCS (OCS-A 0499), Appendix II-K*, (Atlantic Shores Offshore Wind, Atlantic City, NJ), 76. Available at: https://www.boem.gov/renewable-energy/state-activities/atlantic-shores-offshore-wind-construction-and-operations-plan. (Accessed on June 21, 2023).

Azarovitz, T. R. (1981). A brief historical review of the Woods Hole Laboratory trawl survey time series. *Canadian Special Publication in Fisheries and Aquatic Sciences* 58, 62–67.

Bacheler, N. M., and Ballenger, J. C. (2016). Spatial and temporal patterns of black sea bass sizes and catches in the Southeastern United States inferred from spatially explicit nonlinear models. *Mar. Coast. Fisheries: Dynamics Management Ecosystem Sci.* 7, 523–536. doi: 10.1080/19425120.2015.1095826

Bangley, C. W., Whoriskey, F. G., Young, J. M., and Ogburn, M. B. (2020). Networked animal telemetry in the Northwest Atlantic and Caribbean waters. *Mar. Coast. Fisheries* 12, 339–347. doi: 10.1002/mcf2.10128

Bethoney, N. D., and Stokesbury, K. D. (2018). Methods for image-based surveys of benthic macroinvertebrates and their habitat exemplified by the drop camera survey for the Atlantic Sea scallop. *J. Vis. Exp.* 137), e57493. doi: doi: 10.3791/57493

Bonzek, C. F., Gartland, J., Gauthier, D. J., and Latour, R. J. (2017). *Northeast area monitoring and assessment program (NEAMAP) 2016 data collection and analysis in support of single and multispecies stock assessments in the mid-Atlantic: northeast area monitoring and assessment program near shore trawl survey* (Williamsburg, VA: Virginia Institute of Marine Science, William & Mary). doi: 10.25773/7206-KM61

Bureau of Ocean Energy Management (2023)Fishery survey guidelines. Available at: https://www.boem.gov/sites/default/files/documents/about-boem/Fishery-Survey-Guidelines.pdf (Accessed April 16, 2023).

Buyse, J., Hostens, K., Degraer, S., and De Backer, A. (2022). Offshore wind farms affect the spatial distribution pattern of plaice pleuronectes platessa at both the turbine and wind farm scale. *ICES J. Mar. Sci.* 0, 1–10. doi: 10.1093/icesjms/fsac107

Christiansen, N., Daewel, U., Djath, B., and Schrum, C. (2022). Emergence of large-scale hydrodynamic structures due to atmospheric offshore wind farm wakes. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.818501

Coonamessett Farm (2023). Available at: https://www.coonamessettfarmfoundation.org/habcam-surveys (Accessed April 16, 2023).

CVOW (2022a). *Fishery characterization and resource monitoring for black Sea bass (Centropristis striata.) at the coastal Virginia offshore wind (CVOW) development site* (Williamsburg, VA: Virginia Institute of Marine Science), 13.

CVOW (2022b). *Fishery characterization and resource monitoring for whelk (Busycon spp.) at the coastal Virginia offshore wind (CVOW) development site* (Williamsburg, VA: Virginia Institute of Marine Science), 15.

Daewel, U., Akhtar, N., Christiansen, N., and Schrum, C. (2022). Offshore wind farms are projected to impact primary production and bottom water deoxygenation in the North Sea. *Commun. Earth Environ.* 3, 292. doi: 10.1038/s43247-022-00625-0

Degraer, S., Brabant, R., Rumes, B., and Vigin, L. (2021). "Environmental impacts of offshore wind farms in the Belgian part of the North Sea: attraction, avoidance, and habitat use at various spatial scales," in *Memoirs of the marine environment* (Brussels: Royal Belgian Institute of Natural Sciences, OD Natural Environment, Marine Ecology and Management), 104.

Empire Wind (2022). *Empire wind fisheries and benthic monitoring plan.* (Stamford, CT: Empire Wind, LLC), 108. Available at: https://www.empirewind.com/environment-and-sustainability/mariners-and-fisheries/. (Accessed June 21, 2023).

Hare, J. A., Blyth, B. J., Ford, K. H., Hooker, B. R., Jensen, B. M., Lipsky, A., et al. (2022) *NOAA Fisheries and BOEM federal survey mitigation implementation strategy -northeast U.S. region. NOAA technical memorandum 292* (Woods Hole, MA). Available at: https://www.fisheries.noaa.gov/resource/document/federal-survey-mitigation-strategy-northeast-us-region (Accessed April 16, 2023).

Hare, J. A., Morrison, W. E., Nelson, M. W., Stachura, M. M., Teeters, E. J., Griffis, R. B., et al. (2016). A vulnerability assessment of fish and invertebrates to climate change on the northeast U.S. continental shelf. *PloS One* 11 (2), e0146756. doi: 10.1371/journal.pone.0146756

ICES (2022). Working Group on Offshore Wind Development and Fisheries (WGOWDF). *ICES Scientific Reports. 4:98.* (Copenhagen: International Council for the Exploration of the Sea). 18. doi: 10.17895/ices.pub.21750458

ICES (2023). Workshop on a Research Roadmap for Offshore and Marine Renewable Energy (WKOMRE). *ICES Scientific Reports. 5:56.* (Copenhagen: International Council for the Exploration of the Sea), 32. doi: 10.17895/ices.pub.23097404

Lynch, P. D., Methot, R. D., and Link, J. S. (2018). *Implementing a next generation stock assessment enterprise. an update to the NOAA fisheries stock assessment improvement plan* (Silver Spring, MD: U.S. Dep. Commer., NOAA Tech), 127. doi: 10.7755/TMSPO.183

Macpherson, E., and Duarte, C. M. (1991). Bathymetric trends in demersal fish size: is there a general relationship? *Mar. Ecol. Prog. Ser.* 71, 103–112. doi: 10.3354/meps071103

MADMF (2018)Management objectives and research priorities for fisheries in the Massachusetts and Rhode Island-Massachusetts offshore wind energy area. Available at: https://www.mass.gov/doc/management-objectives-and-research-priorities-for-offshore-wind-and-fisheries/download (Accessed April 26, 2022).

Methratta, E. T. (2020). Monitoring fisheries resources at offshore wind farms: BACI vs. BAG designs. *ICES J. Mar. Sci.* 77, 890–900. doi: 10.1093/icesjms/fsaa026

Methratta, E. T., and Link, J. S. (2007). Ontogenetic variation in habitat associations for four flatfish species in the gulf of Maine-georges bank region. *J. fish Biol.* 70 (6), 1669–1688. doi: 10.1111/j.1095-8649.2007.01428.x

Miller, T. J., Das, C., Politis, P. J., Miller, A. S., Lucey, S. M., Legault, C. M., et al. (2010). *Estimation of albatross IV to Henry B. Bigelow calibration factors. northeast fish sci cent ref doc. 10-05* (166 Water Street, Woods Hole, MA: National Marine Fisheries Service), 233. Available at: http://www.nefsc.noaa.gov/nefsc/publications/.

New England Wind (2021a). *New England wind draft benthic habitat monitoring plan*, 25. (Monroe, MA: New England Wind Avangrid).

New England Wind (2021b). New England Wind fisheries monitoring plan. *Construction and Operations Plan Appendix III-U*. (Monroe, MA: New England Wind Avangrid), 42.

NOAA (2023) *Fisheries and offshore wind interactions: synthesis of science. NOAA technical memorandum NMFS-NE-291* (Woods Hole, MA). Available at: https://repository.library.noaa.gov/view/noaa/49151 (Accessed April 16, 2023).

Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the northeast U.S. continental shelf. *Mar. Ecol. Prog. Ser.* 393, 111–129. doi: 10.3354/meps08220

Ocean Wind (2021). *Ocean wind offshore wind farm fisheries monitoring plan*. (Boston, MA: Ocean Wind - Orsted), 76.

Ocean Wind (2022). *Ocean wind offshore wind farm benthic monitoring plan* (Boston, MA: Ocean Wind - Orsted), 76.

Pappal, A. L., Rountree, R. A., and MacDonald, D. G. (2012). Relationship between body size and habitat complexity preference in age-0 and -1 year winter flounder *Pseudopleuronectes americanus*. *J. Fish Biol.* 81, 220–229. doi: 10.1111/j.1095-8649.2012.03328.x

Popper, A. N., and Hawkins, A. D. (2019). An overview of fish bioacoustics and the impacts of anthropogenic sounds on fishes. *J. Fish Biol.* 94, 692–713. doi: 10.1111/jfb.13948

Reubens, J. T., Braeckman, U., Vanaverbeke, J., Van Colen, C., Degraer, S., and Vincx, M. (2013). Aggregation at windmill artificial reefs: CPUE of Atlantic cod (Gadus morhua) and pouting (*Trisopterus luscus*) at different habitats in the Belgian part of the North Sea. *Fisheries Res.* 139, 28–34. doi: 10.1016/j.fishres.2012.10.011

Revolution Wind (2021). *Revolution wind fisheries research and monitoring plan* (Providence, RI: Revolution Wind, LLC and Inspire Environmental), 157. Available at: https://www.boem.gov/renewable-energy/state-activities/revolution-wind-farm-construction-and-operations-plan. (Accessed on June 22, 2023).

Roach, M., Revill, A., and Johnson, M. J. (2022). Co-Existence in practice: a collaborative study of the effects of the westermost rough offshore wind development on the size distribution and catch rates of a commercially important lobster (*Homarus gammarus*) population. *ICES J. Mar. Sci.* 79, 1175–1186. doi: 10.1093/icesjms/fsac040

ROSA (2021) *ROSA offshore wind project monitoring framework and guidelines*. Available at: https://www.rosascience.org/resources (Accessed June 29, 2022).

South Fork Wind (2020). *South fork wind fisheries research and monitoring plan*. (Amagansett, NY: South Fork Wind, LLC), 123. Available at: https://www.boem.gov/renewable-energy/state-activities/south-fork. (Accessed on June 22, 2023).

Sunrise Wind (2021). *Sunrise wind fisheries and benthic research monitoring*. (New York, NY: Sunrise Wind, LLC), 131. Construction and Operation Plan Appendix AA1.

Swain, D. P. (1993). Age-and density-dependent bathymetric pattern of Atlantic cod (*Gadus morhua*) in the southern gulf of st. Lawrence. *Can. J. Fisheries Aquat. Sci.* 50 (6), 1255–1264. doi: 10.1139/f93-142

Thorson, J. T., Adams, G., and Holsman, K. (2019). Spatio-temporal models of intermediate complexity for ecosystem assessments: a new tool for spatial fisheries management. *Fish Fisheries* 20, 1083–1099. doi: 10.1111/faf.12398

Vineyard Wind (2023). *Wind fisheries studies and science*. Available at: https://www.vineyardwind.com/fisheries-science (Accessed April 29, 2023).

Wilber, D. H., Brown, L., Griffin, M., DeCelles, G. R., and Carey, D. A. (2022). Demersal and invertebrate catches relative to construction and operation of North America's first offshore wind farm. *ICES J. Mar. Sci.* 79, 1274–1288. doi: 10.1093/icesjms/fsac051

Willsteed, E. A., Birchenough, S. N. R., Gill, A. B., and Jude, S. (2018b). Structuring cumulative effects assessments to support regional and local marine management and planning obligations. *Mar. Policy* 98, 23–32. doi: 10.1016/j.marpol.2018.09.006

Willsteed, E. A., Jude, S., Gill, A. B., and Birchenough, S. N. R. (2018a). Obligations and aspirations: a critical evaluation of offshore wind farm cumulative impact assessments. *Renewable Sustain. Energy Rev.* 82, 2332–2345. doi: 10.1016/j.rser.2017.08.079

TYPE Original Research
PUBLISHED 12 July 2023
DOI 10.3389/fmars.2023.1217792



OPEN ACCESS

EDITED BY
Kyle Shertzer,
NOAA Southeast Fisheries Science Center,
United States

REVIEWED BY
Gary Nelson,
Massachusetts Division of Marine Fisheries,
United States
Peter Chase,
Northeast Fisheries Science Center
(NOAA), United States

*CORRESPONDENCE
Robert J. Latour
latour@vims.edu

RECEIVED 05 May 2023
ACCEPTED 28 June 2023
PUBLISHED 12 July 2023

CITATION
Latour RJ, Gartland J
and Bonzek CF (2023) Design
and redesign of a bottom trawl
survey in Chesapeake Bay, USA.
*Front. Mar. Sci.* 10:1217792.
doi: 10.3389/fmars.2023.1217792



# Design and redesign of a bottom trawl survey in Chesapeake Bay, USA

Robert J. Latour*, James Gartland and Christopher F. Bonzek

Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, United States

Fisheries-independent surveys that reliably sample a broad size range of exploited and ecologically important species provide valuable data in support of fisheries management and ecosystem science. The operational consistency of surveys over time and space is fundamental to the interpretation of data in the contexts of population dynamics processes, community interactions, policy impacts, and environmental forcing. However, the need to maintain historic sampling protocols over extended time periods limits the utilization of new technologies that could lead to improved data collection. Survey vessel replacements also become inevitable as the maturity of sampling programs becomes multidecadal. This case study describes the motivational origin, initial design, and redesign of a bottom trawl survey operating in Chesapeake Bay, the largest estuary in the United States. Regional aspirations to consider ecosystem principles in fisheries management aided initial development of the survey, and the need to collect specific data types to support that endeavor impacted several early design elements. Following the beginning years of full-scale survey operations, a consistently evolving awareness of potential areas of improvement for the survey grew from formal efforts to engage with scientific and industry partners on trawl gear design, leverage the program for additional survey opportunities, utilize gear testing technology, and analyze extant data. When the delivery of a new, state-of-the-art research vessel forced the transfer of survey operations to a new platform, all potential changes were incorporated simultaneously. A subsequent paired-tow experiment was conducted to build a calibration database that successfully provided estimates of relative selectivity for routinely sampled taxa. This experience yielded several lessons learned that are intended to aid investigators faced with adopting structural changes to fisheries-independent surveys in the future.

KEYWORDS

**trawl survey, Chesapeake Bay, intercalibration, log-Gaussian Cox process, lessons learned**

# 1 Introduction

Fish and aquatic invertebrate populations are routinely surveyed using a variety of sampling gears such as trawls, dredges, gillnets, longlines, traps, seines, hydroacoustics, and video photography (Kimura and Somerton, 2006). Regardless of the gear type used, the primary purpose of a fisheries-independent survey is to obtain representative data that allow estimation of key population quantities. In the context of stock assessment, survey data are typically analyzed to estimate indices of relative abundance over a defined spatiotemporal scale. Depending on the survey sampling design, indices can be estimated using design-based methods based on classic sampling theory (e.g., stratified random sampling, Cochran, 1977; Thompson, 2012) or model-based procedures (e.g., generalized linear models and their extensions, Maunder and Punt, 2004; Venables and Dichmont, 2004). These indices are then used as inputs to a stock assessment model under the assumption that the temporal pattern in the indices reflects that of the overall population. In this respect, survey data can be considered a central component of any fisheries management system (Hilborn and Walters, 1992). However, when survey gear effectively samples species that are not exploited, assessed, or managed, the resulting data can form the basis of valuable biological, ecological, and community analyses. These analyses are usually model-based since they often relate survey catches to synoptically measured abiotic and biotic covariates under the broader theme of ecosystem science. These studies may also be structured to help inform policies that fall along the ecosystem-approaches to ecosystem-based fisheries management gradient (Link, 2010; Link and Marshak, 2022).

Irrespective of the objectives motivating analyses of survey data, an assumption typically required for inference is that survey catches are proportional to total population abundance. This concept is formalized as $C = qEN$, where $C$ is survey catch, $E$ is survey effort, $N$ is total population abundance, and $q$ is the catchability coefficient defined as the fraction of the population captured with one unit of effort (Ricker, 1975). Re-arranging yields $C/E = qN$, which illustrates the proportional linkage between catch-per-unit-effort (*CPUE*) and population abundance. Explicit to the expression for *CPUE* is the notion that $q$ remains constant over the spatiotemporal domain of sampling. To minimize variability in $q$, a high emphasis is placed on maintaining consistency in field protocols across the life of a survey. Gear configuration parameters, deployment procedures, survey vessel in the case of towed gear, and calendar dates of survey expeditions are intentionally held constant over time to avoid influencing $q$. However, despite the most well executed efforts to ensure operational consistency, it is recognized that $q$ can vary temporally due to anthropogenic, environmental, biological, and management processes (Wilberg et al., 2010), or spatially because of heterogeneity in bottom substrate within the sampling domain where towed demersal gear is deployed (Thorson et al., 2013). Examining the constant $q$ assumption should therefore be a continual process throughout the life of a fisheries-independent survey, with information coming from analyses of extant data combined with specific process-oriented field studies designed to investigate factors hypothesized to affect catchability.

The operational consistency of a survey to meet the constant $q$ assumption can be both an asset and a liability. Data collected in the same manner over a defined sampling frame and in accordance with a valid statistical sampling design is arguably the most valuable aspect of a fisheries-independent survey. The accumulated data streams can provide insight into the synergist effects of population dynamics processes, community interactions, fisheries management impacts (for exploited resources), and environmental forcing over short, medium, and long time periods. However, as a survey matures and its longevity becomes multidecadal, the need to maintain historic sampling protocols limits the utilization of new technologies that could lead to improved data collection. Moreover, for surveys with towed gear, vessel refits and eventual replacements can present challenges to maintaining the integrity of data streams since towed gear performance is often tied to specific design and mechanical characteristics of the survey vessel.

The Chesapeake Bay Multispecies Monitoring and Assessment Program (ChesMMAP) is a relatively long-term (2002 – present) bottom trawl survey designed to provide species-specific, community-level, and trophic interactions data for late juvenile/adult fishes and shellfish in Chesapeake Bay (Latour et al., 2003). Recently, the institution responsible for conducting the ChesMMAP survey took delivery of a newly constructed, state-of-the-art research vessel that possesses significantly more capabilities for conducting fisheries-independent surveys when compared to the original vessel. To take advantage of this modern research platform, the ChesMMAP survey was fully revamped, including modifications to the sampling design, gear package, field deployment protocols, and operations were shifted to the new research vessel. This case study describes the inspiration for ChesMMAP, the rationale and process by which the survey was redesigned, field and analytical efforts to maintain interpretability of data streams given the new sampling platform and gear, and lessons learned along the way.

# 2 Inspiration and design of ChesMMAP

## 2.1 The Chesapeake Bay

The Chesapeake Bay is a partially mixed coastal plain estuary located on the U.S. east coast. The bay's watershed covers an expansive area (164,200 km$^2$) and mean depth is relatively shallow (6.5 m, Kemp et al., 2005). Estuarine circulation is driven by freshwater inputs mainly from northern and western tributaries combined with landward-flowing sea water from the Atlantic. Water temperatures in the bay are dynamic intra-annually and can range from as low as 1-4°C in winter (Dec-Mar) to as high as 28-30°C in summer (Jun-Sep). As a result, the bay serves as an important foraging and refuge area for diverse assemblages of both resident taxa and seasonally occurring boreal, temperate, and subtropical fishes (Murdy et al., 1997). Many of those species support economically valuable commercial and recreational fisheries, as well as an array of non-market ecosystem services (Kirkley et al., 2005; Lellis-Dibble et al., 2008; National Marine Fisheries Service [NMFS], 2020).

While the bay remains a highly productive ecosystem, it has experienced significant anthropogenic change since the late 19th century. Eutrophication resulting from increased nutrient inputs has affected water quality, the distribution and density of submerged aquatic vegetation (Orth et al., 2010), hypoxic events (Hagy et al., 2004), and the relative roles of benthic and planktonic processes underlying ecosystem functioning (Kemp et al., 2005). Fishing activities have also had major effects on both resident and seasonally available natural resources in the bay, including cases of stock collapse (Richards and Rago, 1999; Wilberg et al., 2011). Climate change has impacted the bay ecosystem through warming (Ding and Elmore, 2015; Hinson et al., 2022), altered timing of spring phenological events (Thomas et al., 2017), spatiotemporal extent of hypoxic volume (Irby et al., 2018; Tian et al., 2022), and relative habitat utilization among the bay and coastal areas by several taxa (Schonfeld et al., 2022). Additional climate change related effects on the physical, chemical, and biological processes of the bay are expected in the future (Najjar et al., 2010).

## 2.2 Fisheries management and surveys in Chesapeake Bay

Management of fisheries resources important to the Chesapeake Bay region is achieved through a complex jurisdictional framework. State agencies in Pennsylvania, Maryland, and Virginia along with their counterpart in the District of Columbia, and the Potomac River Fisheries Commission (a Maryland-Virginia bi-state agency) each have regulatory authority over fisheries targeting year-round resident species within their respective boundaries. Coastal species that are seasonal bay residents but also inhabit nearshore areas in the Atlantic extending across state boundaries (0-3 nm offshore) are managed by the Atlantic States Marine Fisheries Commission (ASMFC). The home ranges of some ASMFC managed species encompass parts of the exclusive economic zone (EEZ; 3-200 nm offshore) and are co-managed with regional fishery management councils that have federal authority through the Magnuson-Stevens Fishery Conservation and Management Act (Methot et al., 2014). Stock assessments are conducted by various academic and governmental agencies, but regardless of those responsible for the analyses, most assessments incorporate fisheries-independent survey data from the bay.

Surveys of fisheries resources in the Chesapeake Bay have been operating for many decades. The earliest began in 1939 and targeted the eastern oyster (*Crassostrea virginica*; Wilberg et al., 2011). Finfish surveys were initiated in the 1950s and designed to sample juvenile fishes given the importance of the bay as a nursery area for many mid-Atlantic species. Over time, several additional surveys targeting juvenile and adult life stages of diadromous fishes, bivalves, and crustaceans were developed largely in response to emerging management needs. While many of these surveys are longstanding and provide valuable information, the jurisdictional boundary between Maryland and Virginia has historically hindered efforts to develop comprehensive, bay-wide sampling programs.

## 2.3 Ecosystem principles in fisheries management in Chesapeake Bay

During the late 1990s and early 2000s, significant attention was focused on considering ecosystem principles in U.S. fisheries management both federally and across many local sectors. At the national level, the NMFS Ecosystem Principles Advisory Panel (NMFS Panel) produced a report outlining recommendations for implementing ecosystem philosophies, goals, and policies in U.S. fisheries conservation, management, and research (National Marine Fisheries Service [NMFS], 1998). Within the Chesapeake Bay region, similar technical documents were developed that summarized perspectives and rationale for incorporating multispecies and ecosystem considerations into fisheries management (Miller et al., 1996; Fernandez and Leach, 1998). In response to key recommendations from the NMFS Panel, a comprehensive prototype fisheries ecosystem plan (FEP) was developed to provide strategic guidance for ecosystem-based fisheries management in Chesapeake Bay and information on the function and structure of the bay ecosystem (National Oceanic and Atmospheric Administration Chesapeake Bay Fisheries Ecosystem Advisory Panel [NOAA CBFEAP], 2006).

Scientific products that support ecosystem principles require additional data types when compared to those needed for traditional stock assessments. The supporting technical documents and FEP highlighted key data gaps, despite the region having several long-term surveys and a rich understanding of the physical, chemical, and biological processes of the bay. Most notable were data types necessary to develop multispecies and ecosystem models, namely bay-wide information on species abundances, age/size composition, growth and mortality rates, and trophic interactions. The need for these data types along with the regional interest in ecosystem management inspired the design and implementation of ChesMMAP.

## 2.4 ChesMMAP design and sampling protocols

Conceptualization of ChesMMAP began in 2001 with a review of several existing fisheries-independent sampling programs, including fish trawl surveys conducted by ICES in the North and Baltic Seas, the Northeast Fisheries Science Center (NEFSC) in the northwest Atlantic, and the Alaska Fisheries Science Center in the Gulf of Alaska and Bering Sea (Latour et al., 2003). General consideration was given to vessel and trawl gear specifications, temporal and spatial sampling frequency (acknowledging that Chesapeake Bay is geographically much smaller), data types collected, and onboard data collection processes. The intent behind gathering this information was to build familiarity with other successful programs with similar scientific objectives, and to begin shaping design elements for ChesMMAP.

Initial tactical decisions focused on five key areas: identifying a survey vessel, choosing the trawl gear package, sampling design, onboard catch processing logistics, and staffing. All these areas were evaluated with respect to the financial resources available for operations, and choices often reflected tradeoffs among what was considered ideal versus what was practical. After reviewing the specifications of several vessels in the Chesapeake Bay region with research vessel (R/V) designations (length overall, operational parameters, propulsion, electrical service, working deck space, wet and dry lab space, and berthing), the R/V *Bay Eagle* was selected as the survey platform. None of the available research vessels were ideal platforms for conducting trawl operations aimed at sampling larger, more mobile fishes and invertebrates, however, among those in the area, the R/V *Bay Eagle* was the only one that satisfied the very minimum specifications necessary for the survey. Owned and operated by William & Mary's Virginia Institute of Marine Science (VIMS), this vessel is a 19.8 m crewboat with a 400 nm range and 3-4 day endurance that was retrofitted to conduct scientific research. Beam size and available aft deck space did create an upper limit on trawl net size and required a single towing warp deployment with a bridle.

To guide selection of the gear package, information was gathered on the combinations of vessel sizes and gear specifications used by other trawl surveys operating in the region, as well as from those used by the commercial shrimp fishery in the U.S. southeast. Following consideration of several options, the chosen gear package included a 13.7 m four seam bottom trawl with 15.2 cm stretch body mesh and 7.6 cm stretch cod end mesh constructed from twisted nylon twine. The net was equipped with a looped chain sweep for simplicity and to aid adherence to the bottom, and hydroacoustic wing and headrope sensors to provide data for area/volume swept estimation. Accompanying the net was a pair of standard 1 $m^2$ steel vee-doors designed to achieve spreading primarily through ground sheer. The larger body and cod end stretch meshes were intended to mitigate the pressure wave created at faster tow speeds thereby increasing the capture probability of larger, more mobile animals.

The need for a bay-wide survey led to defining the sampling frame for ChesMMAP as the bay mainstem in both Maryland and Virginia (3900 $km^2$ survey area). Although the bay's major tributaries represent important habitat for fishes and invertebrates, cost analyses associated with sampling a spatial area larger than the mainstem against those of increased frequency of cruises per year favored the latter, particularly because of the seasonally dynamic nature of the Chesapeake Bay fish community and the goal to collect information on as many species as possible. Accordingly, survey cruises occurred bimonthly from March to November, and sampling followed a random stratified design with stratification based on region (five 30-minute latitudinal strata) and depth (three strata: 3.0 – 9.1, 9.1 – 15.2, and >15.2 m; Figure 1A). Allocation of sampling effort was proportional to stratum surface area, and 20-min tows were made with the current (initial gear testing revealed the net frequently lost bottom contact when towing against the current, and vessel speed was adjusted when towing with the current to maintain optimal gear geometry based on net mensuration measurements). Wingspread and headrope height were combined with the vessel GPS track to calculate swept area/volume. During the first few years of the survey, a full cruise consisted of 90 sampling sites, however, that target was reduced to 80 based on analyses that indicated such a reduction did not lead to significant losses in precision of estimated relative abundances.

A central philosophy of ChesMMAP is to maximize the data collected at each sampling site. Therefore, each trawl catch is sorted and measured for aggregate weight, count, and individual specimen lengths by species or size-class if distinct classes within a particular species are evident. A subsample of each fish species (excluding bay anchovy, *Anchoa mitchilli*, and striped anchovy, *A. hepsetus*) or size-class is further processed for weight, sex, macroscopic maturity stage, and material for aging and diet composition analysis is preserved and returned to the laboratory for processing. For a few species, additional material is preserved for disease and reproductive biology analysis. More recently, sampling of the benthos and zooplankton community has been added to expand the dimensions of the bay ecosystem for which information is collected. Opportunistic sampling, in the form of additional trawl hauls, biological sample acquisition, and data collection, has been conducted as needed throughout the survey history to support studies conducted by researchers, and particularly students, both within and external to VIMS.

## 3 Motivation for survey changes

As noted above, while institutions typically make every effort to minimize spatiotemporal variation in survey catchability through standardization of sampling protocols, there are times when large changes to survey procedures are unavoidable. The Northeast Fisheries Science Center (NEFSC) Bottom Trawl Survey took delivery of a new survey vessel in 2007 and used the opportunity to implement several new technologies meant to enhance sampling consistency, including improved survey trawl gear (Miller, 2013). Specifically, the NEFSC convened a panel of commercial fishers, trawl manufacturers, and fisheries scientists in 2003 to develop a fishing system designed to maintain a more consistent trawl geometry (i.e., headline height and wingspread), sample a variety of fishes and invertebrates across a broad size range, and be of an appropriate scale for the new research vessel (Johnson and McCay, 2012). The final design was a three-bridle, four-seam bottom trawl that measured 23.3 m along the headline with a 48 m circumference fishing circle. The body of the net was comprised of both 6 cm and 4 cm stretch-mesh polyethylene webbing with a 2.54 cm knotless nylon lined codend and a sweep made of 40.6 cm rubber disks.

At the same time, the ASMFC had partnered with VIMS and several state agencies to develop the Northeast Area Monitoring and Assessment Program (NEAMAP) Inshore Trawl Survey, which was intended to sample the coastal ocean of the Mid-Atlantic Bight given that the new NEFSC vessel would no longer be able to conduct operations in these shallow environments. NEAMAP chose to adopt this three-bridle, four-seam bottom trawl to maintain consistency with the redesigned NEFSC survey, although the sweep was comprised of smaller, 7.6 cm rubber disks since the seafloor in the NEAMAP sampling frame has very few naturally occurring obstructions.

VIMS began sampling with this trawl on NEAMAP in 2007 (Gartland et al., 2023), while the NEFSC Bottom Trawl Survey

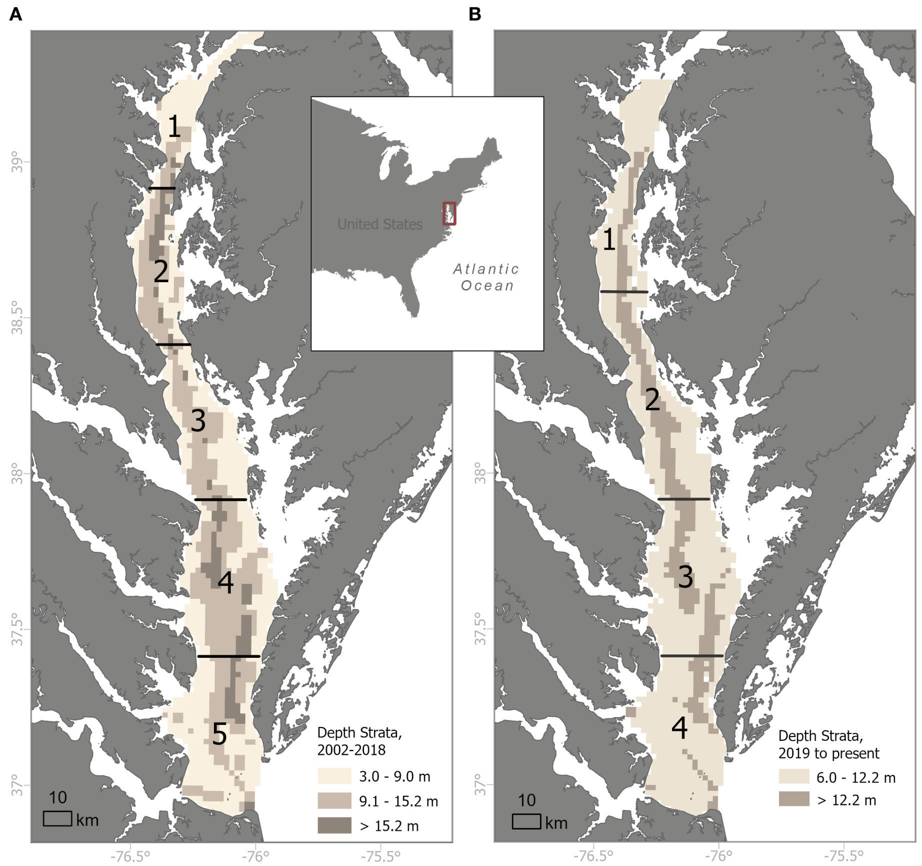



FIGURE 1
Sampling frames and stratification schemes for the Chesapeake Bay Multispecies Monitoring and Assessment Program (ChesMMAP) bottom trawl survey **(A)** historically, 2002-2018 and **(B)** currently, 2019-present. Numbers denote regional strata separated by horizontal lines and the shaded bathymetry shows depth zones.

formally transitioned to this fishing system in 2009 (Miller, 2013). Given the remarkable consistency of the trawl geometry and the diversity and quantity of resulting catch recorded by both surveys, in 2009 ChesMMAP personnel contacted the aforementioned trawl manufacturers and a subset of the commercial fishers referenced above to inquire as to whether a smaller version of this fishing system could be developed for sampling in Chesapeake Bay. The result was a trawl that was identical in design to those used by NEAMAP and the NEFSC, but that measured 11.2 m along the headline with a 24 m circumference fishing circle and a sweep made of 3.8 cm rubber disks, and thus was effectively half of the size.

Prior to conducting field trials with this new trawl net, VIMS commissioned the construction of 1:6 scale models of both the original ChesMMAP trawl and the new trawl, and these model nets were subjected to flume tank trials at Memorial University in St. John's, NL (Figures 2A, B). Given the high costs typically associated with vessel time, flume testing of survey trawls represents a cost-effective approach to evaluating candidate gears prior to conducting sea-trails (Winger et al., 2006). Relative to the model of the original ChesMMAP trawl, the new net model maintained a more stable geometry and higher headline height over a wider range of simulated current speeds, and it also experienced more consistent water pressures across the net body (taught webbing throughout). When combined with the smaller mesh size, these results implied catchability should be higher and more consistent (Winger et al., 2010).

The new net was coupled with a set of 0.88 $m^2$ high-efficiency, cambered trawl doors and limited field-trials were conducted during 2010 and 2011 on the R/V *Bay Eagle*. Members of the commercial fishing industry were instrumental in the execution of these early trials, as they provided valuable advice on the appropriate rigging and deployment of this more complex trawl, and assisted survey personnel with identifying a trawl door configuration that would consistently yield optimal net geometry. Measurements from net mensuration gear during these sea-trials confirmed that headline height and wingspread values were half of those observed for the net used by NEAMAP and the NEFSC, and this new trawl appeared to collect a greater diversity of taxa, a broader size range of animals, and a much larger quantity of catch relative to the original ChesMMAP trawl net (unpubl. data). Although this new trawl appeared to yield a more consistent and robust sampling of the ecological community inhabiting Chesapeake Bay, transitioning the survey to this new fishing system on the R/V *Bay Eagle* would have incurred relatively large

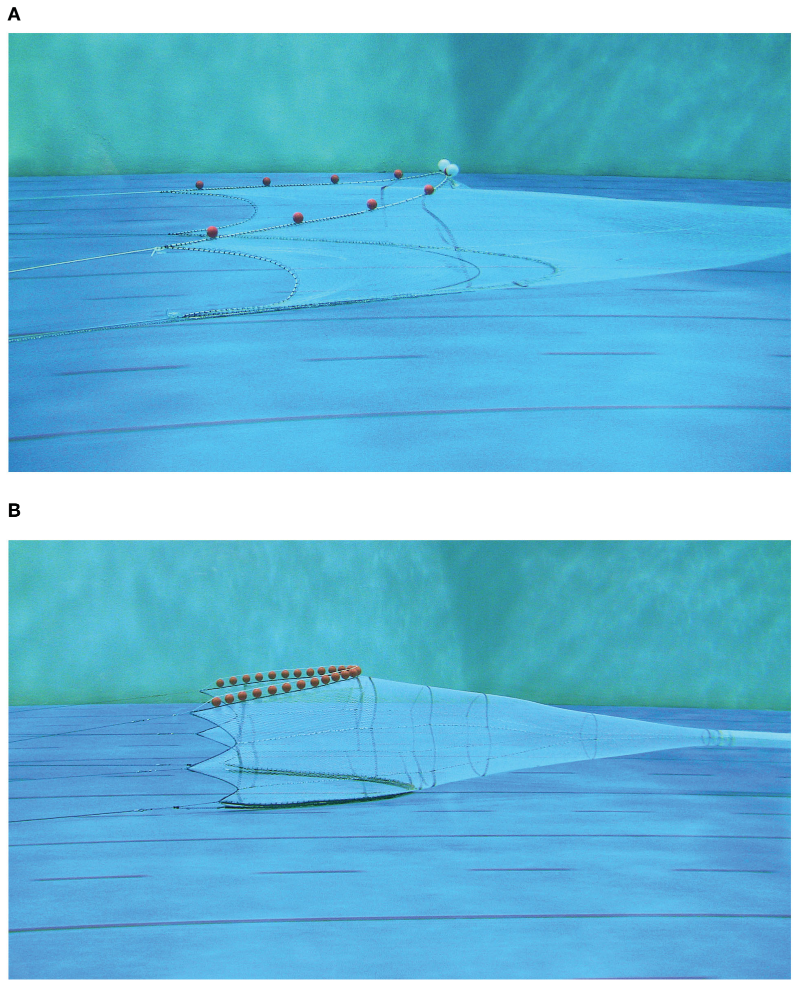


FIGURE 2
Photos of 1:6 scale models of the **(A)** original ChesMMAP survey trawl net (13.7 m four seam bottom trawl with 15.2 cm stretch body mesh and 7.6 cm stretch cod end mesh) and **(B)** the new ChesMMAP survey trawl (three-bridle, four-seam bottom trawl that measures 11.2 m along the headline with a 24 m circumference fishing circle and a 3.8 cm flat sweep) in the flume tank Memorial University in St. John's, NL. December 2009.

costs and posed several logistical challenges. The preferred deployment method for this trawl calls for a dual-warp design, which would have required the acquisition and installation of an additional winch on the vessel, as well as extensive structural modifications to the sampling platform. Given the expected increase in the diversity and quantity of catch, it would also have been necessary to construct and install at least two additional data collection workstations on the aft deck. Taken together, these two modifications would have been very difficult to accommodate, given the already limited deck space. Further, efforts to generate calibration coefficients between the original and new trawls would necessarily have needed to follow a single-vessel, paired-tow design, and the R/V *Bay Eagle* simply did not have the capacity to accommodate both fishing systems onboard simultaneously, let alone to rapidly switch between the two as would be required during a calibration experiment. Fortunately, VIMS acquired funds necessary to construct a new, state-of-the-art research vessel shortly thereafter, and thus implementation of this new trawl was suspended until vessel delivery so that all major changes to the survey could be adopted concurrently.

## 4 Redesigning ChesMMAP

The new VIMS vessel, the R/V *Virginia*, was delivered in October 2018 and is a 28.3 m ship that has a 1500 nm range, 10 day endurance, and was designed and equipped to support a myriad of research activities in Chesapeake Bay and along the U.S. east coast. ChesMMAP sampling operations were transferred to this platform along with the new dual-warp fishing system as soon as it became available for charter in the early summer of 2019. While some sampling programs have decided to defer the implementation of survey changes until after paired-tow experiments were completed and calibration coefficients for key species were estimated (e.g., Miller, 2013), it was decided to make all ChesMMAP changes immediately while concurrently initiating a paired-tow calibration experiment. By implementing changes in this way, the benefits of the improved trawl net (i.e., increases in faunal diversity and size ranges of catch) were able to be realized as soon as was practicable. A drawback, however, was that unanticipated but extended setbacks in completing the calibration experiment (see Section 4.1) prevented the release of data for stock assessments given the need for calibration coefficients to link the time-series collected pre- and post-2019.

Since 2019, trawl sampling in the mainstem of Chesapeake Bay (ChesMMAP) and across the continental shelf in the Mid-Atlantic and New England (NEAMAP, NEFSC) has occurred using a consistent net design. The duration of a standard trawl haul and the target vessel speed over ground during a ChesMMAP tow remained unchanged from the original survey design. The safe operating depth of the R/V *Virginia* is approximately 6 m, however, meaning that it was no longer possible to sample the shallower areas of the original ChesMMAP sampling frame. In response, this logistical challenge was used as an opportunity to redefine and subsequently re-stratify the ChesMMAP survey area. Specifically, analyses (design- and model-based estimation of relative abundance indices and associated uncertainties) of post-stratified extant catch data were conducted to identify four new latitudinal strata that are each subdivided into two depth strata (6.0 – 12.2, > 12.2 m; Figure 1B).

Given the increased daily charter costs associated with the R/V *Virginia*, maintaining the original sampling frequency was not possible. Therefore, analyses of extant data were conducted to examine the spatiotemporal distribution of migratory taxa, and results were compared against available financial resources to evaluate tradeoffs in the annual allocation of sampling effort seasonally and spatially. Efforts were also directed at understanding potential impacts of sampling effort changes on the data streams routinely supplied to stock assessments. Four ChesMMAP survey cruises now occur annually during the months of March, June, September, and November. Sampling during June and September is conducted throughout the mainstem of the bay. March cruises occur in the two northernmost latitudinal strata to sample key anadromous fishes during their spring spawning migrations, while November cruises are limited to the two southernmost strata, as various taxa congregate in the lower bay prior to their migration to overwintering habitats on the continental shelf.

## 4.1 Paired-tow sampling experiment

Data used to support the estimation of calibration coefficients for fishes and invertebrates routinely sampled by ChesMMAP were collected through a series of 15 research cruises (hereafter, calibration cruises) conducted from June 2019 – November 2022. Each calibration cruise occurred after the completion of a survey sampling event, and the sites sampled were those associated with the most recent survey. Thus, the data used to estimate calibration coefficients for the various taxa were collected following a stratified random sampling design. Approximately five to seven days elapsed between the sampling of a site during a survey and a subsequent calibration cruise to minimize any disturbance effects (e.g., Lewy et al., 2004). While a more-costly approach, this temporal separation of survey and calibration cruises ensured that the survey data would not be impacted by the presence of a second vessel sampling in close proximity (Brown et al., 2007).

At a given sampling site, both the R/V *Virginia* and R/V *Bay Eagle* conducted a trawl haul concurrently (i.e., a paired-tow). Both vessels towed in the same direction and were separated by approximately 350 m. Sampling on the R/V *Virginia* occurred using the new fishing system while the original gear was deployed from the R/V *Bay Eagle*. Side-by-side positions of the vessels were randomized and a total of 516 paired-tows were completed. Date and time of sampling was denoted at the outset of each paired-tow, and position (latitude and longitude) was recorded by each vessel throughout the tow. The headline height and wingspread of each trawl were recorded during each tow, and wingspread data were coupled with tow distance to calculate the area swept by each trawl at a given site. For each vessel, resulting catches were sorted by species, and aggregate weight, count, and individual length measurements were recorded for each. Over the course of this field experiment, a total of 97 fishes and 20 invertebrate taxa were collected, where 24 fishes and six invertebrate species were unique to the R/V *Virginia* utilizing the new fishing system and seven fishes and zero invertebrates were unique to the R/V *Bay Eagle*. Three notable events delayed the completion of the field sampling for this calibration experiment by almost a year; two separate, major mechanical failures on the R/V *Bay Eagle* resulted in the loss of approximately seven months of sampling, and the COVID-19 pandemic led to the suspension of field operations for nearly four months.

## 4.2 Statistical framework

Intercalibration of the two vessel-trawl combinations was based on applying log-Gaussian Cox processes to the paired-tow data (following Thygesen et al., 2019). This approach models the size distribution of the population at each sampling site and the size-structured clustering of animals at small temporal and spatial scales to estimate selectivity ratios across the domain of observed size classes. By utilizing a Poisson probability distribution for the catch numbers conditional on latent log-Gaussian variables, the method allows for overdispersion and correlation between catch counts in neighboring size classes. The model structure is as follows:

$$N_{ijk} \mid \Phi, R, S \sim Poisson(A_{ij} \cdot \exp(S_{jk} + \Phi_{ik} + R_{ijk})) \quad ,(1)$$

where for site $i = 1, \ldots, n_s$, gear $j = 1, 2$, and size class $k = 1, \ldots, n_l$, $N_{ijk}$ is the number of individuals captured, $A_{ij}$ is the area swept, $S_{jk}$ is the relative size selectivity (on log scale) such that $S_{1k} = -S_{2k}$, $\Phi_{ik}$ is the log-density that characterizes the population size distribution encountered by both gears, and $R_{ijk}$ is the variability in the size composition at small temporal and spatial scales (independent components unique to each gear). The quantities $S_{jk}$, $\Phi_{ik}$, and $R_{ijk}$ are random variables such that $S_{1k}$ and $\Phi_{ik}$ are modeled as random walks over size classes and $R_{ijk}$ is the sum of a white noise (WN) process and a zero-mean first-order autoregressive (AR) process, $R_{ijk} = R_{ijk}^{WN} + R_{ijk}^{AR}$ (see Thygesen et al., 2019 for details).

The model estimates the fixed effect parameters $\sigma_S^2$, $\sigma_\Phi^2$, $\sigma_{WN}^2$, $\sigma_{AR}^2$, and $\rho$ (correlation coefficient associated with the AR process) along with a large number of random effects: $S$, $\Phi$, and $R$ have $n_l$,

$n_s n_l$, and $n_s 2n_l$ parameters, respectively. Models were fitted using the R package gearcalib (available at github.com/Uffe-H-Thygesen/Intercalibration). This package applies the Laplace approximation to integrate out the unobserved random effects $S$, $\Phi$, and $R$ thus yielding a likelihood function defined by the fixed effect parameters. After the likelihood is maximized and the fixed effects are estimated, the posterior modes of the random effects $S$, $\Phi$, and $R$ are reported, where those associated with $S$ are of primary interest. Optimization of the likelihood and use of the Laplace approximation was accomplished with the Template Model Builder (TMB) package (Kristensen et al., 2016).

## 4.3 Model application

For illustrative purposes, the log-Gaussian Cox processes model was applied to the paired-tow data of four species: Atlantic croaker (*Micropogonias undulatus*), striped bass (*Morone saxatilis*), summer flounder (*Paralichthys dentatus*), and female adult blue crab (*Callinectes sapidus*). These species were chosen because they support valuable fisheries, differ morphologically, and have contrasting habitat characteristics. Prior to modeling, the paired-tow data were filtered to remove samples collected from months and strata that consistently yielded near zero catches by both vessels owing to the notion that not all species are available for sampling during all months of the year or abundantly distributed in all strata. Summaries of the data analyzed indicated that the R/V *Virginia* captured considerably more total animals (except for female adult blue crab likely due to differing trawl sweep configurations between the gears), wider size ranges (except for striped bass due to a few very large animals collected by the R/V *Bay Eagle*), and animals more frequently as evidenced by consistently higher proportion positive tows (probability at least one animal is sampled; Table 1; Figures 3A–H).

Application of the log-Gaussian Cox processes models was generally successful with the caveat that the white noise component of the residuals for all species could not be identified, and as a result, estimates of $\sigma^2_{WN}$ approached zero ($10^{-5}$ order of magnitude). This situation was frequently encountered in a simulation study conducted to verify the model (Thygesen et al., 2019) and is indicative of the broader challenge of estimating variance components in hierarchical models (Auger-Méthé et al., 2016). Within the modeling framework, the random variable $R$ is intended to represent small-scale fluctuations in local abundance. Since the paired tows occur at slightly different locations, it is possible for one gear to encounter or miss an aggregation of animals within the overall sampling space of the two gears. Depending on the information content of the paired-tow data, $R$ could also represent random fluctuations in gear selectivity, which creates a situation where the two effects are confounded. In this case, a high catch in one gear could be the result of encountering an aggregation or because it performed better than average at the sampling site (Thygesen et al., 2019).

This confounding appears to be present in the paired-tow data analyzed herein. On a tow-by-tow basis for the species considered (and several others analyzed but not presented), the gear on the R/V *Virginia* consistently met or outperformed that of the R/V *Bay Eagle* in terms of encounter rates and total catch (Table 1). Thus, fluctuations in local abundance associated with encountering or missing animal aggregations were not distinguishable from the comparative superiority of the R/V *Virginia* fishing system, and consequently $\sigma^2_{WN}$ was not estimable. To further explore this concept, a small simulation study was conducted where the species-specific catches of the two vessels were randomly interchanged and then analyzed with the fully saturated log-Gaussian Cox processes model. For over half of the randomly modified data sets, the $\sigma^2_{WN}$ parameter was estimated well. Therefore, only reduced models that excluded the $\sigma^2_{WN}$ fixed effect were considered for analysis, and all fixed effects parameters associated with the reduced models were generally well estimated (Table 2).

The resultant estimated relative size selectivity curves confirmed trends in the raw data in that most of the estimates exceeded 1.0 for each of the four species (note, the magnitude of the estimates was quite large for Atlantic croaker and summer flounder, Figures 4A–D). While the calibration experiment and data analyses appear to be successful, questions remain about how best to treat the ChesMMAP survey data moving forward, particularly for stock assessments. Developing time-series of indices that span the vessel changes could be accomplished by converting R/V *Bay Eagle* survey data into R/V *Virginia* units using the above relative size selectivity estimates, however, this approach is not without drawbacks. First, since conversion of survey data requires multiplying the size-specific catches by the relative size selectivity estimates, the issue of how to convert the historic R/V *Bay Eagle* zero observations emerges. Due to the superiority of the new trawl net, numerous

TABLE 1 Sampling and catch summaries of the paired-tow data for the four selected species. Italicized subscripts denote the R/V *Bay Eagle* (*BE*) and R/V *Virginia* (*VA*).

| Species | Sampling mos. | No. of tows | $N_{BE}$(count) | $N_{VA}$(count) | $Size_{BE}$(cm) | $Size_{VA}$(cm) | $P_{VA \geq BE}$ |
|---|---|---|---|---|---|---|---|
| Atlantic croaker | Jun, Jul, Aug, Sep, Nov | 413 | 2,445 | 47,316 | 16.3 ± 2.4 | 15.6 ± 1.8 | 0.99 |
| Striped bass | Mar, Nov, Dec | 170 | 1,911 | 4,870 | 33.1 ± 8.8 | 25.9 ± 8.1 | 0.85 |
| Summer flounder | Jun, Jul, Aug, Sep, Nov | 413 | 147 | 873 | 25.6 ± 9.4 | 21.3 ± 7.4 | 0.92 |
| Adult female blue crab | Jun, Jul, Aug, Sep, Nov | 413 | 3,604 | 3,053 | 14.7 ± 1.6 | 14.6 ± 1.5 | 0.73 |

Mean size (± SD) is total length for Atlantic croaker and summer flounder, fork length for striped bass, and carapace width for adult female blue crab. $P_{VA \geq BE}$ denotes the proportion of tows when the total catch of the R/V *Virginia* equaled or exceeded that of the R/V *Bay Eagle*.

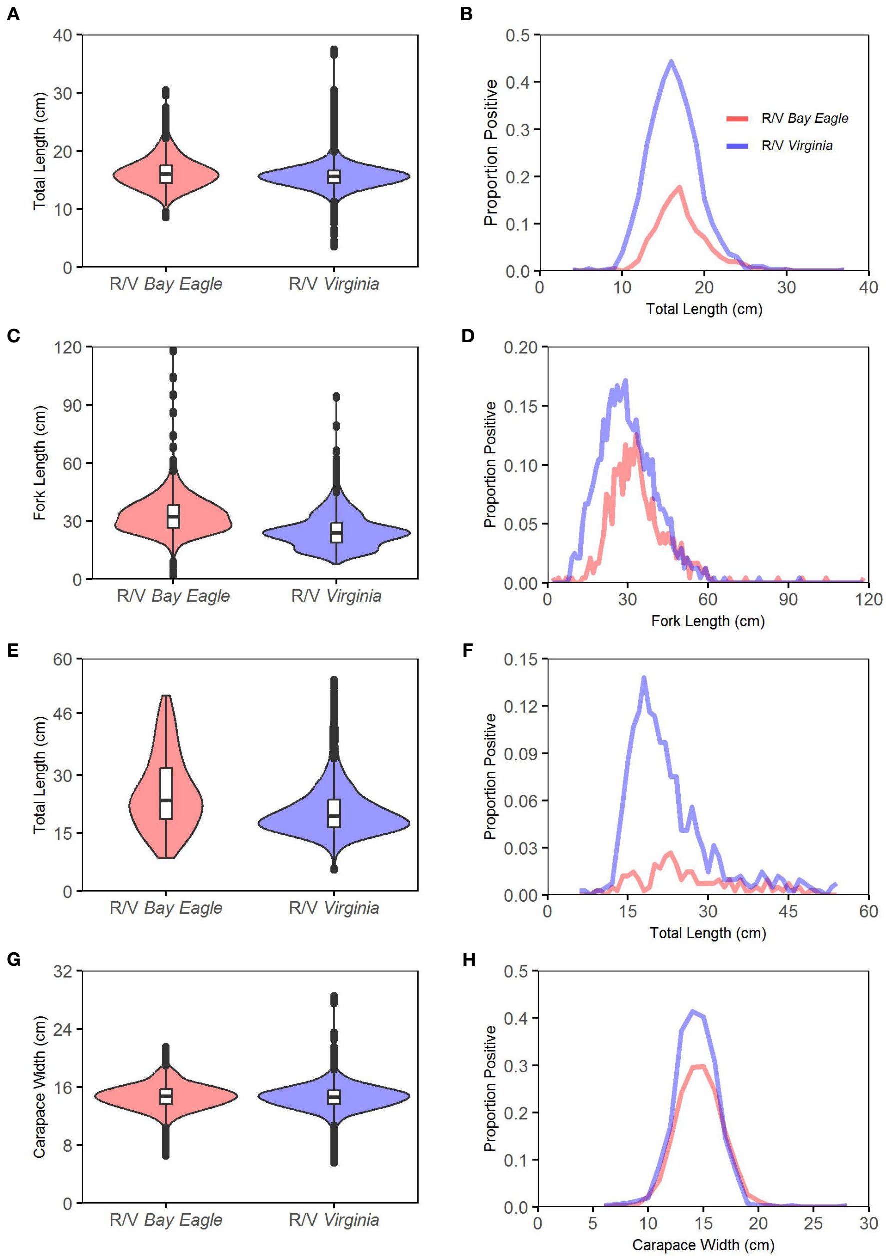



FIGURE 3
Vessel-specific size composition (first column) and proportion positive in relation to size (second column) data summaries from the paired-tow calibration experiment for **(A, B)** Atlantic croaker, **(C, D)** striped bass, **(E, F)** summer flounder, and **(G, H)** adult female blue crab.

**TABLE 2** Parameter and standard error estimates associated with the log-Gaussian Cox processes models fitted to the paired-tow data for the four species.

| Species | log $\sigma_s$ | log $\sigma_\Phi$ | log $\sigma_{AR}$ | $\rho$ |
|---|---|---|---|---|
| Atlantic croaker | 0.34 ± 0.03 | -1.11 ± 0.16 | 0.36 ± 0.06 | 0.91 ± 0.01 |
| Striped bass | -0.29 ± 0.04 | -2.13 ± 0.20 | -0.03 ± 0.09 | 0.98 ± 0.01 |
| Summer flounder | -1.94 ± 0.52 | -1.38 ± 0.16 | 0.20 ± 0.08 | 0.91 ± 0.02 |
| Adult female blue crab | 0.31 ± 0.03 | -2.35 ± 0.30 | 0.03 ± 0.07 | 0.98 ± 0.02 |

10.3389/fmars.2023.1217792

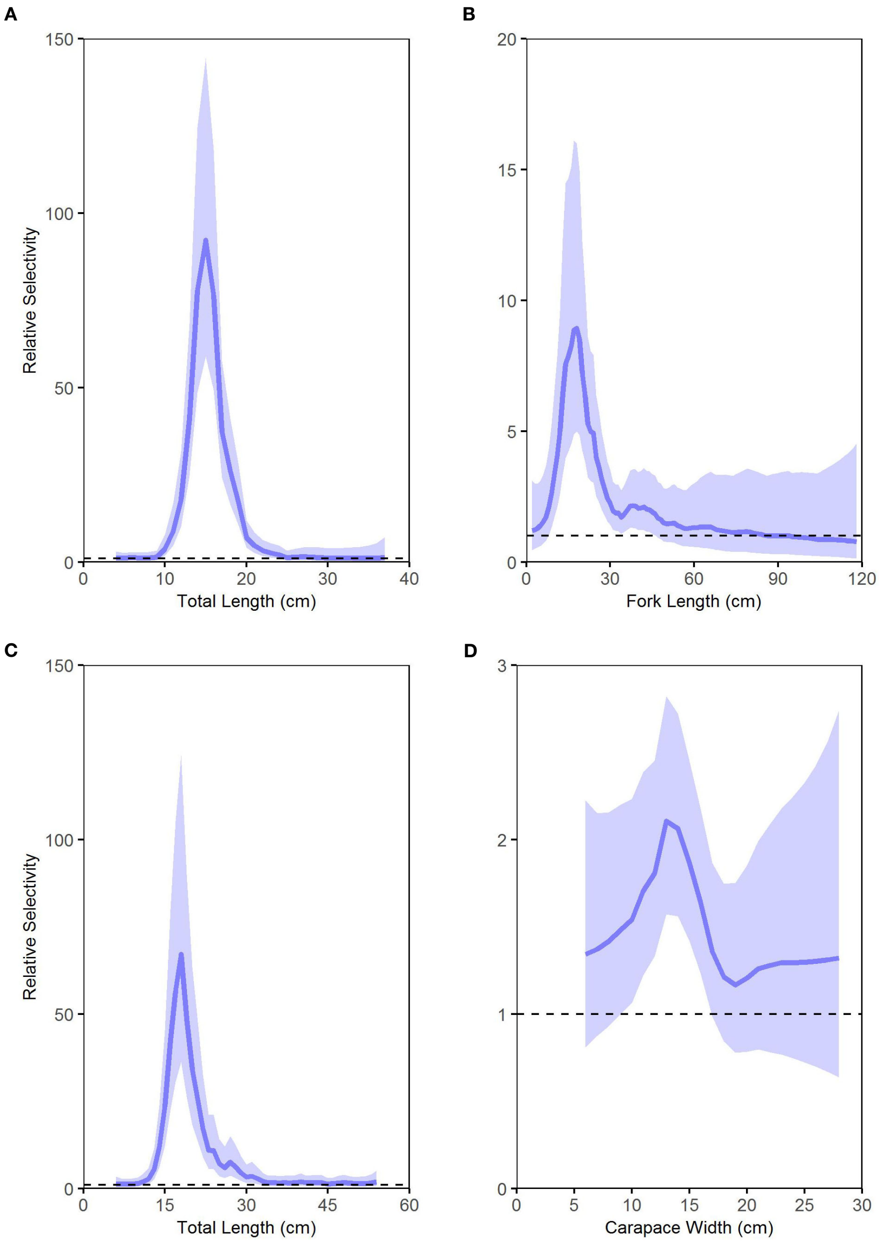



FIGURE 4
Relative selectivity (blue lines) of the trawl gear on the R/V *Virginia* (new fishing system) with that on the R/V *Bay Eagle* (original fishing system) for **(A)** Atlantic croaker, **(B)** striped bass, **(C)** summer flounder, and **(D)** adult female blue crab. For the new fishing system, values above the horizonal lines indicate higher selectivity, values below indicate lower selectivity, and values at the horizonal lines denote no selectivity differences. Shaded areas are 95% confidence intervals.

paired-tows resulted in zero R/V *Bay Eagle* catches and nonzero R/V *Virginia* catches for many species (and vice versa but to a much lesser extent). From the paired-tow data where the R/V *Bay Eagle* catches were zero, a model-based analysis of the associated R/V *Virginia* catches (e.g., generalized additive models) could potentially guide which historic zero catches should remain unchanged or be converted. Second, applying the relative size selectivity estimates (and potentially results from the analyses of zeros) implies the associated estimates of uncertainty need to be incorporated into analyses that yield indices of relative abundance. The confidence intervals around the above relative size selectivity curves were quite wide for some size classes, and when this uncertainty is combined with natural survey observation error (perhaps Monte Carlo or bootstrapping), the overall uncertainty estimates of the relative abundance indices could be quite high.

Other model-based approaches could potentially be utilized to estimate indices of relative abundance that span the full time-period, provided there is sufficient temporal overlap among the vessel-specific data sets. As noted previously, the ChesMMAP calibration sites were selected according to the underlying stratified sampling design, and because the calibration experiment transpired over several years, it is possible to view the paired-tow data as an extension of the R/V *Bay Eagle* data and supplemental to the ongoing R/V *Virginia* time-series. This temporal overlap facilitates application of models that can structurally accommodate distinct catchability and vessel effects (e.g., vector autoregressive spatio-temporal (VAST), Thorson, 2019), or the estimation of vessel-specific indices that could then be reconciled with a time-series approach (e.g., dynamic factor analysis, Peterson et al., 2021). However, both options require a background in advanced statistical modeling and computer coding as the details and implementation of these approaches are not trivial.

Lastly, there are options to bridge vessel changes at the level of the stock assessment model. Time-series of relative abundance could be estimated separately for the two vessels, which could then be inputs to an assessment model with different selectivity patterns and likelihood components (this becomes more practical once the R/V *Virginia* data gain longevity). Or perhaps the stock assessment model could be configured to accommodate the survey data from both vessels and the paired-tow data to internally estimate relative selectivity values along with other assessment parameters (an area of future research). In practice, however, choosing among the analytical options will likely depend on the specific goals and methodological approaches of future applications of the ChesMMAP survey data.

# 5 Lessons learned

The ChesMMAP survey is now in its fifth year of sampling using the new trawl package, vessel, and survey design, and calibration coefficients needed to link the contemporary catch data to those collected prior to the 2019 conversion are available for most of the key species sampled. The new trawl has yielded much larger catches relative to the original fishing system, but more importantly it has sampled a greater diversity of taxa and a broader size range of animals for most species. The R/V *Virginia* has proven to be an ideal platform to accommodate this survey and associated larger catches, while adjustments to the spatiotemporal extent of sampling effort have improved operational efficiency. Indeed, implementing these changes to ChesMMAP incurred some significant costs, however, the data generated now provide a more robust and likely more consistent characterization of the living marine resources in Chesapeake Bay. Further, during the process of redesigning ChesMMAP, our successes, missteps, and reflections on the original survey design revealed five important lessons that could be useful in future situations when it becomes either necessary or desirable to alter survey procedures. Several of these lessons are also applicable when developing a new survey.

## 5.1 Embrace broad collaborations

Network development and open, frequent communication with external partners provided the avenue for the initial discovery and successful implementation of the new trawl gear package adopted by ChesMMAP. While the design, size, and performance of several trawl gears used by successful fisheries-independent surveys and the shrimp fishery in the U.S. southeast were researched when originally designing ChesMMAP, this review was based solely on existing documentation and available expertise within VIMS. Involvement with the NEFSC Trawl Survey Advisory Panel (TSAP) during the early 2000s provided an opportunity for direct engagement with commercial fishers, trawl gear manufacturers, and scientific colleagues conducting fisheries-independent surveys. During these meetings, researchers outlined the desirable characteristics of a scientific survey trawl and the industry members (i.e., fishers and gear manufacturers) designed a gear package to match those criteria as closely as possible. The breadth and depth of industry expertise yielded a trawl net that maintained a very consistent trawl geometry over a broad range of depths and seafloor conditions, and in turn yielded catches that reflected a greater diversity of taxa and a broader size range of animals. Exposure to this gear development process in terms of the relationships established with the industry members and other survey scientists, and the successful implementation of the TSAP trawl on the NEFSC Bottom Trawl Survey and NEAMAP led to inquiries about designing a smaller version of this trawl for ChesMMAP. Without the development of these networks and cultivation of relationships with industry and peers in the scientific community, replacing the original trawl gear with the more efficient fishing system now used by ChesMMAP would have been unlikely.

Including key industry members in the new trawl design resulted in appreciable ‘buy-in’ by fishers on the successful implementation of this gear by all three trawl surveys. Prior to the initial field trials of the new gear package on the R/V *Bay Eagle*, two local fishers donated their time to assist survey personnel with the proper configuration of the wires that connect the net to the trawl doors, given that the more-complicated three-bridle design uses eight wires in total to make these connections. They also provided advice on appropriate setting and hauling procedures, including valuable tips on how to prevent these wires from becoming entangled during these processes. Based on catch rates observed by the NEFSC and NEAMAP with this trawl design, catches with the new ChesMMAP net were expected to be larger than those typically observed with the original gear. The industry members shared approaches that they use to retrieve very large catches and outfitted the ChesMMAP trawl with all rigging materials needed to perform these more-complicated retrievals at no cost. During the first day of field trials, it was clear that the trawl doors were not performing as intended, because wingspread measurements were much lower than expected and examination of the wear patterns on the doors revealed that they were lying flat on the bottom for at least part of each tow. An industry member volunteered to advise operations during the second day of these field trials, and by the third haul had identified the proper adjustments to the trawl warp connections and door backstrap chains to achieve the optimal geometry of the survey trawl. Thus, when selecting and implementing a new sampling gear for a fisheries-independent survey, either as part of a survey transition or development of a new sampling program, partnering with the fishing industry and other survey scientists to benefit from the experience and expertise of both groups is strongly recommended.

## 5.2 Utilize flume tanks

Prior to conducting the initial field trials with the new ChesMMAP trawl on the R/V *Bay Eagle*, scale models of both the new and original nets were constructed, and their performance evaluated in a flume tank located at Memorial University in St. John's, NL. Three survey personnel traveled to this facility in early December 2009, and the performance of the model trawls was documented over a two-day period. Specifically, the geometry of the trawls and the variability in both net wingspread and headline height were measured over a range of simulated towing speeds, rigging configurations, trawl door designs, and codend fullness (representing varying catch sizes). These data clearly demonstrated the superiority of the new net with respect to maintaining a consistent trawl geometry while achieving a greater headline height across the full range of conditions tested. Visual evaluation of the trawl in the flume also revealed that the new net maintained a consistent shape throughout the trials, while the original net was somewhat disfigured at the center of the headline and between the wings of the net and the footrope. Further, the sweep of the new net remained in contact with the bottom across the full range of towing speeds (1.3 – 1.7 m $s^{-1}$), while the original gear would rise off bottom at higher speeds.

The encouraging flume data of the scale model of the new net prompted purchase of two 11.2 m three-bridle, four-seam trawls for field testing in fall 2010. The total expenses associated with the flume trials, including construction of the model trawls, travel, flume rental, and personnel time, were approximately equal to one day of vessel costs. When adding procurement costs of the new trawl package, associated sampling supplies, and personnel compensation, the daily cost of field testing far exceeded the total for the flume trials. Had the new trawl performed poorly, the flume trials would have yielded appreciable cost savings since field testing would have been unnecessary. Moreover, visual evaluation of both net models in the flume provided insight on their overall shape and ability to maintain bottom contact that would have been extremely difficult to acquire through field trials. Given the valuable information gained from the flume, it is recommended that flume testing be conducted on all currently used trawls for performance data and any candidate trawls prior to field operations.

## 5.3 Implement changes simultaneously

Results from the flume trials and initial field testing provided strong support for adopting the new gear package on ChesMMAP. As noted above, transitioning to this gear on the R/V *Bay Eagle* would have incurred significant costs and presented substantial logistical challenges, but doing so was not entirely unachievable. However, shortly after flume testing, VIMS acquired funds to design and build the R/V *Virginia*, and accommodating the new ChesMMAP fishing system influenced many design elements. Given the pending arrival of this vessel and plans to remove the R/V *Bay Eagle* from the VIMS fleet shortly thereafter, it was decided to delay incorporating the new trawl into ChesMMAP until operations could be shifted to the new sampling platform. This conclusion emerged following careful evaluation of both the financial costs associated with a transition and impacts on the time-series of survey data. Had changes been made to ChesMMAP in two steps, first implementing the new trawl on the R/V *Bay Eagle* and then moving survey operations to the R/V *Virginia* once available, two calibration experiments (one at each step) would have been necessary, and financial resources required to maintain linkages across the full time-series would have been approximately doubled. Further, two considerable sources of uncertainty would have been introduced into the time-series, one from each of the respective calibration efforts, which could have unnecessarily diminished the utility of the survey dataset.

Once the decision was made to delay implementation of the new survey trawl until operations transitioned to the new research vessel, all aspects of the ChesMMAP sampling design were more formally evaluated for possible improvements. Although draft restrictions of the R/V *Virginia* forced abandonment of the shallowest sampling locations within the original sampling frame, analyses of extant data supported adjusting the boundaries of the latitudinal regions and depth strata (modest changes in relative abundance patterns and associated uncertainties). Further, the increased daily rate of the new vessel prompted evaluation of the spatiotemporal patterns of the extant catch data to identify redundancies and sampling season and region combinations that historically yielded scant abundance, life history, and trophic information. The resulting stratification changes and associated reallocation of sampling effort were implemented concurrently with the transition of the survey to the new vessel and trawl gear package, and thus the impacts of all adjustments were captured by the calibration coefficients generated from a single paired-tow experiment. Thus, it is recommended to view periods of survey transition as unique opportunities to evaluate and improve as many aspects of the sampling operation as possible, and subsequently implement all changes simultaneously so that the cumulative impacts of these new procedures are reflected in one comprehensive calibration experiment.

## 5.4 Recognize survey calibration costs

Whether changes to sampling procedures are implemented by choice or out of necessity, it is critical that all costs, both financial and non-monetary, be considered. Total costs can be substantial, and if available financial and human resources are not sufficient to meet expected costs, it will be necessary to critically evaluate trade-offs among the proposed survey adjustments and identify those that are both achievable and that provide the greatest benefit to the sampling program. Among the largest costs incurred during a survey transition are those associated with the calibration experiments. In the case where the original vessel is retained but a new trawl package is utilized, data to generate calibration coefficients are collected through a single-vessel, paired-tow design where the original and new gears are hauled in succession at several sampling locations. Because the net that is towed first usually alternates among sampling sites, it is necessary to conduct this experiment separately from routine survey operations because disturbances from towing the alternate trawl first could negatively impact the time-series of survey data. This separation of survey and calibration activities typically requires substantial financial resources to support the additional ship and personnel time.

When both the vessel and survey trawl are to be replaced, the two platforms can conduct simultaneous paired-tows to generate data needed calibration information. Costs associated with this approach can be reduced by coupling the paired-tow experiment with survey operations such that one vessel is conducting the survey while the other samples concurrently and in close proximity at either all or a subset of sites. However, the presence of the second vessel could create disturbance effects that impact survey catches, so whenever possible separating these paired-tow experiments from the survey operations is recommended. Regardless of the approach, personnel costs are approximately double those associated with routine survey operations since it is necessary to employ a science team to process catches on two vessels. At an institutional-level, fleet maintenance costs will increase (at least temporarily) between the time the new vessel is delivered and the previous vessel is retired, as it is imperative that both remain fully operational so that the paired-tow experiment can be completed as quickly as possible. Further, when the decision is made to transition the sampling procedures immediately and conduct paired-tows concurrently, which was the approach adopted herein, it is important to recognize that there can be delays in providing updated indices of relative abundance to stock assessment activities. This information gap represents a cost to the assessment management processes.

Many times, when there is a change in sampling platform, the replacement vessel is often newer, larger, requires additional vessel crew, and therefore is more expensive to operate. For ChesMMAP, funds available to the survey remained relatively constant, which required reexamination of the spatiotemporal distribution of sampling effort. Further, when the vessel is coupled with a new survey trawl package, there are obvious start-up costs associated with the acquisition of the net, trawl doors, and associated rigging materials. If the new trawl is much more efficient than the original gear and average catches increase appreciably, the program will realize increased long-term costs often in one of two ways. Either additional catch processing time will be needed at each sampling site which will lengthen cruises and increase vessel costs, or additional scientific personnel will be needed to process the larger catch volumes. Following an evaluation of the trade-offs between vessel and personnel costs, increasing the scientific crew from four on the R/V *Bay Eagle* to six on the R/V *Virginia* was necessary. While ChesMMAP now generates a more robust, consistent sampling of the Chesapeake Bay ecosystem, the immediate and longer-term costs associated with the transition were quite large, and it is recommended that researchers think broadly and estimate all associated costs prior to initiation changes to survey sampling procedures.

### 5.5 Document and share survey changes

The decision to explore implementation of the 11.2 m three-bridle, four-seam bottom trawl on ChesMMAP was motivated in large part by reports of the gear performance and catch composition generated by the larger version of this gear used by the NEFSC and NEAMAP. Likewise, following successful flume and initial field trials with the new ChesMMAP net, data, results, and experiences rigging, deploying, and retrieving this gear were shared through presentations to regional management councils and commissions, annual progress reports to funding agencies, and informally to colleagues. The ChesMMAP net design has since been adopted by a United States Geological Survey bottom trawl survey in Lake Erie and is under consideration by two additional surveys, one operating in the southeast U.S. Atlantic waters and the other in Southern New England. Using a standardized trawl design on multiple surveys that yields robust, consistent sampling of ecological communities facilities important cross-system comparisons of programmatic datasets, and so as surveys are developed or undergo transitions, broad communication of experiences and results to promote coordination and standardization of sampling procedures where possible is recommended.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The animal study was reviewed and approved by William & Mary Institutional Animal Care and Use Committee.

## Author contributions

All authors contributed to the design of the original ChesMMAP survey, the subsequent redesign following gear and vessel changes, and the design of the paired sampling experiment. CB and JG organized the paired-tow database and RL performed the statistical analyses with input from JG. RL and JG co-wrote the first draft of the manuscript and incorporated comments provided by CB. All authors contributed to the article and approved the submitted version.

## Funding

Financial support over the duration of the ChesMMAP survey was provided by the NOAA Chesapeake Bay Office, the Virginia Environmental Endowment, the U.S. Fish and Wildlife Service, and the Virginia Marine Resources Commission.

## Acknowledgments

The authors wish to acknowledge the field and laboratory efforts of past and present staff and students associated with the ChesMMAP survey at the Virginia Institute of Marine Science. Former and current crews of the R/V *Bay Eagle* and R/V *Virginia* deserve many thanks for their contributions to vessel maintenance and operations. Comments provided by two reviewers and the Associate Editor reviewers helped improve this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References


Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A. E., Lewis, M. A., Jonsen, I. D., et al. (2016). State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. *Sci. Rep.* 6, 26677. doi: 10.1038/srep26677

Brown, R. W., Fogarty, M., Legault, C., Miller, T., Nordahl, V., Politis, P., et al. (2007). *Survey transition and calibration of bottom trawl surveys along the northeastern continental shelf of the united states* (International Council for the Exploration of the Seas). CM 2007/Q:20, Copenhagen.

Cochran, W. G. (1977). *Sampling techniques. 3rd Edition* (New York: Wiley).

Ding, H., and Elmore, A. J. (2015). Spatio-temporal patterns in water surface temperature from landsat time series data in the Chesapeake bay, U.S.A. *Remote Sens. Environ.* 168, 335–348. doi: 10.1016/j.rse.2015.07.009

Fernandez, L., and Leach, S. (1998). *Prospects for multispecies fisheries management in Chesapeake bay* Vol. 79 (Edgewater, MD: Chesapeake Bay Scientific and Technical Advisory Committee [STAC] Publication 98-002).

Gartland, J., Gaichas, S. K., and Latour, R. J. (2023). Spatiotemporal patterns in the ecological community of the nearshore mid-Atlantic bight. *Mar. Ecol. Prog. Ser.* 704, 15–33. doi: 10.3354/meps14235

Hagy, J. D., Boynton, W. R., Keefe, C. W., and Wood, K. V. (2004). Hypoxia in Chesapeake Bay,1950-2001: long-term change in relation to nutrient loading and river flow. *Estuaries* 27, 634–658. doi: 10.1007/BF02907650

Hilborn, R., and Walters, C. J. (1992). *Quantitative fisheries stock assessment* (New York: Chapman & Hall).

Hinson, K. E., Friedrichs, M. A. M., St-Laurent, P., Da, F., and Najjar, R. G. (2022). Extent andcauses of Chesapeake bay warming. *J. Am. Water Resour. Assoc.* 58, 805–825. doi: 10.1111/1752-1688.12916

Irby, I. D., Friedrichs, M. A. M., Da, F., and Hinson, K. E. (2018). The competing impacts of climate change and nutrient reductions on dissolved oxygen in Chesapeake bay. *Biogeosciences* 15, 2649–2668. doi: 10.5194/bg-15-2649-2018

Johnson, T. R., and McCay, B. J. (2012). Trading expertise: the rise and demise of an industry/government committee on survey trawl design. *Marit. Stud.* 11, 14. doi: 10.1186/2212-9790-11-14

Kemp, W. M., Boynton, W. R., Adolf, J. E., Boesch, D. F., Boicourt, W. C., Brush,G., Cornwell, G.J. C., et al. (2005). Eutrophication of Chesapeake bay: historical trends and ecological interactions. *Mar. Ecol. Prog. Ser.* 303, 1–29. doi: 10.3354/meps30301

Kimura, D. K., and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Rev. Fish. Sci.* 14, 245–283. doi: 10.1080/10641260600621761

Kirkley, J. E., Murray, T. J., and Duberg, J. (2005). *Economic contributions of virginia's commercial seafood and recreational fishing industries: a user's manual for assessing economic impacts* (Gloucester Point, VA: Virginia Institute of Marine Science, Marine Resource Report), 2005–2009. doi: 10.21220/m2-q76v-kq84

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). TMB: automatic differentiation and Laplace approximation. *J. Stat. Software* 70, 1–21. doi: 10.18637/jss.v070.i05

Latour, R. J., Brush, M. J., and Bonzek, C. F. (2003). Toward ecosystem-based fisheries management: strategies for multispecies modeling and associated data requirements. *Fisheries* 28, 10–22. doi: 10.1577/1548-8446(2003)28[10:TEFM]2.0.CO;2

Lellis-Dibble, K. A., McGlynn, K. E., and Bigford, T. E. (2008). *Estuarine fish and shellfish species in U.S. commercial and recreational fisheries: economic value as an incentive to protect and restore estuarine habitat. U.S. Dep. Commerce, NOAA technical memorandum NMFS-F/SPO-90* Vol. 102 (Silver Spring, MD: National Oceanic and Atmospheric Administration).

Lewy, P., Nielsen, J. R., and Hovgård, H. (2004). Survey gear calibration independent of spatial fish distribution. *Can. J. Fish. Aquat. Sci.* 61, 636–647. doi: 10.1139/F04-034

Link, J. S. (2010). *Ecosystem-based fisheries management: confronting tradeoffs* (New York: Cambridge University Press).

Link, J. S., and Marshak, A. R. (2022). *Ecosystem-based fisheries management: progress, importance, and impacts in the united states* (New York: Oxford University Press).

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* 70, 141–159. doi: 10.1016/j.fishres.2004.08.002

Methot, J. R.D., Tromble, G. R., Lambert, D. M., and Greene, K. E. (2014). Implementing a science-based system for preventing overfishing and guiding sustainable fisheries in the united states. *ICES J. Mar. Sci.* 71, 183–194. doi: 10.1093/icesjms/fst119

Miller, T. J. (2013). A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. *Can. J. Fish. Aquat. Sci.* 70, 1306–1316. doi: 10.1139/cjfas-2013-0136

Miller, T. J., Houde, E. D., and Watkins, E. J. (1996). *Chesapeake Bay fisheries: prospects for multispecies management and sustainability* Vol. 73 (Edgewater, MD: Chesapeake Research Consortium [CRC] Publication 154B).

Murdy, E. O., Birdson, R. S., and Musick, J. A. (1997). *Fishes of Chesapeake bay* (Washington, DC: Smithsonian).

Najjar, R. G., Pyke, C. R., Adams, M. B., Breitburg, D., Hershner, C., Kemp, M., et al. (2010). Potential climate change impacts on the Chesapeake Bay. *Estuar. Coast. Shelf Sci.* 86, 1–20. doi: 10.1016/j.ecss.2009.09.026

National Marine Fisheries Service [NMFS] (1998). *Ecosystem-based fisheries management: a report to congress by the ecosystem principles advisory panel* Vol. 62 (Silver Spring, MD: U.S. Department of Commerce, NOAA). Available at: https://repository.library.noaa.gov/view/noaa/23730.

National Marine Fisheries Service [NMFS] (2020). *Fisheries economics of the United States 2020. U.S. Dep. Commerce, NOAA tech. memo. NMFS-F/SPO-236.* (Silver Spring, MD: National Oceanic and Atmospheric Administration).

National Oceanic and Atmospheric Administration Chesapeake Bay Fisheries Ecosystem Advisory Panel [NOAA CBFEAP] (2006). *Fisheries ecosystem planning for Chesapeake bay* Vol. 462 (Bethesda, MD: American Fisheries Society).

Orth, R. J., Marion, S. R., Moore, K. A., and Wilcox, D. J. (2010). Eelgrass (*Zostera marina* l.) in the Chesapeake bay region of mid-Atlantic coast of the USA: challenges in conservation and restoration. *Estuar. Coasts.* 33, 139–150. doi: 10.1007/s12237-009-9234-0

Peterson, C. D., Courtney, D. L., Cortés, E., and Latour, R. J. (2021). Reconciling conflicting survey indices of abundance prior to stock assessment. *ICES J. Mar. Sci.* 78, 3101–3120. doi: 10.1093/icesjms/fsab179

Richards, R. A., and Rago, P. J. (1999). A case history of effective fishery management: Chesapeake bay striped bass. *N. Am. J. Fish. Manage.* 19, 356–375. doi: 10.1577/1548-8675(1999)019<0356:ACHOEF>2.0.CO;2

Ricker, W. E. (1975). Computation and interpretation of biological statistics of fish populations. *Bull. Fish. Res. Board Can.* 191, 1–382.

Schonfeld, A. J., Gartland, J., and Latour, R. J. (2022). Spatial differences in estuarine utilization by seasonally resident species in mid-Atlantic bight, USA. *Fish. Oceanogr.* 31, 615–628. doi: 10.1111/fog.12611

Thomas, A. C., Pershing, A. J., Friedland, K. D., Nye, J. A., Mills, K. E., Alexander, M. A., et al. (2017). Seasonal trends and phenology shifts in sea surface temperature on the north American northeastern continental shelf. *Elem. Sci. Anth.* 5, 48. doi: 10.1525/elementa.240

Thompson, S. K. (2012). *Sampling. 3rd Edition* (New York: Wiley).

Thorson, J. T. (2019). Guidance for decision using vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat, and climate assessments. *Fish. Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Thorson, J. T., Clarke, M. E., Stewart, I. J., and Punt, A. E. (2013). The implications of spatially varying catchability on bottom trawl surveys of fish abundance: a proposed solution involving underwater vehicles. *Can. J. Fish. Aquat. Sci.* 70, 294–306. doi: 10.1139/cjfas-2013-0330

Thygesen, U. H., Kristensen, K., Jansen, T., and Beyer, J. E. (2019). Intercalibration of survey methods using paired fishing operations and log-Gaussian cox processes. *ICES J. Mar. Sci.* 76, 1189–1199. doi: 10.1093/icesjms/fsy191

Tian, R., Cerco, C. F., Bhatt, G., Linker, L. C., and Shenk, G. W. (2022). Mechanisms controlling climate warming impact on the occurrence of hypoxia in Chesapeake bay. *J. Am. Water Resour. Assoc.* 58, 855–875. doi: 10.1111/1752-1688.12907

Venables, W. N., and Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish. Res.* 70, 319–337. doi: 10.1016/j.fishres.2004.08.011

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. (2010). Incorporating time-varying catchability into population dynamic stock assessment models. *Rev. Fish. Sci.* 18, 7–24. doi: 10.1080/10641260903294647

Wilberg, M. J., Livings, M. E., Barkman, J. S., Morris, B. T., and Robinson, J.M. (2011).Overfishing, disease, habitat loss, and potential extirpation of oysters in upper Chesapeake bay.*Mar. Ecol. Prog. Ser.* 704, 15–33. doi: 10.3354/meps09161

Winger, P. D., DeLouche, H., and Legge, G. (2006). Designing and testing new fishing gears: the value of a flume tank. *Mar. Technol. Soc J.* 40, 44–49. doi: 10.4031/002533206787353240

Winger, P. D., Eayrs, S., and Glass, C. W. (2010). "Fish behavior near bottom trawls," in *Behavior of marine fishes: capture processes and conservation challenges*. Ed. H. Pingguo (Ames, IA: Wiley-Blackwell), 67–95.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 18 August 2023
DOI 10.3389/fmars.2023.1237549

Check for updates

OPEN ACCESS

EDITED BY
Kyle Shertzer,
NOAA Southeast Fisheries Science Center, United States

REVIEWED BY
Frank Hernandez,
Southeast Fisheries Science Center (NOAA), United States
Norbert Rohlf,
Thünen Institute of Sea Fisheries, Germany

*CORRESPONDENCE
Hsiao-Yun Chang
hsiao-yun.chang@stonybrook.edu

†These authors share first authorship

RECEIVED 09 June 2023
ACCEPTED 31 July 2023
PUBLISHED 18 August 2023

CITATION
Chang H-Y, Sun M, Rokosz K and Chen Y (2023) Evaluating effects of changing sampling protocol for a long-term ichthyoplankton monitoring program. *Front. Mar. Sci.* 10:1237549.
doi: 10.3389/fmars.2023.1237549



# Evaluating effects of changing sampling protocol for a long-term ichthyoplankton monitoring program

Hsiao-Yun Chang[1,2*†], Ming Sun[1,2†], Katrina Rokosz[1,2] and Yong Chen[1,2]

[1]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, United States,
[2]Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY, United States

Abundance indices play a crucial role in monitoring and assessing fish population dynamics. Fishery-independent surveys are commonly favored for deriving abundance indices because they follow standardized or randomized designs, ensuring spatiotemporal consistency in representative and unbiased sampling. However, modifications to the survey protocol may be necessary to accommodate changes in survey goals and logistic difficulty. When the survey undergoes changes, calibration is often needed to remove variability that is unrelated to changes in abundance. We evaluated a long-term monitoring program, the Long River Survey (LRS) in the Hudson River Estuary (HRE), to illustrate the process of calibrating survey data to account for the effects of changing sampling protocol. The LRS provided valuable ichthyoplankton data from 1974 to 2017, but inconsistencies in sampling timing, location, and gears resulted in challenges in interpreting and comparing the fish abundance data in the HRE. Generalized Additive Models were developed for five species at various life stages, aiming to mitigate the impact of sampling protocol changes. Model validation results suggest the consistent performance of the developed models with varying lengths of time series. This study indicates that changes in the sampling protocol can introduce biases in the estimates of abundance indices and that the model-based estimates can improve the reliability and accuracy of the survey abundance indices. The model-estimated sampling effects for each species and life stage provide critical information and valuable insights for designing future sampling protocols.

KEYWORDS

**fishery-independent survey, Hudson River Estuary, ichthyoplankton, survey catchability, model-based abundance index, design-based abundance index**

# 1 Introduction

Abundance indices derived from fishery-dependent and fishery-independent data play a crucial role in stock assessment by providing valuable information on fish population dynamics (Pennino et al., 2016; Maunder et al., 2020). Fishery-independent surveys are commonly favored for deriving abundance indices because they follow standardized or randomized designs, ensuring consistency in gear, effort, and sampling methods across different locations and time periods. However, modifications to the survey protocol may be necessary to accommodate changes in survey goals, such as focusing on specific species or fishery-related measurements. Nominal CPUE, measured as the total catch divided by an observable measure of effort, may not always accurately reflect the true abundance of resources over time and space (Harley et al., 2001), as it can be influenced by various factors such as sampling area, gear used (Chiarini et al., 2022; Ducharme-Barth et al., 2022), and changes in the sampling protocol. In cases where the study area is not uniformly surveyed due to biased sampling, catch rate calibration (Webster et al., 2020) is often employed to eliminate variability that is unrelated to changes in abundance (Walters, 2003).

The Hudson River is an environmentally, economically, and socially important waterbody flowing south through New York, from the Adirondack Mountains through New York City. The Hudson River Estuary (HRE) extends 245 km from Troy, New York to the Battery in New York City, where it drains into the Atlantic Ocean (Figure 1). The estuary is high in nutrients and well-mixed due to tidal mixing, with approximately 1-meter tides, and a salt wedge fluctuating about 100 river km from New York Harbor, depending on freshwater flow and tidal cycles (Cooper et al., 1988). The HRE is home to over 200 fish species (Levinton and Waldman, 2006). The HRE provides critical habitat for freshwater, marine, estuarine, and diadromous fishes, including many key fish species of economic, ecological, and social importance in the northwest Atlantic Ocean (e.g., American Eel *Anguilla rostrata*, American Shad *Alosa sapidissima*, Atlantic Tomcod *Microgadus tomcod*, Striped Bass *Morone saxatilis*, and White Perch *Morone americana*). Habitat restoration and fisheries management are being used to conserve the HRE's ecosystem and restore the HRE's signature fisheries after decades of overharvest and habitat destruction. The success of these efforts depends on understanding how the HRE ecosystem responds to environmental and climate changes and anthropogenic activities.

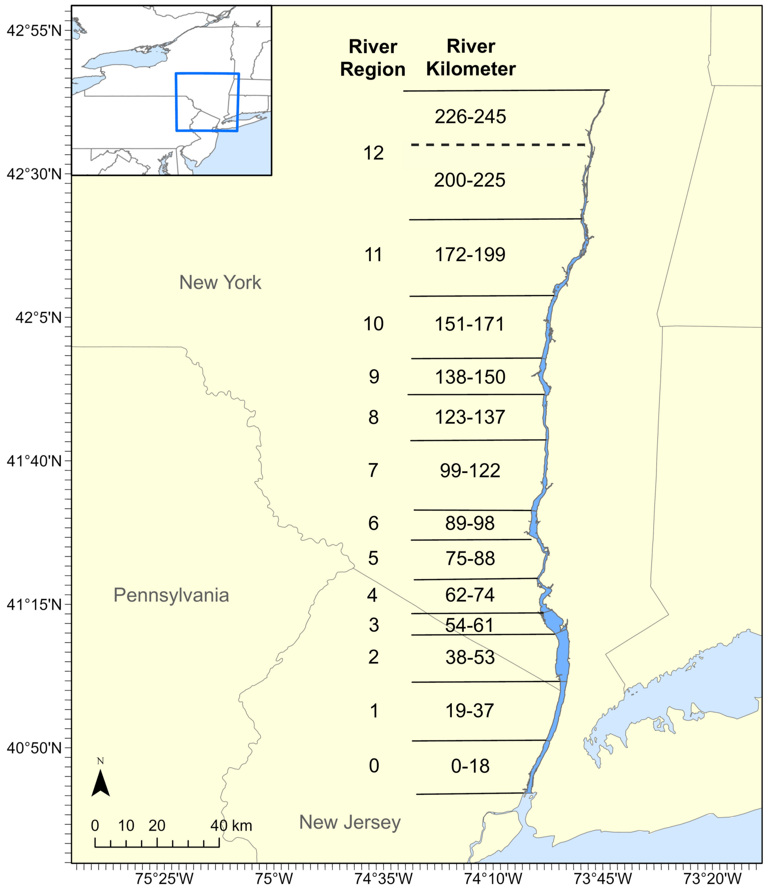



FIGURE 1
The Hudson River Estuary (in dark blue) divided into 13 river regions for the stratified random sampling design of the LRS based on river kilometer. The dashed line indicates the northernmost boundary of the study area prior to the expansion of the study area in 1988 when region 12 expanded to include river kilometers 226-245. Region 0 was not included in the sampling until 1988.

The Long River Ichthyoplankton Survey (LRS) under the Hudson River Biological Monitoring Program (HRBMP) is a fishery-independent ichthyoplankton trawl survey conducted continuously from 1974 through 2017. It was concluded in 2017 with the announcement of the closure of the Indian Point power plant. The LRS collected over 467 thousand observations for over 150 species in the HRE with over 108 thousand tows, targeting a range of fish species and providing critical information on the abundance and spatiotemporal distribution for their early life stages, including eggs, yolk-sac-larvae (YSL), post-yolk-sac-larvae (PYSL), young-of-year (YOY), and older. The LRS provided a unique opportunity to study how different fish species and fish communities respond to climatically and anthropogenically induced changes in environments within the HRE. However, before these data can be used they must be understood and their quality assured. While the LRS followed a stratified random survey design consistently over the survey period, many technical changes occurred during the survey to address various issues encountered in the survey due to logistic limitations, modified survey objectives, and foci on specific topics. These inconsistencies in sampling practices could introduce observation bias in the sampled fish abundance data, raising significant difficulties in using and interpreting survey catch rates and fish dynamics over space and time. This calls for a careful and comprehensive evaluation of the possible impacts of changing survey protocols and the development of approaches to calibrate and standardize the survey data to make them spatiotemporally consistent and comparable, serving as an excellent example for illustrating the process of calibrating fishery-independent data.

The present study aims to use the LRS as an example to develop a data calibration procedure and evaluate the impacts of sampling protocol changes on the estimates of fish abundances and spatiotemporal distributions. Using several representative species and their key early life stages, we aim to: 1) evaluate and identify the influential sampling factors to the LRS dataset, 2) explore appropriate and duplicable statistical approaches to calibrate the data to minimize the sampling bias in fish abundance indices and validate their performance, 3) provide more robust model-based abundance indices, and 4) demonstrate the risks of neglecting sampling bias by comparing discrepancies between the model-based abundance indices and the design-based abundance indices. The formulated data calibration procedure will be widely applicable to not only the entire LRS dataset but also similar fisheries surveys and biological monitoring programs seeking solutions to address sampling bias in data. The findings of this study will also provide insights into optimizing survey designs and analyzing survey data in broader environmental studies. These insights can contribute to improving the reliability and accuracy of abundance estimates in similar fishery-independent surveys.

# 2 Materials and methods

## 2.1 Long river survey

The HRE is the southern portion of the Hudson River, extending 245 rkm (river kilometer) from the Federal Dam in Troy, NY to the Battery in New York City (Figure 1). The upper portion of the HRE is a freshwater ecosystem and the southern 97 rkm is a brackish/marine ecosystem (Daniels et al., 2005). Although the sampling area covers from Albany to Battery Park, NY, the sampling in the Battery Park region and the northernmost reaches of the Albany region did not start until 1988.

The LRS ichthyoplankton data were collected throughout the HRE primarily from April through November, 1974-2017; however, the starting and ending dates varied from year to year (Figure 2). Sampling was done on a weekly basis during May-July, and on a biweekly basis during the other months. Although a stratum-based stratified random sampling design was used for determining sampling locations, the allocation of sampling effort across river regions and strata was adjusted over time based on the projected occurrence and spatial distribution of the target species and life stages (ASA Analysis & Communication (ASAAC), 2016). The sampling strata in the study are divided into 13 longitudinal river regions (Figure 1), ranging from Albany to Battery, and 3 habitat strata, including shoal, channel, and bottom (Heimbuch et al., 1992). A 1 $m^2$ Tucker trawl was used for sampling shoal and channel strata, and a 1 $m^2$ epibenthic sled was used for sampling the shoal and bottom strata. Both gears were fitted with 505 μm mesh plankton nets and were used for all sampling times and areas. In general, the Tucker trawl sampled shallower depths ranging from 0 to 47.3 meters (m) with a mean of 6.5 m, and the epibenthic sled sampled deeper depths ranging from 0 to 60.3 m with a mean of 10.57 m. The sample depth was the distance from the surface of the water to the top of the gear. The sample depth was determined randomly for each tow, based on the strata being sampled.

Sampling was carried out throughout the entire study area, during both daytime and nighttime except 1987-1994. Daytime was defined as the period from 30 minutes after sunrise to 30 minutes before sunset, while nighttime was defined as the period from 30 minutes after sunset to 30 minutes before sunrise. Prior to 1987, surveys were conducted in daylight until early June, after which they were conducted at night to minimize possible gear avoidance by the developing fish (Bowles et al., 1978; Boreman and Klauda, 1988). Gear avoidance by larval fish has been found to relate to visual stimuli (Bridger, 1956) and fish size, with larger fish exhibiting greater gear avoidance (Ahlstrom, 1954). Larval fish have been found to engage in diel migrations, (Haldorson et al., 1993; Murphy et al., 2011; Ospina-Alvarez et al., 2012), including species that are found in the HRE (Noble, 1970). From 1987 through 1994, no daytime sampling was conducted. Sampling intensity was heavily skewed toward nighttime from the years 2000 – 2017. The dates of switching from daytime to nighttime sampling were not consistent over the years (Figure 2).

Abundance data of ichthyoplankton and fishes at several life stages were collected from the LRS, including eggs, YSL, PYSL, and YOY. Analyses were performed using catch data for striped bass, white perch, and American shad (eggs, YSL, and PYSL), Atlantic tomcod (PYSL and YOY), and American eel (YOY and yearling and older (YROL).

## 2.2 Case study species

Five species were selected as species of focus with data from two or three life stages selected for each species for analysis (Table 1).

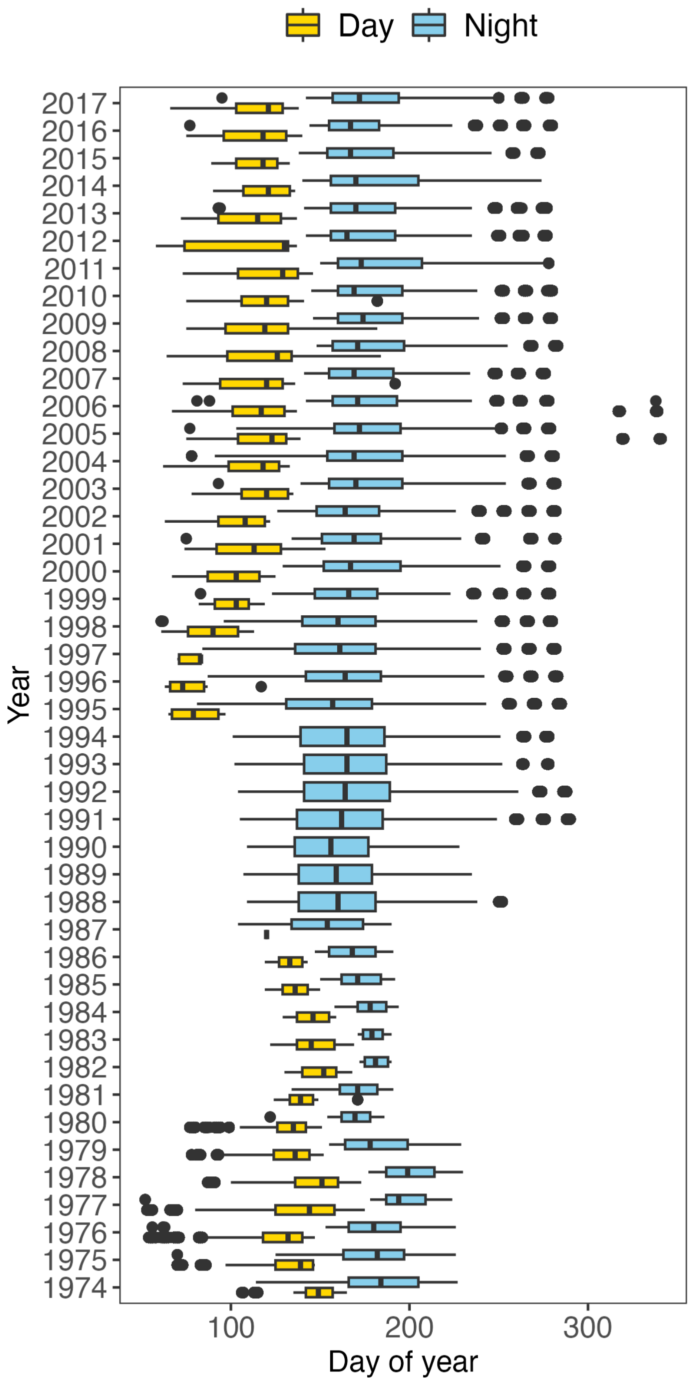



FIGURE 2
Boxplots of the sampling day of year for the Long River Survey in each year. The vertical bars in the boxes are medians. The left and right limits of the boxes are the first (Q1) and third (Q3) quartiles (25th and 75th percentiles). The difference between Q1 and Q3 is the interquartile range (IQR). Potential outliers are defined as observation points that fall outside the range of Q1-1.5*IQR and Q3 + 1.5*IQR. If potential outliers are presented, the whiskers extend to 1.5 times the IQR from Q1 or Q3. If no outliers are presented, the whiskers extend to the minima and maxima of the distributions. Yellow boxes denote daytime sampling, and navy-blue boxes denote nighttime sampling.

These species and their respective life stages were selected considering their representation in the LRS database in terms of ecological roles, targeting status in the survey design, and socio-economic importance.

Striped bass is a significant species in the HRE not only because of their commercial value but also their iconic social-ecological importance (McLaren et al., 1988; Limburg et al., 2006). The Hudson River has been identified as a significant contributor of striped bass to the Atlantic coastal fisheries (McLaren et al., 1981), and the HRE is crucial spawning and nursery ground for striped bass (Nack et al., 2019). Striped bass were initially the single species of focus of the LRS because the utility company was obligated to

TABLE 1 Analyzed species and life stages selected from the Long River Survey database.

| Species | Consideration in survey design | Life stage | Presumed season as DOY | Spatial range as river region as in Figure 1 |
|---|---|---|---|---|
| American eel | never a target species | YOY | 58-338 | 0-12 |
| | | YROL | 56-288 | 0-12 |
| American shad | became a target species since 1982 | egg | 97-188 | 7-12 |
| | | PYSL | 124-211 | 7-12 |
| | | YSL | 117-189 | 7-12 |
| Atlantic tomcod | became a target species since 1975 | PYSL | 52-236 | 0-7 |
| | | YOY | 70-288 | 0-7 |
| Striped bass | target species during the entire period | egg | 100-196 | 3-11 |
| | | PYSL | 122-239 | 2-10 |
| | | YSL | 111-229 | 1-9 |
| White perch | became a target species since 1975 | egg | 97-194 | 7-12 |
| | | PYSL | 111-246 | 7-12 |
| | | YSL | 99-210 | 5-11 |

PYSL, post-yolk-sac-larvae; YOY, young-of-year; YROL, yearling and older; YSL, yolk-sac-larvae.

demonstrate compliance with federal regulations for the construction and operation of the Cornwall pumped-storage facility and thermal power plants with once-through cooling (Barnthouse et al., 1988). Therefore, striped bass eggs, YSL, and PYSL were used in this study.

Later, the Federal Nuclear Regulatory Commission and the U.S. Environmental Protection Agency required the inclusion of white perch and Atlantic tomcod as Representative Important Species to define and assess their long-term population dynamics in relation to power plant operations (Barnthouse et al., 1988; Dew and Hecht, 1994b). The white perch and Atlantic tomcod were then included as target species in the LRS in 1975. Despite the closure of the white perch fishery due to PCB contamination in February 1976, white perch remain crucial in the HRE due to their abundant population, and the HRE serves as vital spawning and nursery grounds for them (Klauda et al., 1988a), thus white perch eggs, YSL, and PYSL were used. Atlantic tomcod stands out as the only abundant species that spawns during winter in the HRE. Their spawning grounds are mainly located in the lower Hudson River, which sets them apart from other fish species in the river in terms of their spatiotemporal distribution (Dew and Hecht, 1994a). Consequently, they are an important species to examine when assessing the impacts of changes on sampling protocols. Due to winter ice conditions in the river, the survey was not able to consistently sample the egg and YSL stages. Therefore, Atlantic tomcod PYSL and YOY stages were used in this study.

The Hudson River American shad fishery has a rich history in New York, existing for over 200 years and was one of the most profitable shad fisheries on the East Coast at one time, before the fishery closed in 2010 due to depletion of the stock (ASMFC (Atlantic States Marine Fisheries Commission), 2020). Despite the historical, economic and cultural importance of American shad, they did not become a target species in the survey until 1982. American shad are anadromous, relying on the HRE as their spawning and nursery habitat, therefore eggs, YSL, and PYSL were used in this study (Limburg, 1996a).

The Hudson River system accommodates several anadromous fish species that use it as a nursery habitat, yet American eel is the only catadromous fish found in the river, which renders them unique (Mattes, 1989). However, the LRS never intended to target the American eel, although the LRS data provided important stock assessment inputs for American eel juvenile and YROL life stages (ASMFC, 2017). Being a catadromous species, the larval eels migrate from the Sargasso Sea to the HRE after hatching, where they spend most of their lives in brackish or freshwater before returning to the Sargasso Sea to spawn (Schmidt, 1923; Mattes, 1989). Therefore, this study focused on American eels in older life stages (YOY and YROL). Additionally, the American eel is the most widely distributed fish species in the Hudson River system (Mattes, 1989), which makes it an excellent candidate for assessing the impact of changes in sampling protocols on non-target species.

## 2.3 Data calibration procedure

### 2.3.1 Variables selection

Changes in LRS sampling protocol had gone through variation in sampling timing in a day, sampling period in a year, sampling locations, gears, and depth. These changes could affect catch rates over the survey period for different species and hence were treated as predictors in our statistical modeling. We then used the catch-per-unit-effort (CPUE) of different species as the measurement of sampling efficiency, which was calculated by dividing the fish abundance by the filtered water volume in $m^3$. Statistical modeling was built to describe the relationship between these variables with the following multivariate formula:

$$CPUE \sim year + f(DOY) + f(hour) + f(rkm) + f(depth) + gear + f(op.inter.terms) + \epsilon, \quad (1)$$

where CPUE is related to all the sampling protocol variables through their respective functions which are specified by different models. In this formula, *CPUE* denotes the CPUE value for certain species and life stage calculated from the LRS database, *year* denotes the data year, *DOY* denotes the sampling day of the year, *hour* denotes the sampling time of the day rounded to the hour, *rkm* denotes the sampling location measured with river kilometer, *depth* denotes the sampling depth, *gear* denotes the gear used in the records, and *op.inter.terms* denotes optional interaction terms between any variables that may be included in candidate models. *Year* and *gear* were modeled as discrete categorical variables while the other variables were modeled as continuous variables.

Multicollinearity among the predictors was evaluated using Variance Inflation Factor (VIF) analysis for all species and life stages prior to the model selection to avoid error inflation and unreliable coefficient estimates (Hair et al., 2010). Outliers in CPUE data were identified as those that were two times larger than the second largest value, which was then excluded from the model development process. As the LRS dataset did not record zero-catch tows from the survey for each species, we added zero-catch tows to the survey stations that did not have a catch record for the study species.

### 2.3.2 Candidate model exploration

We explored different statistical models to calibrate the catch rates for different species and life stages. We restricted our modeling to data collected from their habitats during the seasons each life stage occurred in the HRE, measured with the HRE river region and DOY, respectively (Table 1). The use of a stratum-based stratified random sampling design with 13 river regions by the LRS facilitates the description of the geographical distribution of a given species/life stage. The seasons of occurrence were determined as the ranges of DOY where the first and last non-zero catch was observed over the time series, and their spatial ranges were determined as the river regions they inhabit during certain life stages, through a literature review and preliminary data review. This design could not only ensure the calibrated survey catchability is ecologically reliable in terms of the species' spatiotemporal dynamics but also reduce the potential bias in model fitting using maximum likelihood methods due to the "complete separation" issue (Albert and Anderson, 1984).

Specifically, the habitat of American shad was defined as the upper HRE (river region 7-12), according to their well-reported spawning activities (Limburg 1995; Limburg, 1996b). The occasional observations of American shad eggs in the lower HRE were assumed to be produced by vagrants from different river systems, indicated by their distinct otolith growth rates and Sr : Ca values (Limbrug, 1995), hence not included in our data calibration.

According to Klauda et al. (1988a) in their study on white perch, the upper zone of the freshwater area, particularly in the Saugerties-Albany regions, had a higher incidence of white perch spawning activity. The white perch eggs and YSL were similarly distributed spatially, as they have limited mobility and minimal downstream transport due to their short life stage duration (Klauda et al., 1988a). On the other hand, PYSL was more widely dispersed across the sampling regions in the upper and middle estuary zones (Klauda et al., 1988a). Therefore, for white perch eggs and YSL, the study area was limited to the upper HRE regions 7-12, while for PYSL, the study area was shifted to river regions 5-11.

Striped bass spawn mostly in the middle regions of the HRE (Boreman and Klauda, 1988), and they move downstream as they grow into larval stages (McLaren et al., 1981). This distributional shift by life stages was further adjusted and determined based on their occurrence. Atlantic tomcod is known as a winter spawner in the lower HRE (Klauda et al., 1988b). Accordingly, the inhabiting river regions for their PYSL and YOY were defined as the lower HRE (river regions 0-7) in this study. The American eel YOY and YROL habitats were defined as the entire HRE, considering that they were observed throughout the river, their well-developed mobility, and their catadromous nature (hatch in the ocean and enter the Hudson River Estuary at a later life stage) (Mattes, 1989). Although early life stages such as eggs and larvae are known to have limited mobility, the study areas were determined to cover a sufficient geographical range to include their distributions.

The LRS CPUE data were found to be heavily zero-inflated (Supplementary Materials Figure S-1), which is a common challenge in ecology statistical modeling that needs to be addressed with specified assumptions in distribution and model selections (Zuur et al., 2009). Recognizing the zero-inflated nature of the data, we conducted a preliminary model trial procedure with a suite of generalized models based on several different responsive variable assumptions that were widely used in modeling zero-inflated data in aquatic ecology. Following the standardized multivariate formula (1) and previous modeling practices (specified as references in the parenthesis), the trialed models included:

- Generalized Linear Models (GLMs) with negative binomial distribution (modeling fish abundance from catch sample data, Power and Moser, 1999);
- Generalized Additive Models (GAMs) with negative binomial distribution (modeling the Gulf of Mexico fish community abundance with climatic and oceanographic factors using a fishery-independent dataset, Drexler and Ainsworth, 2013);
- Generalized Additive Models (GAMs) with Tweedie distribution (modeling juvenile fish distribution with environmental variables and prey abundance in the Yellow Sea using a fishery-independent dataset, Xue et al., 2018);
- Generalized Additive Models (GAMs) with zero-inflated (hurdle) Poisson distribution (modeling juvenile crayfish river and stream habitats in New Zealand using an ecology survey database, Jowett et al., 2008);
- Generalized Additive Models (GAMs) with zero-inflated negative binomial distribution (modeling relative abundance indices of silky shark using data collected by observer programs, Lennert-Cody et al., 2019);

- Generalized Linear Models (GLMs) with zero-inflated negative binomial distribution (billfish CPUE standardization using commercial longline fishery data, Walsh and Brodziak, 2015);
- Generalized Linear Mixed-effects Models (GLMMs) with random intercept and slope (modeling Northwest Atlantic shark abundance using fishery-dependent data, Baum and Blanchard, 2010);
- Generalized Additive Mixed-effects Models (GAMMs) with random intercept and spline (modeling capability of two recreational species in West Australia using catch data, Navarro et al., 2020).

Considering some of the examined distributions were only applicable to discrete count data in ecology (such as negative binomial and Poisson distributions), we specifically modified the modeling techniques to deal with the continuous CPUE data in our study by incorporating offset terms (CPUE = catch.abundance/ sampled.volume).

Despite these efforts, the preliminary model trial procedure showed that only the GAM assuming Tweedie distribution could return converged model outputs, while the other models either did not converge or returned extremely poor fitting with an R square of less than 0.01. Tweedie distribution is a generalization of several probability distributions including normal, gamma, inverse-Gaussian, and Poisson distribution, determined by a power parameter theta, which could be estimated *via* maximum likelihood estimation (Tweedie, 1984). With certain values of theta (1<theta<2), the Tweedie distribution can interpret a compound Poisson-gamma distribution (quasi-Poisson and quasi-negative binomial) in the response variable (Tweedie, 1984; Jørgensen, 1987). This characteristic makes it particularly effective in dealing with zero-inflated fisheries and aquatic data such as CPUE and catch volume (Shono, 2008; Arcuti et al., 2013; Berg et al., 2014). The GAMs with Tweedie distribution were developed with the R package "mgcv" version 1.8-40 (Wood and Wood, 2015).

### 2.3.3 Model selection

Three versions of Tweedie GAM variants were developed as the final candidate models following the multivariate formula (1), including a base version without optional interaction terms, a version with an interaction term between depth and year, and a version with an interaction term between depth and gear. The depth-year interaction was evaluated because the LRS tow depths were inconsistent over the surveyed year, with the tows from the more recent years concentrated in shallower water (Supplementary Materials Figure S-2). The depth-gear interaction was evaluated because the two survey gears (Tucker trawl and epibenthic sled) could have different selectivity by depth, which could result in misspecified catchability even in identical depths (Supplementary Materials Figure S-3).

The three final candidate models were developed for the case study species with their respective life stages. Among the candidate models, we aimed to select the single model that best described the survey catchability based on their goodness-of-fit, which were compared with three indicators: Akaike Information Criterion (AIC), Root Mean Square Error (RMSE), and deviance explained.

AIC is a widely used model selection criterion in ecological modeling (Portet, 2020). It measures the goodness-of-fit as well as model complexity of candidate models to a set of data based on the relationship between maximum likelihood and divergence, with a lower AIC value indicating a better fit. RMSE is a commonly used estimator in fisheries stock assessment to measure the difference between the model-fitted values ($\widehat{CPUE_t}$) and the observed value ($CPUE_i$) with the following equation (Wilberg and Bence, 2008; McCormick et al., 2012):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(CPUE_i - \widehat{CPUE_i})^2}{n}}, \quad (2)$$

where $n$ represents the number of CPUE observations. Deviance explained describes how much the fitted model can reduce the deviance compared to a null model that assumes no relationship between predictors and response variables. The values of deviance explained are always strictly between 0 to 1, with higher values representing better model fit.

### 2.3.4 Model validation

Two model validation procedures were implemented to evaluate the models' predictive performance for different species and life stages, including a K-fold analysis and a retrospective analysis. The statistical analyses were conducted using R version 4.0.3 (R Core Team, 2020).

## 2.4 K-fold analysis

For the K-fold analysis, each dataset was divided into five equally sized subsets (folds). The data were randomly resampled within each year without replacement so each year of data is evenly represented in each of the five subsets. Cross-validation was then performed five times per model, with one subset used as a validation set, and the remaining four subsets combined and used as a training set. During each iteration, the model is trained on the training set and evaluated on the validation set.

To evaluate the models' performance, RMSE, root relative square error (RRSE), and Spearman correlation coefficient estimated from each fold were used. The RMSE quantifies the average magnitude of the differences between predicted values and actual values. The RMSE indicates a perfect match between observed and predicted values when it equals zero, with higher RMSE values indicating an increasingly poor match (Kouadri et al., 2021).

The RRSE indicates how well a model performs relative to the average of the true values. Therefore, when the RRSE is lower than one, the model performs better than the simple model. Hence, the lower the RRSE, the better the model. The square root of the relative squared error is used to reduce the error to the same units as the predicted quantity (Kouadri et al., 2021):

$$RRSE = \frac{\sqrt{\sum_{i=1}^{n}(CPUE_i - \widehat{CPUE_i})^2}}{\sqrt{\sum_{i=1}^{n}(CPUE_i - \overline{CPUE})^2}}, \quad (3)$$

where $\overline{CPUE}$ denotes the mean observed CPUE value.

The correlation coefficient assesses the strength and direction of the monotonic relationship between the observations and predictions. Due to a large number of zeros and skewed distribution with extremely large values in the data, the Spearman correlation coefficient (rho) was used (Spearman, 1904).

## 2.5 Retrospective analysis

A retrospective analysis was performed by sequentially removing the data from the most recent years, fitting the best-fit models, and comparing the terminal year estimates (Kell et al., 2021). Retrospective analysis is widely used in quantitative fisheries science to understand the consistency of a statistical model's performance over time, providing key diagnostic evidence for accepting or rejecting a model. The process of sequentially removing the last year's data is called "peeling". In this study, we performed a five-year peel for all the best-fit models and focused on the estimates of year-effects, as they represented the temporal trend in fish abundance, which was the most important stock status indicator. Specifically, we performed the peeling procedure by sequentially removing all data from the terminal year (2017) by a one-year step until five years (when 2012 became the terminal year). The model was then refitted with each set of truncated time series data using the same variable and model structure. We then compared the terminal year estimates of stock abundance to the full model estimate for that year for potential retrospective errors. We used a quantitative indicator, Mohn's rho (Mohn, 1999), to measure the magnitude of the five-year retrospective errors, which is calculated as:

$$Mohn's\ \ rho = \frac{1}{5}\sum_{t=T-5}^{T-1}\frac{y_{(1:t),t} - y_{(1:T),t}}{y_{(1:T),t}} \quad (4)$$

where $T$ is the terminal year of the complete data series, $t$ is the terminal year of the peeled data series, $y_{(1:t),t}$ is the model-based year effect estimated for the terminal year using the peeled data series, $y_{(1:T),t}$ and is the model-based year effect estimated for the terminal year using the full data series. Mohn's rho ranges between -1 to 1, with a value close to 0 representing a negligible retrospective pattern in the model, indicating consistent model performance with different lengths of time-series data. According to the earlier simulation analyses based on integrated, age-structured models with different species (Hurtado-Ferro et al., 2015), a Mohn's Rho is considered reflective of the existence of a retrospective pattern when its value is higher than 0.20 or lower than −0.15 for longer-lived species, or larger than 0.30 or lower than −0.22 for shorter-lived species, though these thresholds may not apply to age-aggregated CPUE estimates as in the presented study.

## 2.6 Model calibration effects

We measured the model calibration effects by comparing the model-based abundance indices (year effects estimated from the Tweedie GAMs) with the design-based abundance indices.

The design-based annual abundance indices have been historically used for evaluating fish abundances for key species. The design-based annual abundance indices ($I$) were calculated as averaged density (number of individuals divided by the volume of water sampled) over all surveyed regions, strata, and weeks (ASA Analysis & Communication (ASAAC), 2016):

$$I = \sum_{w=first\ \ week}^{last\ \ week}\left[\frac{\sum_{i=1}^{12}\sum_{s=1}^{3}V_{i,s}(\sum_j \frac{Ct_{j,i,s,w}}{v_{j,i,s,w}})}{\sum_{i=1}^{12}\sum_{s=1}^{3}V_{i,s}}\right] \quad (5)$$

where $Ct_{j,i,s,w}$ denotes the number of individuals of a species in sample $j$, region $i$, stratum $s$, and week $w$, $v_{j,i,s,w}$ denotes the volume of water sampled for sample $j$ in region $i$, stratum $s$, and week $w$, and $V_{i,s}$ denotes the volume of stratum $s$ in river region $i$, and *first week* denotes the first week of a year in which the accumulative weekly density estimates exceeds 5% of the sum of densities over all weeks of sampling, and *last week* is defined as *first week* + 7 weeks.

To make the annual abundance comparable over time, only river regions 1-12 were used as the Battery (river region 0) was not sampled until 1988. The weeks used for eggs, YSL, and PYSL of striped bass, American shad, and white perch were their proposed peak seasons, assuming an 8-week long duration of spawning season. For Atlantic Tomcod, due to ice conditions in the River, the LRS was unable to consistently sample the YSL stage. However, an abundance index for the period when the transformation from PYSL to the juvenile stage occurred could be calculated for weeks 19-22 (approximately DOY 127-154). This period roughly corresponds to the month of May, and the abundance of Age 0 tomcod was calculated from LRS data for these four weeks (ASA Analysis & Communication (ASAAC), 2016). The annual abundance of American eel was estimated based on data from weeks 18-26 when the survey was conducted throughout the river, assuming that the occurrence of YOY and YROL eels takes place during the spring and early summer (Mattes, 1989). To compare if the proposed weeks (PW) have effects on the estimates, we also estimate abundance indices using all weeks (AW). Pearson's correlation coefficient was used to evaluate the correlation between design-based and model-based abundance indices (Shono, 2008).

Additional analyses were conducted for the Atlantic tomcod PYSL due to their unique spatiotemporal distribution. Although the Atlantic tomcod was included as a Representative Important Species in the survey in 1975 (Barnthouse et al., 1988), the allocation of sampling effort focused on the collection of Atlantic tomcod PYSL was discontinued in 1981 and was not resumed until 1995 (ASA Analysis & Communication (ASAAC), 1996). The missing allocation of sampling effort for tomcod as well as changing survey start dates (DOY) over the years (Figure 2) might have impacts on the estimates of PYSL abundance indices as the tomcod peak season for PYSL was reported from mid-March to mid-April (Klauda et al., 1988b). It is hypothesized that the abundance index would be negatively correlated with the survey start DOY as the early survey start DOY was more likely to include the peak season of tomcod PYSL, considering the survey started in late April or May that could have missed the peak season of high PYSL density. Furthermore, previous studies (Dew and Hecht, 1994a; Dew and Hecht, 1994b) indicated that a significant amount of tomcod 0-age abundance (29-45%) was not accounted for when the survey missed the seaward of the Yonkers region from March through May. Therefore, river regions 0-12 data collected from 1988 to 2017 were used to estimate design-based abundance indices for Atlantic tomcod PYSL and YOY using all weeks and proposed weeks to evaluate the potential impacts of exclusion of the Battery area. It is hypothesized that the inclusion of the Battery area data would improve the estimates of abundance indices, assuming a significant proportion of tomcod postlarvae and juveniles distributed in the Battery area (Klauda et al., 1988b; Dew and Hecht, 1994a; Dew and Hecht, 1994b).

The model-based abundance indices were denoted as $I_M$. The design-based abundance indices using the proposed weeks' data were denoted as $I_{PW}$, and the design-based abundance indices using the proposed weeks' data were denoted as $I_{AW}$ for all species except for Atlantic tomcod PYSL. For the Atlantic tomcod PYSL, the design-based abundance indices using data with the inclusion of Battery area were denoted as $I_{PW\ (0\text{-}12)}$ and $I_{AW\ (0\text{-}12)}$, and the design-based abundance indices using data without the inclusion of Battery area were denoted as $I_{PW\ (1\text{-}12)}$ and $I_{AW\ (1\text{-}12)}$.

# 3 Results

## 3.1 Changes in sampling protocol

The changes in sampling protocol of the LRS have altered the catchability of different ichthyoplankton in the HRE. The duration of sampling varied yearly, with inconsistencies in the start and end dates (Figure 2). The number of days of sampling also varied, ranging from the 32 days of sampling in 1982 to 104 days of sampling in 1995. In addition to varying sampling duration, differences in diel timing of sampling fluctuated. Sampling during the day was conducted in the early weeks of the sampling season, for the years in which there was daytime sampling, starting as early as February and as late as May, and ending as early as March and as late as July (Figure 2). Nighttime sampling was conducted for later river runs, starting as early as February and as late as June, and ended as early as July and as late as October, with little overlap between daytime and nighttime sampling each year (Figure 2).

The Tucker trawl and epibenthic sled were inconsistently used throughout the LRS. The annual sampling intensity for the sled ranged from 270 tows in 2012 to 1591 tows in 1976, and for the trawl ranged from 982 tows in 1982 to 1974 tows in 1976 (Figure 3). The location of sampling also varied by year and by gear. The survey area expanded over the duration of the LRS (Figure 1) and the two gears were disproportionately used in different river regions. The differences in gear and sampling locations also resulted in changes in sampling depths throughout the LRS (Supplementary Materials Figure S-2).

## 3.2 Model fitting

Three versions of Tweedie GAM variants (a base version "tw.gam.base", a version with depth and year interaction "tw.gam.depth.year", and a version with depth and gear interaction "tw.gam.depth.gear") were compared for their goodness-of-fit with AIC, deviance explained, and RMSE (Supplementary Materials Table S-1). The Tweedie GAM with depth and gear interaction performed the best among all cases (except for the Atlantic tomcod PYSL where it did not converge), exhibited particularly by AIC. The base model without interaction always had relatively higher AIC values and lower deviance explained and RMSE. However, the discrepancies in deviance explained and RMSE were mostly negligible. Considering the simplicity of base models and low computational demands (required less than 5% computation time given the large size of LRS), the base model demonstrated a more favorable tradeoff between model complexity and goodness of fit. Therefore, the base model was selected as the optimal calibration model for the following analyses.

Relationships between the survey CPUE and the six considered predictors were modeled using the base Tweedie GAM. The four spatio-temporal factors in the LRS were found to be strongly related to the predicted CPUE values for all species and life stages

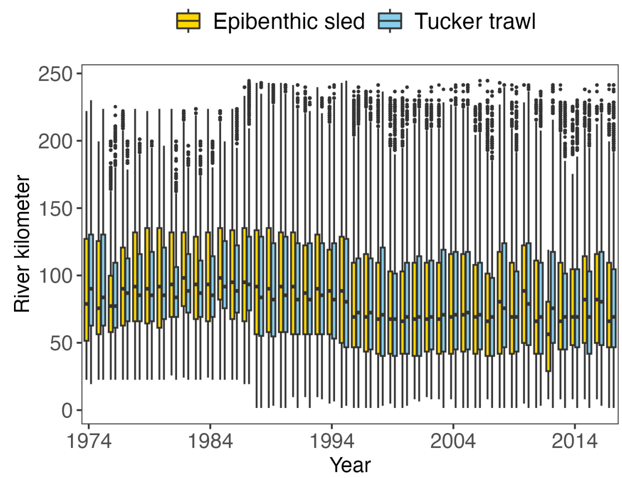



FIGURE 3
Boxplots of the usage of the two gears by river kilometer for all sampling tows for all species in each year of the LRS. The horizontal bars in the boxes are medians. The bottom and top limits of the boxes are the first (Q1) and third (Q3) quartiles (25th and 75th percentiles). The difference between Q1 and Q3 is the interquartile range (IQR). Potential outliers are defined as observation points that fall outside the range of Q1-1.5*IQR and Q3 + 1.5*IQR. If potential outliers are presented, the whiskers extend to 1.5 times the IQR from Q1 or Q3. If no outliers are presented, the whiskers extend to the minima and maxima of the distributions. Purple boxes denote sampling using an epibenthic sled, and green boxes denote sampling using a Tucker trawl.

examined, and the year and gear effects were also significant (Supplementary Materials Figure S-4 to S-16). Peaks in DOY, hour, and Rkm were obvious and unsynchronized among species and life stages, indicating varied seasons and hours of appearance and use of habitat in the HRE. CPUE demonstrated fluctuations over the surveyed years, which could reflect model-based fish abundance (results and analysis see section 3.2). The epibenthic sled consistently displayed significantly higher median CPUE values compared to the Tucker trawl. In only 7 out of 13 cases did their CPUE distributions overlap, indicating considerable gear effects in survey catchability.

## 3.3 Model calibration effects

The model-based and design-based abundance indices were shown in Figure 4. The Pearson's correlation coefficients (*r*) of the model-based and design-based abundance indices were all statistically significant for all species and life stages with a mean r of 0.76 except for Atlantic tomcod PYSL $I_{AW\ (0\text{-}12)}$ and $I_{AW\ (1\text{-}12)}$ (Figure 5 and Supplementary Materials Table S-2). The model-based and design-based abundance indices were generally highly correlated ($r$>0.73) for all species either using the PW data (see methods) or AW data except Atlantic tomcod.

The *r* of $I_M$ and $I_{PW\ (1\text{-}12)}$ was 0.62 (p<0.05, df=42) for tomcod PYSL; however, the *r* of $I_M$ and $I_{AW\ (1\text{-}12)}$ was only 0.20 (p=0.20, df=42). With the inclusion of the Battery river region data, the $I_{PW\ (0\text{-}12)}$ still had a higher correlation coefficient with the $I_M$ ($r$=0.39, p<0.05, df=28) than $I_{AW\ (0\text{-}12)}$ ($r$=0.22, p=0.236, df=28). It should be noted that the estimates using inclusion and exclusion of the Battery data were not directly comparable as the inclusion of Battery data was only available from 1988-2017.

On the other hand, the *r* of $I_M$ and $I_{AW\ (1\text{-}12)}$ was 0.70 (p<0.05, df=42) for tomcod YOY; while the *r* reduced to 0.46 (p<0.05, df=42) for $I_M$ and $I_{PW\ (1\text{-}12)}$ (Table S-2). With the inclusion of the Battery data, the r of $I_M$ and $I_{PW\ (0\text{-}12)}$ for tomcod YOY was 0.72 (p<0.05, df=28) and the *r* of $I_M$ and $I_{AW\ (0\text{-}12)}$ for tomcod YOY was 0.93 (p<0.05, df=28).

Without the inclusion of the Battery data (1974-2017 time series), the $I_{AW\ (1\text{-}12)}$ for tomcod PYSL was negatively correlated with survey start DOY (Figure 6A, $r$=-0.375, p<0.05, df=42) as hypothesized. The $I_{PW\ (1\text{-}12)}$ was positively correlated with survey start DOY, although it was not statistically significant (Figure 6B, $r$=0.14, p=0.365, df=42). Similarly, the $I_M$ (1974-2017) were positively correlated with survey start DOY (Figure 6C, $r$=0.331, p<0.05, df=42), reflecting declining abundance (Figures 3 and 4) when the survey started earlier and resumed the allocation of sampling effort for tomcod PYSL after 1995.

With the Battery data (1988-2017 time series), the $I_{PW\ (0\text{-}12)}$ and $I_{AW\ (0\text{-}12)}$ were negatively (Figure 6D, r=-0.138, p=0.468, df=28) and positively (Figure 6E, $r$=0.132, p=0.485, df=28) correlated with survey start DOY, respectively, yet they were not statistically significant. The $I_M$ (1988-2017) was still positively correlated with survey start DOY (Figure 6F, $r$=0.563, p<0.05, df=28).

## 3.4 Model validation

### 3.4.1 K-fold analysis

The RMSEs varied among species and life stages, depending on the statistics of each dataset. In general, the RMSEs estimated from the 5 trials generate similar RMSEs with low variation (Figure 7A and Supplementary Materials Table S-3). However, a few models (shad eggs and striped bass eggs) showed a wider range of RMSEs as the RMSE is sensitive to extreme values, reflecting the nature of the datasets. The rho between observations and predictions are all significantly different from zero at the significance level of 0.05, suggesting the observations and predictions are significantly correlated (rho around or higher than 0.5) (Figure 7B and Supplementary Materials Table S-3). However, the rho for eel YOY is notably lower (c. 0.25), possibly due to a lack of data during the 1980s. For most models, the RRSEs were below or around 1 (Figure 7C and Supplementary Materials Table S-3), suggesting reduced error compared to the simple models. Nevertheless, the RRSEs were above 1 for shad YSL and white perch eggs, indicating that the model performance was not satisfactory and caution should be taken for the accuracy of the estimates for these two models.

### 3.4.2 Retrospective analysis

A five-year retrospective analysis indicated negligible retrospective errors in the optimal calibration model (base Tweedie GAMs) according to the estimated Mohn's rho (Figure 8). The largest absolute value of Mohn's rho was observed with "American shad egg" at 0.102, which did not indicate noticeable retrospective pattern (<0.3) for such a short-lived species. White perch had the smallest Mohn's rho values for all its life stages compared with other examined species, indicating the most stable model fitting performance over time. Eggs tended to possess the largest absolute Mohn's rho values compared with other life stages, implying a relatively stronger retrospective pattern in the optimal calibration model for egg abundance, despite their extremely low levels.

While the retrospective patterns were not strong for the terminal years (based on which the Mohn's rho values were estimated), there were still some retrospective patterns observed for the intermediate years. Specifically, variabilities between different "peel" models were observed around 2010 for striped bass egg, white perch egg, and white perch YSL. However, the relative range of these variabilities was always smaller than 15%, indicating low risks of retrospective patterns with the optimal calibration models.

# 4 Discussion

## 4.1 Model-based abundance indices can mitigate the spatiotemporal biases in design-based abundance indices

With the growing utilization of long-term data sets for establishing baseline reference points in aquatic environments, it

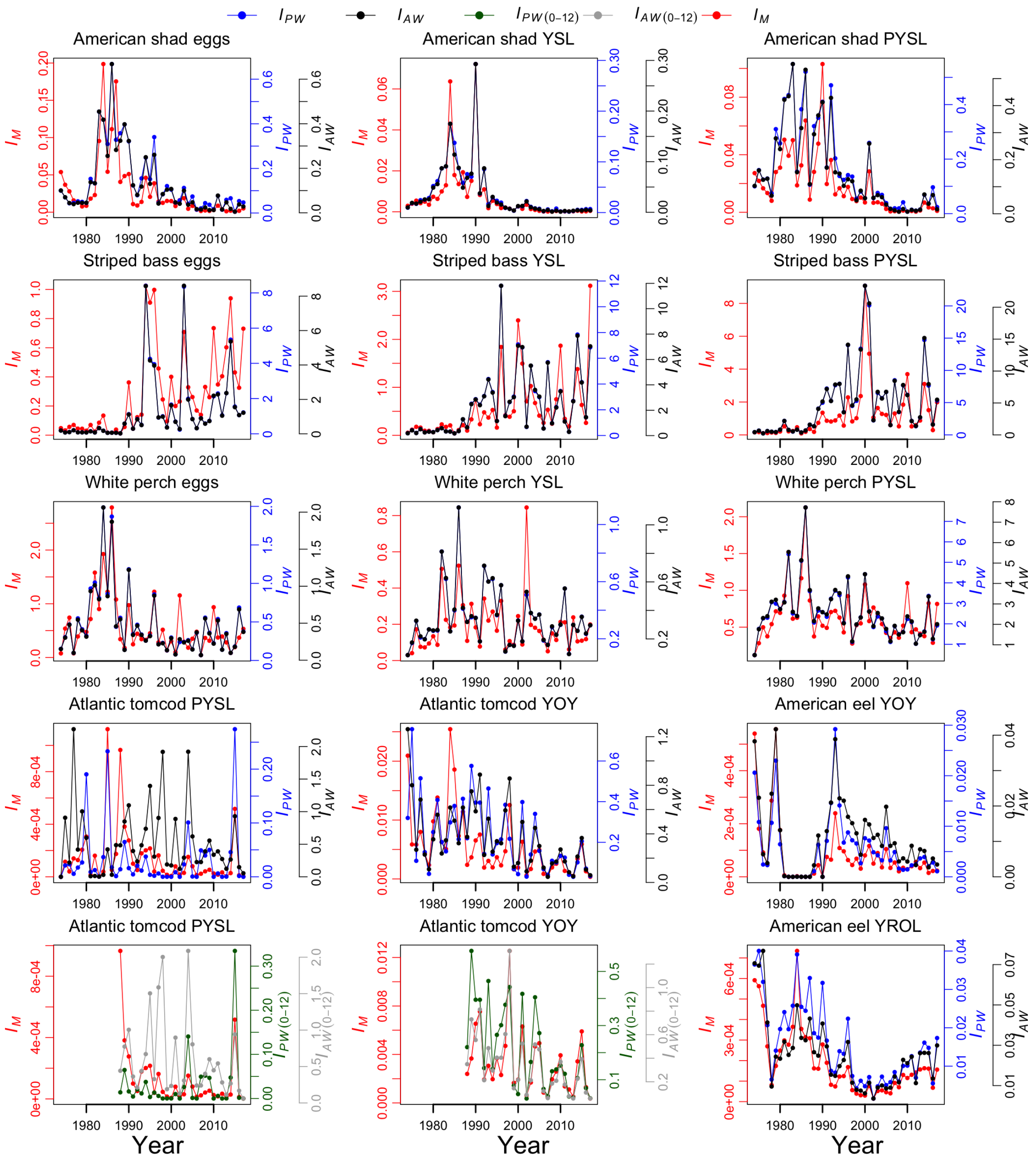



FIGURE 4
Model-based ($I_M$) and design-based abundance indices for each species and life stage. $I_{PW}$ denotes the design-based abundance indices estimated using the proposed weeks' data. $I_{AW}$ denotes the design-based abundance indices estimated using all weeks' data. For the Atlantic tomcod, the $I_{PW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using proposed weeks data with the inclusion of Battery, and the $I_{AW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using all weeks data with the inclusion of Battery. PYSL, post-yolk-sac-larvae; YOY, young-of-year; YROL, yearling and older; YSL, yolk-sac-larvae.

becomes increasingly important to understand any biases or evaluate the uncertainty that may arise from changes in sampling strategies or protocols (Tuckey and Fabrizio, 2013). In this study, the model-based and design-based abundance indices generally exhibited high correlation, despite inconsistencies in the sampling protocol across various areas and time periods resulting in some unsampled areas and inconsistent survey durations and shifts between day and night. This suggests that the design-based abundance indices provide valuable information that is comparable to the model-based abundance indices, even without accounting for sampling effects resulting from changes in the protocol, when considering the annual and river-wide scale.

However, abundance can be underestimated or overestimated when important factors were not considered (e.g., peak season, major spatial distribution). The model-based abundance indices take these factors (e.g., spatial and temporal variables) into account and can address the inconsistent sampling issue. Furthermore, the model estimated sampling effects reflected the observations in other studies.

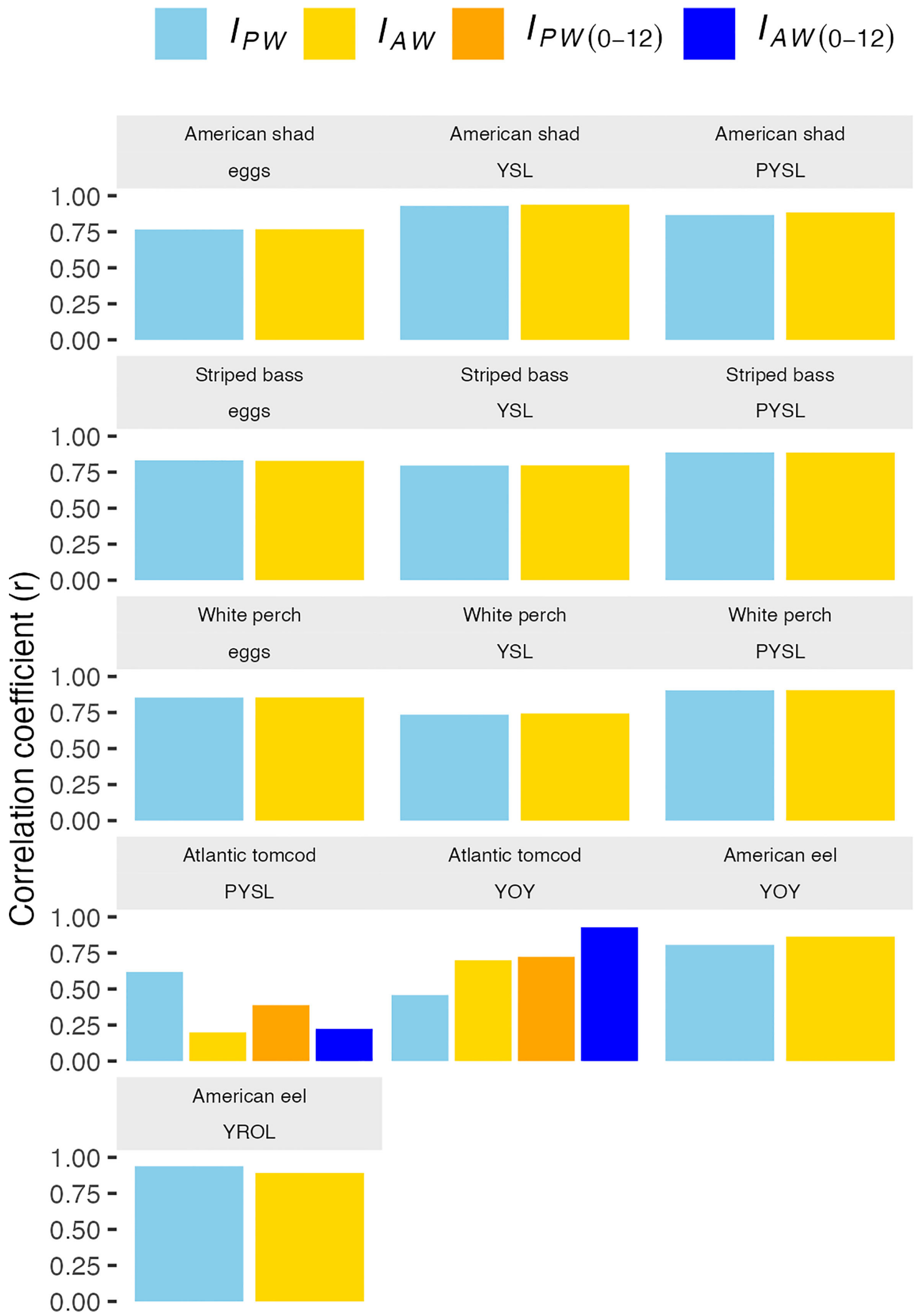



FIGURE 5
Pearson's correlation coefficients between model-based and design-based abundance indices. $I_{PW}$ denotes the design-based abundance indices estimated using the proposed weeks' data. $I_{AW}$ denotes the design-based abundance indices estimated using all weeks' data. For the Atlantic tomcod, the $I_{PW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using proposed weeks data with the inclusion of Battery, and the $I_{AW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using all weeks data with the inclusion of Battery. PYSL, post-yolk-sac-larvae; YOY, young-of-year; YROL, yearling and older; YSL, yolk-sac-larvae.

For example, the model estimated the two spatial peaks for striped bass eggs around rkm 90 and rkm 140 (Supplementary Materials Figure S-11), which corresponded to the observations in Boreman and Klauda (1988). Also, even with the inconsistent survey start DOY, the model estimated the highest density of Atlantic tomcod PYSL occurred during DOY 70-120 (Supplementary Materials Figure S-9), which corresponded to the observations in previous studies (Klauda et al., 1988b; Dew and Hecht, 1994a). On the contrary, the design-based indices may be sensitive to the inconsistent allocation of sampling effort over space and time, especially for early life stages. If the peak of spawning for eggs and YSL occurs in a specific location and time, it can be challenging to capture the maximum abundance of these life stages both spatially and temporally, especially since they last for less than a week (Boreman and Klauda, 1988). The Atlantic tomcod illustrated an example that the estimates could be considerably affected by the inconsistent sampling protocol.

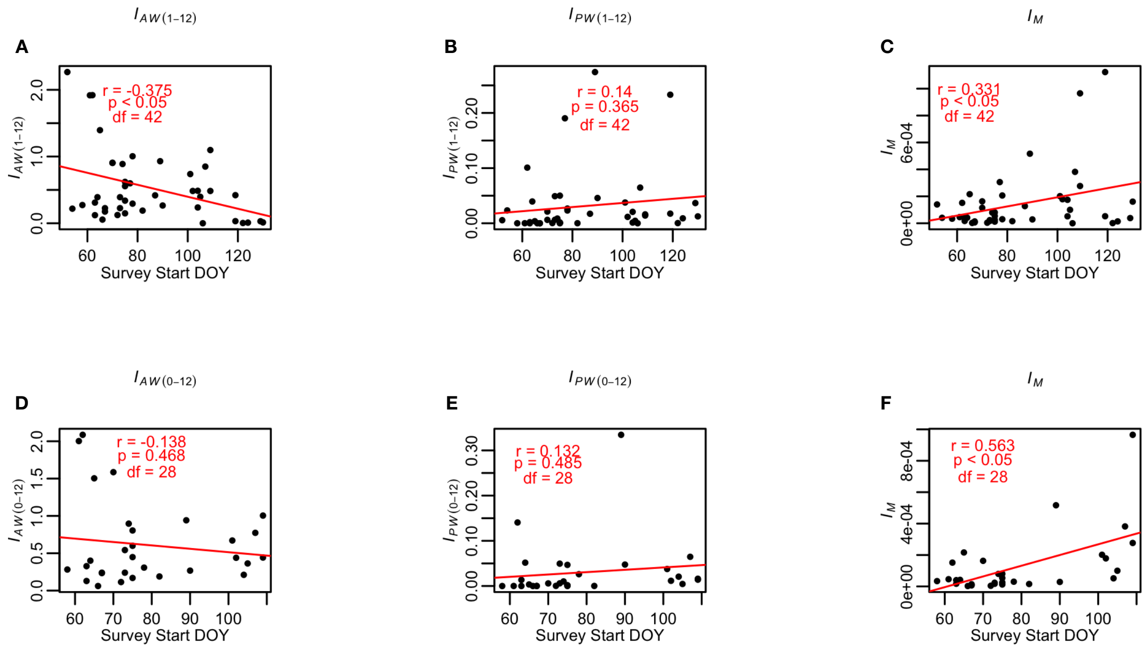



FIGURE 6
Relationships between **(A)** $I_{AW\ (1\text{-}12)}$ and survey start DOY (day of year); **(B)** $I_{PW\ (1\text{-}12)}$ and survey start DOY; **(C)** $I_M$ and survey start DOY, using 1974-2017 time series, and **(D)** $I_{AW\ (0\text{-}12)}$ and survey start DOY; **(E)** $I_{PW\ (0\text{-}12)}$ and survey start DOY; and **(F)** $I_M$ and survey start DOY using 1988-2017 time series data for the Atlantic tomcod PYSL. $I_{PW}$ denotes the design-based abundance indices estimated using the proposed weeks' data. $I_{AW}$ denotes the design-based abundance indices estimated using all weeks' data. $I_{PW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using proposed weeks data with the inclusion of Battery, and the $I_{AW\ (0\text{-}12)}$ denotes the design-based abundance indices estimated using all weeks data with the inclusion of Battery. A trendline using linear regression analysis was added to each of the panels, denoting the trend of the correlation.

Dew and Hecht (1994a; 1994b) pointed out that it is necessary to include the most seaward region of the estuary (Battery) to define a self-contained, measurable population of larval and early juvenile Atlantic tomcod. Our results showed that the inclusion of the Battery region did improve the estimates of abundance for tomcod YOY, suggesting that there is a considerable amount of Atlantic tomcod juveniles distributed in the Battery area over the season. However, the estimates of abundance with the inclusion of Battery may still be biased for the tomcod PYSL if the peak season was missed in several years. In other words, even if the majority of the spatial distribution of the population was covered by the survey, the inconsistent sampling protocol would still have significant impacts on the estimates of abundance if the peak season was missed in several years, especially for early life stages that generally have sharp seasons. Timing in relation to the seasonal cycle and location of the target species, and the fact that only a limited amount of data can be collected, are considered to be two main deficiencies in fishery-independent surveys, which could lead to unrepresentative sampling (Hilborn and Walters, 1992; Pennino et al., 2016). The Atlantic tomcod in this study provides an example where the estimates derived from the fishery-independent survey could be biased.

Although abundance indices were used for evaluating changes in annual abundance for each species, especially for early life stages due to their high mortality rates, it should be noted that having more accurate absolute abundance estimates can provide valuable insights and benefits for fisheries management. Absolute abundance estimates provide information on the size and productivity of the population, which is crucial for setting appropriate fishing quotas or catch limits. Reliable abundance estimates contribute to more effective and sustainable management practices. Furthermore, absolute abundance estimates could be used to identify threatened or endangered populations (e.g. sturgeon species), monitoring population recovery efforts, and assessing the effectiveness of conservation measures. While relative indices provide useful information for assessments, having more accurate absolute abundance estimates adds value from a management perspective.

## 4.2 Sampling efficiencies on the target and non-target species

Long-term fishery-independent survey datasets often involve the addition of new target species, which may require modifications to the sampling protocol. However, the effects of these changes on both target and non-target species are often overlooked, even though data from non-target species can offer valuable insights into population dynamics and ecosystem dynamics. The design-based abundance indices for the eggs and larval stages of most target species (striped bass, American shad, and white perch) included in this study showed a strong correlation with the model-based abundance indices, whether using all weeks or only the proposed week data. This could be due to the fact that early life stages of fishes tend to have shorter seasons compared to juvenile and older stages (Boreman and Klauda, 1988), and the assumption of the design-based abundance indices that the periods of early life stages present

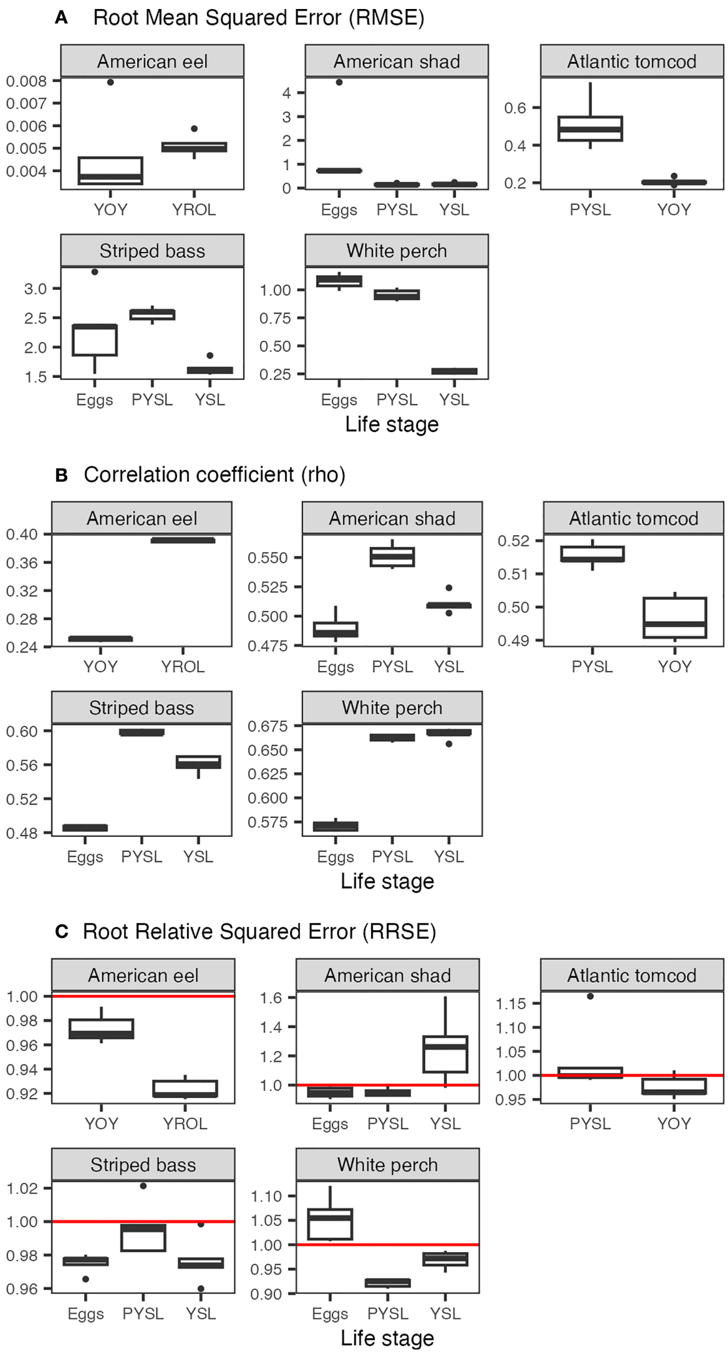



FIGURE 7
**(A)** Root mean squared error (RMSE); **(B)** Correlation coefficient (rho); and **(C)** Root relative squared error (RRSE) of the 5 folds for models of each species and life stage. The red lines indicate 1. PYSL, post-yolk-sac-larvae; YOY, young-of-year; YROL, yearling and older; YSL, yolk-sac-larvae. Note that the y-axis range differs among the panels.

in the river last seven weeks seemed reasonable for the three target species (Heimbuch et al., 1992) when estimating annual abundance indices.

Although both white perch and Atlantic tomcod were included as Representative Important Species since 1975 due to their high abundance and susceptibility to impingement and entrainment (Barnthouse et al., 1988; Klauda et al., 1988a), there was no allocation of sampling effort for tomcod during 1981-1994. The exact reason for over ten years of discontinuation of sampling allocation for the Atlantic tomcod was not clear. It is possibly because of the unique spatiotemporal distribution of the Atlantic tomcod as the only abundant winter spawners in the lower HRE, making it different from other target species (Dew and Hecht, 1994a). This, however, suggests that being considered as a target species did not guarantee better data quality compared to non-target species, particularly when the sampling events were not carried out consistently. As suggested by Dew and Hecht (1994a), a sampling plan that is designed to capture other major species in

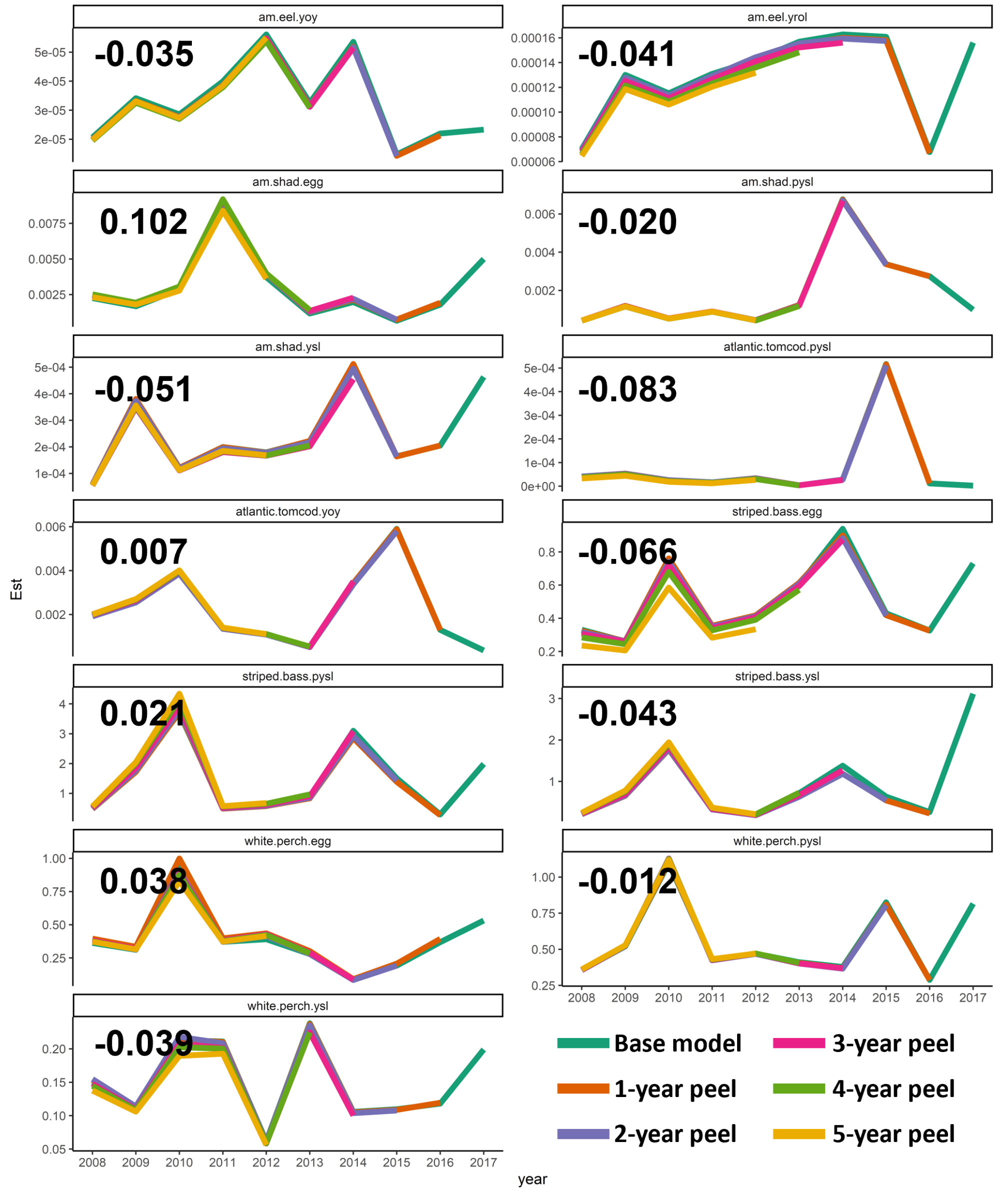



FIGURE 8
Retrospective trajectories of the year effects for the most recent 10 years for the optimal calibration model (The base Tweedie GAM). The calculated Mohn's rho values are shown in the corresponding panel. Est represents the estimated CPUE value. PYSL, post-yolk-sac-larvae; YOY, young-of-year; YROL, yearling and older; YSL, yolk-sac-larvae.

the Hudson River may not be optimal for the Atlantic tomcod due to its unique characteristics.

On the contrary, the abundance indices of American eel showed high correlations ($r > 0.8$, $p < 0.05$, $df = 42$) between the $I_{AW}$ and $I_{PW}$ with $I_M$. Despite never being a target species in the LRS, this finding suggested that the survey had adequately covered the major spatiotemporal distributions of American eel YOY and YROL in most years, even with the changing sampling protocol. However, during the 1980s, several years had no catch data for American eel YOY (1982-1987 and 1990), and only two YOY were observed in 1984, making it challenging to evaluate the effect of survey start date on the abundance index estimates, given that most early survey start dates occurred in the 1980s (Figure 2). It is unclear whether American eel YOY were not observed or were not considered in the survey during those years, as American eel was not regarded as a target species in the LRS.

## 4.3 Spatiotemporal effects in data calibration

For long-term fishery-independent surveys, spatiotemporal scales can be an important factor for assessing the accuracy and uncertainty associated with the estimates. The impact of changes in sampling protocol on estimates varies depending on the scale and purpose of the analysis, as shown in this study. When examining annual and river-wide trends, the design-based abundance indices for most species were consistent with model-based abundance indices, indicating that the major spatiotemporal distributions were well captured by the survey and that the sampling protocol changes did not significantly affect the estimates. However, when analyzing spatiotemporal changes at a finer scale, estimates may be biased or incomparable over time. For instance, the onset of the spawning season for American shad could not be accurately estimated in some years due to a late start in the survey. Additionally, when evaluating distributional shifts over time, the sampling location effects must be taken into account, as some areas were not sampled in early years, which could bias the estimates, especially for species distributed in both ends of the survey area (Dew and Hecht, 1994a; Dew and Hecht, 1994b). Although this study considered several significant factors, there may be other potential factors that can influence the estimates, such as variations in the spawning season due to the lunar phase (Takemura et al., 2004) or climate-related environmental changes (O'Connor et al., 2012). Similarly, distributional shifts over time may be caused by factors such as water quality changes, habitat alterations, invasive species, and anthropogenic activities, which are beyond the scope of this study and require further investigation.

## 4.4 Implications for future survey data calibration and sampling design

This study highlights the importance of identifying target species when designing fishery-independent surveys, as they determine the necessary spatiotemporal coverage of the survey. The results derived from this study indicate that the survey should sufficiently cover the significant spatiotemporal distributions of the target species. This study emphasizes how sampling protocol changes could result in biased estimates of abundance indices (e.g. Atlantic tomcod), providing valuable insights for future sampling protocols. Additionally, the model-estimated spatiotemporal distributions for each species and life stage provide critical information for designing future sampling allocations.

The calibration models developed in this study were effective in removing the effects not directly related to abundance and accounting for changes in the sampling protocol over time. The employed Tweedie GAMs can produce robust data calibration effects with different sample sizes and lengths of time-series data according to the retrospective analysis. However, for a few species and life stages, the uncertainty of the estimates should be taken with caution. For example, the shad YSL and white perch eggs models' performance were not satisfactory, suggesting that there may be other important variables that were not included in the models driving the changes in CPUEs. On the contrary, the retrospective analysis showed that the developed calibration model performed consistently with different lengths of time-series survey data, indicating a relatively stable catchability pattern over the historical years. The K-fold analyses also showed satisfactory prediction performance for most species and life stages, demonstrating relatively consistent calibration effects over varying sample sizes. However, a calibration model update is still required when extreme climatic or environmental events are observed in the HRE ecosystem, as they may drastically affect the spawning dynamics of ichthyoplankton and result in altered survey efficiency.

Some caveats should be noted when interpreting the prediction results over the calibration models. First, the LRS did not have a cross-design sampling scheme which could generate full combinations of all sampling variables at all levels. This could disallow the use of mixed-effect models that assume nested design in data sampling (Schielzeth and Nakagawa, 2013) and result in limitations in predicting the sampling catchability over space and time, although these variables were treated as continuous variables (Webster et al., 2020). In most years, the daytime sampling started first in the year and switched to nighttime sampling to reduce gear avoidance by the PYSL (Boreman and Klauda, 1988), while each species and life stage have varying spawning and growth schedules. Furthermore, there was no or very limited daytime sampling during 1987-1994. Therefore, it should be noted that the daytime and nighttime effects on CPUE might be an artifact resulting from the sampling protocol. Second, the nature of the LRS data poses additional challenges in modeling and predicting sampling catchability. Specifically, the records on ichthyoplankton juvenile abundance (measured with "count") are often in decimal numbers as they were expanded from subsamples collected for laboratory processes, which is a common and standard procedure in collecting juvenile surveys. The dominance of zero tows further adds to the complexity of the data distribution and they appear with various sampling efforts (measured with "water volume filtered"). We chose to use CPUE as an abundance index for the data calibration based on Tweedie GAM, which was the only option that could best address these data issues. However, the smoothing effects in the GAMs still could not perfectly predict the zero CPUE value, which limited its predictive power for extremely low and high catch scenarios.

The identified effects in the sampling designs not only can provide a baseline to calibrate the historical LRS dataset but also offer valuable insights for developing and optimizing future survey designs. The statistical patterns in the sampling factors (such as sampling season, time, and location) highlight improved or reduced survey efficiencies for different species as well as life stages in the HRE. This knowledge allows for more effective sampling for species with emphasized conservation or management demands, while a tradeoff still exists between species-specific and whole-community levels survey objectives. To ensure better data calibration quality, it is recommended to conduct some more standardized samplings following a strict cross-design. This will generate comparable catch records in terms of sampling design and hence enable the evaluation of relative catch efficiencies using more statistical approaches. Gaining a thorough understanding of how to apply available data sets and recognizing their limitations will provide valuable support to scientists and managers who are confronted with uncertainties in research surveys and are tasked with the challenge of effectively managing resources.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author/s.

## Author contributions

Conceptualization: H-YC, MS, KR, YC. Literature review: H-YC, MS, KR. Design and methodology: H-YC, MS, KR, YC. Statistical analysis and interpretation: H-YC, MS. Writing-original draft preparation: H-YC, MS, KR. Writing-review and editing: H-YC, MS, KR, YC. Supervision and securing funds: YC. All authors contributed to the article and approved the submitted version.

## Funding

This study was funded by a Hudson River Foundation research grant (B01/22A).

## Acknowledgments

We express our gratitude to Patrick Sullivan, Gregg Kenney, Richard Pendleton, John Maniscalco, and the NYSDEC and Hudson River Foundation staff for their valuable inputs and discussions. The HRBMP data have been generously gifted to Stony Brook University by the Entergy Corporation. We also extend our appreciation to all those who have contributed to the HRBMP throughout the years.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1237549/full#supplementary-material

## References

Ahlstrom, E. H. (1954). Distribution and abundance of egg and larval populations of the Pacific Sardine. *Fishery. Bull.* 56, 83–140.

Albert, A., and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 71 (1), 1–10. doi: 10.1093/biomet/71.1.1

Arcuti, S., Calculli, C., Pollice, A., D'Onghia, G., Maiorano, P., and Tursi, A. (2013). Spatio-temporal modelling of zero-inflated deep-sea shrimp data by Tweedie generalized additive. *Statistica* 73 (1), 87–101. doi: 10.6092/issn.1973-2201/3987

ASA Analysis & Communication (ASAAC). (1996). *1995 year class report for the Hudson river estuary monitoring program* (Marlboro, NY, USA: Prepared for Entergy Nuclear Operations, Inc).

ASA Analysis & Communication (ASAAC). (2016). *2016 year class report for the Hudson river estuary monitoring program* (Marlboro, NY, USA: Prepared for Entergy Nuclear Operations, Inc).

ASMFC (Atlantic States Marine Fisheries Commission). (2017). *American Eel Stock Assessment Update* (Arlington, VA: ASMFC), 123.

ASMFC (Atlantic States Marine Fisheries Commission). (2020). *American shad benchmark stock assessment and peer review report* (Arlington, VA: ASMFC), 1208.

Barnthouse, L. W., Klauda, R. J., and Vaughn, D. S. (1988). What We Didn't Learn about the Hudson River, Why, and What it Means for Environmental Assessment. *Am. Fisheries. Soc. Monograph.* 4, 329–335.

Baum, J. K., and Blanchard, W. (2010). Inferring shark population trends from generalized linear mixed models of pelagic longline catch and effort data. *Fisheries. Res.* 102 (3), 229–239. doi: 10.1016/j.fishres.2009.11.006

Berg, C. W., Nielsen, A., and Kristensen, K. (2014). Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. *Fisheries. Res.* 151, 91–99. doi: 10.1016/j.fishres.2013.10.005

Boreman, J., and Klauda, R. J. (1988). Distributions of early life stages of striped bass in the Hudson River estuary 1974-1979. *Am. Fisheries. Soc. Monograph.* 4, 53–58.

Bowles, R. R., Merriner, J. V., and Grant, G. C. (1978). *Factors associated with accuracy in sampling fish eggs and larvae* (Washington, DC: Fish and Wildlife Service of US).

Bridger, J. P. (1956). On day and night variation in catches of fish larvae. *Ices. J. Mar. Sci.* 22 (1), 42–57. doi: 10.1093/icesjms/22.1.42

Chiarini, M., Guicciardi, S., Angelini, S., Tuck, I. D., Grilli, F., Penna, P., et al. (2022). Accounting for environmental and fishery management factors when standardizing CPUE data from a scientific survey: A case study for *Nephrops norvegicus* in the Pomo Pits area (Central Adriatic Sea). *PloS One* 17, e0270703. doi: 10.1371/journal.pone.0270703

Cooper, J. C., Cantelmo, F. R., and Newton, C. E. (1988). Overview of the Hudson river estuary. *Am. Fish. Soc Monogr.* 4, 11–24.

Daniels, R. A., Limburg, K. E., Schmidt, R. E., Strayer, D. L., and Chambers, R. C. (2005). "Changes in fish assemblages in the tidal Hudson River, New York," in *Historical changes in large river fish assemblages of the Americas*. Eds. J. N. Rinne, R. M. Hughes and B. Calamusso (Bethesda, Maryland: American Fisheries Society), 471–504.

Dew, C. B., and Hecht, J. H. (1994a). Hatching, estuarine transport, and distribution of larval and early juvenile Atlantic tomcod, *Microgadus tomcod*, in the Hudson River. *Estuaries* 17, 472–488. doi: 10.2307/1352677

Dew, C. B., and Hecht, J. H. (1994b). Recruitment, growth, mortality, and biomass production of larval and early juvenile Atlantic tomcod in the Hudson River Estuary. *Trans. Am. Fisheries. Soc.* 123, 681–702. doi: 10.1577/1548-8659(1994)123<0681:RGMABP>2.3.CO;2

Drexler, M., and Ainsworth, C. H. (2013). Generalized additive models used to predict species abundance in the Gulf of Mexico: an ecosystem modeling tool. *PloS One* 8 (5), e64458. doi: 10.1371/journal.pone.0064458

Ducharme-Barth, N. D., Grüss, A., Vincent, M. T., Kiyofuji, H., Aoki, Y., Pilling, G., et al. (2022). Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application. *Fisheries. Res.* 246, 106169. doi: 10.1016/j.fishres.2021.106169

Hair, J. F., Anderson, R. E., Babin, B. J., and Black, W. C. (2010). Multivariate Data Analysis: A Global Perspective (7th Edition). Upper Saddle River. New Jersey: Pearson, 816

Haldorson, L., Prichett, M. A., Paul, A. L., and Ziemann, D. A. (1993). Vertical distribution and migration of fish larvae in a Northeast Pacific bay. *Mar. Ecol. Prog. Ser.* 101, 67–80. doi: 10.3354/meps101067

Harley, S. J., Myers, R. A., and Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Can. J. Fish. Aquat. Sci.* 58, 1760–1772. doi: 10.1139/cjfas-58-9-1760

Heimbuch, D. G., Dunning, D. J., and Young, J. R. (1992). "Post-yolk sac larvae abundance as an index of year class strength of striped bass in the Hudson River," in *Estuarine research in the 1980s*. Ed. C. L. Smith (Albany, N.Y: SUNY Press), 376–391.

Hilborn, R., and Walters, C. J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics, & uncertainty* (New York: Chapman and Hall), 570.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., et al. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. *ICES. J. Mar. Sci.* 72 (1), 99–110. doi: 10.1093/icesjms/fsu198

Jørgensen, B. (1987). Exponential dispersion models. *J. R. Stat. Soc.: Ser. B. (Methodological).* 49 (2), 127–145. doi: 10.1111/j.2517-6161.1987.tb01685.x

Jowett, I. G., Parkyn, S. M., and Richardson, J. (2008). Habitat characteristics of crayfish (*Paranephrops planifrons*) in New Zealand streams using generalised additive models (GAMs). *Hydrobiologia* 596, 353–365. doi: 10.1007/s10750-007-9108-z

Kell, L. T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., et al. (2021). Validation of stock assessment methods: is it me or my model talking? *ICES. J. Mar. Sci.* 78 (6), 2244–2255. doi: 10.1093/icesjms/fsab104

Klauda, R. J., McLaren, J. B., Schmidt, R. E., and Dey, W. P. (1988a). Life history of white perch in the Hudson River estuary. *Amer. Fish. Soc. Monogr.* 4, 69–88.

Klauda, R. J., Moos, R. E., and Schmidt, R. E. (1988b). *Life History of Atlantic Tomcod, Microgadus tomcod, in the Hudson River Estuary, with Emphasis on Spatio-Temporal Distribution and. Fisheries research in the Hudson River*. (Albany, New York, USA: SUNY Press), 219.

Kouadri, S., Elbeltagi, A., Islam, A. R. M., and Kateb, S.. (2021). Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (*Algerian southeast*). *Appl. Water Sci.* 11, 1–20. doi: 10.1007/s13201-021-01528-9

Lennert-Cody, C. E., Clarke, S. C., Aires-da-Silva, A., Maunder, M. N., Franks, P. J. S., Román, M., et al. (2019). The importance of environment and life stage on interpretation of silky shark relative abundance indices for the equatorial Pacific Ocean. *Fish. Oceanogr.* 28, 1–11. doi: 10.1111/fog.12385

Levinton, J. S., and Waldman, J. R. (2006). *The Hudson river estuary* (Cambridge: Cambridge University Press).

Limbrug, K. E. (1995). Otolith strontium traces environmental history of subyearling American shad *Alosa sapidissima*. *Mar. Ecol. Prog. Ser.* 119, 25–35. doi: 10.3354/meps119025

Limburg, K. E. (1996a). Modelling the ecological constraints on growth and movement of juvenile American shad (*Alosa sapidissima*) in the Hudson River estuary. *Estuaries* 19, 794–813. doi: 10.2307/1352298

Limburg, K. E. (1996b). Growth and migration of 0-year American shad (*Alosa sapidissima*) in the Hudson River estuary: otolith microstructural analysis. *Can. J. Fish. Aquat. Sci.* 53, 220–238. doi: 10.1139/f95-160

Limburg, K., Hattala, K., Kahnle, A., and Waldman, J. (2006). "Fisheries of the Hudson river estuary," in *The Hudson river estuary*. Eds. J. Levinton and J. Waldman (Cambridge: Cambridge University Press), 189–204. doi: 10.1017/CBO9780511550539.016

Mattes, K. C. (1989). *The ecology of the American eel, Anguilla rostrata (LeSueur), in the Hudson River* (New York, USA: Fordham University).

Maunder, M. N., Thorson, J. T., Xu, H., Oliveros-Ramos, R., Hoyle, S. D., Tremblay-Boyer, L., et al. (2020). The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. *Fish. Res.* 229, 105594. doi: 10.1016/j.fishres.2020.105594

McCormick, J. L., Quist, M. C., and Schill, D. J. (2012). Effect of survey design and catch rate estimation on total catch estimates in Chinook Salmon fisheries. *North Am. J. Fisheries. Manage.* 32 (6), 1090–1101. doi: 10.1080/02755947.2012.716017

McLaren, J. B., Cooper, J. C., Hoff, T. B., and Lander, V. (1981). Movements of Hudson River striped bass. *Trans. Am. Fisheries. Soc.* 110 (1), 158–167. doi: 10.1577/1548-8659(1981)110<158:MOHRSB>2.0.CO;2

McLaren, J. B., Klauda, R. J., Hoff, T. B., and Gardinier, M. (1988). "Commercial fishery for striped bass in the Hudson river 1931-80," in *Fisheries research in the Hudson river*. Ed. C. L. Smith (Albany: State University of New York Press), 89–123.

Mohn, R. (1999). The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. *ICES. J. Mar. Sci.* 56 (4), 473–488. doi: 10.1006/jmsc.1999.0481

Murphy, H., Jenkins, G., Hamer, P., and Swearer, S. (2011). Diel vertical migration related to foraging success in snapper *Chrysophys auratus* larvae. *Mar. Ecol. Prog. Ser.* 433, 185–194. doi: 10.3354/meps09179

Nack, C. C., Swaney, D. P., and Limburg, K. E. (2019). Historical and projected changes in spawning phenologies of American shad and striped bass in the Hudson River Estuary. *Mar. Coast. Fisheries.* 11 (3), 271–284. doi: 10.1002/mcf2.10076

Navarro, M., Hailu, A., Langlois, T., Ryan, K. L., and Kragt, M. E. (2020). Determining spatial patterns in recreational catch data: a comparison of generalized additive mixed models and boosted regression trees. *ICES. J. Mar. Sci.* 77 (6), 2216–2225. doi: 10.1093/icesjms/fsz123

Noble, R. L. (1970). Evaluation of the Miller High-Speed Sampler for sampling yellow perch and walleye fry. *J. Fisheries. Res. Board. Canada.* 27 (6), 1033–1044. doi: 10.1139/f70-119

O'Connor, M. P., Juanes, F., McGarigal, K., and Gaurin, S. (2012). Findings on American Shad and striped bass in the Hudson River Estuary: a fish community study of the long-term effects of local hydrology and regional climate change. *Mar. Coast. Fisheries.* 4 (1), 327–336. doi: 10.1080/19425120.2012.675970

Ospina-Alvarez, A., Parada, C., and Palomera, I. (2012). Vertical migration effects on the dispersion and recruitment of European anchovy larvae: From spawning to nursery areas. *Ecol. Model.* 231, 65–79. doi: 10.1016/j.ecolmodel.2012.02.001

Pennino, M. G., Conesa, D., Lo ́ipez-Qu ́ılez, A., Muñoz, F., Fernaíndez, A., and Bellido, J. M. (2016). Fishery-dependent and -independent data lead to consistent estimations of essential habitats. *ICES. J. Mar. Sci.* 73, 2302–2310. doi: 10.1093/icesjms/fsw062

Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infect. Dis. Model.* 5, 111–128. doi: 10.1016/j.idm.2019.12.010

Power, J. H., and Moser, E. B. (1999). Linear model analysis of net catch data using the negative binomial distribution. *Can. J. Fisheries. Aquat. Sci.* 56 (2), 191–200. doi: 10.1139/f98-150

R Core Team (2020). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Schielzeth, H., and Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods Ecol. Evol.* 4 (1), 14–24. doi: 10.1111/j.2041-210x.2012.00251.x

Schmidt, J. (1923). The breeding places of the eel. *Philos. Trans. R. Soc Lond. B. Biol. Scl.* 211, 179–208.

Shono, H. (2008). Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fisheries. Res.* 93 (1-2), 154–162. doi: 10.1016/j.fishres.2008.03.006

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15 (1), 72–101. doi: 10.2307/1412159.JSTOR1412159

Takemura, A., Rahman, M. S., Nakamura, S., Park, Y. J., and Takano, K. (2004). Lunar cycles and reproductive activity in reef fishes with particular attention to rabbitfishes. *Fish. Fisheries.* 5 (4), 317–328. doi: 10.1111/j.1467-2679.2004.00164.x

Tuckey, T. D., and Fabrizio, M. C. (2013). Influence of survey design on fish assemblages: implications from a study in chesapeake bay tributaries. *Trans. Am. Fisheries. Soc.* 142, 957–973. doi: 10.1080/00028487.2013.788555

Tweedie, M. C. (1984). "An index which distinguishes between some important exponential families," in *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference*. (Calcutta, India: is Indian Statistical Institute), 579, 579–604.

Walsh, W. A., and Brodziak, J. (2015). Billfish CPUE standardization in Hawaii longline fishery: Model selection and multimodel inference. *Fisheries. Res.* 166, 151–162. doi: 10.1016/j.fishres.2014.07.015

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Can. J. Fish. Aquat. Sci.* 60, 1433–1436. doi: 10.1139/f03-152

Webster, R. A., Soderlund, E., Dykstra, C. L., and Stewart, I. J. (2020). Monitoring change in a dynamic environment: spatiotemporal modelling of calibrated data from different types of fisheries surveys of Pacific halibut. *Can. J. Fisheries. Aquat. Sci.* 77 (8), 1421–1432. doi: 10.1139/cjfas-2019-0240

Wilberg, M. J., and Bence, J. R. (2008). Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fisheries. Res.* 93 (1-2), 212–221. doi: 10.1016/j.fishres.2008.04.010

Wood, S., and Wood, M. S. (2015). Package 'mgcv'. *R Package version.* 1 (29), 729.

Xue, Y., Tanaka, K., Yu, H., Chen, Y., Guan, L., Li, Z., et al. (2018). Using a new framework of two-phase generalized additive models to incorporate prey abundance in spatial distribution models of juvenile slender lizardfish in Haizhou Bay, China. *Mar. Biol. Res.* 14 (5), 508–523. doi: 10.1080/17451000.2018.1447673

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* Vol. 574 (New York: springer), 574.

TYPE Methods
PUBLISHED 25 August 2023
DOI 10.3389/fmars.2023.1219283

Check for updates

OPEN ACCESS

EDITED BY
Nathan Bacheler,
Beaufort Laboratory, Southeast Fisheries Science Center (NOAA), United States

REVIEWED BY
Sarah Roberts,
University of North Carolina at Chapel Hill, United States
Matthew Vincent,
Southeast Fisheries Science Center (NOAA), United States

*CORRESPONDENCE
Lukas DeFilippo
lukas.defilippo@noaa.gov

†PRESENT ADDRESS
Lukas DeFilippo,
Ecosystem Monitoring and Assessment Program, Alaska Fisheries Science Center, NOAA, Seattle, WA, United States

RECEIVED 08 May 2023
ACCEPTED 07 August 2023
PUBLISHED 25 August 2023

CITATION
DeFilippo L, Kotwicki S, Barnett L, Richar J, Litzow MA, Stockhausen WT and Palof K (2023) Evaluating the impacts of reduced sampling density in a systematic fisheries-independent survey design. *Front. Mar. Sci.* 10:1219283.
doi: 10.3389/fmars.2023.1219283



# Evaluating the impacts of reduced sampling density in a systematic fisheries-independent survey design

Lukas DeFilippo[1*†], Stan Kotwicki[1], Lewis Barnett[1], Jon Richar[2], Michael A. Litzow[2], William T. Stockhausen[3] and Katie Palof[4]

[1]Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA), Seattle, WA, United States, [2]Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA), Kodiak, AK, United States, [3]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration (NOAA), Seattle, WA, United States, [4]Alaska Department of Fish and Game, Division of Commercial Fisheries, Juneau, AK, United States

Fisheries-independent surveys provide critical data products used to estimate stock status and inform management decisions. While it can be possible to redistribute sampling effort to improve survey efficiency and address changing monitoring needs in the face of unforeseen challenges, it is important to assess the consequences of such changes. Here, we present an approach that relies on existing survey data and simulations to evaluate the impacts of strategic reductions in survey sampling effort. We apply this approach to assess the potential effects of reducing high density sampling near St. Matthew Island and the Pribilof Islands in the NOAA eastern Bering Sea (EBS) bottom trawl survey. These areas contain high density "corner stations" that were implemented for finer-scale monitoring of associated blue king crab stocks (*Paralithodes platypus*) which historically supported commercial fisheries but have since declined and are seldom eligible for harvest. We investigate the effects of removing these corner stations on survey data quality for focal *P. platypus* stocks and other crab and groundfish species monitored by the EBS survey. We find that removing the St. Matthew and Pribilof Islands corner stations has negligible effects on data quality for most stocks, except for those whose distributions are concentrated in these areas. However, the data quality for such stocks was relatively low even with higher density sampling, and corner station removal had only minor effects on stock assessment outcomes. The analysis we present here provides a generic approach for evaluating strategic reductions in sampling effort for systematic survey designs and can be applied by scientists and managers facing similar decisions elsewhere.

KEYWORDS
survey design, stock assessment, groundfish, crab, spatiotemporal model

# 1 Introduction

Fisheries-independent surveys provide some of the most important data used to inform stock assessments and management decisions (Gunderson, 1993; Chen et al., 2003). Indices of abundance generated from survey data are typically assumed to be proportional to the true abundance of target populations, a relationship which is expressed via the catchability parameter ($q$) in assessment models (Quinn and Deriso, 1999). The relationship between the survey index and population size is assumed to be constant in most cases, violation of which can lead to spurious estimates of stock dynamics (Pope and Shepherd, 1985; Wilberg et al., 2009; Thorson et al., 2013; Kotwicki et al., 2014). Consequently, changes to survey designs and operations that affect catchability may disrupt the stationarity of this relationship and confound true population trends if not accounted for (Godo, 1994; Kimura and Somerton, 2006; Cadigan et al., 2022). Changes in gear selectivity and effort allocation across space may also shift the composition of ages and sizes sampled, potentially affecting estimates of population age and size structure (Ono et al., 2015; Kotwicki et al., 2017; O'Leary et al., 2020). As such, to the extent possible, fisheries monitoring agencies typically avoid altering survey designs and sampling procedures in order to minimize the impacts such changes could have on data products and population estimates.

While long-term time series of fisheries-independent survey data with minimal interruptions or modifications are desirable (Godo, 1994; Stompe et al., 2020), circumstances may arise where changes are unavoidable or beneficial. For instance, sections of historically surveyed areas may become inaccessible due to marine protected area designations or wind energy development (Field et al., 2006; Methratta et al., 2020; Hare et al., 2022). Similarly, hazardous weather, vessel breakdowns, staffing and budget shortages, as well as pandemics, international conflicts, and economic turmoil can disrupt monitoring agencies' ability to complete surveys (ICES, 2020; Santora et al., 2021; ICES, 2023). Changes may also offer advantages, such as modernizing fishing gear or improving the efficiency of sampling designs (Brown et al., 2007; Oyafuso et al., 2021; Oyafuso et al., 2022). Moreover, marine ecosystems are dynamic, and updating survey designs may be necessary to adapt to shifting species distributions, abundance trends, and management priorities (Dulvy et al., 2008; Pinsky and Mantua, 2014; Maureaud et al., 2021; DeFilippo et al., 2023). While some fisheries-independent survey designs can readily accommodate changes in sampling density (*e.g.*, stratified random), others are less flexible (*e.g.*, systematic) (Cochran, 1977) and may require more detailed evaluation of the consequences of such changes.

The U.S. National Oceanic and Atmospheric Administration (NOAA) eastern Bering Sea (EBS) bottom trawl survey provides critical data in support of stock assessment and management for some of the world's most commercially valuable fish and crab stocks. The EBS survey follows a stratified systematic design composed of 350 rectangular grid cells which are each sampled as part of annual surveys (Lauth et al., 2019). However, near St. Matthew Island and the Pribilof Islands the corners of the grid cells are sampled in addition to the centers. The rationale for finer-scale sampling near St. Matthew and the Pribilof Islands has been to improve monitoring and data products for blue king crab (*Paralithodes platypus*) stocks in these areas, which historically supported valuable commercial fisheries. Blue king crab exhibit a sparse and patchy distribution, resulting in highly uncertain abundance estimates. Sampling these corner stations was instituted to increase the probability of encountering concentrated patches of blue king crab and reduce uncertainty in survey data products. However, both the Pribilof and St. Matthew stocks of blue king crab have declined substantially and are now closed to fishing with little sign of rebuilding to harvestable levels in the near future (Palof et al., 2020; Stockhausen, 2021a). Consequently, it is unclear if the effort and funds required to continue sampling the corner stations is justifiable, or if resources might be better allocated to other priorities, such as improving data quality for other species.

Here, we present an approach for evaluating the consequences of reduced sampling effort in systematic survey designs and apply it to estimate the effects of removing the St. Matthew and Pribilof Island corner stations from the EBS survey. While changes to survey fishing gear and sampling protocols can be resolved by paired fishing and intercalibration studies (*e.g.*, Cadigan and Dowden, 2010; Miller, 2013; Kotwicki et al., 2017; Cadigan et al., 2022), there is less guidance for evaluating changes in sampling effort (but see Zimmermann and Enberg, 2017; ICES, 2020; ICES, 2023), particularly for systematic survey designs. Using existing survey time-series and simulation analyses, our approach estimates the impacts of effort reduction on the precision and accuracy of survey data products, as well as stock assessment output and biological and management reference points. As fisheries management agencies are tasked with monitoring changing marine ecosystems under static or declining budgets, tools for addressing effort reduction and reallocation decisions will be essential for optimizing survey efficiency and ensuring reliable data products (ICES, 2020; ICES, 2023). The approach we present here is generic and flexible and can be widely applied to other species and regions by scientists and managers facing survey effort reduction decisions.

# 2 Methods

## 2.1 Case study background

The NOAA EBS bottom trawl survey occurs southeast of the U.S. – Russian international maritime boundary from Bristol Bay and the Alaska Peninsula to the south, to north of Nunivak and St. Matthew Island (Figure 1). Occurring annually from May to early August, the EBS survey samples a fixed set of 350 rectangular grid cells as part of a systematic design with a minimum grid resolution of 37.04 $km^2$ and a maximum depth of 200m (Lauth et al., 2019). The EBS survey begins in Bristol Bay and proceeds west using two chartered commercial vessels. Each vessel tows a standard 83'-112' eastern otter trawl with 10cm mesh for a duration of 30 minutes at a target speed of 3 knots (Lauth et al., 2019). Survey effort is measured as the area swept by the trawl gear, which is calculated as the product of the distance fished (measured with a GPS and a bottom contact sensor) and net width (measured by an acoustic sensor). Catches of commercially important fishes are identified to species

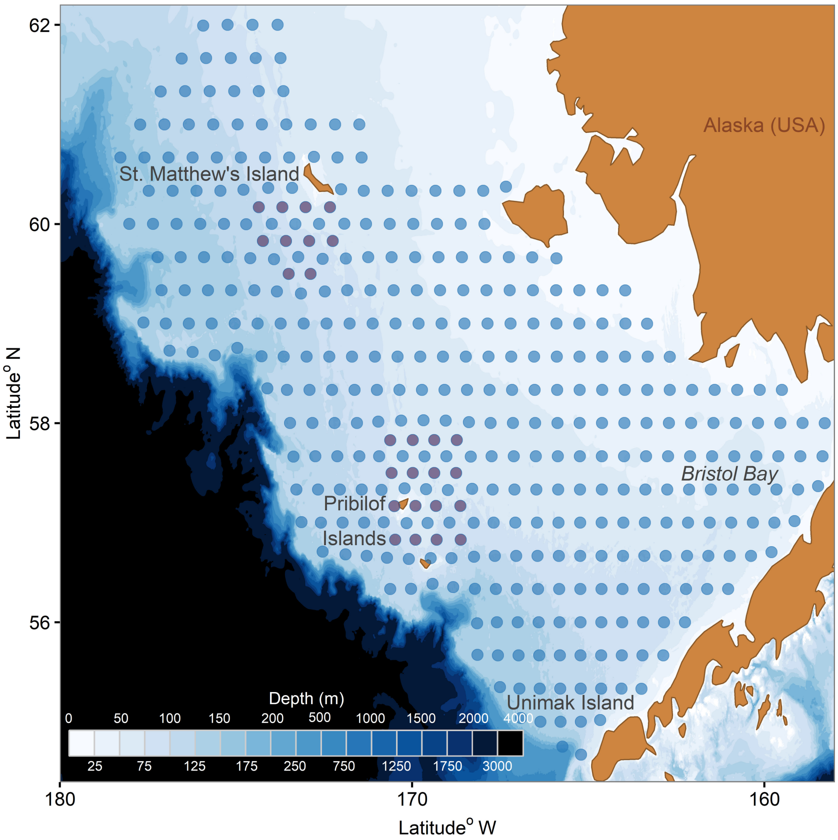



FIGURE 1
Bathymetric map of the southeastern Bering Sea shelf and centroids of the NOAA EBS bottom trawl survey grid. The St. Matthew's and Pribilof Islands corner stations are shown in purple, while the standard grid center stations are shown in blue.

and weighed, with a subset selected for length measurements and otolith extraction following protocols outlined in Lauth et al. (2019) and Stauffer (2004). Catches of commercially important crab species are sexed, assigned a shell condition and maturity status as per classifications in Jadamec et al. (1999), and carapace width (*Chionoecetes* spp.) or carapace length (*Lithodes*/*Paralithodes* spp.) is measured to the nearest 0.1 mm using either Vernier (prior to 2016) or digital (2016 and later) calipers.

Surrounding St. Matthew and the Pribilof Islands, the corners of the EBS survey grid cells are sampled in addition to the centers, leading to higher density sampling in these areas (Figure 1). There are 26 of these "corner stations" in total, which, at an average rate of 4 stations/day, require roughly one week of vessel time to sample. This additional sampling effort was initiated in 1981 for the Pribilof Islands, and 1983 for St. Matthew Island in order to improve the quality of data products for blue king crab stocks in these areas. Blue king crab exhibit a "contagious" distribution, in which individuals tend to be encountered in sparse, concentrated patches. Moreover, blue king crab (particularly females) are often found on rocky substrate and thus are difficult to sample by the EBS bottom trawl survey which primarily occurs in soft bottom habitat (Zacher et al., 2020; Vølstad, 2012). As a result, bottom trawl survey data products for blue king crab are highly uncertain (Stockhausen, 2021a; Palof, 2022). Sampling at a finer spatial resolution via the addition of the corner stations was intended to reduce uncertainty by increasing both the probability of encountering patches of blue king crab, and overall sample sizes. However, following pronounced declines in abundance, the St. Matthew blue king crab stock was declared overfished and closed to harvest in 1999. After a 10-year rebuilding plan, fishing resumed in 2009/2010 but after three years of modest harvest, the fishery was closed again in 2013/2014 due to declining survey abundance estimates and concerns about the productivity of the stock. Fishing resumed in 2014/2015 but fishery catches were poor and harvest has remained closed since 2016; the stock was once again declared overfished in 2018 when it entered into the current rebuilding plan (Palof, 2022). The Pribilof blue king crab stock has been closed to fishing since 1999, was declared overfished in 2002 and subsequently put on a rebuilding plan (Stockhausen, 2021a).

## 2.2 Analysis overview

Our approach to evaluating the effects of survey effort reduction proceeds in three general stages. In the first stage (Empirical Analysis, section 2.2.1), we quantify the retrospective effects of station removal on core survey data products (*i.e.* biomass, compositional estimates) that are used as inputs in stock assessment models. This stage of the analysis is focused on (1) identifying qualitative differences in survey data products as a result of station removal (*i.e.* changes in estimated stock biomass trajectories), and (2) evaluating effects on the uncertainty/precision of the survey data products (*i.e.* changes in the estimated variance of survey biomass indices). This first stage of the analysis uses historical survey data, generating survey data products without any corner station data and comparing them to estimates produced with the data from these stations. The second stage of our analysis focuses on assessing the effects of station removal on the accuracy of the survey data products using simulations (section 2.2.2. Simulation Analysis). Specifically, new survey datasets are simulated from spatiotemporal operating models and we examine how removing the corner stations from these simulated datasets affects the accuracy of estimated biomass indices relative to the true values specified in the operating models. In the third stage of our approach, we propagate the effects of station removal into the stock assessment models (section 2.2.3. Stock Assessment Analysis). This is accomplished in our case study by fitting stock assessment models to the biomass and compositional survey data products generated without corner station data and comparing the output to that from model fits that included the corner station data. Because integrated stock assessment models are complex and time consuming to fit, we do not perform this exercise for the full suite of stocks under consideration in our analysis, but select a representative subset to evaluate. Given that the corner stations were specifically intended to improve data quality for crab, our case study focuses more heavily on crab stocks, with a similar but limited set of analyses performed for groundfish.

### 2.2.1 Empirical analysis

For the first stage of our analyses, we retrospectively evaluated the effects of removing corner station data from the existing survey time-series on derived data products. To do so, we withheld all data collected from the 26 corner stations from 1982-2019 and produced design-based (*i.e.* area-swept catch-per-unit effort expansions) and model-based (*i.e.* standardized indices produced via spatiotemporal models) estimates using only data from the remaining stations. While the corner stations were implemented to improve data quality specifically for blue king crab, it is important to understand the effects of their removal on estimates for other species that occur in these areas as well. As such, in addition to both the St. Matthew and Pribilof Islands blue king crab stocks, we evaluated the effects of corner station removal on Pribilof Islands red king crab (*P. camtschaticus*), snow crab (*Chionoecetes opilio*) and Tanner crab (*C. bairdi*) biomass indices. For all crab biomass estimates, we focused on key size classes for each stock that are particularly important from a management standpoint. Specifically, we examined biomass estimates for the GE103 (GE=greater than or equal to) size class (103 – 189 mm) of male Tanner crab, the GE95 size class (95mm – 178 mm) of male snow crab, the GE120 size class of male Pribilof blue and red king crab (120 – 173 mm and 120 – 209 mm respectively), and both the GE90 (90 – 173 mm) and GE105 (105–173 mm) size classes of St. Matthew blue king crab. Additionally, we evaluated the effects of corner station removal on data products for ten groundfish species that inhabit these areas: Pacific cod (*Gadus macrocephalus*), walleye pollock (*G. chalcogrammus*), yellow Irish lord (*Hemilepidotus jordani*), Alaska plaice (*Pleuronectes quadrituberculatus*), northern rock sole (*Lepidopsetta polyxystra*), yellowfin sole (*Limanda aspera*), Bering flounder (*Hippoglossoides robustus*), flathead sole (*H. elassodon*), Alaska skate (*Bathyraja parmifera*), and Bering skate (*B. interrupta*). While improving data quality for groundfish stocks was not a consideration in implementing the corner stations as part of the EBS survey, groundfish are nonetheless sampled at these locations and it is useful to consider impacts on these stocks' data as well.

Design-based biomass indices for all species in our study were produced following the methods outlined in Wakabayashi et al. (1985) and using standard protocols for EBS groundfish and crab index production (Lauth et al., 2019; Zacher et al., 2020). For the crab stocks in our analysis, we also produced design-based size compositional estimates with and without the corner station data; these results are presented in Appendix A.

Model-based estimators are increasingly being used for index standardization, as they can often improve the precision of indices from fisheries-independent survey data (Thorson et al., 2015; Cao et al., 2017; Thorson and Haltuch, 2019). Model-based estimates are currently included in stock assessments for some of the groundfish species considered in this study, but none of the crab stocks. However, by leveraging spatial autocorrelation to extrapolate to unsampled areas, some model-based estimators may be more robust to reductions in survey effort. As such, we compared model-based biomass indices produced with and without the corner station data to explore how these estimates responded to station removal. Model-based estimates were generated using spatiotemporal models built via the Vector Autoregressive Spatiotemporal (VAST) package (release number 3.9.0) (Thorson and Barnett, 2017) in R-4.1.2 (R Core Team, 2021). The specification of the spatiotemporal models used here is standard for abundance index production, and further details can be found in Thorson (2019). Briefly, VAST models are an extension of a generalized linear mixed model (GLMM) that estimates dependent variable(s) via two linear predictors and a link function. Variation in the response variable(s) over space and time is partitioned into three components: (1) temporal variation ($\beta$), which represents changes from year-to-year that are equal across all locations, (2) spatial patterns ($\omega$), which correspond to variation over space that is constant over time (*i.e.* long-term habitat associations), and (3) spatiotemporal variation ($\epsilon$), which represents changes from year-to-year that are expressed differently across locations. To account for zero-inflated and skewed distributions, a Poisson-link delta modeling approach is used with two estimated linear predictors, $n$ and $w$, which represent expected numerical density and biomass-per-individual, respectively, such that $n_i w_i$ gives the expected biomass density ($d_i$) of sample (survey haul) $i$ (Thorson, 2019):

$$\log(n_i(s_i,\ t_i)) = \beta_1(t_i) + \omega_1(s_i) + \varepsilon_1(s_i,\ t_i) \quad (1)$$

$$\log(w_i(s_i,\ t_i)) = \beta_2(t_i) + \omega_2(s_i) + \varepsilon_2(s_i,\ t_i)$$

where $s_i$ and $t_i$ are the location and year associated with sample $i$. The annual intercepts ($\beta_1(t_i)$, $\beta_2(t_i)$) were specified as fixed effects independent among years, and the spatial variation terms ($\omega$) were estimated as random effects following a multivariate normal distribution:

$$\omega \sim \mathrm{MVN}(0, \sigma_\omega^2 \mathbf{R}(\eta)) \tag{2}$$

where $\sigma_\omega^2$ is the marginal spatial variance and $R(\eta)$ is the correlation matrix among locations ($s$) which is modeled as a Matérn function with decorrelation distance of $\eta$ and a transformation matrix that allows for geometric anisotropy such that decorrelation distance varies with cardinal direction (Thorson et al., 2015). Spatiotemporal effects were specified similarly:

$$\boldsymbol{\varepsilon}(t) \sim \mathrm{MVN}(0, \sigma_\varepsilon^2 \mathbf{R}(\eta)) \tag{3}$$

where $\sigma_\varepsilon^2$ represents the marginal spatiotemporal variance. The predicted density of individuals for each sample $i$ follows a Poisson process with expectation $n_i$ such that the encounter probability ($p_i$) is defined as:

$$p_i = 1 - \exp(-a_i n(s_i,\ t_i)) \tag{4}$$

where $a_i$ is the area swept for bottom trawl sample $i$. For any years in which a given species exhibited a 100% encounter rate across the EBS survey (*i.e.* at least one individual encountered at every station) the encounter probability ($p_i$) was fixed at one. The positive catch rate $r_i$ for sample $i$ is obtained from the numerical density ($n$) and average biomass per individual ($w$) as:

$$r_i = \frac{a_i n(s_i,\ t_i)}{p_i} w(s_i,\ t_i) \tag{5}$$

Given the predicted encounter probability $p_i$ and positive catch rate $r_i$, the probability distribution of the biomass $b_i$ for sample $i$ was specified as:

$$\Pr(b_i = B) = \begin{cases} 1 - p_i, & B = 0 \\ p_i \bullet Gamma(B|\theta^{-2}, r_i\theta^2), & B > 0 \end{cases} \tag{6}$$

where $1 - p_i$ is the probability associated with a biomass of zero, and $\mathrm{Gamma}(B|\theta^{-2}, r_i\theta^2)$ is the probability of biomass $B$ given the expected encounter probability $p_i$ and positive catch rate $r_i$ with estimated Gamma shape and scale parameters $\theta^{-2}$ and $r_i\theta^2$.

The model form described here was fitted individually to each of the ten species of groundfish considered and each of the six crab stocks/size classes, with the exceptions of Pribilof blue and red king crab. The full model configuration described above failed to converge for these latter two stocks, likely due to a limited spatial distribution and low encounter rates. As such, we removed the second spatiotemporal term ($\varepsilon_2(s_i,\ t_i)$) from the Pribilof blue king crab model and the second temporal ($\beta_2(t_i)$ and spatial ($\omega_2(s_i)$),), and both spatiotemporal terms ($\varepsilon_1(s_i,\ t_i)$, $\varepsilon_2(s_i,\ t_i)$) from the Pribilof red king crab model. Additionally, to account for years in which no Pribilof red king crab were observed, the first temporal intercept ($\beta_1(t_i)$) was specified as an independent annual random effect rather than as a fixed effect.

### 2.2.2 Simulation analysis

The empirical analysis described above was focused on assessing the (1) qualitative impacts (*i.e.* changes in biomass trends) and (2) effects on precision (*i.e.* coefficients of variation (CVs) of biomass estimates) of corner station removal on existing survey data products. However, it is not possible from these analyses to evaluate the effects of corner station removal on the accuracy of survey data products. To do so, we conducted a simulation analysis using the VAST model fits for each species as the basis for spatiotemporal operating models (OMs). New fixed and random effects were simulated from the joint precision matrix of the spatiotemporal model fits (*i.e.* conditional on the original data), and new data were then simulated conditional upon these new fixed and random effects via parametric bootstrapping (*e.g.*, Thorson et al., 2021). New observations were simulated at each location and year in which sampling occurred in the original survey data set (including both positive observations, and observations of zero biomass). Data were simulated from model fits that included the corner station data and thus included simulated data points at the corner station locations. These simulated corner station data were then either withheld or retained as design and model-based biomass estimates were obtained using the simulated data. The accuracy of these estimates was measured using the log accuracy ratio (LAR):

$$LAR(t, r) = \log\left(\frac{\hat{B}(t, r)}{B(t, r)}\right) \tag{7}$$

where $\hat{B}(t, r)$ is the estimated (design or model-based) biomass index from year $t$ and simulation replicate $r$, and $B(t, r)$ is the true biomass value specified in the OM. LAR was summarized across years and simulation replicates via the median symmetric accuracy (MSA):

$$\mathrm{MSA} = \exp(\mathrm{Median}(|LAR|) - 1 \tag{8}$$

LAR and MSA offer a number of advantages over other accuracy metrics (*e.g.*, mean absolute percentage error (MAPE), root mean square error (RMSE)), including scale independence, robustness to outliers, symmetry, and interpretability (Morley et al., 2018). LAR and MSA are interpretable such that values of zero represent perfect accuracy, and larger values indicate progressively worse accuracy.

The VAST operating models for Pribilof blue and red king crab and St. Matthew blue king crab were poorly conditioned, likely due to the low abundance levels, restricted ranges, and infrequent encounters for these stocks. Consequently, simulated datasets for Pribilof and St. Matthew king crab generated with the same procedures used for other stocks exhibited unrealistic distributions, with frequent positive encounters outside the Pribilof or St. Matthew management areas. As such, a modified simulation procedure was used for these stocks in which new data were simulated conditional upon the estimated fixed and random effects rather than simulating new fixed and random effects. Additionally, simulated datasets for St. Matthew and Pribilof king crab stocks were generated from operating models with spatial extents that were limited to the St. Matthew and Pribilof management areas respectively, thereby constraining the simulated observations within these boundaries. Given the difficulty of conditioning the operating models, we recommend that the results of the simulation analyses for St. Matthew's blue and Pribilof blue and red king crab be interpreted cautiously.

### 2.2.3 Stock assessment analysis

In addition to examining the effects of corner station removal on the precision and accuracy of survey data products themselves, we also investigated how these effects propagated into stock assessment output. For this phase of the analysis, we focus solely on crab stocks as these were the primary consideration for the implementation of the corner stations. We fitted stock assessment models using survey estimates of biomass (Methods: 2.2.2. Empirical analysis) and size composition (Appendix A) produced without the corner stations. We compared the resulting assessment model predictions and estimated biological and management reference points to those from model fits to data products that included corner station data. For this analysis we focused on Tanner crab and St. Matthew blue king crab. We selected these two stocks to investigate the effects of corner station removal on assessment output for stocks for which the corner stations are (St. Matthew blue king crab) and are not (Tanner crab) focal areas of the stocks' distributions.

The Tanner crab assessment model is a stage/size-based population dynamics model that incorporates sex (male, female), shell condition (new shell, old shell), and maturity (immature, mature) as different categories into which the overall stock is divided on a size-specific basis (Stockhausen, 2021b). The model is fit using a penalized maximum likelihood approach to the design-based survey biomass and size composition time-series, molt increment data, retained catch biomass and size composition time series from the directed fishery, bycatch data (biomass and size composition time series) from the directed fishery (sub-legal males, all females), and bycatch data (biomass and size composition time series) from several other crab and groundfish fisheries. Management quantities are subsequently derived from the maximum likelihood solution using spawner-per-recruit proxies for $F_{MSY}$ and $B_{MSY}$. For this analysis, the 2021 assessment model was fitted using the design-based survey biomass and size composition time series estimated without the corner stations and compared to existing assessment output produced with the corner stations. Model results for the estimated sex/maturity-specific population biomass time series, recruitment time series, management quantities, and rates of natural mortality from the two scenarios were compared.

The St. Matthew's blue king crab stock assessment is a simpler form of the Tanner crab model, using only mature male crab in the size/stage structured model (Palof, 2022). The stock is modeled via the Generalized Modeling for Crustacean Stocks (GMACS) framework in which it is fit using a penalized maximum likelihood approach to design-based survey biomass and size composition time series, retained catch biomass and size composition time series from the directed fishery, bycatch data (biomass and size composition time series) from the directed fishery, and bycatch data (biomass) from several other crab and groundfish fisheries. Estimates of life history parameters, such as natural mortality and growth for this stock are borrowed from other well studied king crab stocks, such as Bristol Bay red king crab. Management quantities are approximated from the long term average of mature male biomass, as directed by the Bering Sea and Aleutian Island (BSAI) crab fishery management plan. For this analysis, the 2021 model was fit using the design-based EBS bottom trawl survey biomass and size composition time series with and without corner stations. Resulting time series of mature male biomass and recruitment, as well as management quantities were compared between model fits.

# 3 Results

## 3.1 Empirical analyses

For the six stocks/size classes of crab that we examined, corner station removal had little qualitative impact on either design or model-based biomass estimates. Biomass trends were generally coherent with one another regardless of corner station inclusion, with some notable transient discrepancies for St. Matthew blue king crab and Pribilof blue and red king crab (Figure 2). Corner station removal had little effect on the precision of design and model-based estimates for both snow crab and Tanner crab (Figure 2; Table 1), with increases in average CV ($\Delta$ CV) of<0.02 for these species. More substantial declines in precision were observed for the other crab stocks, particularly for design-based biomass estimates (Figure 2; Table 1). For the two size classes of St. Matthew blue king crab, average design-based CVs estimates increased by ~ 0.1, while model-based CVs were much less affected by removing the corner stations (Table 1). The average design-based CV increased by 0.185 without the corner station data for Pribilof blue king crab, and 0.071 for Pribilof red king crab, while average the model-based CV for these stocks increased by 0.143 and 0.077 (Table 1). While the precision of model-based estimates was generally more robust to corner station removal compared to that of design-based estimates, we note that this was not the case for Pribilof red king crab; the average model-based CV for this stock increased by more than its design-based counterpart as a result of corner station removal, although the model-based CVs themselves were smaller in all scenarios (Table 1).

For nine of the ten groundfish species that we examined, there was no qualitative effect of corner station removal on the trend or scale of either design or model-based biomass estimates (Figure 3). Similarly, for most of the groundfish species we considered, differences in the precision of biomass estimates produced with and without corner station data were negligible (Figure 3; Table 2). The precision of model-based groundfish biomass estimates was generally more robust to corner station removal than their design-based counterparts (Table 2). The only groundfish species for which corner station removal caused a substantial increase in biomass CVs was yellow Irish lord (Figure 3; Table 2). Average design-based CV

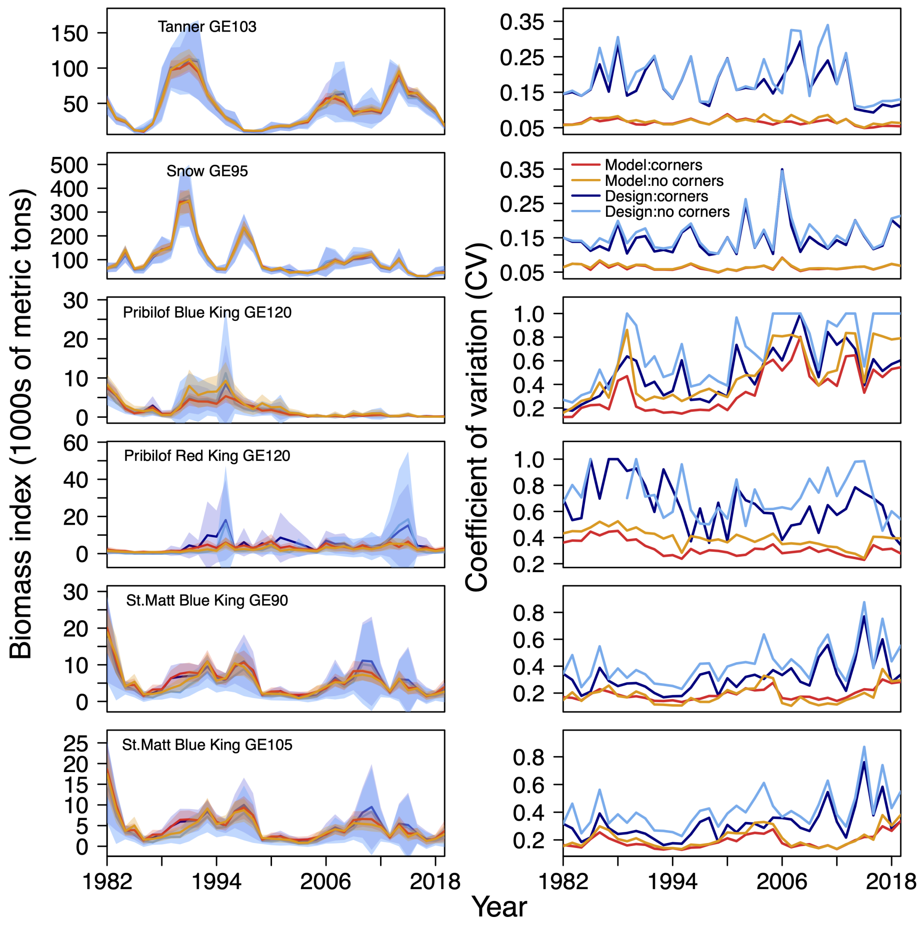



FIGURE 2
Design and model-based survey biomass estimates (left) and associated CVs (right) produced with and without corner station data for crab stocks/ size classes.

for this species increased by 0.096, while model-based CVs were more robust and increased by 0.05 without the corner station data. However, it is worth noting that even with the corner station data included, the biomass CVs for yellow Irish lord were large in many years (*e.g.*, 0.3-0.6), and were the highest of any of the groundfish species we examined (Table 2).

## 3.2 Simulation analyses

For snow and Tanner crab, the accuracy of design and model-based estimates was relatively robust to removal of the corner station data (Figure 4; Table 3). Conversely, the accuracy of biomass estimates for Pribilof blue and red king crab and St. Matthew blue king crab showed more substantial declines resulting from corner station removal (Table 3). The accuracy of model-based biomass estimates for the Pribilof and St. Matthew's king crab stocks was considerably more robust to corner station removal than were their design-based counterparts (Figure 4; Table 3). For instance, removing simulated corner station data for Pribilof red king crab resulted in declines in accuracy that were an order of magnitude lower for the model-based (Δ MSA=0.170) compared to design-based (Δ MSA=2.106) estimates (Table 3).

For nine of the ten groundfish species we examined, the accuracy of both design and model-based biomass estimates was robust to corner station removal (Figure 5; Table 4). There were modest declines (~13%) in the accuracy of design-based estimates for northern rock sole arising from corner station removal (Table 4). The only species for which substantial declines (~20-30%) in accuracy occurred as a result of removing the corner station data was yellow Irish lord (Table 4). However, the accuracy of both design and model-based biomass estimates for yellow Irish lord was limited even with the corner stations included, and the lowest of any groundfish species that we considered (Figure 5; Table 4).

## 3.3 Stock assessment analyses

Predictions of Tanner crab biomass over time were nearly identical with and without the corner station data for immature and mature males and females (Figure 6). Similarly, estimates of Tanner crab recruitment over time were virtually unchanged

TABLE 1 Mean coefficients of variation (CV) of design and model-based biomass estimates for crab species produced with versus without the corner stations.

| Species/Size Class | Estimator | Mean CV (corners) | | Δ CV |
|---|---|---|---|---|
| Tanner GE103 | Design-based | 0.174 | 0.190 | 0.016 |
| | Model-based | 0.065 | 0.069 | 0.004 |
| Snow GE95 | Design-based | 0.147 | 0.158 | 0.011 |
| | Model-based | 0.064 | 0.065 | 0.001 |
| Pribilof blue king GE120 | Design-based | 0.503 | 0.688 | 0.185 |
| | Model-based | 0.352 | 0.495 | 0.143 |
| Pribilof red king GE120 | Design-based | 0.648 | 0.719 | 0.071 |
| | Model-based | 0.318 | 0.395 | 0.077 |
| St. Matthew blue king GE90 | Design-based | 0.320 | 0.415 | 0.095 |
| | Model-based | 0.189 | 0.184 | -0.005 |
| St. Matthew blue King GE105 | Design-based | 0.312 | 0.412 | 0.1 |
| | Model-based | 0.190 | 0.212 | 0.022 |

Δ CV represents the increase in mean CV between estimates produced without the corner stations relative to estimates produced with the corner stations.

(Figure 6), and average recruitment was only marginally greater without the corner station data (Figure 6). Biomass-related reference points were also slightly greater in model runs without the corner station data (Figure 6). The biomass at which maximum spawning potential occurs ($B_{100}$) was estimated to be ~3.5% greater without the corner station data, while estimates of current and projected biomass were ~5% and ~6% greater respectively without the corner station data. Estimates of fishing mortality rates associated with maximum sustained yield ($F_{MSY}$) and the overfishing limit ($F_{OFL}$) were nearly identical regardless of whether the corner station data were included (Figure 6), while estimates of MSY and the overfishing limit (OFL) themselves were ~1.5% and ~4.4% greater without the corner station data. Estimated natural mortality rates for both mature males and females were slightly lower in model runs without the corner station data (Figure S1), likely contributing to the slightly higher estimates of biomass-related reference points.

Stock assessment model runs for St. Matthew blue king crab completed without the corner station data resulted in biomass estimates that were slightly lower than those produced with the corner station data (Figure 7). However, the biomass time-series of these two scenarios were highly correlated with one another (lag-zero cross correlation coefficient = 0.998), exhibiting nearly identical trends (Figure 7). Estimates of $B_{MSY}$ and mature male biomass (MMB) produced without the corner stations were 324 tons and 166 tons lower respectively compared to model fits that included corner station data (Figure 7). However, the estimated ratio of current biomass relative to $B_{MSY}$ ($B/B_{MSY}$) was identical regardless of corner station inclusion (Figure 7). Model predictions of recruitment over time were similar regardless of corner station inclusion, though average recruitment was slightly lower in the absence of the corner station data (Figure 7). The estimated overfishing limit from model runs without the corner station data was 0.01 tons lower compared to model runs that included corner station data (Figure 7).

# 4 Discussion

The crab stocks for which corner station removal had the greatest impacts on data quality were those whose distributions are concentrated around St. Matthew and the Pribilof Islands. However, it is important to note that the uncertainty in these stocks' data products was often substantial even with the corner station data included. The difficulty of producing high quality biomass estimates for the St. Matthew and Pribilof Island king crab stocks may be partially due to these species' associations with hard-bottom habitats that the EBS bottom trawl survey cannot sample effectively (*e.g.*, female blue king crab; Vølstad, 2012; Zacher et al., 2020), and/or limited ranges and low abundance levels (*e.g.*, Pribilof red king crab; Zacher et al., 2020). At such levels of uncertainty, the contributions these data make to stock assessment models are limited and further declines in precision from corner station removal may not represent a meaningful erosion of their information content. For instance, the annual survey biomass estimates for Pribilof blue and red king crab are considered too variable and uncertain for use in stock assessment models even with the corner station data, such that these assessments rely on *post hoc* smoothing of the survey biomass time series (Stockhausen, 2021a; Szuwalski, 2022). In some cases the declines in data precision and accuracy arising from corner station removal were mitigated to some extent by using a model-based versus design-based estimator.

The impacts of corner station removal on both the precision and accuracy of biomass estimates were minor for all groundfish species that we investigated except for yellow Irish lord, which exhibits high

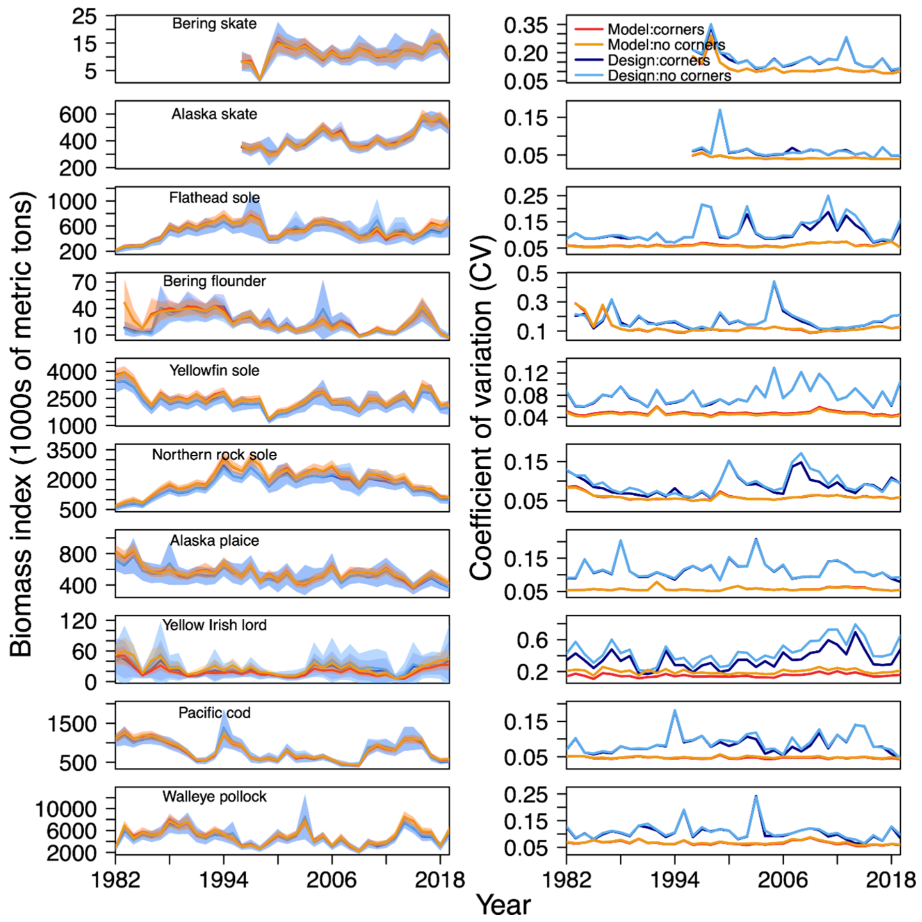



FIGURE 3
Design and model-based survey biomass estimates (left) and associated CVs (right) produced with and without corner station data for groundfish species. Note that Bering skate and Alaska skate were not distinguished on the NOAA EBS bottom trawl survey until 1996, hence why the time-series for those species do not begin until 1996.

concentrations around the Pribilof Islands (Figure S2). The other groundfish species we examined exhibit broader distributions, either due to lack of association with the St. Matthew or Pribilof Island habitat areas or because their distributions are more strongly driven by dynamic environmental conditions than by static habitat characteristics. For instance, Alaska skate are widely distributed throughout the outer EBS shelf area (Figure S3), and yellowfin sole are broadly associated with shallow, inshore areas of the shelf (Figure S4). Moreover, many groundfishes are mobile and alter their distributions in response to environmental conditions. For example, walleye pollock and Pacific cod actively avoid a mass of <2° C subsurface water (the cold pool) that occurs on the EBS shelf (Wyllie-Echeverria and Wooster, 1998; Kotwicki and Lauth, 2013). As such, static habitat associations are likely a less important component of these species' distributions. Importantly, with the exception of yellow Irish lord (which does not have a stock assessment and is of no commercial importance) it is unlikely that the removal of corner stations from the EBS survey would meaningfully impact estimates of stock status or management recommendations for any of the groundfish species we examined.

Our analysis of the effects of corner station removal on stock assessment outcomes showed negligible effects for Tanner crab. The Tanner crab assessment model predictions were nearly identical regardless of whether or not the corner station data were included. The only appreciable changes in model output arising from corner station removal were marginally greater estimates of biomass-related reference points and the OFL, likely attributable to the slightly lower estimates of natural mortality produced without the corner station data. Such minor changes in model predictions and reference points seem unlikely to affect decision-making for this stock, indicating that assessment and management of Tanner crab would be robust to removal of the corner stations.

Not surprisingly, we found somewhat greater effects of corner station removal on stock assessment output for St. Matthew blue king crab. Assessment model predictions made without the corner station for this stock indicated slightly lower mature male biomass and recruitment over time compared to baseline model output that included the corner stations. However, the trends in biomass and recruitment were nearly identical regardless of corner station inclusion. Despite the minor differences between time-series produced with versus without the corner stations, the coherence of trends across data scenarios suggests that ability to detect stock recovery would not be impaired without the corner station data. Similarly, the estimated ratio of current biomass relative to $B_{MSY}$ (B/

TABLE 2 Mean coefficients of variation (CV) of design and model-based biomass estimates for groundfish species produced with versus without the corner stations.

| Species | Estimator | Mean CV (corners) | Mean CV (no corners) | Δ CV |
|---|---|---|---|---|
| Bering skate | Design-based | 0.165 | 0.167 | 0.002 |
| | Model-based | 0.118 | 0.119 | 0.001 |
| Alaska skate | Design-based | 0.061 | 0.062 | 0.001 |
| | Model-based | 0.042 | 0.043 | 0.001 |
| Flathead sole | Design-based | 0.110 | 0.118 | 0.008 |
| | Model-based | 0.061 | 0.060 | -0.001 |
| Bering flounder | Design-based | 0.169 | 0.173 | 0.004 |
| | Model-based | 0.123 | 0.125 | 0.002 |
| Yellowfin sole | Design-based | 0.080 | 0.080 | <0.001 |
| | Model-based | 0.048 | 0.046 | -0.002 |
| Northern rock sole | Design-based | 0.088 | 0.094 | 0.006 |
| | Model-based | 0.060 | 0.059 | -0.001 |
| Alaska plaice | Design-based | 0.115 | 0.116 | 0.001 |
| | Model-based | 0.057 | 0.057 | <0.001 |
| Yellow Irish lord | Design-based | 0.350 | 0.446 | 0.096 |
| | Model-based | 0.151 | 0.201 | 0.05 |
| Pacific cod | Design-based | 0.085 | 0.089 | 0.004 |
| | Model-based | 0.047 | 0.048 | 0.001 |
| Walleye pollock | Design-based | 0.106 | 0.109 | 0.003 |
| | Model-based | 0.066 | 0.068 | 0.002 |

Δ CV represents the increase in mean CV between estimates produced without the corner stations relative to estimates produced with the corner stations.

$B_{MSY}$), the reference point used for tracking stock rebuilding, was not affected by removal of the corner station data. As such, despite the substantial effects of corner station removal on survey data products for St. Matthew blue king crab, it does not appear that there is a correspondingly appreciable effect on assessment outcomes for this stock, at least with respect to monitoring stock rebuilding. This may be partially due to the fact that the EBS bottom trawl survey data is not the only source of fisheries-independent data that are used in the St. Matthew blue king crab stock assessment, the other being the Alaska Department of Fish and Game (ADFG) pot survey (Palof, 2022).

While our analysis focuses on the effects of survey effort reduction on stock assessment inputs (survey data products) and outputs (biological/management reference points), there are other objectives that can be important to fisheries and ecosystem management. For instance, there may be benefits to finer-scale sampling of ecologically important areas – beyond those to stock assessment outcomes – in improving understanding of crucial habitats. Indeed, the Pribilof Islands are important habitat for many fish, invertebrate, seabird and marine mammal species (Craighead and Oppenheim, 1985; Gentry, 1998; Ferrero et al., 2000) and one of the most productive regions in the Bering Sea (Cooney and Coyle, 1982; Coyle and Cooney, 1993), with oceanographic and ecological dynamics that are distinct from the rest of the EBS shelf (Hunt et al., 2008; Ciannelli et al., 2004). However, it is difficult to quantitatively assess the value of finer-scale survey observations for understanding ecological and oceanographic processes. Nonetheless, objectives beyond utility to stock assessment such as ecosystem considerations may be important to managers, and it is important to note that such concerns cannot be assessed using the approach we present here.

We found that the stocks for which corner station removal caused substantial declines in data quality were also those with the lowest quality data to begin with. This result suggests that while sampling the corner stations may improve data quality for some stocks, the extent of improvement may not be sufficient to produce satisfactorily informative data products. The limited data quality for Pribilof and St. Matthew king crab and yellow Irish lord may simply reflect the fact that the EBS bottom trawl survey is not optimized for sampling certain species, a challenge that naturally arises in fisheries-independent survey designs with multispecies objectives (Cochran, 1977; Godo, 1994; Oyafuso et al., 2021). Data quality for these stocks may be more effectively improved by alternative approaches beyond increased bottom trawl sampling density, such as expanded collection of other forms of fisheries-independent data (*e.g.*, pot surveys for king crab; Gish and Vanek, 2010; Palof, 2022), or building capacity for generating and using model-based indices (Thorson, 2019). While we found that

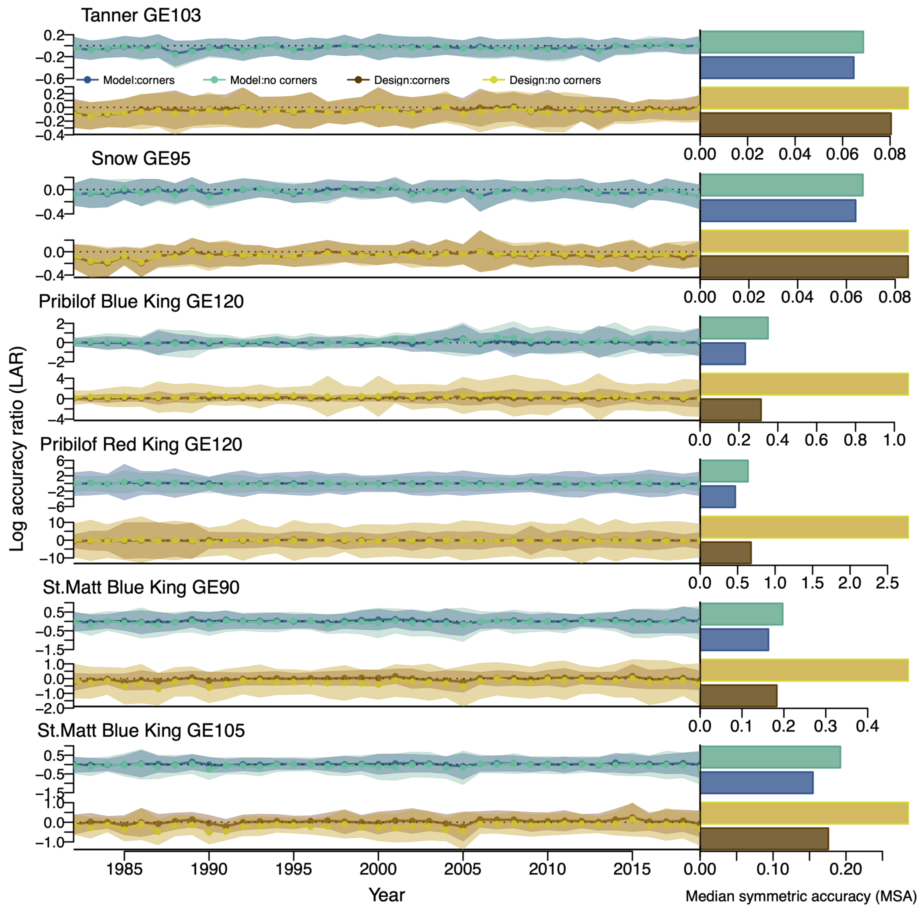



FIGURE 4
Accuracy of design and model-based survey biomass estimates with and without corner station data for crab stocks/size classes. Time-series on the left show the log accuracy ratio (LAR) over time for design and model-based biomass estimates with and without corner station data. The median LAR across simulation replicates over time is shown as solid/dashed lines, while the 95% distribution across simulation replicates is shown as shaded boundaries. Histograms on the right show the median symmetric accuracy (MSA) across years and simulation replicates. A LAR/MSA value of 0 indicates perfect accuracy while larger values indicate progressively worse accuracy.

**TABLE 3 Median symmetric accuracy (MSA) of design and model-based biomass estimates for crab species/size classes produced with versus without the corner stations.**

| Species/Size class | Estimator | MSA (corners) | MSA (no corners) | Δ MSA |
|---|---|---|---|---|
| Tanner GE103 | Design-based | 0.080 | 0.088 | 0.008 |
| | Model-based | 0.065 | 0.069 | 0.004 |
| Snow GE95 | Design-based | 0.085 | 0.086 | 0.001 |
| | Model-based | 0.064 | 0.067 | 0.003 |
| Pribilof blue king GE120 | Design-based | 0.314 | 1.075 | 0.761 |
| | Model-based | 0.232 | 0.350 | 0.118 |
| Pribilof red king GE120 | Design-based | 0.680 | 2.786 | 2.106 |
| | Model-based | 0.469 | 0.639 | 0.170 |
| St. Matthew blue king GE90 | Design-based | 0.183 | 0.499 | 0.316 |
| | Model-based | 0.163 | 0.197 | 0.034 |
| St. Matthew blue King GE105 | Design-based | 0.176 | 0.286 | 0.110 |
| | Model-based | 0.155 | 0.192 | 0.037 |

Δ MSA refers to the difference in MSA between estimates produced with versus without the corner stations.

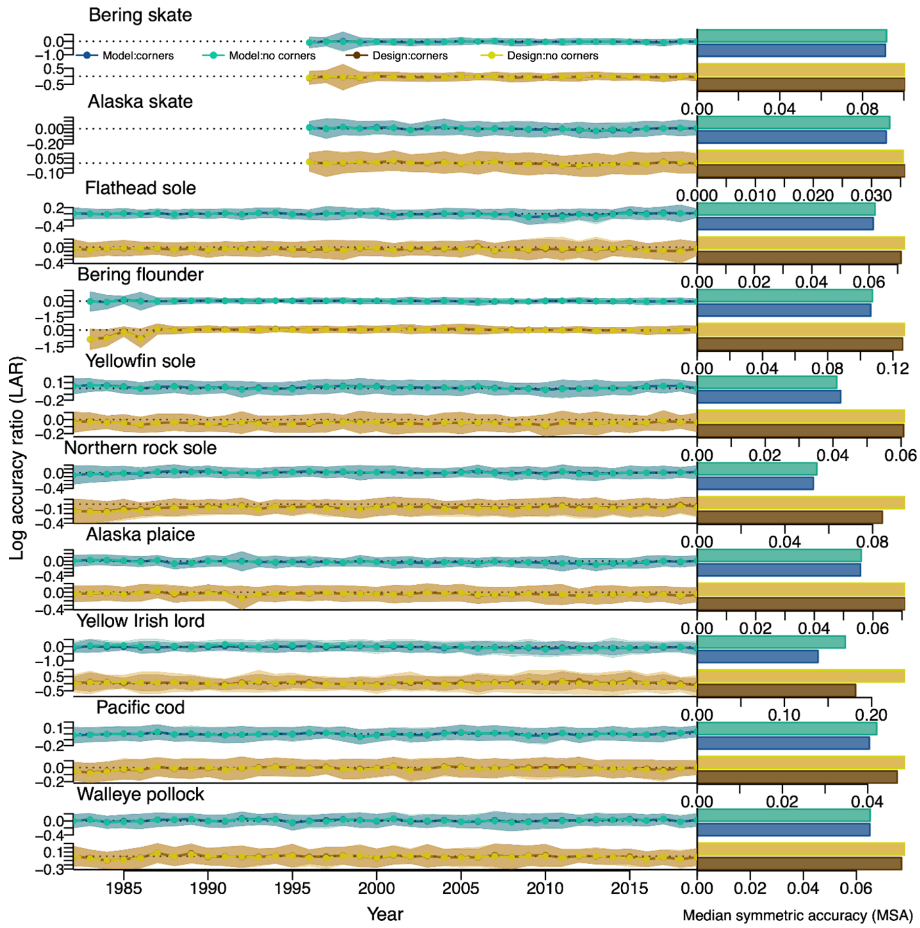



FIGURE 5
Accuracy of design and model-based survey biomass estimates with and without corner station data for groundfishes. Time-series on the left show the log accuracy ratio (LAR) over time for design and model-based biomass estimates with and without corner station data. The median LAR across simulation replicates over time is shown as solid/dashed lines, while the 95% distribution across simulation replicates is shown as shaded boundaries. Histograms on the right show the median symmetric accuracy (MSA) across years and simulation replicates. A LAR/MSA value of 0 indicates perfect accuracy while larger values indicate progressively worse accuracy.

model-based biomass estimates were more robust to corner station removal than their design-based counterparts, adequate model performance was difficult to achieve in some cases (*e.g.*, St. Matthew blue king crab, Pribilof red and blue king crab) and stock assessments for Bering Sea crab still rely entirely on design-based abundance and compositional estimates (Stockhausen, 2014; Szuwalski et al., 2014; Szuwalski and Turnock, 2016; Palof et al., 2020; Stockhausen, 2021a; Stockhausen, 2021b; Szuwalski, 2022). Continued innovation in model-based index production methods for crab, as well as expanded capacity to use model-based indices in crab assessment models may lead to improved inference on stock status and mitigate adverse consequences from potential survey effort reduction. Model-based indices can also incorporate environmental covariates to facilitate extrapolation to unsampled areas (Thorson, 2019; O'Leary et al., 2022) which may also help offset the impacts of reductions in sampling effort.

While stock assessment models are sensitive to the precision of input data, reductions in data quality beyond a certain point may not exert a corresponding effect on their information content. This is due to the integrated design of many contemporary stock assessment models (including those for Tanner and St. Matthew blue king crab), in which multiple data sources contribute to parameter estimates via a joint likelihood (Fournier and Archibald, 1982; Maunder and Punt, 2013). In such integrated models, the greater the uncertainty in one data source, the less information it will contribute to the model, whether by explicit data weighting procedures (*e.g.*, Francis, 2011) or via the joint likelihood itself (*e.g.*, DeFilippo et al., 2021). In our analysis, this effect is best demonstrated by St. Matthew blue king crab, for which corner station removal had a minor effect on assessment model outputs despite exerting a large influence on the input survey data products. This result emphasizes the importance of propagating the impacts of survey effort reduction through stock assessment models to understand the impacts of proposed survey design changes. Indeed, contemporary stock assessment models are complex, with multiple contributing (and often conflicting) sources of information that can interact to affect parameter estimates in unpredictable ways (Maunder and Punt, 2013; Ichinokawa et al., 2014; Peterson et al., 2021).

TABLE 4 Median symmetric accuracy (MSA) of design and model-based biomass estimates for groundfish species produced with versus without the corner stations.

| Species | Estimator | MSA (corners) | MSA (no corners) | Δ MSA |
|---|---|---|---|---|
| Bering skate | Design-based | 0.101 | 0.100 | -0.001 |
| | Model-based | 0.091 | 0.092 | 0.001 |
| Alaska skate | Design-based | 0.036 | 0.035 | -0.001 |
| | Model-based | 0.033 | 0.033 | 0 |
| Flathead sole | Design-based | 0.071 | 0.073 | 0.002 |
| | Model-based | 0.061 | 0.062 | 0.001 |
| Bering flounder | Design-based | 0.126 | 0.128 | 0.002 |
| | Model-based | 0.106 | 0.107 | 0.001 |
| Yellowfin sole | Design-based | 0.061 | 0.061 | 0 |
| | Model-based | 0.042 | 0.041 | -0.001 |
| Northern rock sole | Design-based | 0.084 | 0.095 | .011 |
| | Model-based | 0.053 | 0.055 | 0.002 |
| Alaska plaice | Design-based | 0.071 | 0.071 | 0 |
| | Model-based | 0.056 | 0.056 | 0 |
| Yellow Irish lord | Design-based | 0.182 | 0.238 | 0.056 |
| | Model-based | 0.138 | 0.170 | 0.032 |
| Pacific cod | Design-based | 0.047 | 0.049 | 0.002 |
| | Model-based | 0.040 | 0.042 | 0.002 |
| Walleye pollock | Design-based | 0.077 | 0.078 | 0.001 |
| | Model-based | 0.065 | 0.065 | 0 |

Δ MSA refers to the difference in MSA between estimates produced with versus without the corner stations.

The approach we demonstrate here is flexible and can be used to evaluate the potential effects of removing any set of existing stations from a systematic fisheries-independent survey. The corner stations we considered for removal in our analysis were selected because they represent an area of higher sampling density than the rest of the survey grid and because the stocks that motivated their implementation have remained below harvestable levels for a prolonged period of time. However, there are other situations in which a set of stations may be identified *a priori* as candidates for removal. For instance, a particular area within a survey's boundaries may be under consideration for wind energy development or MPA designation, or become inaccessible due to political (*e.g.*, international strife) or logistical (*e.g.*, prohibitive fuel costs for reaching remote areas) reasons (Field et al., 2006; ICES, 2020; Methratta et al., 2020; Hare et al., 2022; ICES, 2023). In such scenarios, scientists could follow the general approach outlined here to quantify the expected impacts of removing the stations in question on data quality and stock assessment outcomes. These results could be used to understand and predict the impacts of the effort reduction on scientific inference, and/or be used to develop calibration factors to correct for these effects in the survey data time series. Conversely, there may be a need to reduce sampling effort more generally (*e.g.*, due to budgetary and staffing shortages), without a specific subset of stations indicated as leading options for removal (ICES, 2020; ICES, 2023). In response to such a need to reduce sampling effort without obvious candidates for station removal, scientists may use our approach to evaluate a range of alternative scenarios (*e.g.*, thinning the sampling density of the entire survey grid by a specified fraction, removing stations from various areas based on expert opinion), and compare the impacts of each effort reduction scenario on data quality and stock assessment output to inform decision-making.

While evaluating the effects of station removal on all survey data products and stock assessment outcomes is optimal, scientists may prioritize certain stocks to focus effort on (*e.g.*, crab in our case study) and conduct more limited investigations on others for which expected impacts are lower (*e.g.*, groundfish in our case study). As shifting ecosystems and species distributions place growing demands on fisheries-independent monitoring amid rising costs and limited institutional budgets, unavoidable reductions in survey effort are likely to become increasingly common (ICES, 2020; ICES, 2023). Consequently, it is imperative that scientists can evaluate the effects of survey effort reductions to inform decision-making on such actions and understand their consequences. The analysis we present here demonstrates a generic set of steps for evaluating reductions in sampling density for systematic survey designs that can be applied in similar situations elsewhere to provide a quantitative basis for decision-making.

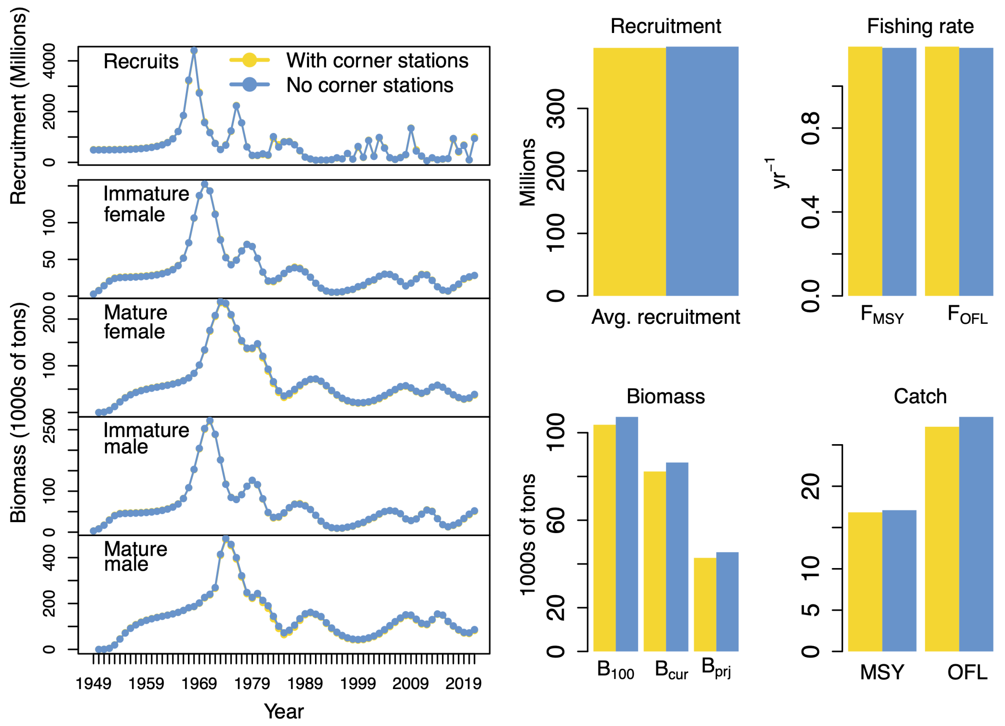



FIGURE 6
Tanner crab stock assessment model output. Left-hand panels show model predictions with (yellow) and without (teal) corner stations for recruits (top row) immature females (second row), mature females (third row), immature males (fourth row), and mature males (fifth row). Barplots on the right show estimated tanner crab reference points produced with (yellow) and without (teal) corner stations. Average recruitment is shown in the top left barplot. Biomass-related reference points are shown in the top-right barplots, including $B_{100}$, current biomass ($B_{cur}$), and projected biomass ($B_{prj}$). Estimates of fishing mortality rates associated with MSY ($F_{MSY}$) and the OFL ($F_{OFL}$) are shown in the bottom barplots. Estimates of MSY and OFL themselves are shown in the bottom right barplot.

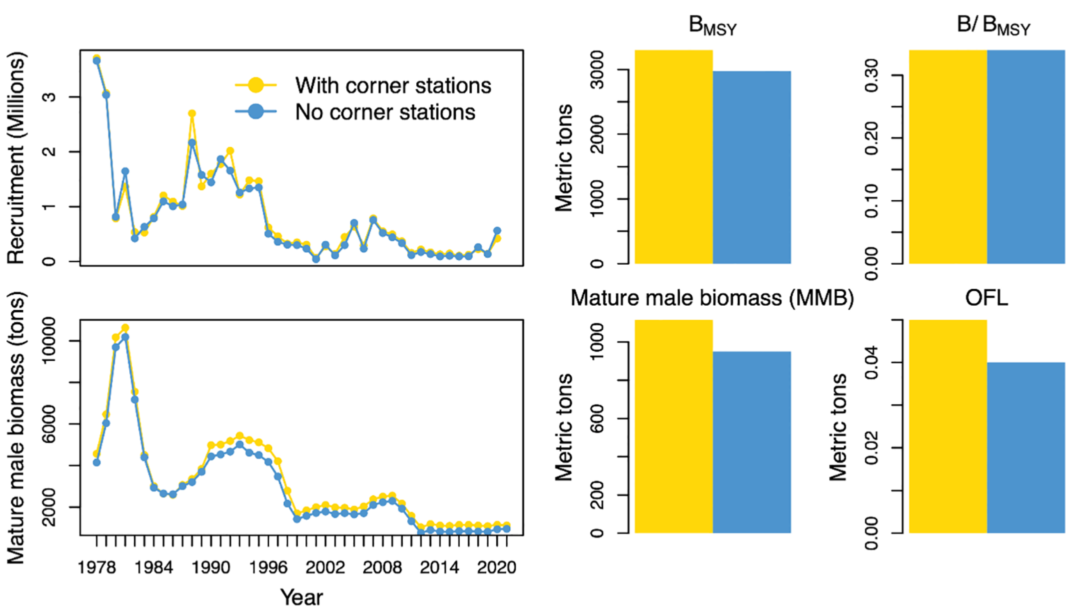



FIGURE 7
St. Matthew blue king crab stock assessment model output. Left-hand panels show model predictions with (yellow) and without (teal) corner stations for recruits (top row) mature male biomass (second row). Barplots on the right show estimated reference points produced with (yellow) and without (teal) corner stations. The biomass at which maximum sustained yield ($B_{MSY}$) is shown in the top left barplot, and the ratio of current biomass to $B_{MSY}$ ($B/B_{MSY}$) is shown in the top right barplot. The current level of mature male biomass is shown in the bottom left barplot, and the estimated overfishing limit (OFL) is shown in the bottom right barplot.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.fisheries.noaa.gov/foss/f?p=215%3A28.

## Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

## Author contributions

SK and LD conceived the study. LD designed and executed the analyses and wrote the paper. WS and KP conducted and helped interpret the stock assessment model runs. JR and ML contributed crab data and input on their use and interpretation. LB, LD, and SK contributed input on interpreting study results. All authors contributed to manuscript writing. All authors contributed to the article and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1219283/full#supplementary-material

## References

Brown, R. W., Fogarty, M., Legault, C., Miller, T., Nordahl, V., Politis, P., et al. (2007). Survey transition and calibration of bottom trawl surveys along the northeastern continental shelf of the United States. *Int. Council Explor. Sea CM.*

Cadigan, N. G., and Dowden, J. J. (2010). Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data. *Fishery Bull.* 108 (1).

Cadigan, N., Yin, Y., Benoît, H. P., and Walsh, S. J. (2022). A nonparametric-monotone regression model and robust estimation for paired-tow bottom-trawl survey comparative fishing data. *Fisheries Res.* 254, 106422. doi: 10.1016/j.fishres.2022.106422

Cao, J., Thorson, J. T., Richards, R. A., and Chen, Y. (2017). Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. *Can. J. Fisheries Aquat. Sci.* 74 (11), 1781–1793. doi: 10.1139/cjfas-2016-0137

Chen, Y., Chen, L., and Stergiou, K. I. (2003). Impacts of data quantity on fisheries stock assessment. *Aquat. Sci.* 65 (1), 92–98. doi: 10.1007/s000270300008

Ciannelli, L., Robson, B. W., Francis, R. C., Aydin, K., and Brodeur, R. D. (2004). Boundaries of open marine ecosystems: an application to the Pribilof Archipelago, southeast Bering Sea. *Ecol. Appl.* 14 (3), 942–953. doi: 10.1890/03-5016

Cochran, W. G. (1977). *Sampling techniques* (Hoboken, New Jersey: John Wiley & Sons).

Cooney, R. T., and Coyle, K. O. (1982). Trophic implications of cross-shelf copepod distributions in the southeastern Bering Sea. *Mar. Biol.* 70 (2), 187–196. doi: 10.1007/BF00397684

Coyle, K. O., and Cooney, R. T. (1993). Water column sound scattering and hydrography around the Pribilof Islands, Bering Sea. *Continental Shelf Res.* 13 (7), 803–827. doi: 10.1016/0278-4343(93)90028-V

Craighead, F. L., and Oppenheim, J. (1985). "Population estimates and temporal trends of Pribilof Island seabirds," in *Outer Continental Shelf Environmental Assessment Program (OCSEAP) (Anchorage Alaska:U.S. Department of Commerce) Final Report*, vol. 30. , 307–356.

DeFilippo, L. B., Buehrens, T. W., Scheuerell, M., Kendall, N. W., and Schindler, D. E. (2021). Improving short-term recruitment forecasts for coho salmon using a spatiotemporal integrated population model. *Fisheries Res.* 242, 106014. doi: 10.1016/j.fishres.2021.106014

DeFilippo, L. B., Thorson, J. T., O'Leary, C. A., Kotwicki, S., Hoff, J., Ianelli, J. N., et al. (2023). Characterizing dominant patterns of spatiotemporal variation for a transboundary groundfish assemblage. *Fisheries Oceanography.* doi: 10.1111/fog.12651

Dulvy, N. K., Rogers, S. I., Jennings, S., Stelzenmüller, V., Dye, S. R., and Skjoldal, H. R. (2008). Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas. *J. Appl. Ecol.* 45 (4), 1029–1039. doi: 10.1111/j.1365-2664.2008.01488.x

Ferrero, R. C., DeMaster, D. P., Hill, P. S., Muto, M., Lopez, A. L., and Brix, K. (2000). Alaska marine mammal stock assessment. *NOAA technical memorandum NMFS-AFSC.* p 119.

Field, J. C., Punt, A. E., Methot, R. D., and Thomson, C. J. (2006). Does MPA mean 'Major Problem for Assessments'? Considering the consequences of place-based management systems. *Fish Fisheries* 7 (4), 284–302. doi: 10.1111/j.1467-2979.2006.00226.x

Fournier, D., and Archibald, C. P. (1982). A general theory for analyzing catch at age data. *Can. J. Fisheries Aquat. Sci.* 39 (8), 1195–1207. doi: 10.1139/f82-157

Francis, R. C. (2011). Data weighting in statistical fisheries stock assessment models. *Can. J. Fisheries Aquat. Sci.* 68 (6), 1124–1138. doi: 10.1139/f2011-025

Gentry, R. L. (1998). *Behavior and ecology of the northern fur seal* Vol. 382 (Princeton, New Jersey: Princeton University Press).

Gish, R. K., and Vanek, V. A. (2010). *Project operational plan for the 2010 St. Matthew Island blue king crab survey* (Kodiak: Alaska Department of Fish and Game, Regional Information Report 4K10-12).

Godo, O. R. (1994). "Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys," in *Marine fish behavior in capture and abundance estimation.* (Hoboken New Jersey:Wiley and Sons)

Gunderson, D. R. (1993). *Surveys of fisheries resources* (Hoboken, New Jersey: John Wiley & Sons).

Hare, J., Blythe, B., Ford, K., Hooker, B., Jensen, B., Lipsky, A., et al. (2022). NOAA fisheries and BOEM federal survey mitigation implementation strategy - northeast U.S. Region. *NOAA Technical Memorandum* 292, 33. doi: 10.25923/jqse-x746

Hunt, G. L. Jr., Stabeno, P. J., Strom, S., and Napp, J. M. (2008). Patterns of spatial and temporal variation in the marine ecosystem of the southeastern Bering Sea, with special reference to the Pribilof Domain. *Deep Sea Res. Part II: Topical Stud. Oceanography* 55 (16-17), 1919–1944. doi: 10.1016/j.dsr2.2008.04.032

ICES (2020). Workshop on unavoidable survey effort reduction (WKUSER). *ICES Sci. Rep.* 2. doi: 10.17895/ices.pub.7453. 72.92pp(ICES, 2020).

ICES (2023). Workshop on unavoidable survey effort reduction 2 (WKUSER2). *ICES Sci. Rep.* doi: 10.17895/ices.pub.22086845.v1

Ichinokawa, M., Okamura, H., and Takeuchi, Y. (2014). Data conflict caused by model mis-specification of selectivity in an integrated stock assessment model and its potential effects on stock status estimation. *Fisheries Res.* 158, 147–157. doi: 10.1016/j.fishres.2014.02.003

Jadamec, L. S., Donaldson, W. E., and Cullenberg, P. (1999). *Biological field techniques for Chionoecetes crabs* (Fairbanks, AK: University of Alaska Sea Grant Program Report 99-02). 80 p.

Kimura, D. K., and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Rev. Fisheries Sci.* 14 (3), 245–283. doi: 10.1080/10641260600621761

Kotwicki, S., Ianelli, J. N., and Punt, A. E. (2014). Correcting density-dependent effects in abundance estimates from bottom-trawl surveys. *ICES J. Mar. Sci.* 71 (5), 1107–1116. doi: 10.1093/icesjms/fst208

Kotwicki, S., and Lauth, R. R. (2013). Detecting temporal trends and environmentally-driven changes in the spatial distribution of bottom fishes and crabs on the eastern Bering Sea shelf. *Deep Sea Res. Part II: Topical Stud. Oceanography* 94, 231–243. doi: 10.1016/j.dsr2.2013.03.017

Kotwicki, S., Lauth, R. R., Williams, K., and Goodman, S. E. (2017). Selectivity ratio: a useful tool for comparing size selectivity of multiple survey gears. *Fisheries Res.* 191, 76–86. doi: 10.1016/j.fishres.2017.02.012

Lauth, R. R., Dawson, E. J., and Conner, J. (2019). Results of the 2017 eastern and northern bering sea continental shelf bottom trawl survey of groundfish and invertebrate fauna. *Natl. Mar. Fish. Serv.*, 396, 1–265. doi: 10.25923/h118-nw41

Maunder, M. N., and Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Res.* 142, 61–74. doi: 10.1016/j.fishres.2012.07.025

Maureaud, A., Frelat, R., Pécuchet, L., Shackell, N., Mérigot, B., Pinsky, M. L., et al. (2021). Are we ready to track climate-driven shifts in marine species across international boundaries?-A global survey of scientific bottom trawl data. *Global Change Biol.* 27 (2), 220–236. doi: 10.1111/gcb.15404

Methratta, E. T., Hawkins, A., Hooker, B. R., Lipsky, A., and Hare, J. A. (2020). Offshore wind development in the Northeast US Shelf large marine ecosystem. *Oceanography* 33 (4), 16–27. doi: 10.5670/oceanog.2020.402

Miller, T. J. (2013). A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. *Can. J. Fisheries Aquat. Sci.* 70 (9), 1306–1316. doi: 10.1139/cjfas-2013-0136

Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather* 16 (1), 69–88. doi: 10.1002/2017SW001669

O'Leary, C. A., DeFilippo, L. B., Thorson, J. T., Kotwicki, S., Hoff, G. R., Kulik, V. V., et al. (2022). Understanding transboundary stocks' availability by combining multiple fisheries-independent surveys and oceanographic conditions in spatiotemporal models. *ICES J. Mar. Sci.* 79 (4), 1063–1074. doi: 10.1093/icesjms/fsac046

O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. (2020). Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (Gadus chalcogrammus) stock assessment. *Fisheries Oceanography* 29 (6), 541–557. doi: 10.1111/fog.12494

Ono, K., Licandeo, R., Muradian, M. L., Cunningham, C. J., Anderson, S. C., Hurtado-Ferro, F., et al. (2015). The importance of length and age composition data in statistical age-structured models for marine species. *ICES J. Mar. Sci.* 72 (1), 31–43. doi: 10.1093/icesjms/fsu007

Oyafuso, Z. S., Barnett, L. A., and Kotwicki, S. (2021). Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. *ICES J. Mar. Sci.* 78 (4), 1288–1300. doi: 10.1093/icesjms/fsab038

Oyafuso, Z. S., Barnett, L. A. K., Siple, M. C., and Kotwicki, S. (2022). *A flexible approach to optimizing the Gulf of Alaska groundfish bottom trawl survey design for abundance estimation* (Seattle, WA: U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-434). 142 p.

Palof, K. (2022). *Saint matthew island blue king crab stock assesssment* (Anchorage, AK: Alaska Department of Fish and Game).

Palof, K., Zheng, J., and Ianelli, J. (2020). "Saint matthew island blue king crab stock assessment, (2020)," in *Stock assessment and fishery evaluation report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands regions* (Alaska: North Pacific Fishery Management Council, Anchorage). 2020 Final Crab SAFE.

Peterson, C. D., Courtney, D. L., Cortés, E., and Latour, R. J. (2021). Reconciling conflicting survey indices of abundance prior to stock assessment. *ICES J. Mar. Sci.* 78 (9), 3101–3120. doi: 10.1093/icesjms/fsab179

Pinsky, M. L., and Mantua, N. J. (2014). Emerging adaptation approaches for climate-ready fisheries management. *Oceanography* 27 (4), 146–159. doi: 10.5670/oceanog.2014.93

Pope, J. G., and Shepherd, J. G. (1985). A comparison of the performance of various methods for tuning VPAs using effort data. *Ices J. Mar. Sci.* 42 (2), 129–151. doi: 10.1093/icesjms/42.2.129

Quinn, T. J., and Deriso, R. B. (1999). *Quantitative fish dynamics* (New York, New York: Oxford university Press).

R Core Team (2021). *R: A language and environment for statistical computing* (Vienna, Austria: R foundation for Statistical Computing). Available at: https://www.R-project.org/.

Santora, J. A., Rogers, T. L., Cimino, M. A., Sakuma, K. M., Hanson, K. D., Dick, E. J., et al. (2021). Diverse integrated ecosystem approach overcomes pandemic-related fisheries monitoring challenges. *Nat. Commun.* 12 (1), 1–10. doi: 10.1038/s41467-021-26484-5

Stauffer, G. D. (2004). *NOAA protocols for groundfish bottom trawl surveys of the nation's fishery resources, March 16, 2003.* (Seattle WA:U.S. Department of Commerce)

Stockhausen, W. T. (2014). *2014 stock assessment and fishery evaluation report for the tanner crab fisheries of the bering sea and aleutian islands regions.* (Anchorage AK: North Pacific Fishery Management Council)

Stockhausen, W. T. (2021a). *2021 stock assessment and fishery evaluation report for the pribilof islands blue king crab fisheries of the bering sea and aleutian islands regions.* (Anchorage AK:North Pacific Fishery Management Council)

Stockhausen, W. T. (2021b). *2021 stock assessment and fishery evaluation report for the tanner crab fisheries of the bering sea and aleutian islands regions* (Anchorage, AK: North Pacific Fishery Management Council). Available at: https://meetings.npfmc.org/CommentReview/DownloadFile?p=55d56b05-4f74-4c5d-9372-28ced4a2bfc2.pdf&fileName=C1%20Tanner%20Crab%20SAFE.pdf.

Stompe, D. K., Moyle, P. B., Kruger, A., and Durand, J. R. (2020). Comparing and integrating fish surveys in the San Francisco Estuary: why diverse long-term monitoring programs are important. *San Francisco Estuary Watershed Sci.* 18 (2). doi: 10.15447/sfews.2020v18iss2art4

Szuwalski, C. S. (2022) *2022 assessment for Pribilof Islands red king crab*. Available at: https://meetings.npfmc.org/CommentReview/DownloadFile?p=4783b2a2-e8a2-4efe-bdfb-37f89216c1b3.pdf&fileName=PIRKC%20SAFE%202022%20Final.pdf.

Szuwalski, C. S., Foy, R. J., and Turnock, B. J. (2014). "2014 Stock assessment and fishery evaluation report for the Pribilof Islands red king crab fishery of the Bering Sea and Aleutian Islands Regions," in *Stock assessment and fishery evaluation report for the king and tanner crab fisheries of the bering sea and aleutian islands regions* (605 W. 4th Avenue, no. 306, Anchorage, AK 99501: North Pacific Fishery Management Council), 546–605.

Szuwalski, C., and Turnock, J. (2016). A stock assessment for eastern Bering Sea snow crab 167–251. Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. *North Pacific Fishery Management Council, Stock Assessment and Fishery Evaluation Report*

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Thorson, J. T., Barbeaux, S. J., Goethel, D. R., Kearney, K. A., Laman, E. A., Nielsen, J. K., et al. (2021). Estimating fine-scale movement rates and habitat preferences using multiple data sources. *Fish Fisheries* 22 (6), 1359–1376. doi: 10.1111/faf.12592

Thorson, J. T., and Barnett, L. A. (2017). Comparing estimates of abundance trends and distribution shifts using single-and multispecies models of fishes and biogenic habitat. *ICES J. Mar. Sci.* 74 (5), 1311–1321. doi: 10.1093/icesjms/fsw193

Thorson, J. T., Clarke, M. E., Stewart, I. J., and Punt, A. E. (2013). The implications of spatially varying catchability on bottom trawl surveys of fish abundance: a proposed solution involving underwater vehicles. *Can. J. Fisheries Aquat. Sci.* 70 (2), 294–306. doi: 10.1139/cjfas-2012-0330

Thorson, J. T., and Haltuch, M. A. (2019). Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. *Can. J. Fisheries Aquat. Sci.* 76 (3), 401–414. doi: 10.1139/cjfas-2018-0015

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES J. Mar. Sci.* 72 (5), 1297–1310. doi: 10.1093/icesjms/fsu243

Vølstad, J. H. (2012). *CIE independent peer review report.* (Bergen Norway:Center for Independent Experts) https://www.st.nmfs.noaa.gov/science-quality-assurance/cie-peer-reviews/cie-review-2012.

Wakabayashi, K., Bakkala, R. G., and Alton., M. S. (1985). "Methods of the U.S.-Japan demersal trawl surveys," in *Results of cooperative U.S.-Japan groundfish investigations in the Bering Sea during May-August 1979*. Eds. R. G. Bakkala and K. Wakabayashi (Vancouver, CA: Int. North Pac. Fish. Comm. Bull. 44), 7–29.

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. (2009). Incorporating time-varying catchability into population dynamic stock assessment models. *Rev. Fisheries Sci.* 18 (1), 7–24. doi: 10.1080/10641260903294647

Wyllie-Echeverria, T. I. N. A., and Wooster, W. S. (1998). Year-to-year variations in Bering Sea ice cover and some consequences for fish distributions. *Fisheries Oceanography* 7 (2), 159–170. doi: 10.1046/j.1365-2419.1998.00058.x

Zacher, L. S., Richar, J. I., and Foy, R. J. (2020). *The 2019 eastern and northern Bering Sea continental shelf trawl surveys: Results for commercial crab species* (Anchorage, AK: U.S. Department of Commerce NOAA Technical Memorandum NMFS-AFSC-400). 234 p.

Zimmermann, F., and Enberg, K. (2017). Can less be more? Effects of reduced frequency of surveys and stock assessments. *ICES J. Mar. Sci.* 74 (1), 56–68. doi: 10.1093/icesjms/fsw134

TYPE Original Research
PUBLISHED 04 September 2023
DOI 10.3389/fmars.2023.1176696

Check for updates

OPEN ACCESS

EDITED BY
Fabio Campanella,
Centre for Environment, Fisheries and Aquaculture Science (CEFAS), United Kingdom

REVIEWED BY
Jorge Paramo,
University of Magdalena, Colombia
Serdar Sakinan,
Wageningen Marine Research IJmuiden, Netherlands

*CORRESPONDENCE
Allison L. White
awhit100@fiu.edu

RECEIVED 28 February 2023
ACCEPTED 16 August 2023
PUBLISHED 04 September 2023

CITATION
White AL, Sullivan PJ, Binder BM and Boswell KM (2023) An evaluation of survey designs and model-based inferences of fish aggregations using active acoustics. *Front. Mar. Sci.* 10:1176696. doi: 10.3389/fmars.2023.1176696



# An evaluation of survey designs and model-based inferences of fish aggregations using active acoustics

Allison L. White[1]*, Patrick J. Sullivan[2], Benjamin M. Binder[1] and Kevin M. Boswell[1]

[1]Marine Ecology and Acoustics Laboratory, Department of Biology, Florida International University, North Miami Beach, FL, United States, [2]Department of Natural Resources, Cornell University, Ithaca, NY, United States

"Star" survey designs have become an increasingly popular alternative to parallel line designs in fisheries-independent sampling of areas with isolated fish aggregations, such as artificial reefs, seamounts, fish aggregating devices, and spawning aggregation sites. In this study, we simulated three scenarios of fish aggregating around a feature of interest with variations in the size and complexity of aggregations as well as their location relative to the habitat feature. Simulated and empirical data representing goliath grouper (*Epinephalus itajara*) spawning aggregations at artificial reefs were utilized as a case study, and scenarios were generated in relation to both a single habitat feature and a reef complex with multiple structures. Seven variations of survey design using both star and parallel transects were examined and compared by geostatistical and generalized additive models (GAMs) to identify the most robust approach to quantify fish aggregations in each scenario. In most scenarios, precision in the mean and variability of backscatter estimates is not significantly affected by the number of transects passing over the habitat feature as long as at least one pass is made. Estimation error is minimized using the GAM approach, and is further reduced when sampling variance is high, which was better accomplished by parallel designs overall. These results will help inform surveyors on the best overall approach to improve precision in quantifying fish aggregations given basic knowledge of their behavior around an established habitat feature and help them to adapt their survey designs based on common difficulties in sampling these populations simulated below.

KEYWORDS

**fish aggregations, active acoustics, survey design, star surveys, model-based inference, generalized additive models, geostatistics, fisheries-independent surveys**

## 1 Introduction

Many fish species aggregate to spawn and/or associate with conspicuous habitat features such as seamounts and artificial reefs (Doonan et al., 2003). These sites are of great interest among fisheries scientists and managers, as fish which aggregate in highly localized and predictable areas (especially when associated with established bottom features) may be susceptible to higher catchability (Hieu et al., 2014). Continuous mobile surveys, such as those performed by towed active acoustics or camera systems, are often the optimal method to provide fisheries-independent inferences about the size of such highly localized fish aggregations. Although mobile surveys allow for ensured sampling over the discrete aggregation of interest, this inherently negates randomness in the survey design. Design-based methods of population inference, which are based on the distribution of all possible estimates within a survey design, are therefore not ideal, and surveyors of such aggregations often rely on model-based estimations of population size (Neyman, 1934; Jolly and Hampton, 1990; Gregoire, 1998; Rivoirard et al., 2000; Chang et al., 2017).

Two commonly used survey designs for mobile surveys of highly localized fish aggregations are parallel line and star designs (Figures 1A, B). Parallel line designs consist of either randomly- or evenly-spaced parallel transects. This design allows for stratified randomization by incorporating random-spacing between transects and/or randomly selecting the starting point (Jolly and Hampton, 1990). In the case of highly localized aggregations, however, evenly-spaced parallel transects are generally preferred in order to maximize the number of passes over the aggregation and starting points are selected in relation to the location of the aggregation or habitat feature (e.g. Taylor et al., 2006; Boswell et al., 2010; Kang et al., 2011).

Though parallel line surveys offer better coverage of the area surrounding a fish aggregation and less spatial autocorrelation between transect nodes, they often involve a greater number of transects and present several practical difficulties in maneuvering tight turns. Star surveys involve fewer transects which are arranged in alternating directions and which all cross at the center of the aggregation site. Star designs may be easier to maneuver and provide a higher sampling of the targeted aggregation per survey effort, but they have an inherently large spatial autocorrelation between transect nodes (i.e., the point at which transects bisect one another). They also offer poor coverage of the area surrounding the habitat feature of interest, which results in a decreased ability to measure variability in population estimates. Despite these downfalls, star designs have increased in popularity in recent years due to the reduction in time required to obtain multiple passes over the aggregation of interest. Many surveyors believe that they can reduce the cost of vessel time and/or maximize the number of surveys conducted by employing star designs over parallel designs. Here, we weigh the advantage of cost/time reduction against decreased precision in population estimates through comparison of model-based inferences from both survey designs.

Model-based inferences do not require random sampling, and are therefore less heavily influenced by the high spatial sampling autocorrelation inherent in both parallel line and star surveys. Unlike design-based approaches, however, model-based approaches are strongly dependent on assumptions about the underlying distribution and structure of a population (Gregoire, 1998). Geostatistical models, for example, assume that the spatial distribution of the population is stationary and isotropic. In stationary populations, the statistical properties of the population do not change across time or space, and in isotropic populations the correlation between any two observations depends solely on the distance between them regardless of their relative orientation. Live fish aggregations may not fulfill either of these assumptions, resulting in biased estimations of population size. A further complication with continuous data of fish aggregations is the high

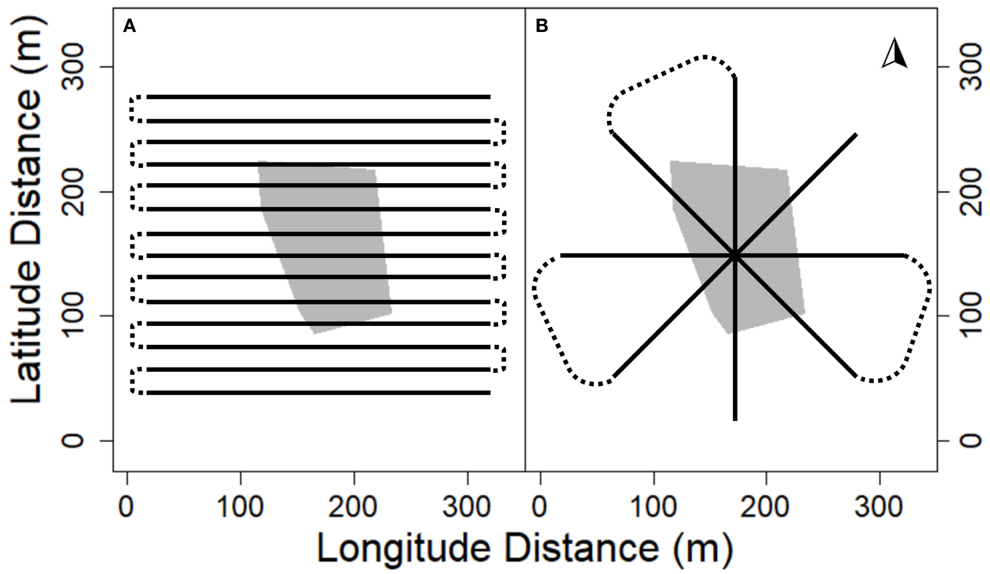



FIGURE 1
**(A)** Parallel line survey designs and **(B)** star survey designs plotted over the sunken barge (grey shaded area) at MG111. Dashed lines represent turns made between transects (solid lines) which were excluded from analysis.

proportion of zero observations. Most models, geostatistical or non-geostatistical, do not perform well on highly skewed data without extraneous methods.

In this study, we evaluate the relative biases associated with different survey designs and commonly used inference approaches to quantifying goliath grouper (*Epinephalus itajara*) spawning aggregations over artificial reefs in Jupiter, FL. Three scenarios of the spatial distribution of goliath grouper acoustic backscatter at two artificial reefs are surveyed by multiple variations of parallel line and star designs and estimated by a geostatistical and a non-geostatistical modeling approach as well as a design-based approach (Figure 2). Understanding the spatial distribution of the underlying population is an important first step in constructing sampling designs (Gunderson, 1993). Initial surveys should be considered exploratory and aim to give context about the spatial distribution of the targeted population. The purpose of this study is to provide guidance on least-biased survey designs and inference approaches given previous knowledge about the spatial trends of fish aggregations. We also address common issues involved with surveying populations that are assumed to be densely aggregated over or around a discrete and predictable location.

# 2 Methodology

## 2.1 Acoustic data collection and processing

Goliath grouper spawning aggregations were surveyed at two artificial reef complexes in Jupiter, FL. The northernmost reef complex, colloquially known as MG111, consists of the remains of a 59 m long barge resting at a depth of 18 m. A field of ~3 m tall concrete columns standing upright on the seafloor stretches north of the barge over roughly 100 x 50 m of sand. The second reef complex, colloquially known as Wreck Trek (WRT), is located ~2 km south of MG111 and includes a collection of three shipwrecks arranged north to south with a ~170 m distance between the northernmost and southernmost wrecks. The 45 m long Esso Bonaire oil tanker, 17 m long Miss Jenny barge, and the stern of the 50 m long Zion Train cargo ship all lie in 27 m of water. Both artificial reef complexes were selected for this study based on their consistent use by goliath grouper as spawning aggregation sites (Koenig et al., 2017).

Active acoustic surveys (n=10) were conducted over these reefs during goliath grouper spawning months (August through November) in 2017 and 2018. Surveys were conducted with a 38 kHz (10°) split-beam Simrad EK80 echosounder towed at the surface ~15 m behind a 7 m research vessel. The echosounder was calibrated following the standard sphere method (38.1 mm tungsten carbide sphere with 6% cobalt binder; Demer et al., 2015). Two transect designs were conducted sequentially at each survey event: 1) a parallel line survey which consisted of 15-20 east-west parallel transects that were spaced ~20 m apart (Figure 1A); and 2) a star survey consisting of four radial transects separated by ~45° (Figure 1B). Transects in both survey designs were 300 m in length and centered over the main structure at each of the reef complexes.

Raw acoustic data were visualized and processed in Echoview 12.0 (Echoview Software Pty. Ltd.). Data were visually inspected to remove turns between transects and dropout from rapid speed

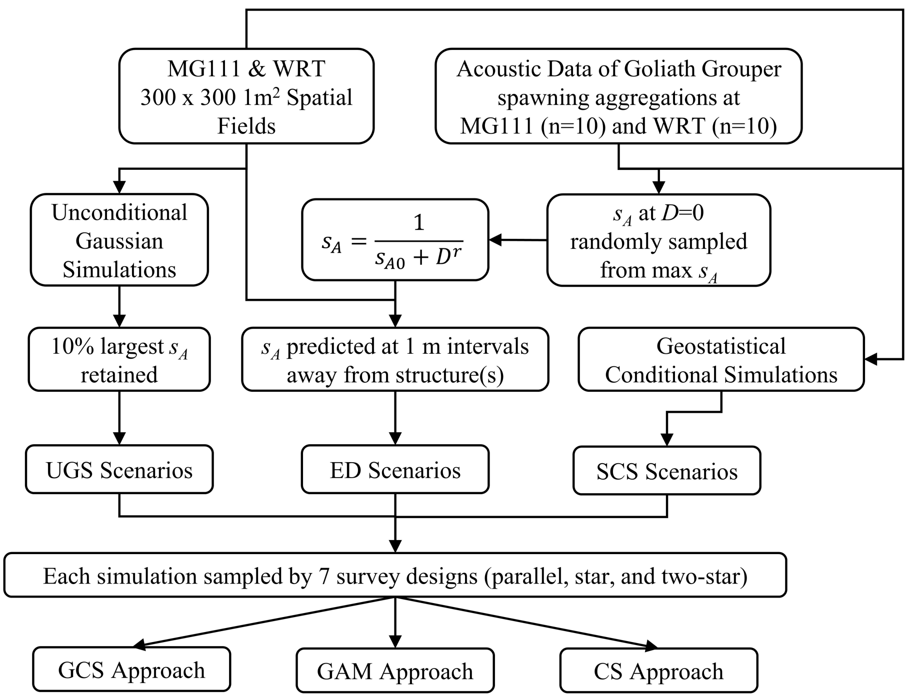



FIGURE 2
Flow chart showing the steps used to simulate, sample, and make inferences about $s_A$ of goliath grouper aggregations at the MG111 and WRT artificial reefs.

changes. A best bottom candidate detection algorithm with an offset of 0.5 m was applied to exclude the acoustic dead zone and seafloor from analysis. Data within 5.5 m of the transducer face were excluded to eliminate transducer ringdown and near-field effects. Volume-backscattering strength ($S_v$; dB re 1m$^{-1}$) data were thresholded at -60 dB re 1 m$^{-1}$ to eliminate sources of backscatter that did not originate from swim-bladdered fish. Backscatter representing large-bodied goliath grouper was isolated by applying a -40 dB re 1 m$^2$ threshold to the target strength (TS; dB re 1 m$^2$) data (Binder, 2022). The areas in the TS echogram attributed to goliath grouper were then masked over the $S_v$ echogram of swim-bladdered fish, which was used to calculate the nautical area scattering coefficient ($s_A$; m$^2$ nmi$^{-2}$) from echo integrals in 5 m along-track x 5 m depth intervals from the best bottom candidate exclusion line (MacLennan et al., 2002). Resulting echo integrals near the surface less than 2 m in maximum height were excluded from further analysis.

The convex hull of the mean coordinates for all 5 m transect distances throughout the water column was used to generate a spatial field with a 300 m x 300 m grid of 1 m$^2$ cells for simulation. Three scenarios of the underlying distribution of fish aggregations at the two artificial reefs were simulated across these spatial fields: an unconditional Gaussian scenario, an exponential decay scenario, and a stochastic conditional scenario (Figure 3). Each scenario included ten simulations of goliath grouper $s_A$ over each the spatial fields of MG111 and WRT.

## 2.2 Unconditional Gaussian scenario

The first scenario was unconditional to the data and aimed to reflect the assumptions inherent in most modelling approaches that the underlying distribution of fish exhibits spatial independence. In this scenario, fish are randomly distributed throughout the spatial field with no correlation to any promontories. A Gaussian random field with a Matérn covariance structure was generated over the spatial fields in each simulation (Cressie, 1993). Variability in the randomness of $s_A$ distributed across the spatial field and in the size

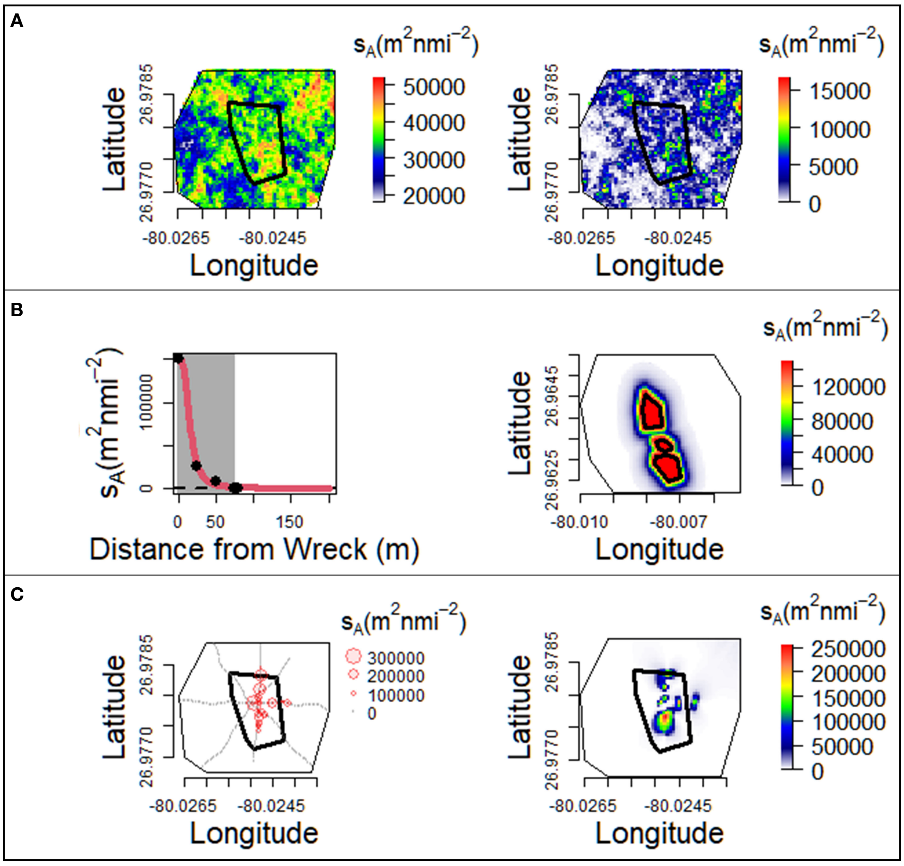



FIGURE 3
Construction of the three scenarios of the underlying distribution of goliath grouper aggregations. Black polygons indicate the structures at each artificial reef. Unconditional Gaussian simulations **(A)** were generated over the spatial fields of each wreck (left) and transformed to simulate a highly skewed population (right; MG111 shown). The exponential decay scenario **(B)** was built by modeling $s_A$ as a function of distance from the boundary of each structure at the two reefs (left). Maximum $s_A$ observed from acoustic surveys of goliath grouper spawning aggregations and area of influences (shown by the grey shaded region) were used to generate datasets (black filled points) fit by a non-linear regression of the exponential decay function (red line). Predicted values at 1 m distance intervals were applied to concentric polygons around the structures at each artificial reef (right; WRT shown). Stochastic conditional simulations **(C)** were built using $s_A$ predicted by GCS (right) of $s_A$ observed during 2017 and 2018 acoustic surveys of goliath grouper spawning aggregations (left; MG111 shown).

of fish aggregations were incorporated into these simulations following the method described in Chang et al. (2017). To simulate the highly skewed distribution of data present in most water column sampling methods (including active acoustics), the $s_A$ values generated by each unconditional Gaussian simulation (UGS) were subtracted by their mode and a 100 m$^2$ nmi$^{-2}$ offset. All adjusted $s_A$ values below 100 m$^2$ nmi$^{-2}$ were assigned zero values, resulting in spatial fields with 44-55% $s_A$=0 (Figure 3A).

## 2.3 Exponential decay scenario

The second scenario was simulated to reflect the assumption that fish backscatter is highest directly over promontories and exhibits an exponential decay in every direction away from the boundaries of the structure. In this scenario, goliath grouper $s_A$ was modelled at increasing distances away from the boundaries of the main structures in each reef complex following the exponential decay (ED) function:

$$s_A = \frac{1}{s_{AO} + D^r} \quad \text{(Eq. 1)}$$

where goliath grouper $s_A$ is a function of the reciprocal of the goliath grouper backscatter directly over the structure $s_{A0}$ as it decays at rate $r$ with increasing distance $D$ (m) away from the boundary of the structure. A dataset was generated for each simulation with eight observations: 1) the maximum $s_A$ at $D$ = 0 (randomly sampled from maximum goliath grouper $s_A$ observed from all processed acoustic surveys); 2) the minimum distance where $s_A$ = 0 to represent the area of influence around the artificial structure (randomly sampled between 35 and 179 m as estimated by White et al., 2022); and 3-8) six sequential distances at 10 m intervals greater than the area of influence where $s_A$ = 0 to represent the horizontal asymptote in goliath grouper backscatter. Nonlinear least squares regression models were used to fit the exponential decay function (Eq. 1) to these generated datasets (Bates and Watts, 1988). The predicted $s_A$ at each 1 m distance away from the structures were then attributed to concentric polygons drawn at 1 m intervals away from the boundary of the main structure across the spatial field of $s_A$ at MG111. For the three wrecks at WRT, predicted $s_A$ was attributed to concentric polygons which were drawn at 1 m intervals away from the boundaries of each of the three structures (Figure 3B). Goliath grouper $s_A$ simulated by this method had 50-78% of the observations where $s_A$=0.

## 2.4 Stochastic conditional scenario

The last scenario was built upon real world acoustic data of goliath grouper spawning aggregations and was designed to incorporate the stochasticity of live fish aggregations. This stochastic conditional simulation (SCS) scenario was made to reflect the behavior of fish around a habitat feature at any given point in time, such as spawning behaviors or utilizing structures to shelter from currents, predator avoidance, varying light levels around a structure, etc. Geostatistical conditional simulations (GCS) were used to generate realizations of goliath grouper spawning events observed at MG111 and WRT (Figure 3C). These simulations produced random fields which were conditional to the $s_A$ observed during goliath grouper spawning events and highly zero-inflated (40-94% $s_A$=0). See section 2.6 below for a more detailed description of how GCS were performed.

## 2.5 Survey designs

Goliath grouper $s_A$ was sampled from each simulation in the above three scenarios via seven variations in survey designs of parallel line and star transect methods (Figure 4). These designs were:

1. Ideal transects: Evenly spaced 250 m long parallel line transects (n=6 spaced 40 m apart for MG111 and n=12 spaced 20 m apart for WRT) centered so that four transects crossed over the structure at MG111 and four transects crossed over each of the Esso Bonaire and Zion Train wrecks at WRT. The ideal star design was composed of four 250 m long transects separated by 45° which converged over the center point of the barge at MG111 and over the center point of the three shipwrecks at WRT. An additional transect method with two stars of four 250 m long transects each was constructed at WRT with one star centered over the center point of the Esso Bonaire and one star centered over the Zion Train wreck.
2. Single offset transect: As (1) but with one transect offset over the reefs so that only three transects passed over each structure instead of the four in the ideal transects.
3. Double offset transects: As (1) but with two transects offset over the reefs so that only two transects passed over each structure.
4. Triple offset transects: As (1) but with three transects offset over the reefs so that only one transect passed over each structure.
5. Structure avoidance transects: The same number and length of transects as described in (1), but with no transects passing over the structures at each reef. For the parallel method, transects were arranged as in (1) but with transects which would have passed over the structure ending at the boundary of the structure. For the star and two-star methods, transects were arranged around the borders of the structure in order to maximize the distance along each transect which passed close to the structure (while avoiding other structures at WRT).
6. Shorter transects: Transects were arranged as in (1) but were only 150 m in length.
7. Non-uniform transects: Selected from real transects driven over MG111 and WRT. Like (1), the selected real-world transects were centered over the structures at each reef so that four transects passed over the barge at MG111 and the Esso Bonaire and Zion Train wrecks at WRT in all designs. Unlike (1), the spacing between transects varies along the

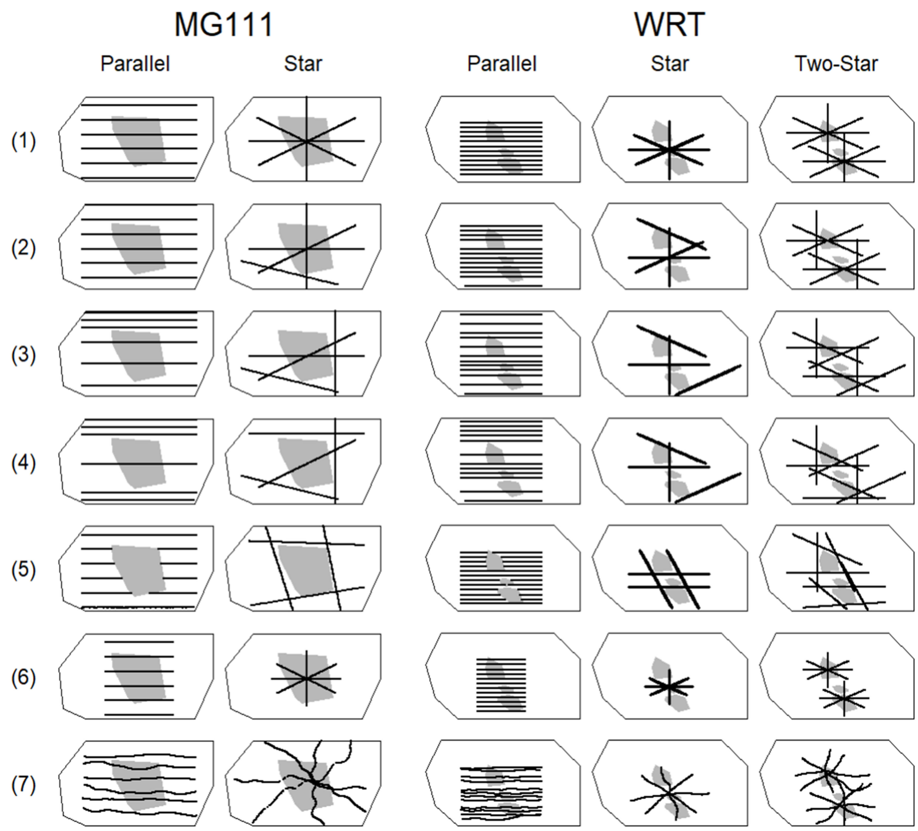



FIGURE 4
The seven survey designs sampled from goliath grouper $s_A$ simulated within the spatial field (outer polygon) at MG111 (left) and WRT (right). Solid black lines represent survey transects and grey shaded areas show the structure(s) at each artificial reef. Ideal transects (Design 1), single offset transects (Design 2), double offset transects (Design 3), triple offset transects (Design 4), structure avoidance transects (Design 5), shorter transects (Design 6), and non-uniform transects (Design 7) are shown for the parallel, star, and two-star methods.

transect and transect length is not exactly 250 m. Rather, all transects were cutoff at 125 m distance away from the center point of each design in every direction.

The ideal transects (1) were designed to represent a scenario where the surveyor has perfect control over the placement of all transects (including vessel behavior). As this is not always possible, the remaining survey designs were compared to the ideal case in order to examine potential biases resulting from various non-ideal scenarios. Designs 2, 3, and 4 were compared to the ideal in order to estimate how much a surveyor should prioritize driving over the structure. Design 5 represents a scenario in which the surveyor is unable to drive over the aggregation of interest due to the structure extending out of the water (such as an oil rig) or high boat traffic directly over the feature of interest (which was often the case at MG111 and WRT during daylight hours). The shorter transects (6) are designed to estimate the importance of transect length away from the structure of interest. Lastly, the non-uniform transects (7) were designed to mimic situations in which the surveyor is not able to drive transects in straight lines. When towing equipment over a fish aggregation many factors may influence the path of each transect, such as strong currents, visibility, sea state, presence of other vessels, etc. Each simulation of the UGS, ED, and SCS scenarios was sampled by the seven survey designs for each of the three transect methods and input to the following modeling- and design-based approaches to estimate goliath grouper $s_A$ at the two reefs.

## 2.6 Inference approaches

We tested the performance of two model-based methods for spatial interpolation of goliath grouper $s_A$: geostatistical conditional simulations (GCS) and generalized additive models (GAMs). Conditional simulations of ordinary kriging (a common geostatistical model) were selected for this study due to their ability to generate the spatial variability of population estimates while honoring the data at observed locations. While kriging allows for interpolation of data at unsampled locations (such as between transects), it also minimizes error variance (Isaaks and Srivastava, 1989). By generating multiple realizations (or simulations) of the spatial structure of the population (captured by characteristics such as the histogram and variogram), uncertainty in the global estimate of population size can be obtained. Uncertainty in global estimates was of particular concern for star survey designs, where the areas not sampled between transects increases towards the outer regions of the survey. Conditional simulations are made on the Gaussian random function model, which is not applicable to highly skewed data such as in our case. Therefore, a Gaussian anamorphosis of $s_A$ with a Gibbs sampler to iteratively simulate the points where $s_A = 0$ was performed for all GCS following the methods described by Woillez et al. (2009; 2016). All GCS and Gaussian anamorphoses were conducted using the R package "RGeostats" (MINES ParisTech/ARMINES, 2022).

GAMs are a non-parametric regression approach capable of predicting non-linear relationships (Hastie and TibshIrani, 1990). In these models, splines are utilized to estimate smooth functions between predictor and response variables. Spatial inferences can be made by including a smooth functional relationship of the exploratory variable as a response of the interaction between latitudes and longitudes of observations. GAMs are more flexible towards non-Gaussian data than geostatistical approaches. In this study, we modelled goliath grouper $s_A$ as a function of the thin plate regression spline interaction between latitude and longitude assuming a Tweedie distribution (Wood, 2003). The Tweedie dispersion model has a non-negative support and a discrete mass at zero that makes it useful in modeling datasets with a mixture of zero and positive observations (Dunn and Smyth, 2005). GAMs were performed in R using the "mgcv" package (Wood, 2017).

In addition to the two model-based approaches described above, one design-based approach was considered. Mean goliath grouper $\bar{s}_A$ was calculated in the surveyed space $j$ using a cluster sampling (CS) formula (Scheaffer et al., 2012):

$$\bar{s}_{Aj} = \frac{\sum_{k=1}^{K} n_k \bar{s}_{Ak}}{\sum_{k=1}^{K} n_k} \qquad \text{(Eq. 2)}$$

where $\bar{s}_{Ak}$ is the average $s_A$ in transect $k$ with $n$ observations across all transects $K$. A global estimate of $\bar{s}_A$ was interpolated by dividing the sum of observations across all transects from $\bar{s}_{Aj}$ multiplied by the area of the convex hull of each survey design. Calculation of the standard deviation was adapted from Scheaffer et al. (2012) as:

$$\mathrm{sd}(\bar{s}_A) = \sqrt{\frac{1}{K(K-1)} \sum_{k=1}^{K} (\bar{s}_{Ak} n_k - \bar{s}_{Aj} n_k)^2} \qquad \text{(Eq. 3)}$$

## 2.7 Survey design and model comparison

The precisions of global estimates of the mean and coefficient of variation (CV) of $s_A$ predicted by each combination of survey design and inference approach under the three scenarios of underlying fish distribution were compared using relative mean absolute error (RMAE):

$$\mathrm{RMAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(\hat{z} - z)}{z} \right| \qquad \text{(Eq. 4)}$$

where $\hat{z}$ is the predicted global estimate of the mean or CV of $s_A$ and $z$ is the true $s_A$ value. Global estimates of $\hat{z}$ were calculated from $s_A$ interpolated over the convex hull of the survey design for each inference approach, and $z$ was calculated from the true $s_A$ generated by each simulation within the same space. Five null hypotheses ($H_{01\text{-}5}$) were formally tested to compare RMAE resulting from the seven survey designs using a two-way analysis of variance (ANOVA) and Tukey's honest significant difference (TukeyHSD) for pairwise comparisons. One ANOVA was performed for each of the three inference approaches at the two wrecks under each of the three scenarios for RMAE of the mean and of the CV, resulting in 36 models of RMAE as a function of the interactions between transect method (parallel, star, or two-star) and design (1:7). $H_{01}$ was tested by comparing the difference in mean RMAE of each transect method (Designs 1-7) at each of the two artificial reefs. $H_{02}$ compared errors among survey designs with differing number of transects passing over the feature of interest (Designs 1-4) for each transect method. $H_{03}$ addressed the scenario in which surveying directly over the feature of interest is not possible via comparisons between Design 5 in each transect method. Precision in longer (Design 1) and shorter (Design 6) transects was compared in $H_{04}$, and in $H_{05}$ the precision in uniform (Design 1) and non-uniform (Design 7) transects was compared. All simulations, survey design generations, and inference approaches were performed in R (R Core Team, 2022).

## 3 Results

Among the three inference approaches, the design-based approach (CS) had the poorest fit (Tables 1, 2). Error in the mean and variability of $s_A$ estimated by the design approach was at least 50x larger than either of the modeling-based approaches, and is therefore not reported in the results of *post hoc* analyses. Both modeling approaches reproduced highly skewed distributions of $s_A$ similar to the simulated values, including a large number of zeros (Figure 5). Precision in mean $s_A$ was highest in the GAM approach, although estimation of variability in $s_A$ was equivocal between the GAM and GCS approaches (Table 1). A two-way ANOVA revealed that the relative performances of GCS versus GAMs was somewhat

TABLE 1 Average RMAEs of the mean and CV $s_A$ estimated by the three inference approaches (geostatistical conditional simulations, generalized additive models, and cluster sampling).

| Scenario | RMAE (mean) | | | RMAE (CV) | | |
|---|---|---|---|---|---|---|
| | GCS | GAM | CS | GCS | GAM | CS |
| UGS | 0.30 | 0.09 | 55 | 0.55 | 0.55 | 36 |
| ED | 0.31 | 0.12 | 41 | 0.44 | 0.46 | 43 |
| SCS | 0.57 | 0.39 | 37 | 0.47 | 0.36 | 52 |

TABLE 2 Mean and CV of the true $s_A$ compared to $s_A$ estimated by the geostatistical conditional simulations, generalized additive models, and cluster sampling.

| Scenario | Statistic | Approach | MG111 | | WRT | | |
|---|---|---|---|---|---|---|---|
| | | | Parallel | Star | Parallel | Star | Two-Star |
| UGS | mean | True | 2205 | 2189 | 4000 | 4108 | 4050 |
| | | GCS | 2741 | 2881 | 4970 | 4697 | 5024 |
| | | GAM | 2086 | 2171 | 3971 | 3746 | 3886 |
| | | CS | 86287 | 112156 | 141844 | 378420 | 260356 |
| | CV | True | 145 | 145 | 143 | 143 | 141 |
| | | GCS | 62 | 61 | 63 | 72 | 63 |
| | | GAM | 69 | 68 | 61 | 62 | 63 |
| | | CS | 6302 | 5891 | 5542 | 3697 | 4749 |
| ED | mean | True | 37978 | 42543 | 12778 | 15776 | 12485 |
| | | GCS | 40267 | 52211 | 14676 | 17242 | 13538 |
| | | GAM | 39409 | 42863 | 12017 | 14354 | 11560 |
| | | CS | 877228 | 1668315 | 225132 | 856404 | 868508 |
| | CV | True | 145 | 137 | 173 | 151 | 172 |
| | | GCS | 108 | 113 | 181 | 176 | 172 |
| | | GAM | 78 | 63 | 106 | 81 | 99 |
| | | CS | 11981 | 4090 | 7451 | 7948 | 2944 |
| SCS | mean | True | 1692 | 1927 | 2333 | 2622 | 2411 |
| | | GCS | 1987 | 2503 | 3135 | 2737 | 3393 |
| | | GAM | 1172 | 1618 | 2062 | 2164 | 2358 |
| | | CS | 29301 | 10134 | 77195 | 129120 | 94423 |
| | CV | True | 526 | 509 | 458 | 427 | 453 |
| | | GCS | 256 | 279 | 225 | 182 | 223 |
| | | GAM | 354 | 374 | 275 | 294 | 328 |
| | | CS | 36419 | 2246 | 16114 | 14174 | 24545 |

Values shown represent averages across simulations for each transect method at the MG111 and WRT wrecks.

dependent on transect method for the estimation of $s_A$ (Table 3). Pairwise comparisons from TukeyHSD tests revealed that GAMs significantly outperformed GCS models in the estimation of mean $s_A$ at both wrecks and in the estimation of variability at MG111 regardless of transect method under the UGS scenario. In the ED scenario, however, there was no significant variation in RMAE to suggest improved performance of one model over the other when employing the two-star design. Variability of $s_A$ at MG111was better predicted by GCS than GAM, but only by the parallel transect method. In the SCS scenario, the only significant evidence of GAMs outperforming GCS models was provided when using the star design at MG111.

Average residuals for each transect method ($\bar{x}$) showed that mean $s_A$ at both wrecks was overestimated by the GCS approach and underestimated by the GAM approach for the UGS and SCS scenarios (Figure 6A; Table 2). In the ED scenario, mean $s_A$ at MG111 was overestimated by both GCS and GAM, while mean $s_A$ at WRT was underestimated by both modeling approaches. Variability in $s_A$ was underestimated by both modeling approaches in all three scenarios, with one exception (Figure 6B; Table 2). The CV of $s_A$ at WRT was overestimated by the GCS approach in the ED scenario, which was driven by an overestimation of CV in Design 5 (structure avoidance) in all transect methods.

Within survey designs, the selection of transect method (parallel, star, or two-star) was the only significant influence on model-based precision of $s_A$ estimated under all three scenarios ($H_{A1}$, Table 4). In the UGS scenario, precision of variability in $s_A$ predicted by GCS was highest for the star design in comparison to both parallel and two-star designs at WRT. Parallel designs produced the least-biased estimation of mean $s_A$ at the same wreck using GAMs. Precision in the mean and CV of $s_A$ did not differ significantly between transect methods at the single structure reef using either modeling approach under this scenario. This was

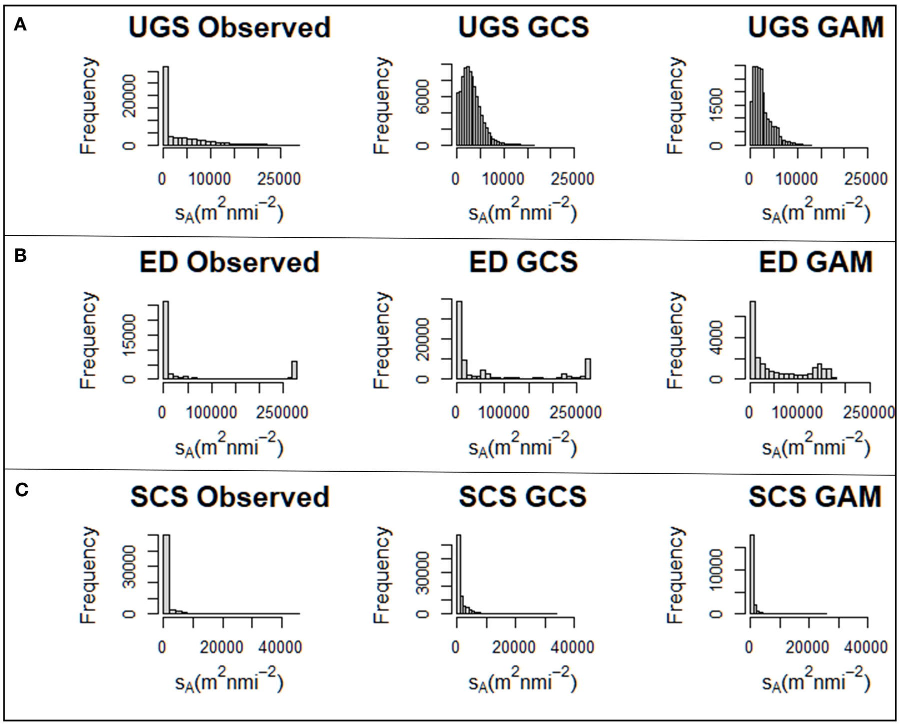



FIGURE 5
Example distributions of $s_A$ observed at MG111 predicted by geostatistical conditional simulations and generalized additive models in the unconditional Gaussian simulation **(A)**, exponential decay **(B)**, and stochastic conditional simulation **(C)** scenarios.

**TABLE 3 Average RMAEs of the mean and CV $s_A$ estimated for each transect method by the geostatistical conditional simulations and generalized additive models.**

| Scenario | Transect | Wreck | RMAE (mean) | | | RMAE (CV) | | |
|---|---|---|---|---|---|---|---|---|
| | | | p-value | GCS | GAM | p-value | GCS | GAM |
| UGS | Parallel | MG111 | <0.01 | 0.30 | 0.08 | <0.01 | 0.57 | 0.52 |
| | Star | MG111 | <0.01 | 0.34 | 0.09 | <0.01 | 0.58 | 0.53 |
| | Parallel | WRT | <0.01 | 0.29 | 0.03 | 0.94 | 0.55 | 0.57 |
| | Star | WRT | 0.01 | 0.25 | 0.10 | 0.06 | 0.50 | 0.56 |
| | Two-Star | WRT | <0.01 | 0.31 | 0.13 | 0.99 | 0.55 | 0.54 |
| ED | Parallel | MG111 | <0.01 | 0.27 | 0.08 | <0.01 | 0.26 | 0.47 |
| | Star | MG111 | <0.01 | 0.43 | 0.11 | 0.12 | 0.48 | 0.56 |
| | Parallel | WRT | <0.01 | 0.25 | 0.07 | 0.92 | 0.29 | 0.41 |
| | Star | WRT | <0.01 | 0.33 | 0.07 | 0.24 | 0.73 | 0.48 |
| | Two-Star | WRT | 0.87 | 0.28 | 0.25 | 0.97 | 0.45 | 0.37 |
| SCS | Parallel | MG111 | 0.35 | 0.56 | 0.40 | <0.01 | 0.50 | 0.33 |
| | Star | MG111 | <0.01 | 0.76 | 0.39 | <0.01 | 0.43 | 0.29 |
| | Parallel | WRT | 0.86 | 0.40 | 0.21 | 0.91 | 0.44 | 0.40 |
| | Star | WRT | 0.99 | 0.43 | 0.41 | 0.07 | 0.55 | 0.45 |
| | Two-Star | WRT | 0.98 | 0.69 | 0.57 | 0.17 | 0.43 | 0.34 |

P-values indicate results of TukeyHSD pairwise comparisons made in the two-way ANOVA at the MG111 and WRT wrecks.

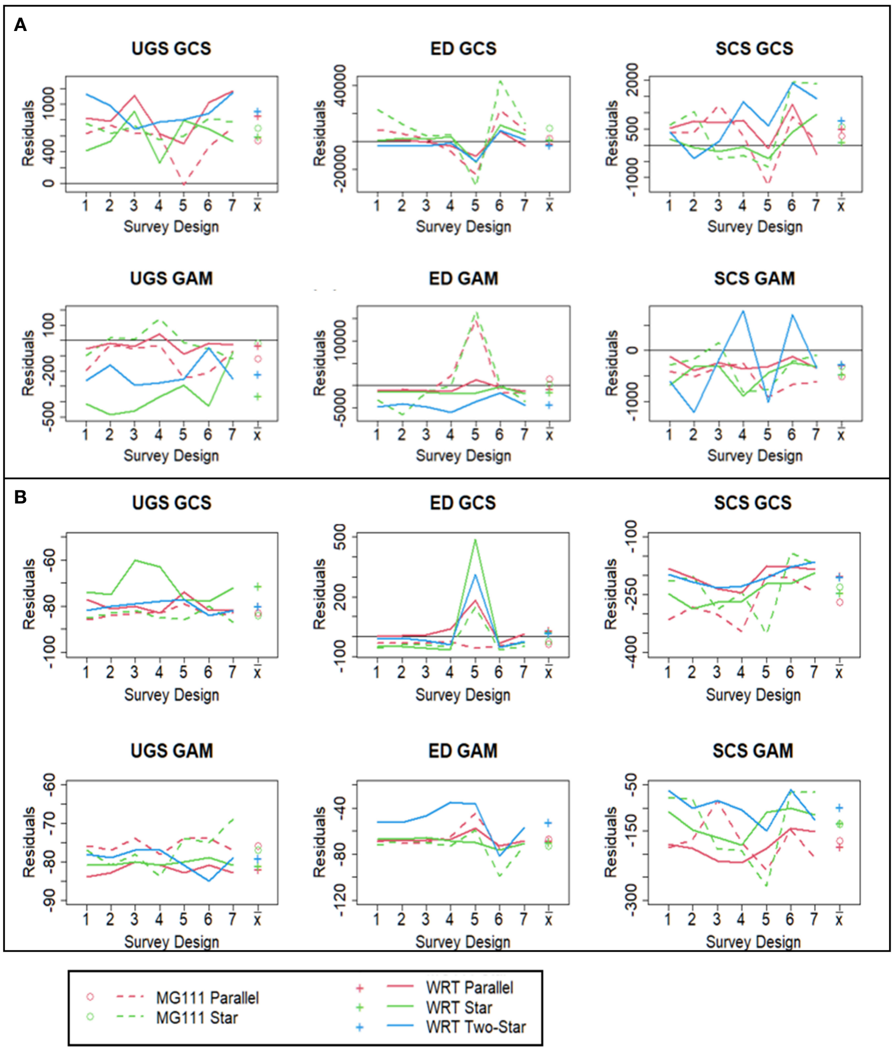



FIGURE 6
Average residuals of the mean **(A)** and CV **(B)** estimated by geostatistical conditional simulations and generalized additive models in each of the three scenarios sampled by the seven survey designs and the mean of each transect method ($\bar{x}$).

**TABLE 4 Significant results of the two-way ANOVAs in the unconditional Gaussian, exponential decay, and stochastic conditional scenarios addressing $H_{01}$: no difference in average RMAE of the mean or CV estimated by geostatistical conditional simulations and generalized additive models among parallel, star, and two-star transect methods.**

| Scenario | Approach | Wreck | Statistic | Transects | p-value | Parallel | Star | Two-Star |
|---|---|---|---|---|---|---|---|---|
| UGS | GCS | WRT | CV | P:S | <0.01 | 0.55 | 0.50 | |
| | GCS | WRT | CV | P3:S3 | 0.02 | 0.56 | 0.42 | |
| | GCS | WRT | CV | P4:S4 | 0.04 | 0.58 | 0.44 | |
| | GCS | WRT | CV | S:SS | <0.01 | | 0.50 | 0.55 |
| | GCS | WRT | CV | S3:SS3 | 0.04 | | 0.42 | 0.55 |
| | GAM | WRT | mean | P:S | <0.01 | 0.03 | 0.10 | |
| | GAM | WRT | mean | P:SS | <0.01 | 0.03 | | 0.13 |
| | GAM | WRT | mean | P6:SS6 | <0.01 | 0.05 | | 0.28 |
| ED* | GCS | MG111 | mean | P:S | <0.01 | 0.27 | 0.43 | |
| | GCS | MG111 | CV | P:S | <0.01 | 0.26 | 0.48 | |

*(Continued)*

TABLE 4 Continued

| Scenario | Approach | Wreck | Statistic | Transects | p-value | Parallel | Star | Two-Star |
|---|---|---|---|---|---|---|---|---|
| | GCS | WRT | mean | P:S | <0.01 | 0.25 | 0.33 | |
| | GCS | WRT | CV | P:S | <0.01 | 0.29 | 0.73 | |
| | GCS | WRT | mean | S:SS | 0.04 | | 0.33 | 0.28 |
| | GCS | WRT | CV | S:SS | 0.04 | | 0.73 | 0.45 |
| | GAM | MG111 | mean | P:S | <0.01 | 0.08 | 0.11 | |
| | GAM | MG111 | CV | P:S | <0.01 | 0.47 | 0.56 | |
| | GAM | WRT | CV | P:S | <0.01 | 0.41 | 0.48 | |
| | GAM | WRT | mean | P:SS | <0.01 | 0.07 | | 0.25 |
| | GAM | WRT | mean | S:SS | <0.01 | | 0.07 | 0.25 |
| | GAM | WRT | CV | S:SS | <0.01 | | 0.48 | 0.37 |
| SCS | GCS | MG111 | mean | P7:S7 | 0.02 | 0.37 | 1.67 | |
| | GCS | WRT | CV | P:S | <0.01 | 0.44 | 0.55 | |
| | GCS | WRT | CV | S:SS | <0.01 | | 0.55 | 0.43 |
| | GCS | WRT | mean | S6:SS6 | 0.03 | | 0.27 | 2.38 |
| | GAM | WRT | mean | P:SS | <0.01 | 0.21 | | 0.57 |
| | GAM | WRT | mean | P6:SS6 | 0.01 | 0.14 | | 1.33 |
| | GAM | WRT | CV | S:SS | 0.03 | | 0.45 | 0.34 |

Pairwise comparisons from the TukeyHSD are made between averages of all designs and between individual survey designs (1-7) under each transect method at the MG111 and WRT wrecks. *Significant comparisons between individual survey designs under each transect method are not shown for the exponential decay scenario.

also true at MG111 in the SCS scenario, with the exception of the non-uniform design (7), in which the parallel method provided a better fit than the star method in the GCS estimation of mean $s_A$. Under the SCS scenario, precision of variability in $s_A$ predicted by both GAM and GCS was lowest for the star design in comparison to both parallel and two-star designs at WRT. Mean $s_A$ from the parallel transect method was less biased than from the two-star method at WRT when using GAMs.

Alterations of survey design were most influential to the model-based estimation of $s_A$ within the ED scenario, as all null hypotheses tested by the two-way ANOVA ($H_{01:5}$) were rejected under this scenario (Tables 4–7). In this scenario, parallel transects were significantly less biased than star transects at MG111 for mean and variability estimated by both modeling designs (Table 4). Star transects also generated the largest error in mean and CV estimated at WRT by the GCS approach. When using the GAM approach, only precision in the variability of $s_A$ was lowest for star transects at WRT. Precision in mean $s_A$ was lowest for two-star transects under this approach.

The null hypothesis of no difference in estimation error of mean $s_A$ among the number of transects that pass over the feature of interest ($H_{02}$) was rejected at the single structure reef under both GCS and GAM (Table 5). Reducing the number of transects which pass over the structure resulted in an increase in precision for both the parallel and star transects at MG111 using the GCS approach. This trend was also significant in the GAM approach, but only for the star transect method. The number of transects passing over structures was less influential in the presence of a field of structures, where the only significant differences in precision were found between one pass and four or three passes for the parallel transects in the GCS approach and two-star transects in the GAM approach, respectively. We found no evidence that variability in $s_A$ is significantly affected by the number of transects passing over the habitat feature of interest, regardless of the number of features present.

In the case where passing over the feature of interest is not possible, the null hypothesis that the transect method does not influence mean or variability of $s_A$ ($H_{03}$) predicted by GCS was rejected (Table 6). Parallel transects outperformed both star and two-star designs, irrespective of the number of structures present at the reef. At WRT, the two-star design generated less bias than the single star method in prediction of the mean. When using the GAM approach, error in the means predicted from the parallel and star designs were equal, and both were significantly less than the error predicted from the two-star design. However, error in the variability predicted from the star design was larger than the two-star design in this case.

Transect length ($H_{04}$) had significant impacts on the mean and variability of $s_A$ predicted by both modeling approaches (Table 7). Under the GCS approach, precision in mean $s_A$ decreased with transect length for all transect methods at both wrecks. Under the GAM approach, precision in the CV of $s_A$ was affected more than the mean, although the mean $s_A$ at WRT had significantly larger bias from shorter transects in two-star surveys. Variability was better estimated by longer transects in all survey methods at the two wrecks when using GAM, except for the parallel method at WRT (p-value=0.09). Transect uniformity ($H_{05}$) was the least impactful survey design alteration on precision of $s_A$, and was only significant in the prediction of mean $s_A$ by GCS (Table 7). The results of this

TABLE 5 Significant results of the two-way ANOVA in the exponential decay scenario addressing $H_{02}$: average RMAE of the mean or CV estimated by geostatistical conditional simulations and generalized additive models is not influenced by the number of transects (1-4) which pass over the feature of interest.

| Wreck | Statistic | Designs | p-value | 4 Passes | 3 Passes | 2 Passes | 1 Pass |
|---|---|---|---|---|---|---|---|
| GCS | | | | | | | |
| MG111 | mean | P1:P3 | <0.01 | 0.22 | | 0.05 | |
| MG111 | mean | P2:P3 | <0.01 | | 0.15 | 0.05 | |
| MG111 | mean | P3:P4 | <0.01 | | | 0.05 | 0.20 |
| MG111 | mean | S1:S3 | <0.01 | 0.53 | | 0.16 | |
| MG111 | mean | S1:S4 | <0.01 | 0.53 | | | 0.16 |
| MG111 | mean | S2:S3 | <0.01 | | 0.26 | 0.16 | |
| MG111 | mean | S2:S4 | <0.01 | | 0.26 | | 0.16 |
| WRT | mean | P1:P4 | 0.02 | 0.09 | | | 0.27 |
| GAM | | | | | | | |
| MG111 | mean | S2:S3 | <0.01 | | 0.14 | 0.02 | |
| MG111 | mean | S2:S4 | <0.01 | | 0.14 | | 0.02 |
| WRT | mean | SS2:SS4 | 0.02 | | 0.25 | | 0.38 |

Pairwise comparisons from the TukeyHSD are made between averages of individual survey designs (1-4) under each transect method at the MG111 and WRT wrecks.

comparison suggest that when one structure is present, non-uniform transects produce less bias than uniform transects. When multiple structures are present, however, GCS results suggest that uniform transects produce less bias.

# 4 Discussion

## 4.1 Unconditional Gaussian scenario

In the UGS scenario, fish aggregations are small and randomly dispersed around the fixed point of interest (the shipwrecks in this study). This scenario may be representative of fish aggregations at expansive and structurally complex habitat features, such as natural reefs or fields of closely-placed artificial structures. These fish aggregations are weakly stationary, and were constructed in this case with an isotropic covariance structure (Figure 7). Of the three scenarios, the geostatistical modeling approach performed best in estimating the mean of fish aggregations simulated with these patchy and randomly dispersed distributions, as this scenario met the assumptions of stationarity and isotropy best (Table 1).

The relative arrangement of transects when surveying such populations is arbitrary, unless large-scale trends are present in the distribution of fish. Three simulations at WRT exhibited a latitudinal

TABLE 6 Significant results of the two-way ANOVA in the exponential decay scenario addressing $H_{03}$: no difference in average RMAE of the mean or CV estimated by geostatistical conditional simulations and generalized additive models among transect methods when no transects pass over the feature of interest.

| Wreck | Statistic | Transects | p-value | Parallel | Star | Two-Star |
|---|---|---|---|---|---|---|
| GCS | | | | | | |
| MG111 | mean | P5:S5 | <0.01 | 0.63 | 0.93 | |
| MG111 | CV | P5:S5 | <0.01 | 0.41 | 0.92 | |
| WRT | mean | P5:S5 | <0.01 | 0.58 | 0.95 | |
| WRT | CV | P5:S5 | <0.01 | 1.15 | 3.00 | |
| WRT | mean | P5:SS5 | <0.01 | 0.58 | | 0.88 |
| GAM | | | | | | |
| WRT | mean | P5:SS5 | 0.02 | 0.08 | | 0.21 |
| WRT | mean | S5:SS5 | 0.02 | | 0.08 | 0.21 |
| WRT | CV | S5:SS5 | <0.01 | | 0.47 | 0.26 |

Pairwise comparisons from the TukeyHSD are made between averages of individual Design 5 under each transect method at the MG111 and WRT wrecks.

TABLE 7 Significant results of the two-way ANOVA in the exponential decay scenario addressing $H_{04}$: transect length has no influence on the average RMAE of the mean or CV estimated by geostatistical conditional simulations and generalized additive models.

| Approach | Wreck | Statistic | Designs | p-value | RMAE | |
|---|---|---|---|---|---|---|
| $H_{04}$ | | | | | Longer | Shorter |
| GCS | MG111 | mean | P1:P6 | <0.01 | 0.22 | 0.42 |
| GCS | MG111 | mean | S1:S6 | <0.01 | 0.53 | 0.63 |
| GCS | WRT | mean | P1:P6 | <0.01 | 0.09 | 0.32 |
| GCS | WRT | mean | S1:S6 | <0.01 | 0.05 | 0.55 |
| GCS | WRT | mean | SS1:SS6 | 0.04 | 0.18 | 0.35 |
| GAM | MG111 | CV | P1:P6 | <0.01 | 0.47 | 0.66 |
| GAM | MG111 | CV | S1:S6 | <0.01 | 0.54 | 0.97 |
| GAM | WRT | CV | S1:S6 | <0.01 | 0.44 | 0.64 |
| GAM | WRT | mean | SS1:SS6 | <0.01 | 0.28 | 0.06 |
| GAM | WRT | CV | SS1:SS6 | <0.01 | 0.35 | 0.68 |
| $H_{05}$ | | | | | Uniform | Non-Uniform |
| GCS | MG111 | mean | S1:S7 | <0.01 | 0.53 | 0.36 |
| GCS | WRT | mean | S1:S7 | <0.01 | 0.05 | 0.27 |

$H_{05}$ is also shown: no significant difference in precision of uniform versus non-uniform transects at the MG111 and WRT wrecks. Pairwise comparisons from the TukeyHSD are made between averages of longer (Design 1) and shorter (Design 6) transect methods for $H_{04}$ and between averages of uniform (Design 1) and non-uniform (Design 7) transect methods for $H_{05}$.

gradient in the concentration of fish towards the edge of the survey area (offset from the three shipwrecks around which designs were centered). The distribution of fish at these simulations was non-stationary and anisotropic, and both properties were intensified for the sampled distribution in the application of star and two-star transects in comparison to parallel transects. The parallel transect method produced the most precise estimates of mean aggregation size even when using the GAM approach. Similar simulations of scallop populations have previously illustrated that geostatistical models produce more bias in comparison to non-geostatistical approaches (including GAMs) in the presence of large-scale spatial trends when sampling with parallel transects (Chang et al., 2017). Our results

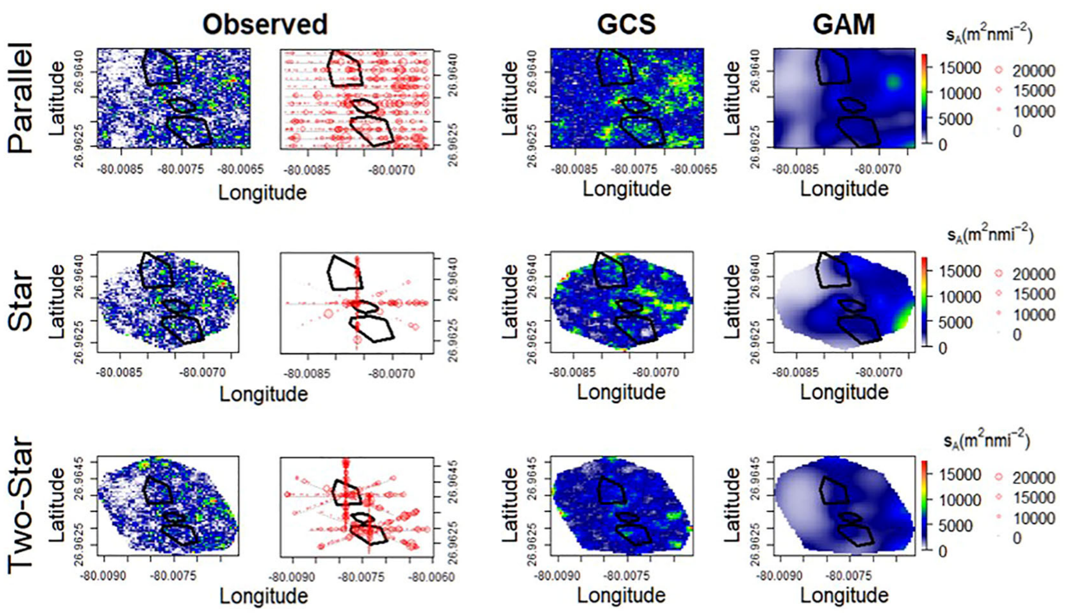



FIGURE 7
Simulated (left) $s_A$ interpolated by geostatistical conditional simulations and generalized additive models (right) in the unconditional Gaussian simulation scenario at WRT. Bubble plots (red circles with radii proportional to $s_A$) show observations sampled by the ideal survey design (Design 1). Black polygons indicate the locations of the Esso Bonaire, Miss Jenny, and Zion Train shipwrecks.

indicate that GAMs are also at least partially influenced by non-stationarity and anisotropy, and that the associated error is minimized in parallel transect designs compared to star transects.

## 4.2 Exponential decay scenario

The influence of disparities in stationarity and isotropy on precision of estimated fish quantities surveyed by different transect methods was best exemplified in the ED scenario (Figure 8). This scenario represents aggregations centered above an established location, such as a seamount or fish aggregating device (FAD). Density is highest in the center of the aggregation and decays exponentially with increasing distance in every direction. Samples of such aggregations only exhibit isotropy if the transects do not extend beyond the densest part of the aggregations (represented by survey Design 6 in the current study). For all transect methods in this case, variance in the sampled data distribution was low and the geostatistical model was more precise in measuring variability than GAMs, although error in both models was higher than when transects extended beyond the aggregation (Table 7).

Unlike the UGS and SCS scenarios, selection of transect starting points has a significant effect on bias in this scenario for aggregations fixed to a single point (Table 5). As long as at least one transect passes over the point of interest, increased sampling of the surrounding area where transects cross the edges of the aggregation can reduce bias in parallel and star designs when using GCS, and in star designs when using GAMs. Residual sampling autocorrelation is reduced in this case, which likely also contributes to improved precision. Doonan et al. (2003) simulated a similar scenario of orange roughy aggregations over seamounts and found that star surveys with transects offset from the fixed point of the seamount, but which still mostly pass over the aggregation, produced less-biased geostatistical estimates of biomass than stars with all transects intersecting at the center. Doray et al. (2008) utilized the star method to sample tuna aggregations around moored FADs and found that abundance estimation variance from universal kriging was reduced with increased number of transects over the FAD. Their results were purely based on the number of transects present in a traditional star survey design, and did not consider transects which do not cross the center of the aggregation, but rather provide higher sampling variance by increasing the number of samples

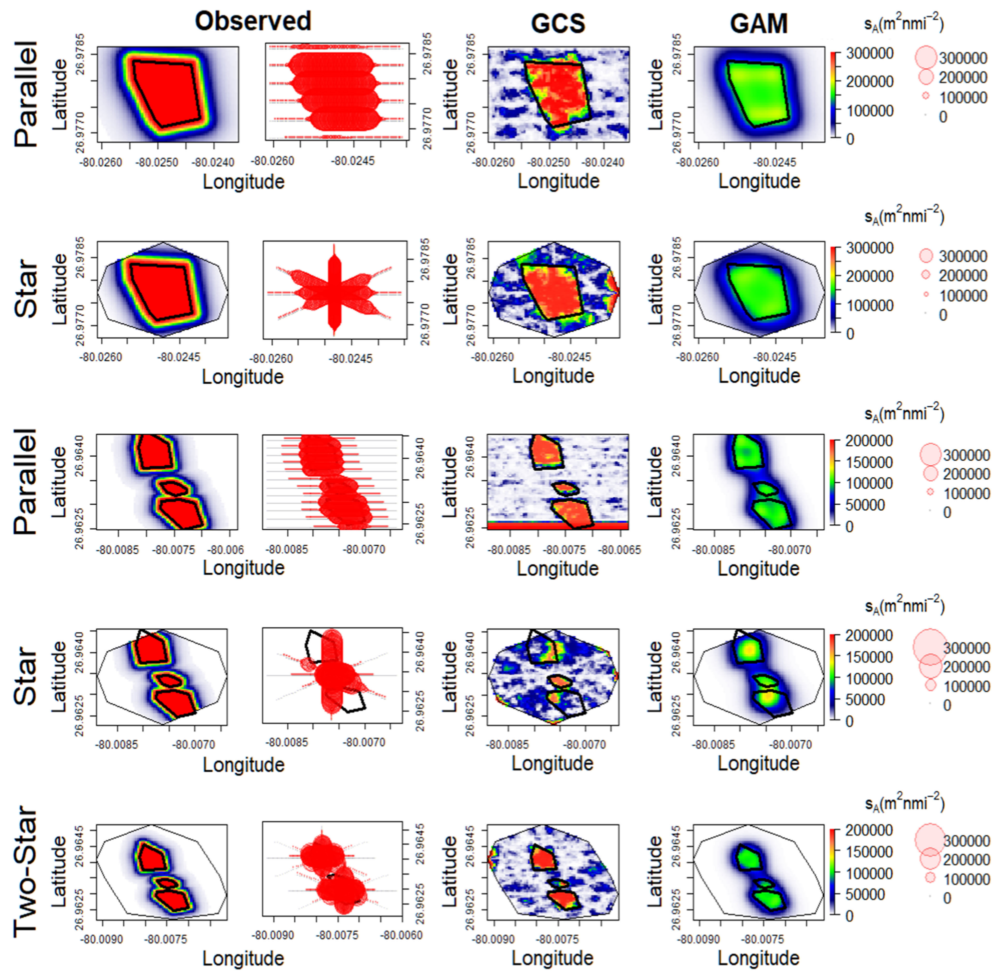


FIGURE 8
Simulated (left) $s_A$ interpolated by geostatistical conditional simulations and generalized additive models (right) in the exponential decay scenario at MG111 (top two rows) and WRT (bottom three rows). Bubble plots show $s_A$ observations sampled by the ideal survey design in each transect method (Design 1). Black polygons indicate the locations of each shipwreck.

around the edges of the aggregation. We propose that a more robust transect method for dense fish populations aggregated around a fixed location could include a random selection of transect starting points with a more even ratio of centered and offset transects driven in various directions. Further investigation is required to determine if such a "dropped sticks" survey design (similar to star Designs 3 or 4 at MG111 in this study) could minimize bias in the quantification of fish aggregations through reduced sampling autocorrelation and increased measure of variance when compared to parallel and star designs.

When considered as one population, fish aggregations fixed to multiple points (such as at WRT) are highly anisotropic under this scenario (Figure 8). There is no clear transect method which significantly increases precision in this case, although parallel designs produced less error in mean estimates (Table 4). In the situation where sampling over fish aggregations fixed to either single or multiple locations is not possible (Design 5 in this study), estimations will have large errors regardless of survey design. A prominent example of this situation in the nGOM are surveys of fishes at oil and gas platforms, where information is reliant on transects within close proximity to the structure but not within the structure frame which supports a large fish community. Our findings suggest that geostatistical estimations of mean and variability will be highly biased in this design, and are therefore not recommended (Table 6). GAM-based errors in the mean were much lower, but equal between parallel and star designs. Only bias from the two-star design at WRT was significantly greater in this case, indicating that there is no benefit from conducting separate surveys around each structure in a complex of multiple habitat features spaced within 100 m of each other.

## 4.3 Stochastic conditional scenario

The SCS scenario was simulated from acoustic data of live goliath grouper spawning aggregations at artificial reefs. This scenario is most representative of fish aggregations associated with established locations, but which exhibit fine-scale stochasticity around habitat features. Unlike the majority of UGS simulations considered here, these fish aggregations display significant spatial association to a structure without necessarily occurring directly over it. This scenario is most applicable to fish aggregations at artificial reefs and some seamounts. Unlike the ED scenario, the number, complexity, and exact location of aggregations is highly variable. For this reason, the occurrence of transects passing over fish aggregations is much less predictable than in either the UGS or ED scenarios (Figure 9).

Star surveys of aggregations fixed to a single point reflect isotropy only if they intersect above the center of the aggregation (as observed by the star method at MG111 in the ED scenario, Figure 8), although this isotropy is highly reliant on the shape of the horizontal footprint of the aggregation. In their simulation of orange roughy aggregations around seamounts, Doonan et al. (2003) found that precision in kriging-based estimates of biomass was lower for star surveys which were not centered over the aggregations, and decreased further with increasing number and complexity of schools. Parallel transects were also considered in this study, although only for simulations where the centers of the aggregation and seamount overlapped (similar to the ED scenario in this study). The authors concluded that star designs outperform parallel designs for such aggregations only when the number of transects is low (<3), and that even when designs are centered over fixed-point aggregations there is no significant difference in the performance of either transect method in the

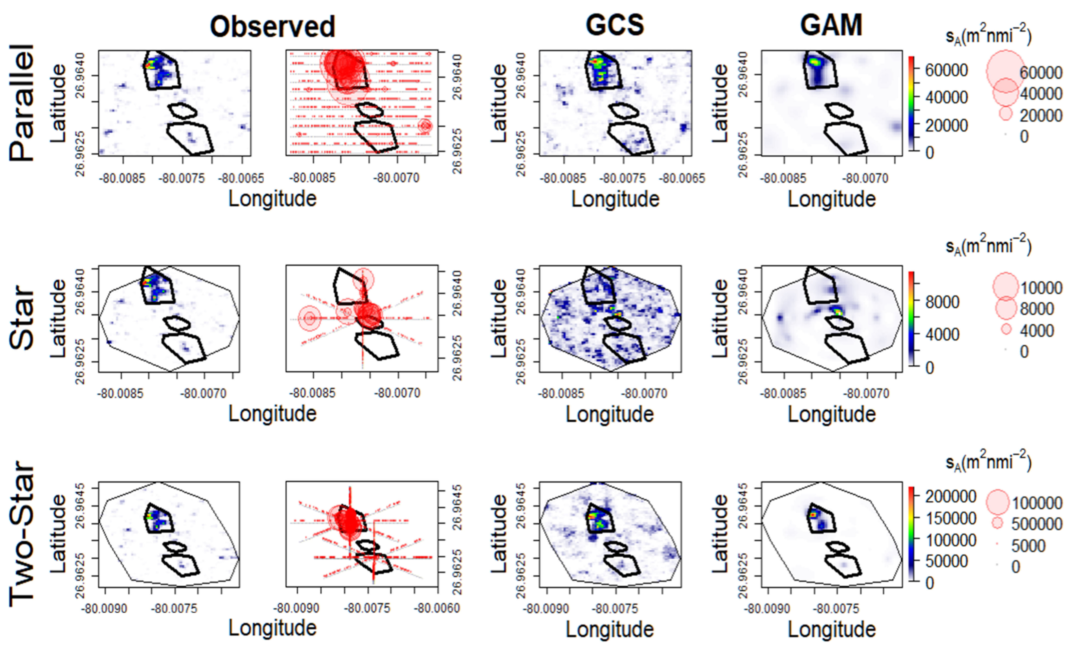



FIGURE 9
Simulated (left) $s_A$ interpolated by geostatistical conditional simulations and generalized additive models (right) in the stochastic conditional simulation scenario at WRT. Bubble plots show $s_A$ observations sampled by the ideal survey design in each transect method (Design 1). Black polygons indicate the locations of the three shipwrecks.

presence of six or more transects. Our parallel transects outnumbered star transects at both the single structure reef (parallel=6, star=4) and the multi-structure reef complex (parallel=12, star=4, two-star=8). Mean estimates from parallel transects were only more precise than mean estimates from two-star transects at WRT and mean estimates from star designs in the non-uniform design (7) at MG111 (Table 4). However, parallel transects were more precise in estimation of variability, and are therefore recommended over star designs when the feature of interest and center of the fish aggregations are not necessarily coincident. The "dropped sticks" transect method described above could provide a compromise between sampling effort and estimation bias in this scenario, as we observed no significant influence on precision in designs where some transects in the star design did not cross the feature of interest (Designs 1-4). It should be noted that the parallel design is ultimately still preferable in this scenario, as stars and dropped sticks are both prone to over-sampling some locations and under-sampling others.

In all scenarios, there was little evidence to suggest that transect uniformity has a significant impact on precision. Transects performed by smaller vessels in non-ideal conditions (such as rough sea states, low boat visibility, impeding boat traffic, etc.) are equally as capable of providing robust quantitative estimates of fish aggregations as perfectly driven transects as long as trends in the underlying fish distribution are not impacted by the complicating conditions (as may be the case with current or moon phase for surveys conducted at night). Surveyors should not be deterred by the inability to drive perfect transects due to equipment limitations or poor weather conditions unless the quality of data collected is impacted (e.g. excessive dropout in split-beam echosounder data due to rough sea state). Variability in estimates was higher in parallel designs than in star designs under all scenarios except for the UGS scenario, where precision of variability in fish aggregations with large-scale trends offset from the point of interest was greater in the star design (but only using the GCS approach).

### 4.4 Conclusions

The evidence provided in this study supports the use of parallel survey designs over stars in most cases. In the few cases where error was equivalent between parallel and star designs, variability was still better estimated by the parallel design. In cross-habitat studies where fish aggregations may be represented by combinations of the three scenarios examined in this study, the star design generates more error in observations of fish aggregations similar to the UGS and ED scenarios, and is not recommended. In studies of fish aggregations at artificial reefs which mimic the patterns shown here by the SCS scenario, where there was no significant difference in error from either design, the lack of significant error reduction by performing one, two, three, or four passes over the habitat feature negates the preconceived benefit of maximizing passes over the feature. In instances where surveys are limited by time or cost of vessel operations, we recommend focusing effort on obtaining better measurements of variability as long as at least one pass is made over the habitat feature. As the residual spatial autocorrelation after modeling approaches have been implemented is still higher in star designs than in parallel designs and variability is better measured in the parallel design, we still recommend parallel designs over stars in this case.

We support the conclusion drawn from previous research of aggregated populations that GAMs are more robust than geostatistical approaches in the presence of both fine- and large-scale spatial trends which often result in non-stationary and anisotropic data (e.g. Yu et al., 2013; Chang et al., 2017). We add that these properties still have a strong influence on the precision of GAMs, but that precision can be improved with transect designs given some basic knowledge of how fish are aggregated around a point of interest. The survey designs and model approaches presented here will inform fisheries-independent sampling of the least-biased transect methods for quantifying fish aggregations when their underlying distribution is well-documented or understood, and that they will maximize the efficiency of adaptive sampling based on initial data when they are not.

## Data availability statement

Most data was simulated. Acoustic dataset has not yet been published. Requests to access these datasets should be directed to BB, bbind002@fiu.edu.

## Author contributions

AW, KB, and PS contributed to the conception and design of the study. KB and BB were awarded funding for data collection, which was conducted by BB, AW, and KB. Data processing was performed by BB. AW completed statistical analysis with guidance from PS. AW wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Funding for the acoustic data collection was provided by NOAA MARFIN awarded to James Locascio and KB (NA15NMF4330152). This material is based upon work supported by the National Science Foundation under Grant No. HRD-1547798 and Grant No.HRD-2111661. These NSF Grants were awarded to Florida International University as part of the Centers of Research Excellence in Science and Technology (CREST) Program.

## Acknowledgments

We greatly appreciate the assistance of Nicholas Tucker, Savannah LaBua, Olivia Odum, and Daniel Correa in conducting fieldwork for this study. This is contribution #1607 from the Coastlines and Oceans Division of the Institute of Environment at Florida International University.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications* (New York, NY: Wiley).

Binder, B. (2022). *The patterns of occurrence, management, and behavioral ecology of fish spawning aggregations in southeast Florida* (Miami, FL: Doctoral dissertation, Florida International University).

Boswell, K. M., Wells, R. J. D., Cowan, J. H., and Wilson, C. A. (2010). Biomass, density and size distributions of fishes associated with a large-scale artificial reef complex in the Gulf of Mexico. *Bull. Mar. Sci.* 86 (4), 879–889. doi: 10.5343/bms.2010.1026

Chang, J. H., Shank, B. V., and Hart, D. R. (2017). A comparison of methods to estimate abundance and biomass from belt transect surveys. *Limnol. Oceanogr.: Methods* 15 (5), 480–494. doi: 10.1002/lom3.10174

Cressie, N. (1993). *Statistics for Spatial Data: Wiley Series in Probability and Statistics* (New York, NY: John Wiley and Sons).

Demer, D. A., Berger, L., Bernasconi, M., Bethke, E., Boswell, K., Chu, D., et al. (2015). Calibration of acoustic instruments. *ICES Cooperative Res. Rep.* 326, 133. doi: 10.25607/OBP-185

Doonan, I. J., Bull, B., and Coombs, R. F. (2003). Star acoustic surveys of localized fish aggregations. *ICES J. Mar. Sci.* 60 (1), 132–146. doi: 10.1006/jmsc.2002.1331

Doray, M., Petitgas, P., and Josse, E.. (2008). A geostatistical method for assessing biomass of tuna aggregations around moored fish aggregating devices with star acoustic surveys. *Can. J. Fish. Aquat. Sci.* 65, 1193–1205. doi: 10.1139/F08-050s

Dunn, P., and Smyth, G. (2005). Series evaluation of Tweedie exponential dispersion models. *Stat Comput.* 15, 267–280. doi: 10.1007/s11222-005-4070-y

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Can. J. For. Res.* 28, 1429–1447. doi: 10.1139/x98-166

Gunderson, D. R. (1993). *Surveys of Fisheries Resources* (New York, NY: John Wiley and Sons, Inc.), 248.

Hastie, T. J., and TibshIrani, R. J. (1990). *Generalized Additive Models* (London: Chapman and Hall).

Hieu, N. T., Brochier, T., Tri, N. H., Auger, P., and Brehmer, P. (2014). Effect of small versus large clusters of fish school on the yield of a purse-seine small pelagic fishery including a marine protected area. *Acta Biotheoretica* 62, 339–353. doi: 10.1007/s10441-014-9220-1

Isaaks, E. H., and Srivastava, R. M. (1989) (Oxford), 561.

Jolly, G. M., and Hampton, I. (1990). A stratified random transect design for acoustic surveys of fish stocks. *Can. J. Fish. Aquat. Sci.* 47, 1282–1291. doi: 10.1139/f90-147

Kang, M., Nakamura, T., and Hamano, A. (2011). A methodology for acoustic and geospatial analysis of diverse artificial-reef datasets. *ICES J. Mar. Sci.* 68 (10), 2210–2221. doi: 10.1093/icesjms/fsr141

Koenig, C. C., Bueno, L. S., Coleman, F. C., Cusick, J. A., Ellis, R. D., Kingon, K., et al. (2017). Diel, lunar, and seasonal spawning patterns of the Atlantic goliath grouper, Epinephelus itajara, off Florida, United States. *Bull. Mar. Sci.* 93 (2), 391–406. doi: 10.5343/bms.2016.1013

MacLennan, D. N., Fernandes, P. G., and Dalen, J. (2002). A consistent approach to definitions and symbols in fisheries acoustics. *J. Mar. Sci.* 59, 365–369. doi: 10.1006/jmsc.2001.1158

MINES ParisTech/ARMINES. (2022). *RGeostats: The Geostatistical R Package*. Available at: http://cg.ensmp.fr/rgeostats.

Neyman, J. (1934). On the two different aspects of the representation method - the method of stratified sampling and the method of selection. *J. R. Statist. Soc* 97, 558–606. doi: 10.2307/2342192

R Core Team. (2022). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https//www.R-project.org/.

Rivoirard, J., Simmonds, J., Foote, K. G., Fernandes, P., and Bez, N. (2000). *Geostatistics for Estimating Fish Abundance* (Oxford, UK: Blackwell Science, Oxford), 206.

Scheaffer, R. L., Mendenhall, W. III, Ott, R. L., and Gerow, K. G. (2012). *Elementary Survey Sampling. 7th edn* (Boston, MA: Cengage Learning).

Taylor, J. C., Eggleston, D. B., and Rand, P. S. (2006). "Nassau grouper (*Epinephelus striatus*) spawning aggregations: hydroacoustic surveys and geostatistical analysis," in *In Emerging Technologies for Reef Fisheries Research and Management*. Ed. J. C. Taylor (Tortola, British Virgin Islands: NOAA Professional Paper), 18–25.

White, A. L., Patterson, W. F. III, and Boswell, K. M. (2022). Distribution of acoustic backscatter associated with natural and artificial reefs in the Northeastern Gulf of Mexico. *Fisheries Res.* 248, 106199. doi: 10.1016/j.fishres.2021.106199

Woillez, M., Rivoirard, J., and Fernandes, P. G. (2009). Evaluating the uncertainty of abundance estimates from acoustic surveys using geostatistical simulations. *ICES J. Mar. Sci.* 66, 1377–1383. doi: 10.1093/icesjms/fsp137

Woillez, M., Walline, P. D., Ianelli, J. N., Dorn, M. W., Wilson, C. D., and Punt, A. E. (2016). Evaluating total uncertainty for biomass-and abundance-at-age estimates from eastern Bering Sea walleye pollock acoustic-trawl surveys. *ICES J. Mar. Sci.* 73 (9), 2208–2226. doi: 10.1093/icesjms/fsw054

Wood, S. N. (2003). Thin-plate regression splines. *J. R. Stat. Soc. B* 65 (1), 95–114. doi: 10.1111/1467-9868.00374

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R. 2nd ed.* (Boca Raton, FL: Chapman and Hall/CRC).

Yu, H., Jiao, Y., and Carstensen, L. W. (2013). Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. *Fisheries Res.* 147, 18. doi: 10.1016/j.fishres.2013.06.002

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 September 2023
DOI 10.3389/fmars.2023.1214798

Check for updates

OPEN ACCESS

EDITED BY
Ben Scoulding,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Joshua Lawrence,
Heriot-Watt University, United Kingdom
Elizabeth A Babcock,
University of Miami, United States

*CORRESPONDENCE
Derek G. Bolser
derek.bolser@noaa.gov

†PRESENT ADDRESS
Derek G. Bolser,
Assessment and Monitoring Division, Office of Science and Technology, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, Silver Spring, MD, United States

‡These authors share senior authorship

RECEIVED 30 April 2023
ACCEPTED 21 August 2023
PUBLISHED 11 September 2023

CITATION
Bolser DG, Berger AM, Chu D, de Blois S, Pohl J, Thomas RE, Wallace J, Hastie J, Clemons J and Ciannelli L (2023) Using age compositions derived from spatio-temporal models and acoustic data collected by uncrewed surface vessels to estimate Pacific hake (*Merluccius productus*) biomass-at-age.
*Front. Mar. Sci.* 10:1214798.
doi: 10.3389/fmars.2023.1214798



# Using age compositions derived from spatio-temporal models and acoustic data collected by uncrewed surface vessels to estimate Pacific hake (*Merluccius productus*) biomass-at-age

Derek G. Bolser [1*†], Aaron M. Berger [2], Dezhang Chu [3], Steve de Blois [3], John Pohl [3], Rebecca E. Thomas [3], John Wallace [3], Jim Hastie [3‡], Julia Clemons [2‡] and Lorenzo Ciannelli [1,4‡]

[1]Cooperative Institute for Marine Ecosystem and Resources Studies, Oregon State University, Newport, OR, United States, [2]Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Newport, OR, United States, [3]Fisheries Research Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, United States, [4]College of Earth, Ocean, and Atmospheric Studies, Oregon State University, Corvallis, OR, United States

Generating biomass-at-age indices for fisheries stock assessments with acoustic data collected by uncrewed surface vessels (USVs) has been hampered by the need to resolve acoustic backscatter with contemporaneous biological (e.g., age) composition data. To address this limitation, Pacific hake (*Merluccius productus*; "hake") acoustic data were gathered from a USV survey (in 2019) and acoustic-trawl survey (ATS; 2019 and eight previous years), and biological data were gathered from fishery-dependent and non-target (i.e., not specifically targeting hake) fishery-independent sources (2019 and eight previous years). To overcome the lack of contemporaneous biological sampling in the USV survey, age class compositions were estimated from a generalized linear mixed spatio-temporal model (STM) fit to the fishery-dependent and non-target fishery-independent data. The validity of the STM age composition estimation procedure was assessed by comparing estimates to age compositions from the ATS in each year. Hake biomass-at-age was estimated from all combinations of acoustics (USV or ATS in 2019, ATS only in other years) and age composition information (STM or ATS in all years). Across the survey area, proportional age class compositions derived from the best STM differed from ATS observations by 0.09 on average in 2019 (median relative error (MRE): 19.45%) and 0.14 across all years (MRE: 79.03%). In data-rich areas (i.e., areas with regular fishery operations), proportional age class compositions from the STM differed from ATS observations by 0.03 on average in 2019 (MRE: 11.46%) and 0.09 across years (MRE: 54.96%). On average, total biomass estimates derived using STM age compositions differed from ATS age composition-based estimates by approximately 7% across the study period (~ 3% in 2019) given the same source of acoustic data. When biomass estimates from different sources of acoustic data (USV or ATS) were compared given the same source of age composition data, differences were nearly ten-fold greater (22% or 27%, depending on if ATS or STM age compositions were used). STMs fit to non-contemporaneous data may provide suitable information for assigning population structure to acoustic backscatter in data-rich areas, but advancements in acoustic data processing (e.g., automated echo classification) may be needed to generate viable USV-based estimates of biomass-at-age.

KEYWORDS

**USV, acoustic-trawl survey, spatio-temporal model, VAST, Pacific hake, biomass estimation, fishery-independent survey, age composition**

# 1 Introduction

Autonomous vessels have shown great promise for enhancing ocean observation programs. Uncrewed surface vessels (USVs) are particularly useful for missions of long duration in harsh environments (Liu et al., 2016; Mordy et al., 2017; Meinig et al., 2019), and are well suited to carry out survey operations in circumstances that would limit or prevent the operation of crewed surveys (e.g., De Robertis et al., 2021). Therefore, USVs have been used to understand physical oceanography (Wills et al., 2021; Nickford et al., 2022; Zhang et al., 2022), animal distribution and behavior (De Robertis et al., 2019b; Verfuss et al., 2019; Levine et al., 2021), and collect data in service of fishery resource survey programs (Chu et al., 2019; De Robertis et al., 2021; Sepp et al., 2022). Incorporating USVs into fishery-independent survey programs is of particular interest given their potential to increase the efficiency of ship-based survey effort and mitigate the effects of unexpected circumstances (e.g., funding shortfalls, vessel unavailability).

Fishery-independent survey programs that generate indices for stock assessment typically rely on two types of data: (1) abundance or biomass and (2) composition (e.g., age, length, species composition). One gear can provide both types of data in some situations (e.g., trawls, stereo-video), but many survey programs employ multiple gears to collect sufficient data of each type. One of the most common combinations of gears is an acoustic-trawl survey (Simmonds and MacLennan, 2005), where trawls provide the much more broadly sampled acoustic backscatter with point samples of composition data that are extrapolated to estimate the biomass-at-age of target species. USVs can collect acoustic data in support of fishery resource survey programs but are ill-equipped to collect composition data, which has thus far limited their operational use. The study of De Robertis et al. (2021), which used an empirically derived relationship between acoustic backscatter and biomass to estimate the total biomass of Walleye Pollock (*Gadus chalcogrammus*) in Alaska, was a significant advancement towards providing data for stock assessments with USV surveys. However, as most modern stock assessments are age- or length-structured, providing a biomass index with a USV survey that is equivalent to one generated with a survey vessel requires biomass estimates to be resolved by age or length.

Spatio-temporal models (STMs) are statistical models that make highly resolved predictions based on spatial, temporal, and spatio-temporal effects. Accordingly, they may be suitable for estimating composition data for acoustic surveys when contemporaneous biological sampling is not possible. One type of STM, the vector autoregressive spatio-temporal (VAST) model, is particularly well-suited for integrating data from multiple sources, multi-variate applications, and generating robust indices in a variety of situations (Grüss and Thorson, 2019; Thorson, 2019; Brodie et al., 2020). Therefore, VAST models are commonly used to generate indices of biomass or abundance distribution (Thorson and Barnett, 2017; Godefroid et al., 2019; Thorson et al., 2021) and estimate diet and age-length-sex composition (Thorson and Haltuch, 2019; Grüss et al., 2020; O'Leary et al., 2020). If estimates of composition data derived from STMs such as VAST can be applied to acoustic data collected by USVs, the utility of USVs would be greatly expanded. Adding USV data collection could facilitate a higher level of spatial and temporal resolution in fishery resource surveys than adding additional crewed data collection given the same operational constraints (e.g., funding, person hours). The additional high-resolution data would support the next generation of spatially-resolved stock assessment and management strategies (Berger et al., 2017).

We aimed to determine if a combination of acoustic data from USVs and age compositions derived from a STM could generate viable estimates of Pacific hake (*Merluccius productus;* "hake") biomass-at-age. Hake are the most abundant groundfish in the California Current Large Marine Ecosystem and support one of the largest fisheries on the U.S. West Coast south of Alaska (Hamel et al., 2015; NOAA, 2015; Johnson et al., 2021). Hake undertake a seasonal northward migration in the spring, where the extent of the migration is determined in part by age, size, and oceanographic conditions (Dorn, 1995; Hamel et al., 2015; Malick et al., 2020). The Joint U.S.-Canada Integrated Ecosystem and Pacific Hake Acoustic Trawl Survey (ATS) is conducted biennially in the summer months to estimate the biomass-at-age of the entire stock, which is managed and surveyed jointly by the U.S. and Canada. In 2019, a USV (Saildrone) acoustic survey was conducted in conjunction with the hake survey (de Blois, 2020).

The specific objectives of the present study were the following: (1) estimate hake age class composition with STMs fit to a combination of fishery-dependent and non-target fishery-independent age data, (2) estimate hake biomass-at-age from all combinations of acoustic (USV or ATS) and age composition information (STM or ATS), (3) compare the age compositions estimated in objective 1 with those of the ATS to evaluate the validity of age compositions estimated with STMs, and (4) compare total biomass estimates derived in objective 2 to determine the relative effects of differing sources of age composition and acoustic data on total biomass estimates. We completed these objectives with data from 2019, when the USV and ATS sampled the survey area in tandem, and for eight previous survey years (ATS vessel acoustics only) to examine if the performance of generating age compositions using STMs was reasonable and stationary under different conditions. If USV surveys are to be useful for age-resolved indices of abundance for stock assessment, there needs to be an understanding of the benefits, limitations, and efficiencies available to best leverage this technology and inform management decisions.

# 2 Methods

## 2.1 Study domain

This study took place in waters off the U.S. West Coast from 34.4 to 55.5° latitude and -135.5 to -120.6° longitude (Figure 1). Some datasets employed in this study covered this entire area, while others covered smaller portions of the survey domain. (Figure 1; described in greater detail below). For spatio-temporal modelling, we modified the spatial grid used to generate kriged biomass estimates from data collected in the ATS by Northwest Fisheries Science Center (National Marine Fisheries Service – National Oceanic and Atmospheric Administration; NWFSC) personnel (Chu et al., 2017). This grid was clipped to match the latitudinal extents of the U.S. West Coast and reached the 1500 m isobath or 35 nmi offshore, whichever was furthest offshore (Chu et al., 2017; de Blois, 2020). Grid cells were generally 21.4 km$^2$ except in areas where they conformed to the coastline or shelf contour. In some analyses, we made comparisons between predictions in geographic strata defined by the International Pacific Fisheries Council (INPFC), which are shown in Figure 1.

While 2019 was the focal year of this study as it was the year in which the USV survey was conducted in tandem with the ATS, we also analyzed data from previous years in which the ATS was conducted using modern protocols (i.e., protocols that closely resemble current protocols; 2003, 2007, 2009, 2011, 2012, 2013, 2015, and 2017). We did this to examine the performance of our method for generating estimates of hake age composition with a STM fit to non-contemporaneous biological sampling data under different stock population structures (age composition, size/status, and distribution).

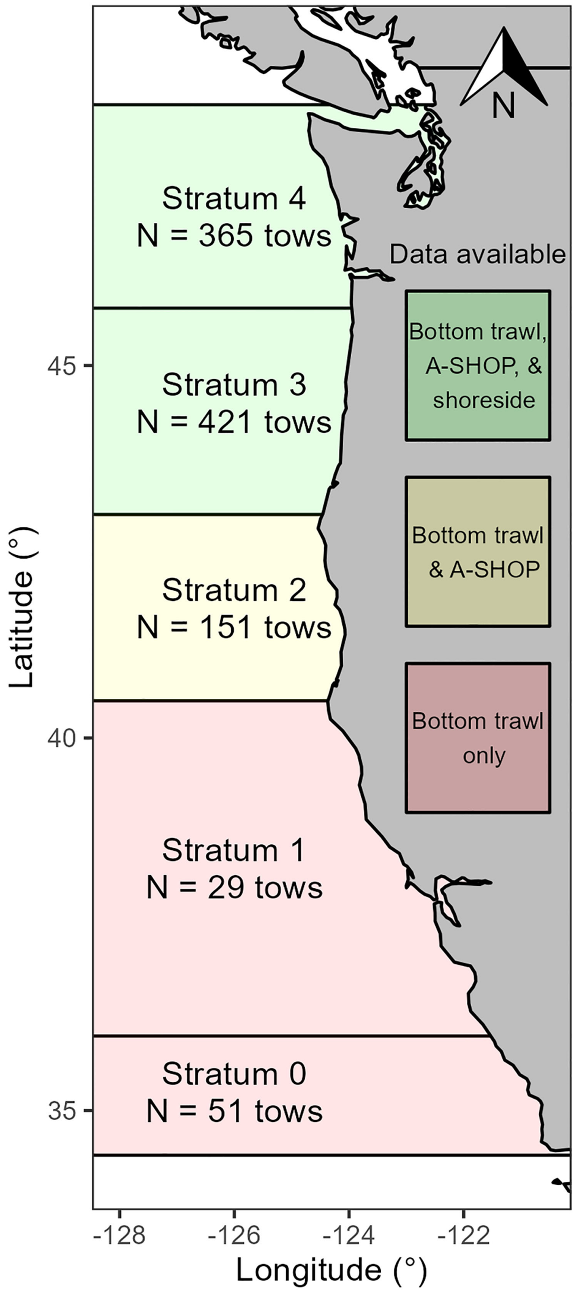



FIGURE 1
Study area with International Pacific Fisheries Commission Geographic Strata (0-4), type of data generally available in each stratum, and number of tows by strata for 2019. A-SHOP refers to at-sea hake observer program data, shoreside refers to the hake fishery in which catches are processed on shore, and bottom trawl refers to the NOAA Northwest Fisheries Science Center fishery-independent bottom trawl survey. We note that while the figure depicts the general lack of shoreside data in stratum 2, 4 shoreside tows were conducted in stratum 2 across years and included in this study.

## 2.2 Acoustic trawl survey and data processing

The U.S. portion of the 2019 ATS began near Point Conception, California and proceeded north along the west coast of the United

States to the Canadian border. Acoustic transects were oriented east-west and ranged from the 50 m isobath to either the 1,500 m isobath or a location 35 nmi west of the inshore waypoint. Transects were spaced 10 nmi apart. Transects were traversed sequentially and were surveyed acoustically between sunrise and sunset when hake form identifiable aggregations. A Simrad scientific echosounder system collected raw acoustic data from up to five split-beam transducers operating at 18, 38, 70, 120, and 200 kHz (pulse duration: 1 ms; ping rate: 1-4 per second depending on bottom depth). Echosounder calibrations were performed pre- and post-survey according to standard procedures (Demer et al., 2015). Data from the 38 kHz echosounder (the primary frequency used for generating biomass estimates) were post-processed for hake based on scattering properties and behavioral cues (e.g., school morphology; see Supplementary Material S1 and Chu et al. (2017) for further details on echogram scrutiny). Daytime trawling was used to classify observed backscatter layers to species and size composition and to collect specimens of hake and other organisms (see Supplementary Material S1 and Chu et al. (2017) for further details on calculation of hake backscatter in mixed aggregations). The number and locations of trawls were not pre-determined, but instead depended on the occurrence and pattern of backscattering layers observed at the time of the survey. Highest priority for trawling was given to sampling distinct layers of intense backscatter that are indicative of high densities of hake. While southern and offshore extents, transect spacing, and acoustic frequencies employed (namely, the inclusion of 18 kHz) varied moderately over the study period (2003-2019), the purpose of the ATS (generating a biomass index for the hake stock) and core methodology remained constant. Notably, the ATS biomass estimates did not include INPFC stratum 0 in 2003-2011, and no biological data were collected in stratum 0 in 2003-2007.

## 2.3 Saildrone (USV) survey and data processing

The U.S. portion of the 2019 ATS was conducted on the fishery survey vessel (FSV) *Bell M. Shimada* and ran from June 15 to August 21, with concurrent survey operations by four to five USVs (Saildrones from Saildrone, Inc.) operating from June 17 to August 25. The USVs were equipped with the company's standard package of oceanographic and atmospheric sensors, including Simrad scientific echosounders mounted in the USV's keelpod. The echosounder system consisted of an EK80 Wide Band Transceiver Mini with a 38 kHz split beam transducer and a 200 kHz single beam transducer (ES38-18/200-18C; pulse length: 1.024 ms; ping rate: 1 per two seconds). Echosounder calibrations were performed pre- and post-mission (June 10 and September 27, respectively) off the Saildrone, Inc. dock in Alameda, CA according to standard procedures (Demer et al., 2015).

At-sea operations of the USVs were governed in two ways. Pre-mission planning of sampling protocols established mission parameters at the outset (for example, only surveying acoustically from sunrise to sunset). In-field decisions based on navigation and weather data delivered via a graphical user interface (the mission portal), were operationalized through near real-time satellite-driven command and control of the vessel's navigation. USV operators – all pilots and engineers employed by Saildrone, Inc. – piloted the vessels in collaboration with NWFSC scientists aboard the research vessel monitoring the USVs through a web-based mission portal.

The USVs surveyed the same parallel transects as the ATS in the same direction along the U.S. west coast, from south to north. Each transect was either surveyed entirely by one USV, or by a pair of USVs operating in conjunction to cover either the inshore or offshore halves. This practice of using multiple USVs for a single transect helped mitigate the difference in operational speed between the USVs and the research vessel. In the 2019 mission, the USVs matched the FSV in completing 85 transects over their 70 operational days. USV survey speeds varied across the five vehicles, depending on weather, currents, and operational idiosyncrasies inherent to each vehicle. In general, the speed-over-ground for the USVs ranged from 0 knots when becalmed, to nearly 5 knots when transecting in 20 plus knots of wind from favorable directions. In the three instances where the ATS protocol extended a transect to follow suspected hake backscatter (hake designation methodology briefly explained below, and in further detail in Supplementary Material S1 and Chu et al. (2017)), the USVs likewise extended. Operationally, 68% of the USV transects were within +/- 3 days of the ATS transects, while 84% were within +/- 5 days.

The post-cruise judging team for the USV transects used Echoview 11 (a commercial software developed by Echoview Software Pty.; https://echoview.com). The team was organized to minimize initial recollection bias from any FSV echograms they might have viewed in the previous cruise legs. This was done by tasking two analysts to judge those USV transects where they themselves had not also been on board the ATS when the ship sampled those particular lines. Furthermore, these analysts did not review the archived trawl or processed acoustic data from these ATS transects prior to reviewing the ATS collections in their charge. Lastly, a procedural, final review of all echograms by the survey's chief scientist, though conducted for the ATS echograms, was withheld from the USV echograms.

Insofar as possible, given the lack of validating trawl data and the reduced number of acoustic frequencies available for the USVs versus the ATS (38 and 200 kHz versus the 18, 38, 70, 120, and 200 kHz), acoustic data review was done in a similar fashion to that of the ATS judging. The "Impulse Removal" and "Background Noise Removal" processing modules within Echoview software were used for the 38 and 200 kHz echograms, to better account for signal attenuation from inclement weather. The 38 and 200 kHz echograms were scrutinized simultaneously, with identifying regions drawn around suggestive backscatter in the 38 kHz echogram. Regions drawn in one echogram automatically appear in all others, allowing comparisons. In drawing regions, analysts relied on morphometric cues (shape, size of the aggregations), behavioral cues (depth, relative positioning to other schools, proximity to the 200-meter shelf break), and frequency response (when possible), to make their determinations.

Regions were drawn tightly around areas judged as likely to be "hake" (i.e., 100% hake), "CPS" (i.e., coastal pelagic species), "zooplankton," or "unclassified." Unlike with ATS judging, no species mixes or biological information was linked for these regions. In most instances determined to be hake, judges relied on the evidence offered by the 38 kHz alone, as hake typically occurs at depths greater than range of good data from the 200 kHz. Lastly, the two USV reviewers met after all assigned transects were completed to hold their own procedural review. Each EV file was jointly scrutinized so as to make a shared, consistent decision about the shape and assignation of each backscatter region within. Supplementary Material S1 and Chu et al. (2017) provide further detail on echogram scrutiny.

## 2.4 Biological data for spatio-temporal models

STMs were fit to data from three sources: the NWFSC's Pacific Coast Groundfish Bottom Trawl Survey (hereafter 'bottom trawl'), the At-Sea Hake Observer Program ('A-SHOP'), and observer data from each state's shoreside hake fishery (i.e., fishery in which catches are processed on shore; 'shoreside'). We describe these data briefly below.

Bottom trawl data were available across the study area from May-October in each year of the study, and were collected at random sites across the U.S. West Coast at depths from 55-1,280 m from chartered fishing vessels (Keller et al., 2017). There were 164 tows in 2019 and 2,140 tows across all years that were retained for analysis. Sites were selected via a stratified random grid-based design in which percentages of sampling effort were allocated to each INPFC geographic stratum (Keller et al., 2017). Tows were conducted on trawlable habitat within the selected grid cell for 15 minutes (plus liftoff lag, Wallace and West, 2006) and catch weights were recorded for each species caught. A subsample of hake in each trawl was weighed, measured, and later aged. We allocated total hake weight by age class for each haul based on the age class proportions recorded in the haul's subsample and overall median weight-at-age-class from all hauls across the survey area.

A-SHOP data were available in INPFC strata 2-4 (Figure 1) in each year in May, June, September, October, and November, although some years had data from July and August as well (n = 284 tows in July and August across years). There were 775 tows in 2019 and 5,323 tows across all years that were retained for analysis. These data were recorded by National Marine Fisheries Service (NMFS) observers aboard at-sea processing vessels (catcher-processors and motherships) that generally fish offshore of Oregon and Washington but occasionally set nets in Northern California (catcher motherships only). Observers recorded the haul weight for each species observed and a subsample of hake in each trawl was weighed, measured, and then later aged by shore-based personnel (NWFSC, 2022). In the same manner as the bottom trawl data, we allocated total hake weight by age class for each haul based on the age class proportions recorded in the haul's subsample and overall median weight-at-age-class from all hauls across the survey area.

Shoreside data were generally available in INPFC strata 3-4 (Figure 1) in each year of the study from May-October, and were recorded by NMFS observers aboard fishing vessels that deliver their catch to shore-based processing plants. Only a very small number of tows (n = 4 across years) were conducted south of stratum 3 (i.e., in stratum 2). There were 69 shoreside tows in 2019 and 476 tows across all years that were retained for analysis. Shoreside vessels generally fish offshore of Oregon and Washington but often at closer distances to shore than at-sea vessels (Saelens and Jesse, 2007). Observers recorded the haul weight of each species observed and a subsample of hake from each trip was weighed, measured, and later aged. Because hake were subsampled onshore after trips were completed, and more than one haul was conducted on some trips, it was necessary to assign age class compositions to the centroid of the broader area that was fished on a given trip by a given vessel. After this was done, we allocated total hake weight by age class for each haul based on the age class proportions recorded in the haul's subsample and overall median weight-at-age-class from all hauls across the survey area in the same manner as the other data sources.

## 2.5 STM specifications and age class proportions

Spatially and temporally resolved estimates of hake biomass-at-age class, which were subsequently converted to proportions of biomass at age class, were derived from STMs built with the *VAST* R package (ver. 3.7.1; Thorson and Barnett, 2017) in the R Studio environment (R ver. 4.04). We provide an overview of our VAST-based STMs here but refer the reader to Thorson (2019) for additional details about VAST models in general. Our STMs were configured to make predictions over three groups of months: May-June, July-August, and September-November. These delineations were made based on data availability and to capture the seasonal northward migration of hake (Hamel et al., 2015). For 2019, we employed a 3x3 factorial design with three age class configurations (three, four, and five age classes; Table 1) and three configurations for spatial and spatio-temporal terms (independent, identically distributed factors (IID), single-factor, multi-factor; Table 2). For the other study years, we employed a 3x1 factorial design and tested the three model term configurations with the age class configuration that performed best in 2019. Age class configurations were chosen based on hake life history (e.g., differences in weight-at-age, scale of migration between ages), available data, and the distribution of samples in 2019. Candidate age class configurations differed in the resolution of older adult (age 4 and older) ages (Table 1), which was limited given that data availability declined with age. Model term configurations were chosen to represent situations in which the spatio-temporal biomass distribution of age classes (1) varied

TABLE 1 Candidate STM age class configurations.

| Name | Age classes |
|---|---|
| 5 age class | 2, 3, 4-6, 7-9, 10+ |
| 4 age class | 2, 3, 4-6, 7+ |
| 3 age class | 2, 3, 4+ |

TABLE 2 Candidate STM term configurations.

| Name | STM Term Configuration |
|---|---|
| IID | Spatial and spatio-temporal variation specified as independent, identically distributed random effects among age classes |
| Single-factor | One random effect each for spatial and spatio-temporal variation across all age classes. |
| Multi-factor | 1< n< g random effects for spatial and spatio-temporal variation across age classes, where n is the number of spatial or spatio-temporal terms and g is the number of age classes in the model. The magnitude of the effect of spatial and spatio-temporal terms differs between age classes. |

independently from one another (IID configuration), (2) was affected by the same spatial and spatio-temporal variables (single-factor configuration), or (3) was affected by multiple spatial and spatio-temporal variables at different levels of impact between age classes (multi-factor configuration) (Table 2).

Specifically, STMs fit in the VAST R package were developed using a delta generalized mixed model framework. We describe the models by starting with the basic formulation of temporal, spatial, and spatio-temporal effects and building to the incorporation of a spatially-varying catchability term and the alternative incorporation of effects via model term configurations. Linear predictors for encounter/non-encounter ($p_1$) and biomass conditional on encounter ($p_2$) for age class $c_i$, month group $t_i$, and spatial knot $s_i$ were each defined as follows:

Eq. 1,

$$p(i) = \beta(c_i, t_i) + \omega(s_i, c_i) + \epsilon(s_i, c_i, t_i)$$

where $\beta$ is a fixed effect for temporal variation (month group). Spatial ($\omega$) and spatio-temporal ($\epsilon$) variation were modelled as random effects with a first-order autoregressive correlation structure between month groups. Since we fit the model to a combination of fishery-dependent and fishery-independent data, it was necessary to add a catchability ratio to the model to account for differences in fishing power between datasets (Thorson et al., 2012; Grüss et al., 2023). We specified the catchability ratio as a spatially-varying effect because spatial domain and selectivity at age can vary between fishery-independent and fishery-dependent data (Grüss et al., 2023; Thorson et al., 2023). Additionally, hake spatial distribution is known to vary with age (Hamel et al., 2015). With the addition of a spatially-varying catchability ratio in each linear predictor, the formulation becomes:

Eq. 2,

$$p(i) = \beta(c_i, t_i) + \omega(s_i, c_i) + \epsilon(s_i, c_i, t_i) + \sum_{m=1}^{n_m} \xi(s_i, m) M(i, m)$$

where $\xi(s_i, m)$ is is the additive, spatially-varying impact of data source $m$ at location $s_i$. This impact is set to 0 for the fishery-dependent data and is estimated for the fishery-independent data as a random effect following a multivariate normal (MVN) distribution:

Eq. 3 $\zeta(i) \sim \mathrm{MVN}(\lambda \mathbf{1}, \sigma_\lambda^2(m)\mathbf{R}(k))$

where $\lambda 1$ is a matrix of fixed effects (average log-ratio between fishery-dependent and fishery-independent data), $\sigma_\lambda^2$ is the estimated pointwise variance of the spatially varying response to fishery-dependent data $m$, and $\mathbf{R}$ is a matrix of spatial correlations given an estimated decorrelation distance $k$. Because the spatially-varying data source effect is set to 0 for the fishery-dependent data, it then follows that the specified spatially-varying data source effect on expected catch rates enables the estimation of a fishing-power ratio for the fishery-independent data relative to the fishery-dependent data. Fishery-dependent data were set as the 'reference' data because of (1) relative similarity in fishing practices between the fishery and ATS when compared to the non-target fishery-independent bottom trawl data and (2) richness (i.e., number of observations, spatial and temporal coverage) of the fishery-dependent data relative to the non-target fishery-independent bottom trawl data.

Eq. 2 was the final formulation for the 'IID' model term configuration, in which the distribution of hake within age classes varies independently from other age classes. In the 'single-factor' model configuration, we included loading vectors $L$ to add the correlation of single spatial and spatio-temporal effects between age classes:

Eq. 4,

$$p(i) = \beta(c_i, t_i) + L_\omega(c_i)\omega(s_i) + L_\epsilon(c_i)\epsilon(s_i, t_i) + \sum_{m=1}^{n_m} \xi(s_i, m) M(i, m)$$

In the 'multi-factor' model configuration, we must sum across multiple spatial and spatio-temporal random factors $f$ to capture the aggregate spatial and spatio-temporal effects, and the loading vector $L$ becomes a matrix with dimensions $n_c \times n_f$:

Eq. 5,

$$p(i) = \beta(c_i, t_i) + \sum_{f=1}^{n_\omega} L_\omega(c_i, f)\omega(s_i, f) + \sum_{f=1}^{n_\epsilon} L_\epsilon(c_i, f)\epsilon(s_i, f, t_i) + \sum_{m=1}^{n_m} \xi(s_i, m) M(i, m)$$

In some cases, the variance or effect of spatial and spatio-temporal terms approached zero. These terms were iteratively removed such that the final models contained only terms with quantifiable, non-zero effects and variances. In the multi-factor configuration, the number of factors $n_f$ in the initial model was set equal to the number of categories $n_c$. To ensure that only influential predictors were retained in the final model, models were re-fit iteratively after removing factors with proportions of explained variance, as estimated in the loadings matrix, lower than 10%.

In all models, a gamma distribution was specified for the error of the second linear predictor. All models employed 500 spatial knots, which provided a suitable balance between resolution and run times based on preliminary analyses. It was not possible to directly include an effort offset due to incomplete recording of effort metrics across datasets. For this stock, it is common practice to assume biological sampling effort from at-sea vessels (tows) and shoreside vessels (trip) are approximately equivalent (Berger et al., 2023).

Each linear predictor was transformed using a conventional logit or exponential power link function to predict sample data as follows:

Eq. 6,

$$r_1(i) = logit^{-1}(p_1(i))$$

$$r_2(i) = exp(p_2(i))$$

where the subscripts *1* and *2* indicate encounter/non-encounter and biomass conditional on encounter, respectively.

A probability density function to predict biomass density $B$ was specified as:

Eq. 7,

$$f(B = b_i) = \begin{cases} 1-r_1(i) & if\ B=0 \\ r_1(i) * Gamma\{B=b_i|r_2(i),\sigma_m^2(c_i)\} & if\ B>0 \end{cases}$$

where $b_i$ is biomass density for sample $i$, $\sigma_m^2(c_i)$ is the residual variance in positive catch rates, and $f(B = b_i)$ is the data likelihood function.

Given the above, biomass density ( $d(s, c, t)$) was predicted for each location, category, and time by transforming linear predictors and removing terms that affect catchability:

Eq. 8,

$$d(s,c,t) = logit^{-1}(\beta_1(c_i,t_i) + \sum_{f=1}^{n_{\omega1}} L_{\omega1}(c_i,f)\omega_1(s_i,f) + \sum_{f=1}^{nn_{\epsilon1}} L_{\epsilon1}(c_i,f)\epsilon_1(s_i,f,t_i))$$
$$* \quad exp(\beta_2(c_i,t_i) + \sum_{f=1}^{n_{\omega z}} L_{\omega2}(c_i,f)\omega_2(s_i,f) + \sum_{f=1}^{n_{\epsilon z}} L_{\epsilon2}(c_i,f)\epsilon_2(s_i,f,t_i))$$

in all cases, biomass density estimates were corrected for retransformation bias using the epsilon estimator (Thorson, 2019).

Proportions at age-class (i.e., age class compositions) were calculated by dividing the biomass estimate for a given age class by the overall biomass estimate across age classes at each location and time point. We used this proportion at age class estimate as an input to the standard ATS biomass estimation process (a replacement for proportions at age measured with biological sampling in the ATS), which also incorporates acoustic data and is described in a subsequent section. Since STM biomass estimates were not derived from acoustic data and were only used to calculate proportion-at-age class, we did not draw any conclusions from the STM biomass estimates themselves.

## 2.6 Comparisons between age composition estimates

After proportional age class compositions were calculated from STM biomass-at-age class predictions, they were compared with observed proportional age class compositions from the ATS across the study area, within INPFC geographic strata (Figure 1), and across the area in which fishery-dependent data were available (INPFC strata 2-4). Comparisons were made in each of the study years. Notably, the ATS did not sample INPFC stratum 0 in some years (2003, 2007), so statistics reported for stratum 0 are based on fewer years than statistics in other strata. The model configuration that produced age class composition estimates that were most similar to age class composition observations from the ATS in 2019 was selected for use in biomass estimation.

Differences in age class compositions were represented as both the average absolute difference in proportion-at-age class (across strata, age classes, or both) and median relative error, which was calculated by

Eq. 9,

$$\frac{p_{ATS} - p_{STM}}{p_{ATS}} * 100$$

Where $p_{ATS}$ is the proportion at age class observed in the ATS and $p_{STM}$ is the proportion at age class estimated by the STM. Relative error was represented with a median value rather than a mean as infinite values were calculated when the ATS proportion-at-age-class was zero.

## 2.7 Biomass estimation and comparisons

For biomass estimation, it was necessary to refine age class compositions derived from STMs (Table 1) into exact age compositions. For each stratum, the STM-estimated proportion at age class was multiplied by the proportion of exact ages within that age class, as calculated from the raw model input data (A-SHOP, shoreside, and bottom trawl). Empirical weight-at-length and weight-at-age relationships were also calculated from raw model input data across the entire survey area for biomass calculations.

To be consistent with the standard ATS biomass estimates, biomass estimates for age-2 and older hake were calculated using standard procedures in the ATS (Chu et al., 2017), with three notable exceptions: 1) biomass estimates were generated with INPFC geographic strata instead of with strata developed within the ATS procedure based on similarity in trawl composition, 2) biomass was only estimated in U.S. waters (excluding Alaska; not the full extent of the stock into Canadian waters), and 3) the ATS age composition had age-1 hake removed prior to the biomass calculation. Therefore the acoustic backscatter that would have been converted to age-1 hake under standard protocols was instead allocated to age-2 and older fish, since it was not possible to differentiate age-1 hake from 2 and older hake in the USV dataset due to the lack of contemporaneous biological sampling. Age-1 hake are treated in a multi-factor manner by the survey and inclusion would confound the results.

Briefly, acoustic backscatter attributed to hake was apportioned based on age composition data, scaled to biomass using the hake target strength-length relationship (Traynor, 1996) and empirical weight-length and weight-age relationships, then kriged over the study area to generate a biomass estimate resolved by space and age (Chu et al., 2017). We note that the biomass estimates presented in this study differ from those used in the stock assessment for management of hake due to the differences described above, and caution against the use of biomass estimates presented in this study for anything other than a research purpose.

For all years, we estimated biomass-at-age with 38 kHz acoustic data collected in the ATS paired with (1) age compositions from the ATS, and (2) age composition estimates from a STM. Based on preliminary observations of poor STM performance in stratum 0, we also estimated biomass-at-age with age composition estimates from the best STM for strata 1-4 paired with an average age composition from the ATS in stratum 0 (across study years, 2009-2019, excluding the estimation year) as a sensitivity analysis. Results for this sensitivity analysis were presented in Supplementary Material S2. For 2019, we generated biomass estimates by pairing the three sources of age composition data described above with two sources of 38 kHz acoustic data: 1) the ATS, and 2) the USV. In the main text, we report comparisons between total biomass estimates derived from different combinations of data. We report biomass-at-age estimates derived from each combination of data in Supplementary Material S3.

Differences in total biomass estimates derived from STM age composition estimates and observed ATS age composition data were qualitatively examined to investigate the robustness of results at different life-stages and population structures. This was a qualitative analysis given the low number of study years and high number of plausible influential factors.

## 2.8 Model evaluation

To further evaluate models fit to 2019 data, we conducted simulation testing and k-fold cross-validation procedures. Here, we note that although the operational product of the STMs was a biomass-at-age class composition (proportion), the models themselves predicted biomass-at-age class (from which we subsequently calculated age class proportions). So, model evaluations were conducted with biomass-at-age class estimates as response variables, which have substantially higher dimensionality than age class composition estimates. Given the indirect nature of these evaluation procedures, we presented further details and results in Supplementary Material S4.

# 3 Results

## 3.1 Comparisons between ATS and STM age compositions

In 2019, The 5-age class model configuration (Table 1) produced estimates of age class composition that were most similar to observed age class compositions from the ATS (see Supplementary Material S5 for results from other age class configurations). Differences between model term configurations were small in 2019. Across the study area, the average absolute difference in proportion-at-age class was 0.09 (median relative error: 19.45%) for the single-factor model configuration, and 0.11 for the multi-factor and IID model configurations (median relative errors: 26.50, 26.76%, respectively). In the areas where both fishery-dependent and fishery-independent data were available (INPFC strata 2-4), the average absolute difference in proportion-at-age class was 0.03 (median relative error: 11.46%) for the single-factor model configuration, 0.04 (median relative error: 15.77%) for the multi-factor configuration, and 0.05 (median relative error: 15.85%) for the IID configuration. The single-factor term configuration of the 5-age class model was therefore selected for further examination and biomass estimation. We refer to this model configuration as the 'best STM' hereafter. Proportion at age class values for other term configurations of 2019 STMs and the 2019 ATS by strata were reported in Supplementary Material S6.

In 2019, the best STM produced age class compositions that were most similar to observed age class compositions in the ATS in stratum 2 (Figure 2). Predicted proportions at age class in strata 1, 2, 3, and 4 were all within 0.07 (median relative errors: 52.79, 6.03, 12.02, 19.45%, respectively) of proportions in the ATS across age classes, while differences in proportions across age classes in stratum 0 exceeded 0.3 (median relative error: 546.29%) (Figure 2). The high magnitude of differences in stratum 0 were driven by proportions of age-2 hake in the ATS data that exceeded 0.9, which no model configuration was able to predict (Figures 2, S4.2; S4.3).

Across study years,the single-factor 5-age class model produced estimates of age composition that were slightly more similar to ATS observations than other term configurations (Supplementary Material S7), validating our designation of the single-factor 5-age class model as the best STM. Detailed comparisons in individual years other than 2019 for all term configurations were presented in Supplementary Material S8. Across the entire study area and time period, the average difference in proportion-at-age class between the best STM and the ATS was 0.14 (median relative error: 79.03%; Figure 3). In the area where both fishery-dependent and fishery-independent data were available (INPFC strata 2-4), the average difference was 0.09 (median relative error: 54.96%; Figure 4). Relative error in proportion-at-age class estimates were highest and most variable for older age classes (7-9, 10+, Figures 3, 4), likely due in part to low relative abundance magnifying slight deviations in proportion-at-age estimates. On average, the best STM produced estimates of age class composition that were most similar to observed age class compositions in the ATS in strata 3 and 4 (0.08 average absolute difference; median relative errors: 61.1, 44.1%, respectively). The average absolute difference in proportion at age class across age classes and years was 0.11 (median relative error: 68.84%) in stratum 2, 0.16 (median relative error: 79.55%) in stratum 1, and 0.26 (median relative error: 510.27%) in stratum 0. Similar to results in 2019, the high magnitude of differences in stratum 0 across years was driven by high proportions of age-2 hake in ATS data – in some years representing 100% of the catch – which no model configuration was able to replicate. We note that the ATS did not collect biological samples in stratum 0 in 2003 and 2007, so results presented above for stratum 0 are based on fewer years than other strata.

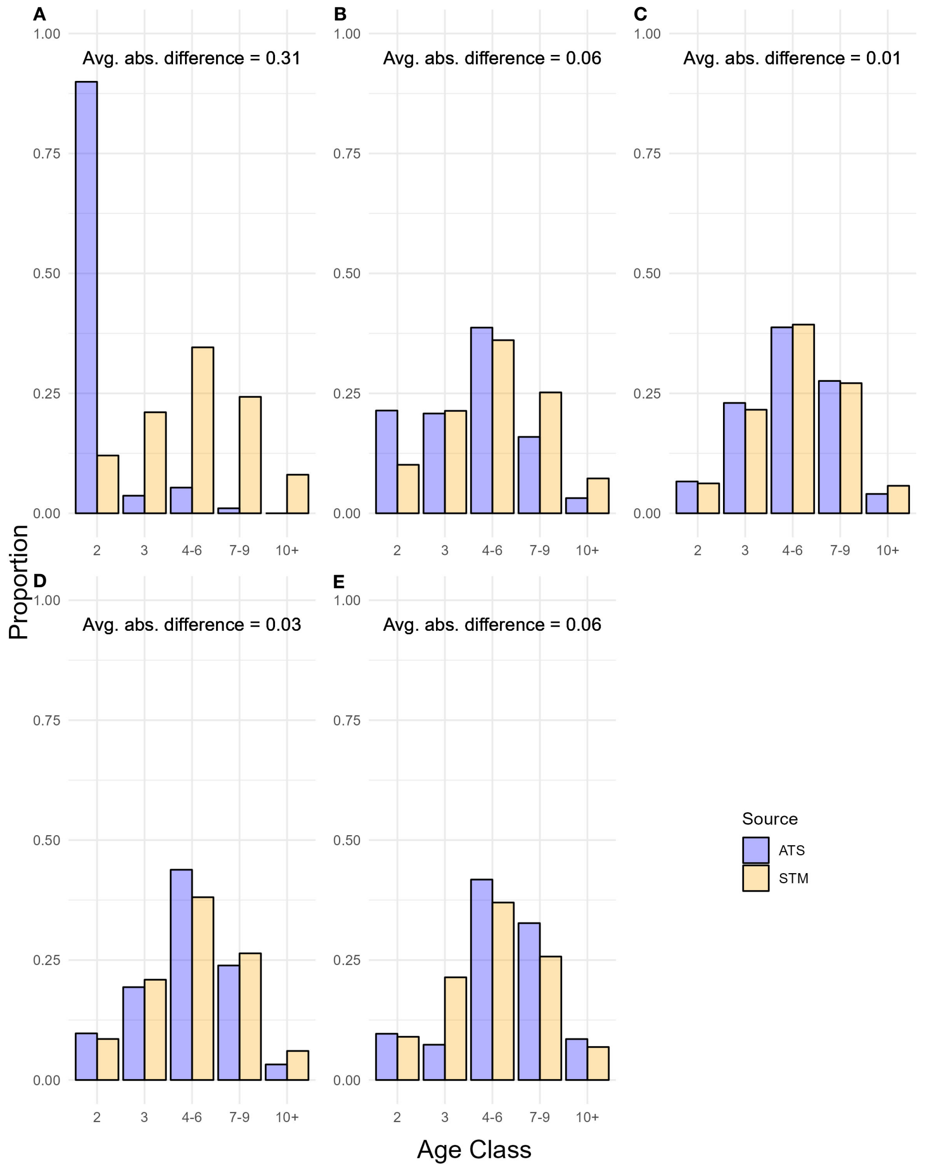



FIGURE 2
Age class composition in 2019 from the best spatio-temporal model (STM) predictions and hake acoustic-trawl survey (ATS) data in **(A)** INPFC stratum 0, **(B)** stratum 1, **(C)** stratum 2, **(D)** stratum 3, and **(E)** stratum 4. Avg. abs. difference refers to the average absolute value of differences between STM and ATS age compositions across age classes.

## 3.2 Comparisons between biomass estimates

For 2019, changing the source of acoustic data (ATS or USV) used for biomass estimation had a nearly ten-fold higher impact on biomass estimates than changing the source of age composition data (ATS or STM) (Table 3). Total biomass estimates were not substantially impacted if ATS average age compositions were substituted for STM predictions in stratum 0 (Supplementary Material S2). For 2019, holding the source of age composition constant and changing the source of acoustic data for biomass estimates yielded differences in biomass of 22.4% (when STM age compositions were used in both cases) and 26.8% (when ATS age compositions were used in both cases), while holding the source of acoustic data constant and varying the source of age composition data yielded differences in biomass of 2.6% (when USV acoustic data were used in both cases) and 3.2% (when ATS acoustic data were used in both cases) (Table 3).

Overall, the average difference between biomass estimates derived from ATS and STM age composition data was 5.6% (7.2% absolute difference). Differences between biomass estimates derived from ATS and STM age composition data were least pronounced in 2003, 2013, and 2019 (-3.9%, 2.4%, and -3.4%, respectively; Figure 5). Differences were between 7.2% and 11.1% in 2007-2012, 2015, and 2017 (Figure 5). From 2007-2017, biomass estimates derived from STM age composition estimates exceeded estimates derived from the observed ATS age composition data, while less biomass was estimated in 2003 and 2019 (Figure 5). Differences between biomass estimates derived from STM age composition estimates and observed

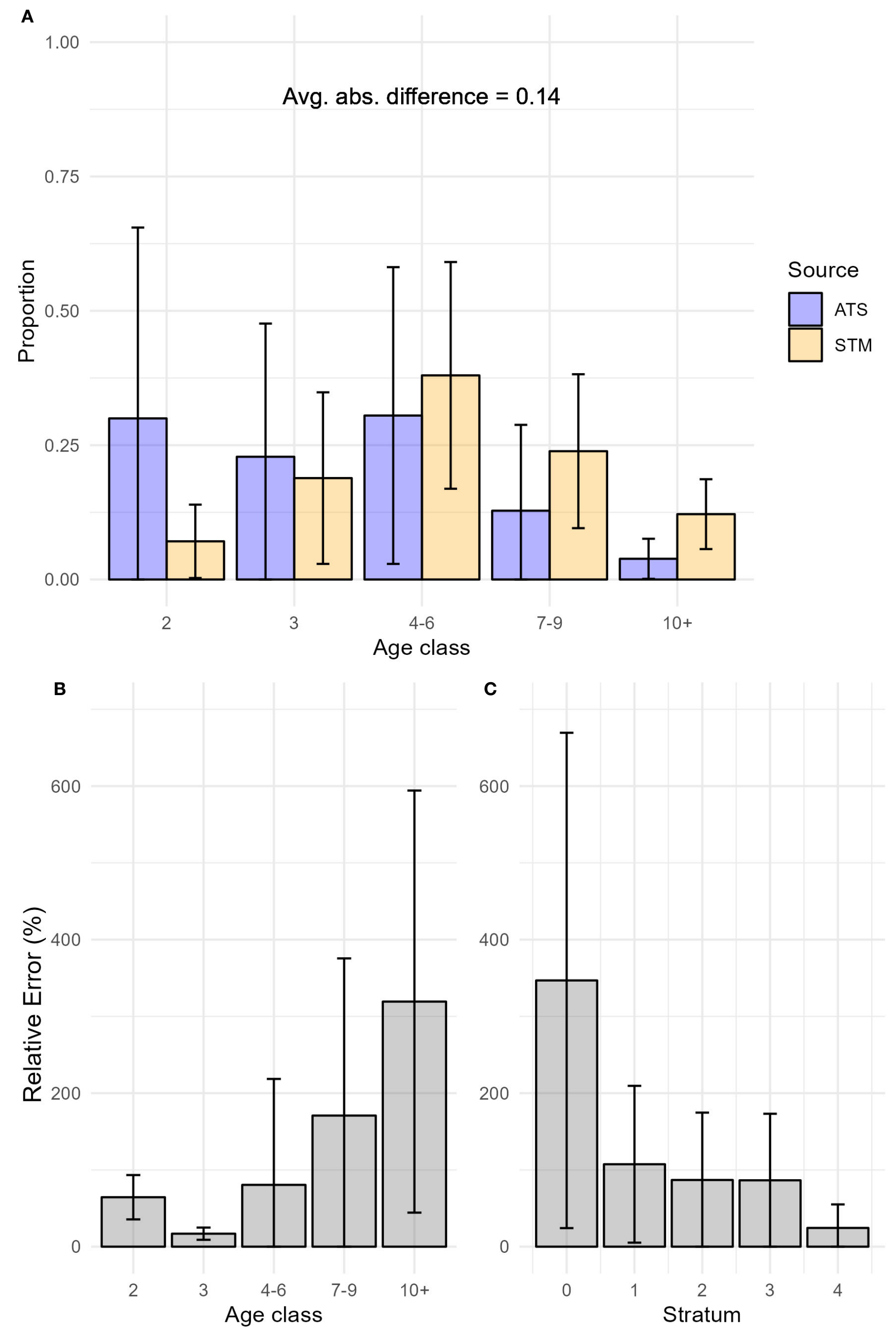



FIGURE 3
**(A)** Proportion at age class from the best spatio-temporal model (STM) model predictions and the hake acoustic trawl survey (ATS) data from 2003-2019, **(B)** Average relative error of STM proportion at age class predictions by age class from 2003-2019, and **(C)** Average relative error of STM proportion at age class predictions by strata from 2003-2019-. Bars indicate the standard deviation of the quantity plotted. Avg. abs. difference refers to the average absolute value of differences between STM and ATS age compositions across age classes.

ATS age composition data appeared to be qualitatively associated with population structure (i.e., total biomass and the proportion of young, maturing age-2 and 3 hake in the ATS biological sampling; Figure 5). With notable exceptions, biomass estimates derived from STM age composition estimates were most comparable to those derived from observed ATS age composition data when the proportion of age-2 and 3 hake were low and total hake biomass was high (Figure 5).

# 4 Discussion

The inability of USVs to collect biological composition data (e.g., age, length) has thus far hampered their use in fishery resource survey programs. We developed an approach that utilizes non-contemporaneous biological sampling data (a combination of fishery-dependent and non-target fishery-independent data) fit to

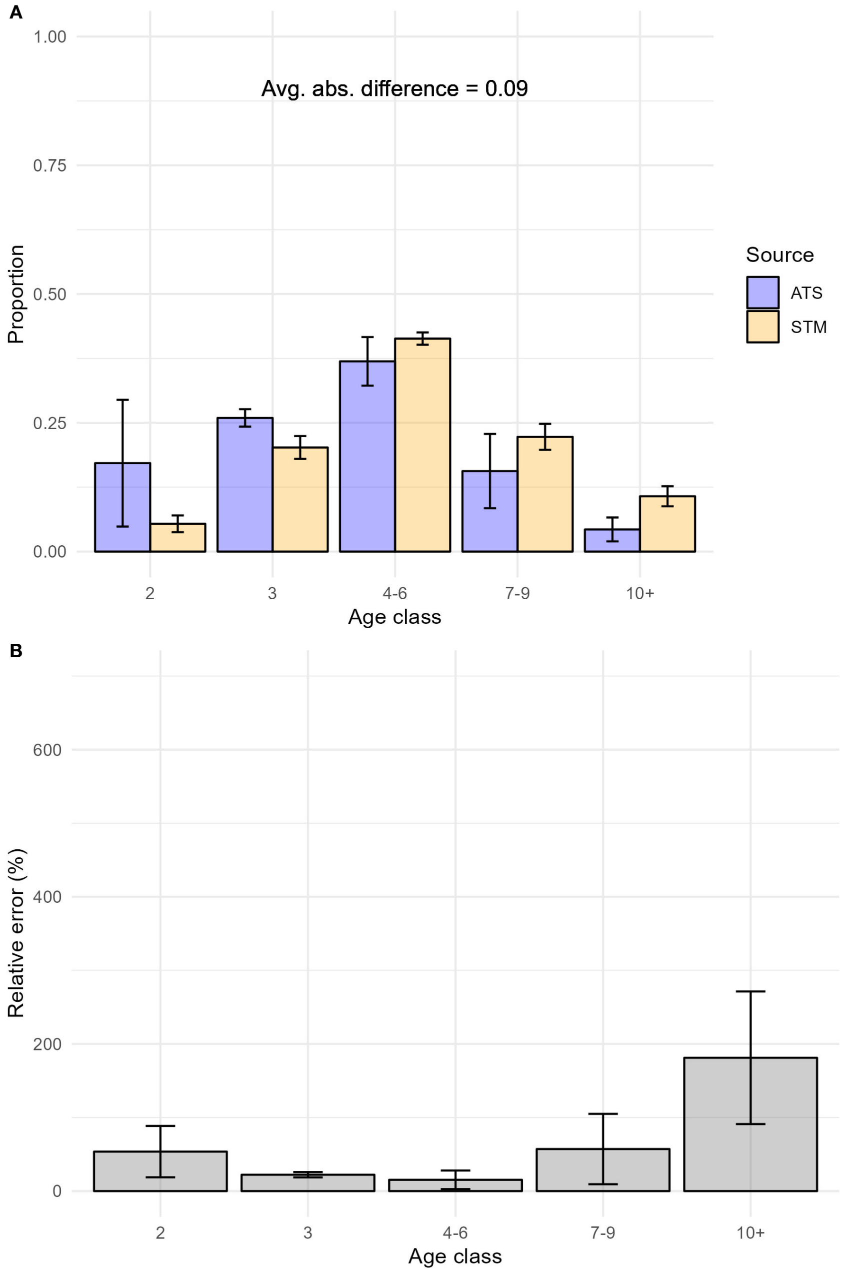



FIGURE 4
**(A)** Proportion at age class from the best spatio-temporal model (STM) model predictions and the hake acoustic trawl survey (ATS) data from 2003-2019 in strata 2-4, and **(B)** Average relative error of STM proportion at age class predictions by age class from 2003-2019 in strata 2-4. Bars indicate the standard deviation of the quantity plotted. Avg. abs. difference refers to the average absolute value of differences between STM and ATS age compositions across age classes.

**TABLE 3** Total hake biomass estimates (kilotonnes; kt) derived from different sources of data in 2019.

| | ATS age composition data | STM age composition data | Difference between biomass estimates (age composition data difference) |
|---|---|---|---|
| ATS acoustic data | 1523.59 kt | 1473.91 kt | -3.2% |
| USV acoustic data | 1114.88 kt | 1143.80 kt | +2.6% |
| Difference between biomass estimates (acoustic data difference) | -26.8% | -22.4% | |

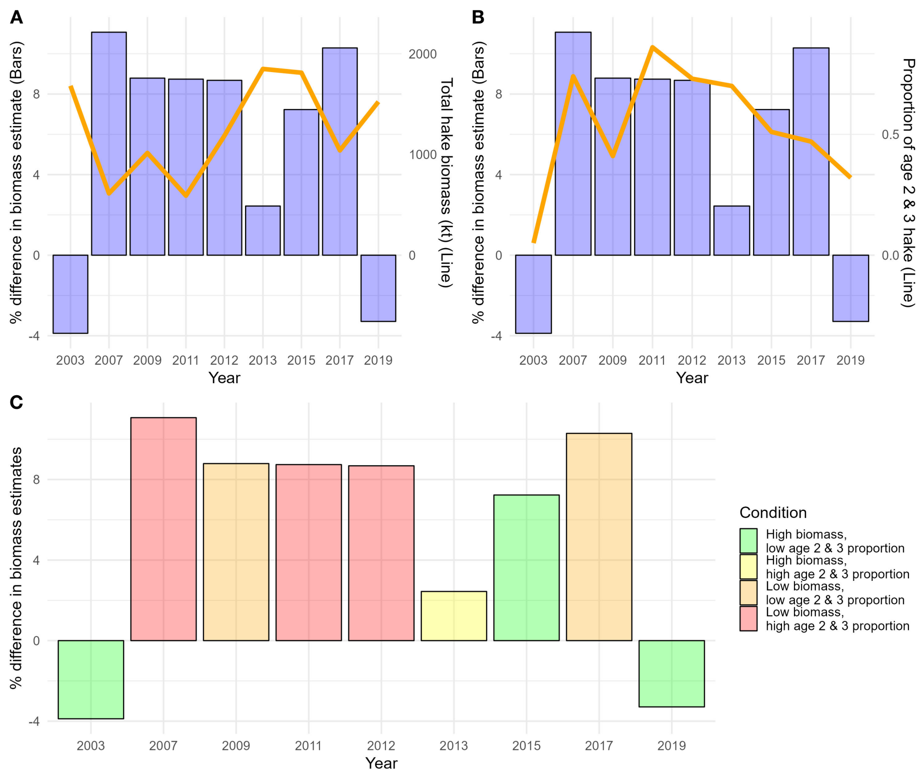



FIGURE 5
Differences in total biomass estimates derived from the best spatio-temporal model (STM) age composition estimates and observed hake acoustic trawl survey (ATS) age composition data across years. **(A)** Percentage difference in biomass estimates (bars) and total age-2 and older hake biomass (line). Percentages reflect differences in biomass estimates derived from ATS only (age composition and acoustic) and STM predictions (age composition) paired with ATS acoustics relative to ATS biomass estimates. **(B)** Same as A, but with the proportion of age-2 and 3 hake sampled in the ATS (line). **(C)** Same as A, but years are color coded according to different population structure conditions: high or low total biomass and biomass of young, maturing (age 2 and 3) fish. Low and high were demarcated by the overall average (biomass: 1256.93 kt; age 2 and 3 proportion: 0.53).

a STM to address this limitation. In the areas where fishery-dependent data were available (INPFC strata 2-4; Figure 1), this approach produced estimates of proportional age class composition that were on average within 0.03 of proportional age class compositions observed in the hake ATS in the study focal year of 2019 (median relative error: 11.46%), and 0.09 on average across the study period (2003-2019; median relative error: 54.96%) (Figures 2, 3). While the magnitude of differences between modelled and observed age compositions was greater across the entire study area (0.09 difference on average in 2019, median relative error: 19.45%; 0.14 difference on average across all years over the entire survey area, median relative error: 79.03%), total biomass estimates that were derived from STM age composition estimates were within approximately 7% of those derived from ATS composition data given the same source of acoustic data across study years (~ 3% in 2019; Table 3). These levels of change in the overall survey biomass estimate are well below the total variability (coefficient of variation of 36%) associated with the ATS in recent stock assessments (Berger et al., 2023). While differences between STM- and ATS-produced age compositions did not translate into large discrepancies in total biomass, accurate estimates of age structure remain particularly important for management operations. In particular, age-2 and 3 compositions in the most recent survey provide a critical first observation related to recent recruitment levels for the stock assessment, and recent recruitment estimates considerably impact current and near-term harvest specifications for Pacific Hake. Future work should evaluate how alternative survey protocols ultimately influence existing management procedures (e.g., stock assessment formulation and harvest policy) by integrating specific results from this work into the existing Pacific Hake management strategy evaluation tool (Jacobsen et al., 2021).

When biomass was estimated from USV (Saildrone) acoustic data, estimates differed from those derived from ATS acoustic data by more than 22%, regardless of whether STM or ATS age composition information was used (Table 3). Thus, the primary limitation for operationalizing USV data for hake biomass-at-age estimation appears to be rectifying discrepancies between USV and ATS acoustic data, whereas the lack of contemporaneous age composition sampling associated with USV surveys appears secondary. There are several explanations for the differences between acoustic backscatter recorded by USVs and the acoustic backscatter recorded by the ATS in this study. The most prominent of those include signal attenuation due to bubble injection (Simmonds and MacLennan, 2005; Shabangu et al., 2014; Ryan et al., 2015), other effects of inclement weather, which are more pronounced for USVs (Jech et al., 2021), calibration uncertainty

(Demer et al., 2015; De Robertis et al., 2019a), differences in vessel response (De Robertis and Handegard, 2013; De Robertis et al., 2019b), and differences in the ability of analysts to identify hake from other species in acoustic data. The final consideration is one of the most plausible source of the observed differences in backscatter, as while there are well-established procedures for identifying hake in acoustic data, analysts of USV data made most determinations based on 38 kHz data alone (or with limited information from 200 kHz) rather than based on several frequencies (18, 38, 70, 120, 200 kHz) and in conjunction with trawl data.

Biological sampling is crucial for ground-truthing acoustic surveys. While we showed that reasonable age compositions can be estimated for hake without it in some areas, contemporaneous biological sampling remains the most viable method for validating estimated species compositions in the ATS. The lack of biological data for analysts to verify species classifications in the acoustic data and the limited suite of acoustic frequencies available led to more subjective designations by analysts. Results indicated that the increased subjectivity led to analysts taking a precautionary approach to apportioning backscatter to hake. The high relative importance of biological sampling and multiple frequencies in the ATS echogram judging procedures may be reflected by comparing our findings with others who compared research vessel and USV data. We observed substantially less hake backscatter in USV data relative to research vessels, while others observed more backscatter in USV data (Swart et al., 2016; Chu et al., 2019; De Robertis et al., 2019b). Given the performance of STMs for estimating age compositions and the capability for other USVs to be equipped with a suite of echosounders that are more comparable (or exactly the same) to the research vessel suite, reliably classifying acoustic backscatter by species is the most significant obstacle to using USV data to produce viable biomass-at-age estimates. It is possible that STMs fit to fishery and non-target fishery-independent data could be useful in overcoming this obstacle. However, automated methods for species classification in acoustic data (e.g., artificial intelligence/machine learning, Sarr et al., 2021; inversion methods, Urmy et al., 2023) and analysis of frequency modulated (i.e., broadband) acoustic data are likely to have the most utility.

Using STMs to provide age composition data worked reasonably well because of an important property of the data we used to fit our models and hake survey protocols. In general, fishers target larger, older individuals (Birkeland and Dayton, 2005). However, the boom-and-bust nature of hake recruitment results in the stock being supported primarily by two-to-four strong age classes in any given year (Horne and Smith, 1997; Hamel et al., 2015; Johnson et al., 2021). As a result, hake fishers target the larger and older hake less than they would if fishing a species with more constant recruitment. This means that differences in selectivity at age between fishery-dependent and fishery-independent data are not pronounced – at least for hake older than three (Figure S9.1; Berger et al., 2023). Further, the ATS does not trawl randomly along survey transects and instead targets suspected aggregations of hake. Thus, trawling locations in the ATS are decided using criteria that are largely similar to those the fishery uses, although the diffuse aggregations of hake are likely targeted more frequently in the ATS than fishery. Accordingly, the approach of the present study may work similarly well for fishes with similar patterns of recruitment and survey practices (e.g., clupeids, gadoids), but perhaps not as well for fishes with different recruitment patterns and survey practices (e.g., serranids, *Sebastes* spp.). Additional research on a diversity of species will be necessary to determine the applicability of the approach described in this study beyond hake.

One important limitation of our approach to providing age composition estimates is the paucity of small, young hake in the fishery-dependent and non-target fishery-independent data. The ATS generates a biomass index for age-1 hake and their biomass is removed from age-2 and older biomass estimates in the estimation procedure. Since we did not have enough data to include age-1 hake in our STMs, records of age-1 hake were removed from ATS biological data to avoid confounding biomass comparisons. While this was reasonable in a research context, age-1 hake must be represented in assessment and management contexts. For age-2 and older hake, the influence of differences in the relative abundance of small, young hake between the ATS and STM predictions was reflected in biomass estimates. Biomass estimates derived from STM age composition data were generally higher than those derived with ATS age composition data, which was likely due to lower relative abundance of age-2 and 3 hake in STM predictions. When the same amount of acoustic backscatter is attributed to larger (older) fish, more biomass is estimated. This limitation and finding underscore the importance of targeted fishery-independent sampling for scarce size and age classes to supplement ancillary data in similar applications, and to fishery-independent survey programs generally.

In general, STM age composition-derived biomass estimates were most similar to those derived from ATS age compositions when total biomass was above average (Figure 5). The combination of above average biomass and above average proportions of age-2 and 3 hake in the ATS was associated with the smallest difference in biomass estimates, although this condition only occurred once in the study period so the magnitude of difference could be coincidental (Figure 5). In other years in which the difference in biomass estimates was below average, the proportion of age-2 and 3 hake was also below average (Figure 5). Fishery selectivity of age-2 and 3 hake is relatively low in a typical year (Figure S9.2; Berger et al., 2023) and age-2 and 3 hake were more relatively abundant in strata 0 and 1, where our dataset was most sparse. Despite the qualitative associations described above, population structure did not appear to completely explain differences in biomass estimates. Other phenomena that plausibly contribute to explaining why our approach for modelling age composition data worked better in some years than others include variability in oceanographic conditions (e.g., temperature at depth and subsurface flow; Agostini et al., 2006; Hamel et al., 2015; Malick et al., 2020) and ecological dynamics (e.g., Humboldt squid and krill distributions; Litz et al., 2011; Thomas et al., 2011; Phillips et al., 2022) that affected hake distribution at age, fishing practices in the hake fishery, and uncertainty in the hake survey.

Hake conduct a seasonal northward migration, and their distribution patterns have been explained by variation in water temperature (Dorn, 1995; Hamel et al., 2015; Malick et al., 2020), sub-surface flow and bottom depth (Smith, 1990; Agostini et al., 2006; Agostini et al., 2008), and age (Beamish and McFarlane, 1985; Dorn, 1995, Hamel et al., 2015). We did not include distinct environmental or ecological covariates in our models, and instead used latent spatial and spatio-temporal variables to predict variation in biomass distribution over space and time. The 'single-factor' configuration of the 5-age class model produced age composition estimates that were most similar to those observed in the acoustic trawl survey in 2019, but it only performed marginally better than other configurations (Supplementary Material S6-8). Interpreting the ecological effects of latent variables in spatio-temporal models can be challenging. Future work should integrate the hypothesis-driven work cited above into expanded STM frameworks (e.g., mechanistic species distribution models) to advance the predictive capabilities necessary to plan for emerging challenges (e.g., climate change).

In future work, it would be advantageous to directly model composition data (e.g., Thorson and Haltuch, 2019; Grüss et al., 2020; Thorson et al., 2022) so that the product of models could be evaluated directly with commonly used statistical procedures (e.g., k-fold cross validation). The dimensionality of biomass estimates is significantly higher than the age compositions (proportion) calculated posthoc. Differences in preferential sampling between fishery-dependent and independent data, which can be impactful in STMs (Conn et al., 2017; Alglave et al., 2022), likely also had a greater influence on biomass estimates than on proportional age class compositions. These influences likely contributed to our finding of generally unfavorable statistical evaluations of biomass-at-age class predictions (Supplementary Material S4) despite generally favorable comparisons between observed and estimated age compositions. In essence, however, the goal of statistical evaluation procedures is to test if an approach produces reasonable predictions in different scenarios for the underlying data. We treated analysis of data in years prior to 2019 as a 'practical' evaluation method, and since total biomass and age class distribution and strength varied substantially between years, we were confident that our approach produced reasonable estimates of age-class composition in areas where sufficient data were present.

The disparate performance of STMs between areas where the fishery operates and areas where it does not operate illuminates important considerations for future sampling designs involving USVs. Our non-target fishery-independent bottom trawl data captured age compositions that were far more static than ATS midwater trawl age compositions across the survey area. Fishery-independent midwater trawls appear uniquely equipped to capture the high relative abundance of small, young hake in the southern extent of their range (e.g., INPFC stratum 0). While truly contemporaneous collection of age composition data may not be necessary for species such as hake in most areas, it may be necessary to pair USV acoustic surveys with regional chartered fishing vessel surveys in areas such as INPFC stratum 0, where only sparse non-target fishery-independent data were available. Such a protocol could be an important bridge between entirely FSV based surveys and USV-only surveys while platform and species classification issues are addressed. Though more resource-intensive than a USV-only design, USV-chartered vessel or USV-research vessel surveys could be considerably less resource-intensive than research vessel only surveys and could facilitate expansions of survey spatio-temporal coverage with lower monetary and operational burden.

Fitting STMs to a combination of fishery-dependent and -independent data produced estimates of hake age-class composition that were largely comparable to those observed in the ATS in areas where the fishery operates. While further research is necessary before data from USVs like Saildrones are incorporated into biomass-at-age estimates that are suitable for the hake stock assessment, using our approach to provide age composition data appears to be largely viable where data are abundant. This research opens the door for the use of acoustic data without contemporaneously-collected age composition data – provided that differences in acoustic data between platforms are understood quantitatively, age data from other sources are available, and selectivity differences between datasets are well described. Given other successful USV surveys of fish stocks (e.g., De Robertis et al., 2021), we believe that platform issues (e.g., differences in echosounder configurations) will be relatively straightforward to overcome. Species classification remains the primary obstacle that impedes the use of non-ground-truthed acoustic data in generating indices for stock assessment. Looking to the future, we believe that broadband acoustic data, machine learning, and inversion methods will be increasingly useful for naïve species classification, and pairing such analyses with STMs of species and age composition could eventually be a viable approach for generating biomass-at-age estimates that are suitable for many stock assessments with acoustic data collected by USVs.

## CDFW disclaimer

CDFW collects data from various sources for fisheries management purposes, and data may be modified at any time to improve accuracy and as new data are acquired. CDFW may provide data upon request under a formal agreement. Data are provided as-is and in good faith, but CDFW does not endorse any particular analytical methods, interpretations, or conclusions based upon the data it provides. Unless otherwise stated, use of CDFW's data does not constitute CDFW's professional advice or formal recommendation of any given analysis. CDFW recommends users consult with CDFW prior to data use regarding known limitations of certain data sets.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Confidential fisheries data were used in this article and are not publicly available. Fishery-independent data are available through data requests to the NOAA Northwest Fisheries Science Center. Requests to access these datasets should be directed to derek.bolser@noaa.gov.

## Author contributions

DB, AB, and LC conceived of the study. DB, AB, DC, JC, JH, and LC refined the approach and scope. DB, DC, SB, JP, RT, JW, and JC analyzed data. DB wrote the first draft of the manuscript. All authors contributed to the final submitted materials.

## Acknowledgments

We thank the hake fishers, observers, and survey personnel who collected the data used in this study. We are grateful to the California, Oregon, and Washington Departments of Fish and Wildlife and the Pacific States Fisheries Information Network for their role in providing the shoreside data, and Vanessa Tuttle of the At-Sea Hake Observer Program for her role in providing the A-SHOP data. We recognize Kelli Johnson, Maxime Olmos, Jim Thorson, and Arnaud Grüss for their helpful feedback at various stages of this project. We are especially grateful to Jim Thorson for a pre-submission review of the materials. We also recognize the contributions of Canadian Department of Fisheries and Oceans personnel to the hake acoustic trawl survey, although Canadian data was not analyzed in this study. This study was funded by an internal grant from the Northwest Fisheries Science Center, National Oceanographic and Atmospheric Administration, U.S. Department of Commerce that was administered by the Oregon State University Cooperative Institute for Marine Resources Studies.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1214798/full#supplementary-material

## References

Agostini, V. N., Francis, R. C., Hollowed, A. B., Pierce, S. D., Wilson, C., and Hendrix, A. N. (2006). The relationship between Pacific hake (*Merluccius productus*) distribution and poleward subsurface flow in the California Current System. *Can. J. Fish. Aquat. Sci.* 63, 2648–2659. doi: 10.1139/f06-139

Agostini, V. N., Hendrix, A. N., Hollowed, A. B., Wilson, C. D., Pierce, S. D., and Francis, R. C. (2008). Climate–ocean variability and Pacific hake: A geostatistical modeling approach. *J. Mar. Syst.* 71, 237–248. doi: 10.1016/j.jmarsys.2007.01.010

Alglave, B., Rivot, E., Etienne, M.-P., Woillez, M., Thorson, J. T., and Vermard, Y. (2022). Combining scientific survey and commercial catch data to map fish distribution. *ICES J. Mar. Sci.*, 79 (4), 1133–1149. fsac032. doi: 10.1093/icesjms/fsac032

Beamish, R. J., and McFarlane, G. A. (1985). Pacific whiting, Merluccius productus, stocks off the west coast of Vancouver Island, Canada. *Mar. Fish. Rev.* 47 (2), 75–81.

Berger, A. M., Goethel, D. R., Lynch, P. D., Quinn, T., Mormede, S., McKenzie, J., et al. (2017). Space oddity: The mission for spatial integration. *Can. J. Fish. Aquat. Sci.* 74, 1698–1716. doi: 10.1139/cjfas-2017-0150

Berger, A. M., Grandin, C. J., Johnson, K. F., and Edwards, A. M. (2023). Status of the Pacific hake (whiting) stock in U.S. and Canadian waters in 2023. Prepared by the Joint Technical Committee of the U.S. and Canada Pacific hake/whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada. 208

Birkeland, C., and Dayton, P. K. (2005). The importance in fishery management of leaving the big ones. *Trends Ecol. Evol.* 20, 356–358. doi: 10.1016/j.tree.2005.03.015

Brodie, S. J., Thorson, J. T., Carroll, G., Hazen, E. L., Bograd, S., Haltuch, M. A., et al. (2020). Trade-offs in covariate selection for species distribution models: a methodological comparison. *Ecography* 43, 11–24. doi: 10.1111/ecog.04707

Chu, D., Parker-Stetter, S., Hufnagle, L. C., Thomas, R., Getsiv-Clemons, J., Gauthier, S., et al. (2019). 2018 Unmanned Surface Vehicle (Saildrone) acoustic survey off the west coasts of the United States and Canada. in. *OCEANS 2019 MTS/IEEE SEATTLE*, 1–7. doi: 10.23919/OCEANS40490.2019.8962778

Chu, D., Thomas, R. E., de Blois, S. K., and Hufnagle, J. L. C. (2017) *Pacific hake integrated acoustic and trawl survey methods – DRAFT*. Available at: https://media.fisheries.noaa.gov/2021-11/2017-acoustic-survey-methods.pdf.

Conn, P. B., Thorson, J. T., and Johnson, D. S. (2017). Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. *Methods Ecol. Evol.* 8, 1535–1546. doi: 10.1111/2041-210X.12803

de Blois, S. (2020). The 2019 joint U.S.–Canada integrated ecosystem and Pacific hake acoustic-trawl survey: cruise report SH-19-06. U.S. Department of commerce, NOAA processed report NMFS-NWFSC PR-2020-03. doi: 10.25923/p55a-1a22

Demer, D. A., Berger, L., Bernasconi, M., Bethke, E., Boswell, K., Chu, D., et al. (2015). Calibration of acoustic instruments. *Int. Council Explor. Sea (ICES)*. pp. 133. doi: 10.25607/OBP-185

De Robertis, A., Bassett, C., Andersen, L. N., Wangen, I., Furnish, S., and Levine, M. (2019a). Amplifier linearity accounts for discrepancies in echo-integration measurements from two widely used echosounders. *ICES J. Mar. Sci.* 76, 1882–1892. doi: 10.1093/icesjms/fsz040

De Robertis, A., and Handegard, N. O. (2013). Fish avoidance of research vessels and the efficacy of noise-reduced vessels: a review. *ICES J. Mar. Sci.* 70, 34–45. doi: 10.1093/icesjms/fss155

De Robertis, A., Lawrence-Slavas, N., Jenkins, R., Wangen, I., Mordy, C. W., Meinig, C., et al. (2019b). Long-term measurements of fish backscatter from Saildrone unmanned surface vehicles and comparison with observations from a noise-reduced research vessel. *ICES J. Mar. Sci.* 76, 2459–2470. doi: 10.1093/icesjms/fsz124

De Robertis, A., Levine, M., Lauffenburger, N., Honkalehto, T., Ianelli, J., Monnahan, C. C., et al. (2021). Uncrewed surface vehicle (USV) survey of walleye pollock, Gadus chalcogrammus, in response to the cancellation of ship-based surveys. *ICES J. Mar. Sci.* 78, 2797–2808. doi: 10.1093/icesjms/fsab155

Dorn, M. W. (1995). The effects of age composition and oceanographic conditions on the annual migration of Pacific whiting, Merluccius productus. *Calif. Coop. Ocean. Fish. Investig. Rep.* 97–105. doi: 10.1093/icesjms/fsab155

Godefroid, M., Boldt, J. L., Thorson, J. T., Forrest, R., Gauthier, S., Flostrand, L., et al. (2019). Spatio-temporal models provide new insights on the biotic and abiotic drivers shaping Pacific Herring (Clupea pallasi) distribution. *Prog. Oceanography* 178, 102198. doi: 10.1016/j.pocean.2019.102198

Grüss, A., and Thorson, J. T. (2019). Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. *ICES J. Mar. Sci.* 76, 1748–1761. doi: 10.1093/icesjms/fsz075

Grüss, A., Thorson, J. T., Anderson, O. F., O'Driscoll, R., Heller-Shipley, M., and Goodman, S. (2023). Spatially varying catchability for integrating research survey data with other data sources: case studies involving observer samples, industry-cooperative surveys, and predators-as-samplers. *Can. J. Fish. Aquat. Sci.* doi: 10.1139/cjfas-2023-0051

Grüss, A., Thorson, J. T., Carroll, G., Ng, E. L., Holsman, K. K., Aydin, K., et al. (2020). Spatio-temporal analyses of marine predator diets from data-rich and data-limited systems. *Fish Fisheries* 21, 718–739. doi: 10.1111/faf.12457

Hamel, O. S., Ressler, P. H., Thomas, R. E., Waldeck, D. A., Hicks, A. C., Holmes, J. A., et al. (2015). Biology, fisheries, assessment and management of Pacific hake (*Merluccius productus*). *Hakes: Biol. exploitation* 234–262. doi: 10.1002/9781118568262.ch9

Horne, J. K., and Smith, P. E. (1997). Space and time scales in Pacific hake recruitment processes: latitudinal variation over annual cycles. *California Cooperative Oceanic Fisheries Investigations Rep.* 38, 90–102.

Jacobsen, N. S., Marshall, K. N., Berger, A. M., Grandin, C. J., and Taylor, I. G. (2021). doi: 10.25923/X9F9-9B20. Management Strategy Evaluation of Pacific Hake: Exploring the Robustness of the Current Harvest Policy to Spatial Stock Structure, Shifts in Fishery Selectivity, and Climate-Driven Distribution Shifts. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-168.

Jech, J. M., Schaber, M., Cox, M., Escobar-Flores, P., Gastauer, S., Haris, K., et al. (2021). Collecting quality echosounder data in inclement weather. *ICES Cooperative Res. Rep.* 352, 108. doi: 10.17895/ices.pub.7539

Johnson, K. F., Edwards, A. M., Berger, A. M., and Grandin, C. J. (2021). *Status of the Pacific Hake (whiting) stock in U.S. and Canadian waters in 2021* (Lisbon, Portugal: Prepared by the Joint Technical Committee of the U.S. and Canada Pacific Hake/Whiting Agreement, National Marine Fisheries Service and Fisheries and Oceans Canada) 269.

Keller, A. A., Wallace, J. R., and Methot, R. D. (2017). The Northwest fisheries science center's west coast groundfish bottom trawl survey: survey history, design, and description. U.S. Department of commerce, NOAA technical memorandum, NMFS-NWFSC-136. pp. 47

Levine, R. M., Robertis, A. D., Grünbaum, D., Woodgate, R., Mordy, C. W., Mueter, F., et al. (2021). Autonomous vehicle surveys indicate that flow reversals retain juvenile fishes in a highly advective high-latitude ecosystem. *Limnology Oceanography* 66, 1139–1154. doi: 10.1002/lno.11671

Litz, M. N., Phillips, A. J., Brodeur, R. D., and Emmett, R. L. (2011). Seasonal occurrences of Humboldt squid (*Dosidicus gigas*) in the northern California Current System. *CalCOFI Rep.* 52, 97–108.

Liu, Z., Zhang, Y., Yu, X., and Yuan, C. (2016). Unmanned surface vehicles: An overview of developments and challenges. *Annu. Rev. Control* 41, 71–93. doi: 10.1016/j.arcontrol.2016.04.018

Malick, M. J., Hunsicker, M. E., Haltuch, M. A., Parker-Stetter, S. L., Berger, A. M., and Marshall, K. N. (2020). Relationships between temperature and Pacific hake distribution vary across latitude and life-history stage. *Mar. Ecol. Prog. Ser.* 639, 185–197. doi: 10.3354/meps13286

Meinig, C., Burger, E. F., Cohen, N., Cokelet, E. D., Cronin, M. F., Cross, J. N., et al. (2019). Public–private partnerships to advance regional ocean-observing capabilities: A saildrone and NOAA-PMEL case study and future considerations to expand to global scale observing. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00448

Mordy, C. W., Cokelet, E. D., De Robertis, A., Jenkins, R., Kuhn, C. E., Lawrence-Slavas, N., et al. (2017). Advances in ecosystem research: Saildrone surveys of oceanography, fish, and marine mammals in the bering sea. *Oceanography* 30, 113–115. doi: 10.5670/oceanog.2017.230

Nickford, S., Palter, J. B., Donohue, K., Fassbender, A. J., Gray, A. R., Long, J., et al. (2022). Autonomous wintertime observations of air-sea exchange in the gulf stream reveal a perfect storm for ocean CO2 uptake. *Geophysical Res. Lett.* 49, e2021GL096805. doi: 10.1029/2021GL096805

NOAA (2015) *Fisheries economics of the United States* (National Oceanic and Atmospheric Administration). Available at: http://www.st.nmfs.noaa.gov/economics/publications/feus/fisherieseconomics2015/index (Accessed 12 July 2019).

Northwest Fisheries Science Center (NWFSC), At-Sea Hake Observer Program. (2022). Sampling manual. Fishery resource snalysis and monitoring, At-Sea hake observer program. NWFSC, 2725 Montlake Blvd. East, Seattle, Washington 98112.

O'Leary, C. A., Thorson, J. T., Ianelli, J. N., and Kotwicki, S. (2020). Adapting to climate-driven distribution shifts using model-based indices and age composition from multiple surveys in the walleye pollock (Gadus chalcogrammus) stock assessment. *Fisheries Oceanography* 29, 541–557. doi: 10.1111/fog.12494

Phillips, E. M., Chu, D., Gauthier, S., Parker-Stetter, S. L., Shelton, A. O., and Thomas, R. E. (2022). Spatiotemporal variability of euphausiids in the California Current Ecosystem: insights from a recently developed time series. *ICES J. Mar. Sci.* 79, 1312–1326. doi: 10.1093/icesjms/fsac055

Ryan, T. E., Downie, R. A., Kloser, R. J., and Keith, G. (2015). Reducing bias due to noise and attenuation in open-ocean echo integration data. *ICES J. Mar. Sci.* 72, 2482–2493. doi: 10.1093/icesjms/fsv121

Sarr, J.-M. A., Brochier, T., Brehmer, P., Perrot, Y., Bah, A., Sarré, A., et al. (2021). Complex data labeling with deep learning methods: Lessons from fisheries acoustics. *ISA Trans.* 109, 113–125. doi: 10.1016/j.isatra.2020.09.018

Saelens, M. R., and Jesse, L. K. (2007). Shoreside hake observation program: 2006 annual report. marine resources program oregon department of fish and wildlife hatfield marine science center newport. 97365

Sepp, E., Vetemaa, M., and Raid, T. (2022). Trends in maritime technology and engineering: *Proceedings of the 6th International Conference on Maritime Technology and Engineering* (MARTECH 2022, Lisbon, Portugal, 24-26 May 2022) (1st ed.). (London: CRC Press).doi: 10.1201/9781003320289

Shabangu, F. W., Ona, E., and Yemane, D. (2014). Measurements of acoustic attenuation at 38kHz by wind-induced air bubbles with suggested correction factors for hull-mounted transducers. *Fisheries Res.* 151, 47–56. doi: 10.1016/j.fishres.2013.12.008

Simmonds, J., and MacLennan, D. N. (2005). *Fisheries acoustics: Theory and practice.* Second edition. Blackwell Science, (Oxford: John Wiley & Sons). 1–252

Smith, S. J. (1990). Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Can. J. Fish. Aquat. Sci.* 47, 894–903. doi: 10.1139/f90-103

Swart, S., Zietsman, J., Coetzee, J., Goslett, D., Hoek, A., Needham, D., et al. (2016). Ocean robotics in support of fisheries research and management. *Afr. J. Mar. Sci.* 38, 525–538. doi: 10.2989/1814232X.2016.1251971

Thomas, R., Stewart, I., Chu, D., Pohl, J., Cooke, K., Grandin, C., et al. (2011). Acoustic biomass estimation and uncertainty of Pacific hake and Humboldt squid in the Northern California current in 2009. *J. Acoustical Soc. America* 129, 2691–2691. doi: 10.1121/1.3589026

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Thorson, J. T., Arimitsu, M. L., Levi, T., and Roffler, G. H. (2022). Diet analysis using generalized linear models derived from foraging processes using R package mvtweedie. *Ecology* 103, e3637. doi: 10.1002/ecy.3637

Thorson, J. T., Barnes, C. L., Friedman, S. T., Morano, J. L., and Siple, M. C. (2023). Spatially varying coefficients can improve parsimony and descriptive power for species distribution models. *Ecography* n/a, e06510. doi: 10.1111/ecog.06510

Thorson, J. T., and Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES J. Mar. Sci.* 74, 1311–1321. doi: 10.1093/icesjms/fsw193

Thorson, J. T., Clarke, M. E., Stewart, I. J., and Punt, A. E. (2012). The implications of spatially varying catchability on bottom trawl surveys of fish abundance: a proposed solution involving underwater vehicles. *Can. J. Fisheries Aquat. Sci.* 70 (2), 294–306. doi: 10.1139/cjfas-2012-0330

Thorson, J. T., Cunningham, C. J., Jorgensen, E., Havron, A., Hulson, P.-J. F., Monnahan, C. C., et al. (2021). The surprising sensitivity of index scale to delta-model assumptions: Recommendations for model-based index standardization. *Fisheries Res.* 233, 105745. doi: 10.1016/j.fishres.2020.105745

Thorson, J. T., and Haltuch, M. A. (2019). Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. *Can. J. Fish. Aquat. Sci.* 76, 401–414. doi: 10.1139/cjfas-2018-0015

Traynor, J. J. (1996). Target-strength measurements of walleye pollock (*Theragra chalcogramma*) and Pacific whiting (*Merluccius productus*). *ICES J. Mar. Sci.* 53 (2), 253–258.

Urmy, S. S., De Robertis, A., and Bassett, C. (2023). A Bayesian inverse approach to identify and quantify organisms from fisheries acoustic data. *ICES J. Mar. Sci.* fsad102. doi: 10.1093/icesjms/fsad102

Verfuss, U. K., Aniceto, A. S., Harris, D. V., Gillespie, D., Fielding, S., Jiménez, G., et al. (2019). A review of unmanned vehicles for the detection and monitoring of marine fauna. *Mar. pollut. Bull.* 140, 17–29. doi: 10.1016/j.marpolbul.2019.01.009

Wallace, J. R., and West, C. W. (2006). Measurements of distance fished during the trawl retrieval period. *Fisheries Res.* 77, 285–292. doi: 10.1016/j.fishres.2005.10.014

Wills, S. M., Cronin, M. F., and Zhang, D. (2021). Cold pools observed by uncrewed surface vehicles in the central and eastern tropical pacific. *Geophysical Res. Lett.* 48, e2021GL093373. doi: 10.1029/2021GL093373

Zhang, C., Levine, A. F., Wang, M., Gentemann, C., Mordy, C. W., Cokelet, E. D., et al. (2022). Evaluation of surface conditions from operational forecasts using *in situ* saildrone observations in the pacific arctic. *Monthly Weather Rev.* 150, 1437–1455. doi: 10.1175/MWR-D-20-0379.1

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 05 January 2024
DOI 10.3389/fmars.2023.1199076

Check for updates

OPEN ACCESS

EDITED BY
Kyle Shertzer,
NOAA Southeast Fisheries Science Center, United States

REVIEWED BY
Marcel Reichert,
Retired, Charleston, SC, United States
Jerald Ault,
University of Miami, United States

*CORRESPONDENCE
Meagan N. Schrandt
Meagan.Schrandt@MyFWC.com

RECEIVED 02 April 2023
ACCEPTED 21 August 2023
PUBLISHED 05 January 2024

CITATION
Schrandt MN, MacDonald TC, Winner BL, Blewett DA, Paperno R, Onorato GV, Gandy DA and Switzer TS (2024) Sampling design modifications to a fishery-independent monitoring survey balance the maintenance of long-term data with emerging management needs and funding limitations.
*Front. Mar. Sci.* 10:1199076.
doi: 10.3389/fmars.2023.1199076



# Sampling design modifications to a fishery-independent monitoring survey balance the maintenance of long-term data with emerging management needs and funding limitations

Meagan N. Schrandt[1]*, Timothy C. MacDonald[2], Brent L. Winner[1], David A. Blewett[3], Richard Paperno[4], Gregory V. Onorato[1], David A. Gandy[5] and Theodore S. Switzer[1]

[1]Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, St. Petersburg, FL, United States, [2]Nashotah House Theological Seminary, Nashotah, WI, United States, [3]Charlotte Harbor Field Laboratory, Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, Port Charlotte, FL, United States, [4]Indian River Field Laboratory, Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, Melbourne, FL, United States, [5]Apalachicola Bay Field Laboratory, Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, East Point, FL, United States

Development of Florida's marine Fisheries-Independent Monitoring (FIM) program began in 1985, and it initiated long-term monitoring in Tampa Bay in 1989 with the purpose of providing timely fishery-independent data and analyses to fishery managers for the conservation and protection of the species that support Florida's fisheries. Over time, the program expanded sampling in other Florida estuaries. Data from this monitoring program are integral to the assessment and management of numerous state and federally managed fishes, so long-term consistency is of paramount importance, but sampling design modifications have been necessary over time. This review presents three case studies in which the estuarine component of the FIM program was changed to address emerging data needs in management of Florida's fishery resources, while maintaining standardization and consistency with long-term surveys statewide. In the first case study, survey changes increased the amount of data on juvenile common snook, *Centropomus undecimalis*, on the Gulf of Mexico and Atlantic coasts and improved indices of abundance, survival estimates, and age-length keys for state management purposes. In the second case study, the FIM program improved estimates of abundance of juvenile reef fishes by initiating a complementary survey that expanded FIM sampling to new regions of the Florida Gulf coast and targeted a unique habitat (polyhaline seagrass beds along estuarine shoal habitats) not previously sampled in the standard long-term survey. In the third case study, the FIM program addressed a data limitation for regional management of spotted seatrout, *Cynoscion nebulosus*, by initiating seasonal sampling in western Florida. In each case study, the standard long-term survey design was modified to include sampling of new and unique areas and habitats, providing valuable data on estuarine fish assemblages that support analyses at the broader, ecosystem level. Survey amendments were designed to maintain standardization and consistency, all with limited additional funding. The success of these survey modifications was the result of several key factors: mission-focused programmatic goals, geographically dispersed laboratories, standardized protocols, ongoing critical analysis of the data, grant award success, and high-level data management. Although each case study originated with a survey expansion, all were followed by survey reductions or streamlining, so expansion and reduction scenarios are presented in this review. Regardless of these modifications, the mission of the FIM program remains the same: to provide timely data and analysis for the use of fishery managers, including state and federal partners.

KEYWORDS

***Centropomus undecimalis*, *Cynoscion nebulosus*, data management, fishery management, funding limitations, long-term monitoring, sampling design**

## 1 Introduction

Saltwater fishing in Florida employed almost 90,000 people and had an economic impact of almost $13 billion in 2020 (https://myfwc.com/conservation/value/saltwater-fishing/). Appropriate and timely management is critical to sustaining not only the saltwater fishery but also the economy that prospers from it. Researchers and managers have traditionally monitored changes in fish stocks with catch and effort data derived from commercial and recreational fisheries. Analysis of these fishery-dependent data provide valuable information on the status of fish stocks, but changes, including changes in vessel types, fleet size, fishing gear, or methods of operation, can make fishery-dependent data difficult to interpret (Ultang, 1977). Management actions (e.g., closed seasons, changes in size or bag limits, fluctuations in market values) and changes in fishing behavior further bias the utility of fishery-dependent catch data in assessing trends in abundance through time (Bryan and McCarthy, 2015; Smith et al., 2015; SEDAR, 2018). Long-term fishery-independent monitoring, which targets juvenile and subadult fishes that have not been subjected to fishing pressure, can provide less biased estimates of trends in fish stocks than fishery-dependent sampling (Myers and Cadigan, 1993). Changes in juvenile abundance within a season can be attributed to natural mortality, immigration, emigration, or recruitment. Shifts in juvenile abundance can also be used to forecast changes in the adult stock, allowing implementation of necessary modifications to harvest regulations before the fish have fully recruited to the fishery (Goodyear, 1985). Multispecies, multihabitat, long-term monitoring programs are also valuable in documenting ecosystem changes, evaluating the effects of natural and anthropogenic disturbances, and making management decisions (e.g., Coull, 1985; Wolfe et al., 1987; Stevens et al., 2016; Schrandt and MacDonald, 2020; Schrandt et al., 2021a).

In 1985, staff from what would become the Florida Fish and Wildlife Conservation Commission's (FWC) Fish and Wildlife Research Institute (FWRI) began planning for the marine Fisheries-Independent Monitoring (FIM) program after suspecting declines in the population of an important recreational and commercially harvested species, red drum, *Sciaenops ocellatus*, in Gulf of Mexico estuarine and offshore waters. Originally, the idea was to focus the program on the juvenile life stage of red drum. However, early leaders in the program recognized the importance of ecological surveys and sampling was instead designed to assess multispecies population trends in multiple estuarine habitats across a range of life history stages. From its inception until 1996, the program underwent extensive gear and procedural developments to refine the sampling design. During that period, many gear types (e.g., drop nets, entangling nets, purse seines, beach seines, haul seines, roller frame, and otter trawls) and sampling approaches (i.e., fixed-station vs. directed sampling vs. stratified-random, day vs. night, seasonal vs. monthly sampling) were assessed and kept, modified, or discarded. In 1996, the program finalized its main sampling-gear types (21.3-m center-bag seines, 183-m haul seines, 6.1-m otter trawls) and sampling design (stratified-random, monthly, and daytime), which have remained consistent through time and geographic expansion. The final sampling design was reviewed and approved that year by a team of outside scientific experts in fishery management from the American Fisheries Society (https://fisheries.org/).

Dedicated funding for the FIM program began with a $400,000 special appropriation by the Florida Legislature in 1986 and a recurring, small ($260,000) federal Sportfish Restoration grant in 1987. In 1988, additional funding became available from a second, recurring special appropriation through the Florida Legislature. Since 1995, the program has been largely funded through the sale of Florida's saltwater fishing licenses and by competitive grant awards (e.g., National Estuary Program, National Park Service, Florida water management districts, U.S. Army Corps of Engineers) that have allowed it to address knowledge gaps related to habitats (e.g., inshore reefs, polyhaline seagrasses, tidal tributaries), management actions (e.g., water withdrawals, dredging and filling), stock boundaries (Northwest Florida), and geographic locations (e.g., Sarasota Bay, Everglades National Park, Florida Bay). With additional funding from sales of the Florida saltwater fishing license, the program rapidly expanded into geographically dispersed estuaries (Figure 1). Sampling began in Tampa Bay (TB) and Charlotte Harbor (CH; 1989), then expanded to the northern Indian River Lagoon (NIRL; 1990), Cedar Key (CK; 1996), the southern IRL (SIRL; 1997), Apalachicola Bay (AP; 1998), and Northeast Florida (JX; 2001). The FIM program also expanded to offshore habitats in 2008 to address key limitations in available data on groundfish assemblages (Matheson et al., 2017; Christiansen et al., 2022b; see Pollack et al. (submitted) in this issue for trawling surveys) as well as reef fish assemblages and their habitats (Keenan et al., 2022; Thompson et al., 2022; Switzer et al., 2023).

Although the FIM program has successfully expanded (e.g., geography, habitats) and continues to provide timely, accurate data integral to the assessment and management of fisheries, there have been challenges and some facets of the program had to be scaled back to free up resources for necessary expansions. Like many long-term monitoring programs, it has struggled with long-term funding and has had to respond to deficits due to stagnant funding coupled with inflation. Thus, the program relies on grants, most of which are short-term and competitive, for approximately half of its funding. The FIM program, therefore, has had to address specific grant objectives while maintaining broad programmatic consistency. Additionally, the program has had to address maintaining statewide standardization, new and changing technologies, evolving management needs, and staffing shortages and retention issues. Regardless of the challenges, the program has maintained its original mission of providing timely data and analysis for fishery managers, including state, federal, and nongovernment partners. Given these challenges, we will discuss three case studies, each aiming to improve size-structured abundance estimates and other data for various species, from the estuarine component of the FIM program that exemplify processes used to inform sampling survey modifications, expansions, and contractions to meet emerging fishery management needs for the state of Florida. The FIM program's mission-focused programmatic goals, geographically dispersed field laboratory model, standardized protocols, continual analysis of data, successful grant award record, and stable, yet adaptable, database design, have been significant features in the program's ability to adapt to changing fishery and ecological priorities and budgetary deficiencies to maintain long-term data collections.

## 2 Standard FIM sampling design

The three case studies highlighted herein leveraged the standard FIM sampling protocols, database design, and logistics, ensuring data comparability with the long-term FIM program while reducing start up and continuation costs by 50-70%. The standard FIM design samples 7 estuaries, Apalachicola Bay (AP), Cedar Key (CK), Tampa Bay (TB), and Charlotte Harbor (CH) on the Gulf coast, and northeast Florida (JX), northern Indian River Lagoon (NIRL), and southern Indian River Lagoon (SIRL) (Figure 1) on the Atlantic coast, with 3 gear types. A center-bag seine (21.3-m x 2.0-m with 3.2-mm mesh) is deployed to sample in both bay (~140 $m^2$/set) and river (~68 $m^2$/set) habitats to a water depth of 1.8-m. A haul seine (183-m x 3.0-m with 38-mm mesh) is set in bay habitats (~4120 $m^2$/set) in depths ≤2.5 m. An otter trawl (6.1-m with a 3.2-mm mesh

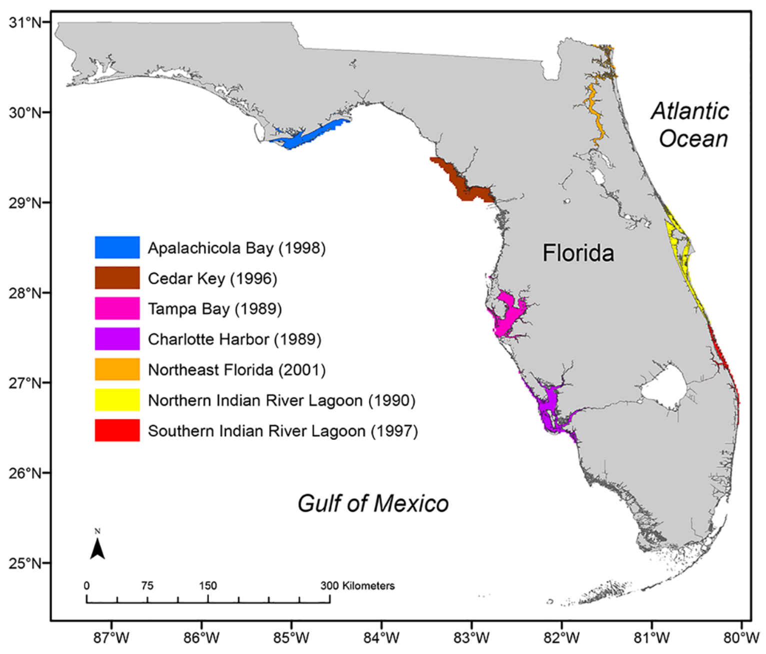



FIGURE 1
The six Florida estuaries with long-term Fisheries-Independent Monitoring (FIM) surveys through 2023. The Indian River Lagoon is one estuarine system but is sampled separately for logistics, so it is depicted as two colors (yellow and red). The first year of long-term monitoring for each estuary is noted in parentheses.

bag liner) is deployed to sample bay (~1440 $m^2$/set) and river (~720 $m^2$/set) habitats between 1.0 and 7.6 m deep. The sampling universe within each estuary generally encompasses waters from the mouth of the estuary through the tidally influenced portion of major rivers. The FIM program is a multispecies survey to assess the diverse number of economically important species in Florida waters and, therefore, the design is not optimized around a specific metric. Strata are defined based on space (approximately homogenous biological and water quality 'zones' within each estuary) and habitat (i.e., presence or absence of submerged aquatic vegetation, and presence or absence of overhanging shoreline vegetation like mangroves). Early data analyses suggested species assemblages differed among these habitat strata.

The FIM sampling design is stratified-random, has a monthly periodicity, and samples are collected during daylight hours. Monthly sampling effort in each stratum for each estuary is proportional to the total number of potential sampling sites available but weighted toward important habitats like vegetated habitats. The total number of samples collected each month is ultimately a function of funding. The primary objective is to monitor annual, not monthly, trends so statistical power to detect annual trends is greater than power to detect monthly trends. Site-selection for the standard FIM stratified-random sampling design is cell-based. The sampling universe within each estuary is divided into cells (185 m × 185 m), with each cell assigned binary values for whether each combination of gear and strata is likely to be available for sampling. Cells are then selected in a stratified random manner, without replacement, for sampling each month.

# 3 Case study 1: Juvenile common snook

## 3.1 Rationale

Common snook, *Centropomus undecimalis* (hereafter referred to as snook), are one of Florida's most sought-after game fishes for inshore anglers. While commercial harvest of snook in Florida has been illegal since 1957, recreational fishing is not, and pressure and landings in the recreational fishery have increased dramatically in recent decades (Munyandorero et al., 2020). Snook populations in Florida have been subject to several stressors, including extreme cold-weather events, degradation of habitat, and northward expansion of the Gulf coast population, warranting diligent management and stringent regulations to maintain populations and a sustainable inshore fishery statewide (Stevens et al., 2016; Adams et al., 2019; Purtlebaugh et al., 2020). Since 1996 the FIM program has monitored snook abundance for fishery management using the 21.3-m center-bag seine to collect small juvenile snook (<100 mm SL, standard length) and the 183-m haul seine to collect larger snook (>300 mm SL). Data from the FIM program were previously used in FWC stock assessments and updates (e.g., 2012, 2013, 2015) but stock-assessment analysts desired better estimates of juvenile snook (100–300 mm SL) abundance, survival, and growth for age-based models used in the assessment. In early 2013, the program was tasked with developing a sampling design to improve collections of juvenile snook that would refine Florida's snook stock assessment. Challenges in addressing this need included the species' protracted spawning period, its highly variable growth rate, and its geographic distribution and identified stock differences (Taylor et al., 2000; Trotter et al., 2012; Young et al., 2014). Therefore, the main objective for this modification to the statewide program was to provide improved data (abundance, age, and distribution) on juvenile snook for stock assessments, while complementing and enhancing existing fishery monitoring efforts.

## 3.2 Evaluation, reconnaissance, and gear testing

The objective to provide improved data on juvenile snook was addressed in three phases—evaluation, reconnaissance, and gear testing—before the modified sampling design was implemented, in 2016.

*Evaluation phase.*—In the evaluation phase, pertinent scientific articles on juvenile snook (i.e., life history, ecology, distribution) were reviewed, and existing FIM survey data (from short- and long-term studies) were analyzed and summarized to identify gaps in the data already being provided for management purposes. Before selecting gear types, the age and approximate size range were first defined for snook that were already being collected. Snook of >300 mm SL were well represented in the haul seine collections and those <100-mm SL were well represented in the center-bag seine collections, but those of 100–300 mm SL were not as well represented in either gear, especially in the southern estuaries (Charlotte Harbor and Southern Indian River Lagoon). Combining FIM data with those from other FWRI surveys, it was then determined that the general size range of age-0, age-1, and age-2+ snook was approximately <100 mm, 100–300 mm, and >300 mm SL, respectively, but multiple ages were present within the age-1 and age-2+ size ranges (Figure 2). This information was used to design a gear-testing study in which the results would be used in modifying the long-term survey design. Gear-testing included a range of gear types that could effectively sample the range of sizes of juvenile snook.

The evaluation phase also indicated that, while snook occur almost statewide, the distribution closely approximates that of mangroves (Marshall, 1958), i.e., they are primarily distributed along Florida's southern coastlines, from Cape Canaveral on the Atlantic coast to Tarpon Springs on the Gulf coast (Gilmore et al., 1983; Rivas, 1986; Winner et al., 2010; Stevens et al., 2016; Purtlebaugh et al., 2020). Therefore, gear-testing focused on TB and CH on the Gulf coast, and the IRL on the Atlantic coast, regions in which the FIM program already had an existing long-term monitoring survey. Snook spawning has been documented from April through December in Florida waters, and FIM data indicate that juvenile snook are present year-round in the sampled estuaries (Winner et al., 2010), so gear could be tested at any time of year. Snook geospatial distributions were also evaluated to determine which areas were used by juvenile snook in each estuary. Tidal tributaries (i.e., tidal creeks and rivers) represent primary nursery

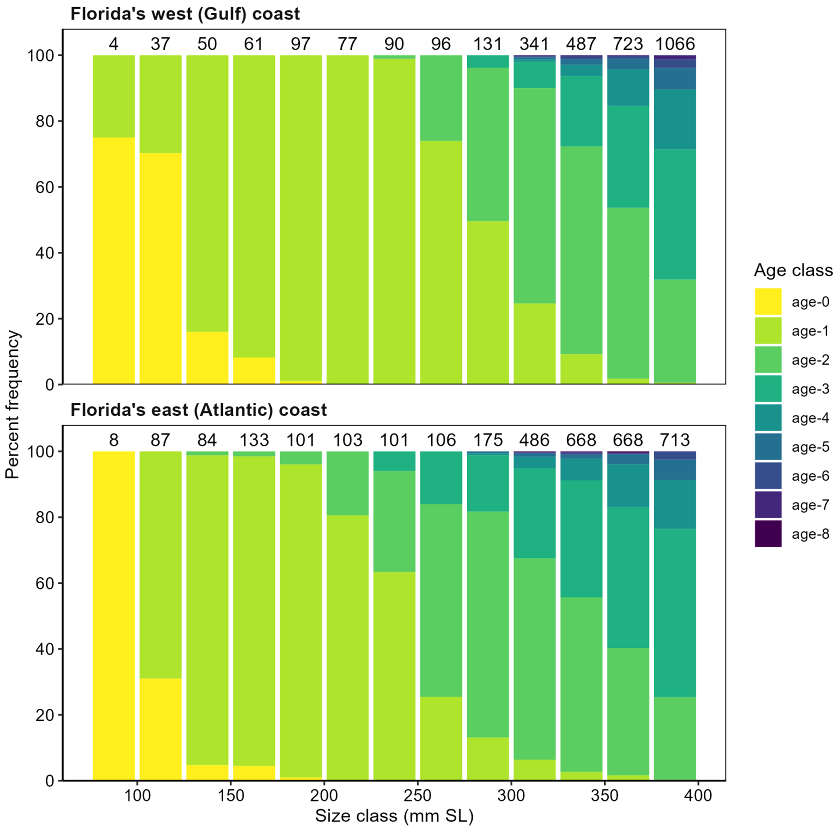



FIGURE 2
Percent frequency of size at age for juvenile common snook, *Centropomus undecimalis*, from all FWC-FWRI studies (i.e., life history 1986–1996, 183-m haul seine 1996–2012, electrofishing 2011–2012, 21.3-m and 40-m seines 2014–2015) on the Gulf and Atlantic coasts of Florida. The number of individuals aged in each size class is listed above each bar. The horizontal axis has been truncated at 400 mm SL to emphasize the sizes-at-age for juvenile snook; the entire database contains fish to 1200 mm SL.

areas for snook and other species in many of the estuarine systems in Central and South Florida (Peters et al., 1998; Wilson et al., 2022), and their proximity to snook spawning grounds (i.e., ocean inlets and passes) influences juvenile recruitment into tidal tributaries (Wilson et al., 2022). Main-stem and backwater tidal tributaries (i.e., smaller tributaries, embayments, and coastal ponds) bordered by mangrove and salt marsh provide juveniles ample forage and protection from predators (McMichael et al., 1989; Brame et al., 2014).

*Reconnaissance phase.*—Using information from the evaluation phase, we conducted reconnaissance trips using various gear types and sampling a variety of regions and habitat types. Reconnaissance was crucial for establishing the sampling universe and gear types for the final, gear-testing, phase. Haul seines of different lengths (9–183 m) and mesh sizes (3.2–38.1 mm stretch mesh) were evaluated during these reconnaissance trips. Sampling extended into geospatial areas not included in the standard FIM long-term survey design, including the bays and tidal tributaries of TB, CH, and the IRL (Figure 3). Reconnaissance trips identified juvenile snook nursery areas in waterbodies with small geospatial footprints (backwater areas and tidal tributaries); these areas, although included in the FIM universe, were undersampled because of their small geospatial extent in comparison to the rest of the sampling universe. Additionally, nursery areas were identified in two riverine systems (St. Lucie and Loxahatchee rivers) of the SIRL that had not been included in the standard FIM design.

*Gear-testing phase.*—Based on findings from the evaluation and reconnaissance phases, two seine types (21.3-m and 40-m) were selected for the gear-testing phase. The 21.3-m center-bag seine has been used in the standard FIM survey since 1989 and provides the majority of FIM data on juvenile snook. The 40-m center-bag seine (with 25-mm stretch mesh) was an experimental seine configuration that covered more shoreline per sample and had a larger mesh size (i.e., different size selectivity) than the 21.3-m seine. The seines were tested from July 2014 through June 2015 in all three estuarine systems (TB, CH, IRL). A sampling universe that included these backwater embayments, tidal tributaries, and the two riverine systems in the SIRL (Figure 3) was created for each estuarine system. Sampling sites were randomly selected, and effort was stratified by waterbody type (main stem, backwater, or tidal tributary/creek). Both seine types were set at each randomly selected sampling site. Each seine was deployed from the stern of the boat, arched out along the shoreline, with the center bag falling 5–6 m from the shore. All fish and selected invertebrate species captured were identified to the lowest possible taxonomic level, counted, and a random sample of at least 10 individuals were measured.

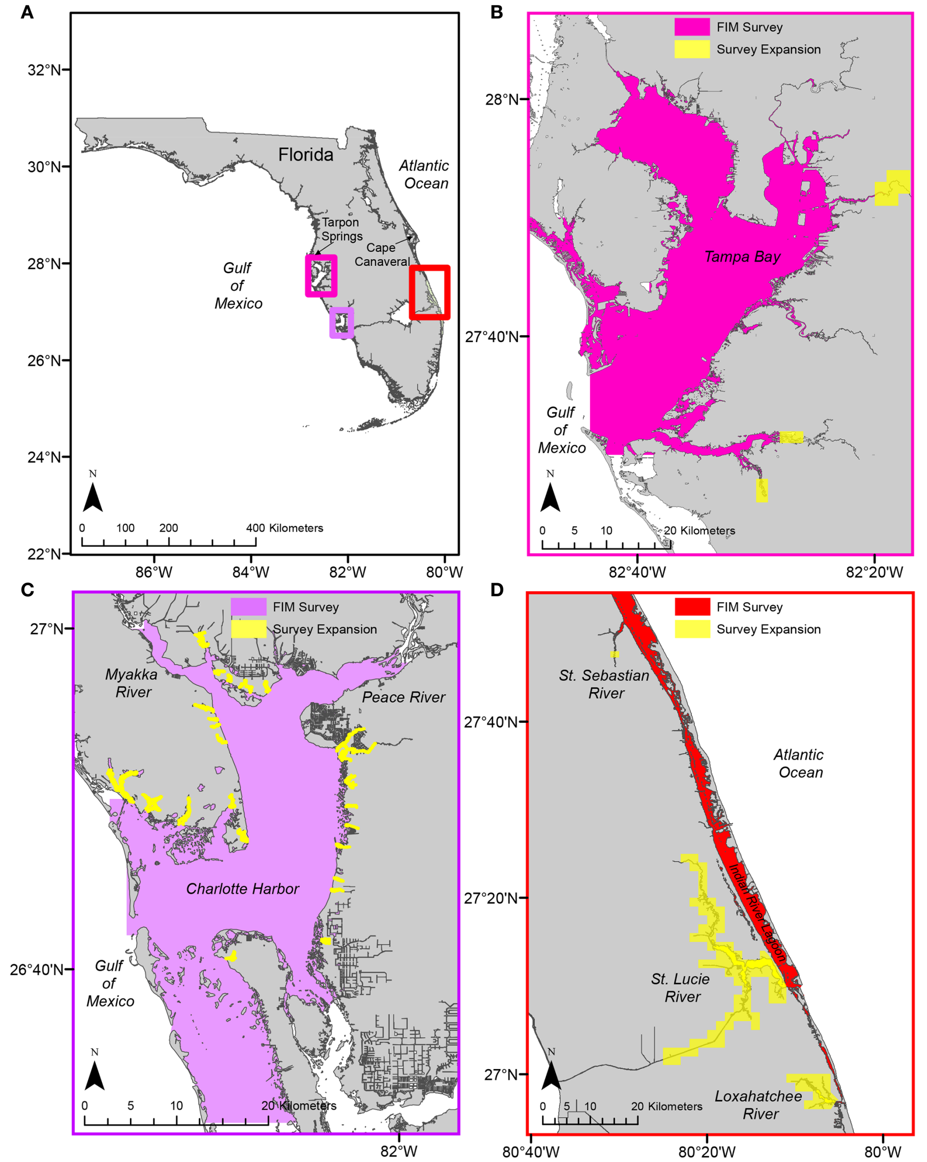



FIGURE 3
The Fisheries-Independent Monitoring (FIM) program modified its long-term monitoring survey in three estuaries, represented by the small rectangles in **(A)** and individual panels in **(B–D)**, to increase data collection of juvenile common snook, *Centropomus undecimalis*. Survey expansion efforts focused on backwater and tidal tributaries of rivers in Tampa Bay **(B)**, tidal tributaries in Charlotte Harbor **(C)**, and three rivers in the Indian River Lagoon **(D)**.

During the gear-testing phase, 1,872 seine samples were collected from the three estuaries, capturing 2,315 juvenile snook: 1,285 age-0, 714 age-1, and 316 age-2+. In all three estuaries, catch-per-unit-effort of age-1 and age-2 snook was similar between the 21.3-m and 40-m seines (Figures 4, 5). The 21.3-m seine collected age-0 snook, which were largely absent from the 40-m seine collections, supplementing and improving age-0 indices of abundance available for stock assessments. In addition, the 21.3-m seine accounted for more than 95% of the overall catch and had a greater taxonomic diversity of fishes than did the 40-m seine (TB: 21.3-m seine, 93 taxa; 40-m seine, 50 taxa; CH: 21.3-m, 70 taxa; 40-m, 43 taxa; IRL: 21.3-m, 163 taxa; 40-m, 94 taxa). These findings indicated that the 21.3-m seine could provide data for juvenile snook, while improving data available for age-0 snook and for a variety of other estuarine fish species.

The gear-testing phase also confirmed juvenile snook preference for these areas with small geospatial footprints (backwater and tidal tributaries). Within rivers, both age-0 and juvenile snook used similar areas; they were both collected at higher abundance in backwater riverine areas and smaller tidal tributaries than in main-stem areas. Juvenile snook were collected with the 21.3-m seine during all months, confirming the efficacy of monitoring the youngest life-history stages of this species with a year-round monthly sampling design. Collectively, results from the evaluation, reconnaissance, and gear-testing phases indicated that a geographic expansion of sampling with the 21.3-m seine into

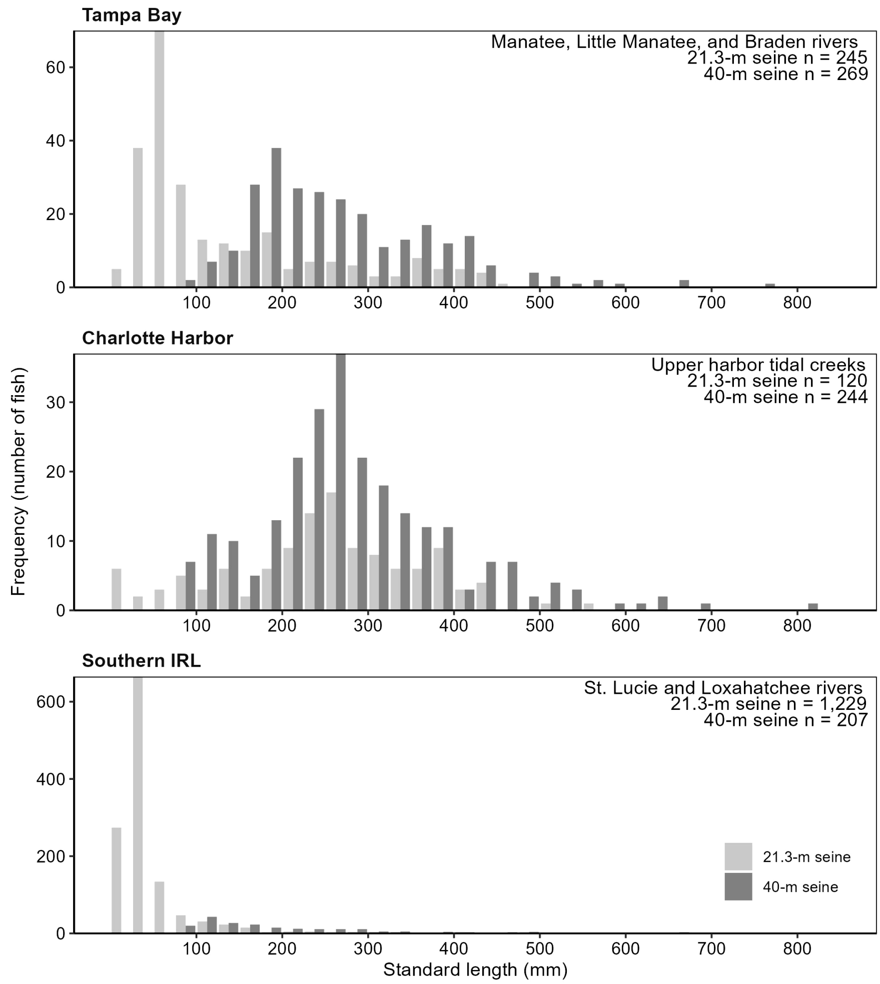



FIGURE 4
Length frequencies of common snook, *Centropomus undecimalis*, by estuary and gear type, collected during Florida's Fisheries-Independent Monitoring (FIM) gear testing, 2014–2015.

backwater riverine areas and smaller tidal tributaries would provide the necessary juvenile snook data, and improve the age-0 snook data that stock assessment needed.

## 3.3 Implementation of the snook survey

In January 2016 FIM modified the design of its long-term monitoring survey to facilitate collection of data that better address the needs of juvenile snook as identified by stock assessment analysts. Additional monthly sampling effort with the 21.3-m seine was allocated in each of the three estuarine systems, establishing these previously undersampled water bodies with small geospatial footprints (i.e., backwater and tidal tributaries; Figure 3) as a geospatial stratum within the FIM sampling universe. Sampling effort was apportioned among the estuarine systems based on results from the gear-testing phase and the data already provided by the standard long-term FIM survey design. In Tampa Bay, data on juvenile snook were provided from the existing river sampling, but to refine these data and supply adequate biological samples the sampling effort was increased in backwater and smaller tidal tributaries by eight seine hauls per month. In Charlotte Harbor, the existing tidal tributary sampling in the standard FIM program had not been providing sufficient data, so sampling effort was increased by 30 seine hauls per month, distributed among 27 tidal creeks that had been undersampled in the standard FIM design. In the Indian River Lagoon, existing sampling was not providing sufficient data, so sampling effort was increased by 24 seine hauls per month, which increased sampling in the St. Sebastian River and added two river systems (St. Lucie and Loxahatchee rivers) that had not been included in the standard FIM design. To improve age-and-growth estimates for juvenile snook used in state stock assessments, randomization procedures for retaining snook ≥100 mm SL for biological sample collection were also established with this modified survey design.

## 3.4 Implications

Stock assessment analysts' desire for juvenile snook abundance and age data were met by expanding sampling with an existing gear type (21.3-m seine) into undersampled waterbodies that had very

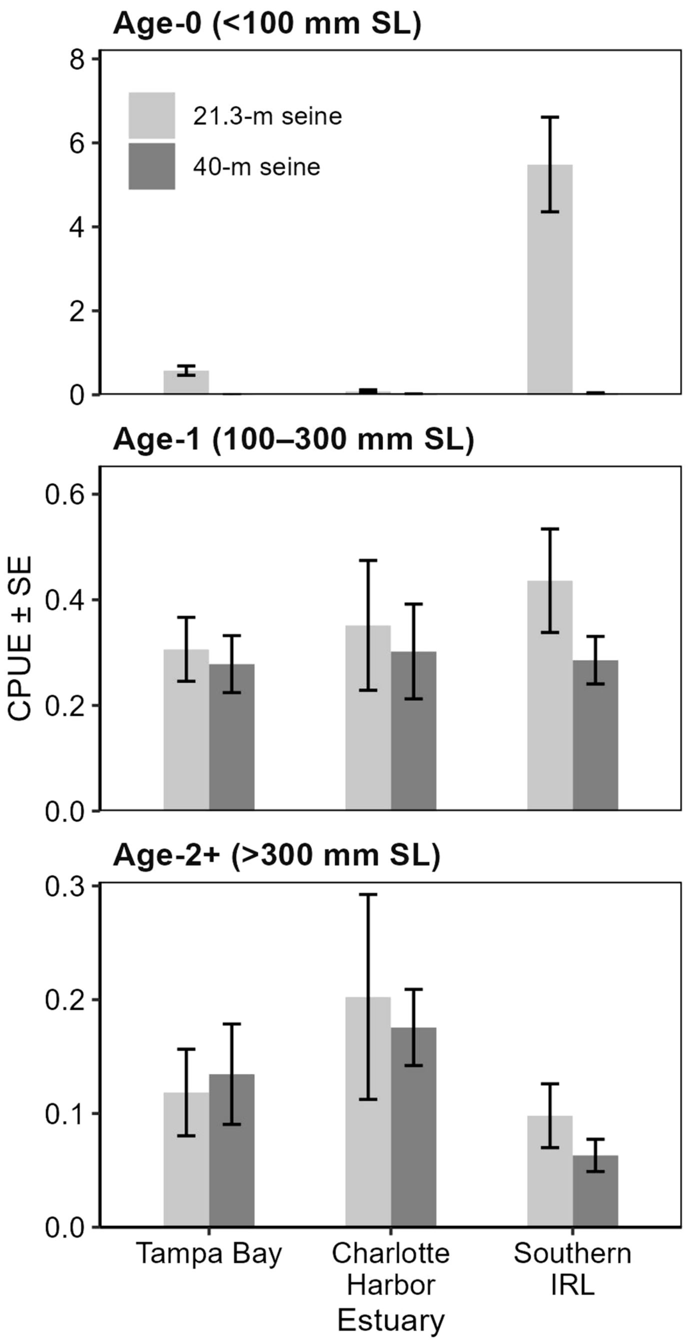



FIGURE 5
Mean catch-per-unit-effort (CPUE; fish per 100 $m^2$) ± SE of common snook, *Centropomus undecimalis*, by estuary and gear type, collected during Florida's Fisheries-Independent Monitoring (FIM) gear testing, 2014–2015. Approximate ages were assigned based on snook length.

small geospatial extents. Data from the reconnaissance and gear-testing phases of the present study were immediately used in snook stock assessments (Munyandorero et al., 2020). The modified sampling design also improved indices of abundance of age-0 snook and strengthened snook age–length keys with the collection of additional individuals for age-and-growth analyses. Furthermore, valuable fish community data, including other managed species, have been collected under the modified design. In addition to supporting stock-assessment needs, these data are being consulted for restoration and conservation efforts in Florida (Stevens et al., 2022; Wilson et al., 2022).

The geographically inclusive scope of the FIM program was instrumental to the success of this amended sampling design. And without the earlier data from each of the estuaries, developing and implementing the snook survey modification would have been cost-prohibitive. The two rivers added into the IRL are outside the standard FIM survey design; without the earlier survey in the IRL and the existence of a field laboratory adjacent to these rivers, they could not have been included in the amended sampling design. The annual sample size (n=744) for this survey is greater than the number of collections made annually at the average FIM field laboratory. The funding for the development, deployment, and maintenance of this survey was provided by a continuing grant from Florida's Snook Stamp program. Because of the FIM program's extensive network of field laboratories, existing data, standardized procedures, database, and infrastructure, this survey was implemented for less than 1/3 of the funding that would have been necessary to establish and maintain a new field laboratory.

### 3.5 Modifying the snook survey

After five years (2016–2020) of level funding that did not account for increased operational costs, the FIM program revisited the juvenile snook survey design in 2020. The overall goal was to examine spatial and temporal data in hopes of identifying ways in which sampling might be reduced without compromising critical management data needs, thereby reducing costs. Data analyses indicated that spatial sampling reductions could not be implemented without affecting data quality and the long-term data sets.

Survey data were also examined for monthly trends in each of the estuarine systems to identify months during which abundance of juvenile snook was low, such that effort might be reduced or eliminated. Abundance trends of other economically important species (e.g., red drum and spotted seatrout, *Cynoscion nebulosus*) were also examined to ensure that any reduction of effort would not compromise data on those species. Four sampling scenarios were examined to assess possible effects on survey data: 6-month (Sep–Feb), 7-month (Sep–Mar), 8-month (Aug–Mar), and 9-month (no sampling during March, May, or July, which results in bimonthly sampling during months when snook abundance is low in 21.3-m seines). The 9-month scenario minimized negative effects on the long-term data while providing an adequate reduction (25%, 333 annual net hauls) in sampling effort and was implemented in 2021. The 9-month scenario encompassed the traditional recruitment window for snook and covered recruitment windows for red drum and spotted seatrout.

## 4 Case study 2: Juvenile estuarine-dependent reef fish species

### 4.1 Rationale

Reef fishes in the Gulf of Mexico and western Atlantic Ocean support multibillion-dollar recreational and commercial fisheries. Many of these species are estuarine-dependent (e.g., Koenig and Coleman, 1998; Nagelkerken et al., 2001; Nagelkerken et al., 2002; review by Gillanders et al., 2003; Casey et al., 2007; Switzer et al., 2012; Lefcheck et al., 2019), juveniles occupying estuarine habitats, and mature fish occupying offshore reef habitats. Managing these fisheries is complex, and management is more effective when indices of juvenile abundance are available for use in predicting recruitment to the fishery. Predicting the strength of recruitment to the fishery, in turn, allows managers to better assess outcomes from management actions (Hansen et al., 2015). Under its standard long-term sampling design, the FIM program captured highly variable numbers of juveniles of estuarine-dependent reef species, resulting in highly variable indices of juvenile abundance (Switzer et al., 2012; Flaherty-Walia et al., 2015; Switzer et al., 2015). Therefore, the standard long-term estuarine survey undersampled a preferred juvenile habitat: polyhaline (salinity >18) seagrass beds, and researchers and managers recommended improving indices of juvenile abundance for reef species (Switzer et al., 2012; Flaherty-Walia et al., 2015). The FIM program needed a survey that better sampled juvenile (estuarine-dependent) reef fishes to carry out its mission of providing timely and accurate data for fishery management. Toward that end, results from existing FIM data for the eastern Gulf of Mexico were used to inform the design of a complementary FIM survey aiming to improve the ability to characterize the abundance of juvenile gag, *Mycteroperca microlepis*, (Casey et al., 2007; Switzer et al., 2015) and other estuarine-dependent and seagrass-associated reef fishes by targeting the preferred polyhaline seagrass habitat. Work in Charlotte Harbor was instrumental in developing the survey, as Casey et al. (2007) documented that juvenile gag were collected mainly between April and December in habitats with ≥50% seagrass cover. Relative abundance of gag was also about 2.9 times as great on shoals as that near mangrove and beach shorelines (Casey et al., 2007). This information was used, in part, in the successful proposal for a much larger award for monitoring reef fishes along Florida's Gulf coast. The FIM program in 2008 initiated this complementary survey (hereafter, the West Florida Shelf Inshore [WI] survey) to extend its standard, long-term monitoring survey to deep seagrass habitats found in estuaries already sampled by the FIM program and in adjacent estuaries not sampled by the FIM program (Figure 6). The main objective of the WI survey was to provide data to improve estimates of abundance of juvenile estuarine-dependent reef fish in the eastern Gulf of Mexico to inform federal reef fish assessments and management decisions.

### 4.2 Implementing the complementary WI survey

Implementing the WI survey in 2008 did not change the standard long-term estuarine monitoring survey, but rather complemented it by expanding the sampling universe to include habitats that had been undersampled in all estuaries. Specifically, the complementary WI survey targets polyhaline seagrass habitats within estuaries already sampled by the FIM program (AP, TB, CH) and in Saint Andrew (SA) and Big Bend (BB), estuaries adjacent to AP and CK, respectively. Within these estuaries, the WI sampling universe was defined by sampling cells (185 m × 185 m) with polyhaline seagrass habitats. Cells were then stratified by space, and sites were randomly selected. Within estuaries already sampled by the FIM program (AP, TB, CH), the WI universe covers ~224 sq km (ranging from 5–10% of the standard FIM estuarine universe) that were previously undersampled by the standard long-term survey. In

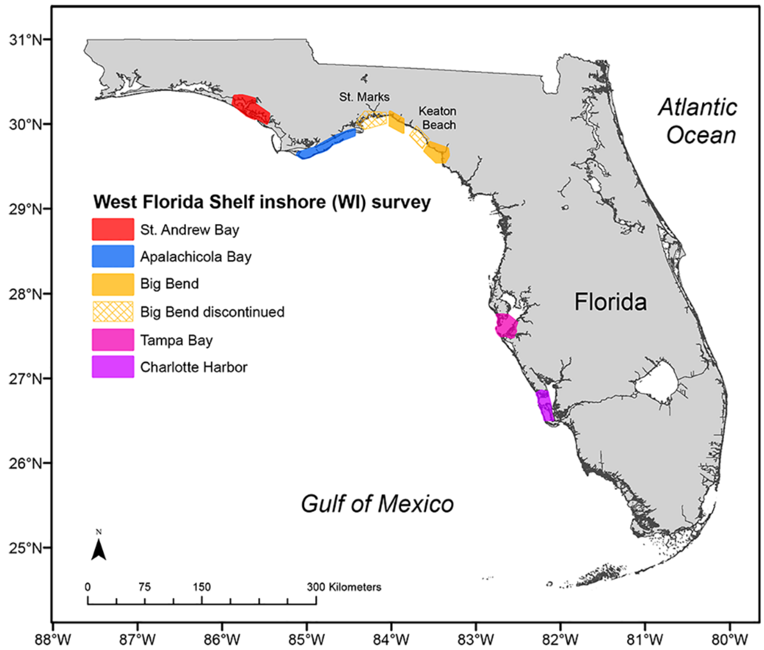



FIGURE 6
Five estuaries sampled for the West Florida Shelf Inshore (WI) survey to collect data for estimating juvenile reef fish abundances. The survey began in 2008 in all color-coded geographic areas; the two areas with hashed fill are Keaton Beach and St. Marks, where sampling for the WI survey was discontinued after 2008 and 2018, respectively.

the new, adjacent estuaries (SA, BB), the WI universe is ~221 square kilometers. Implementing the WI survey also required the development of new sampling protocols. The WI survey initially used two types of standard inshore FIM sampling gear: the 6.1-m otter trawl and the 183-m haul seine. The 6.1-m otter trawl procedures for WI sampling follow standard FIM otter trawl protocols with two important modifications. First, the trawl must sample a path with ≥50% cover of submerged aquatic vegetation (confirmed before gear deployment). Second, the trawl is towed for half the distance (0.1 nautical mile at 1.2 kts; ca. 5-min tow) of a standard FIM bay trawl (0.2 nautical miles) to effectively sample smaller seagrass beds, reduce algae bycatch, and ensure standardized trawl samples in the WI survey. Additional details regarding 6.1-m otter trawl sampling in the WI survey can be found in Switzer et al. (2012); Flaherty-Walia et al. (2015) and Schrandt et al. (Schrandt et al., 2018; Schrandt et al., 2021b).

The 183-m haul seine deployments for the WI survey were also modified from the standard FIM inshore sampling survey. Under standard FIM protocols, the 183-m haul seines are deployed by boat and set in a rectangular shape along shorelines, where the seine's wings are pulled together along the water's edge or along a shore type that prevents reaching the water's edge (e.g., vegetation). Under WI sampling protocols, haul seines are deployed by boat in a rectangular shape along shallow shoals (in 0–1.0 m of water) usually >100 m from a persistent shoreline where the difference between the wing depth (depth at which the ends of the net were pulled together along the shoal, the shallowest portion of the deployment) and the bag depth (depth at the bag of the haul seine, the deepest portion of the deployment) is at least 0.5 m (De Angelo et al., 2014).

Other than these procedural changes, WI sampling followed standard FIM sampling procedures practiced statewide, wherein, for all, all fish and macroinvertebrates are identified and counted, providing as much continuity as possible between the complementary WI survey and the standard long-term survey. This also ensures that data are available for assessment of other taxa that are not estuarine-dependent reef fish and for ecosystem management-type assessments.

## 4.3 Evaluating and modifying the WI survey

After the WI survey was initiated in 2008, the survey underwent a series of assessments and amendments to improve efficiency, reduce variability stemming from sampling or observation error, and improve statistical power to detect changes in abundance over time. With the completion of the initial grant and loss of the original funding source, these assessments and subsequent modifications were paramount to balancing workloads with reduced funding and the critical data needs for stock assessments and fisheries managers. The first changes, in 2009 and 2016, focused on reducing variability in the abundance estimates of various reef species by discontinuing sampling 1) during the months in which reef fishes do not actively recruit to the estuaries and habitats of interest and 2) in areas in which reef fish are not recruiting to the seagrass beds in numbers large enough to provide statistically powerful indices of juvenile abundance. This meant changing the sampling months from May–November to June–November, corresponding to the approximate months of the seagrass growing season (Zieman and Zieman, 1989), and discontinuing sampling in an unproductive area (Keaton Beach) in the Big Bend (Table 1; Figure 6). In 2019, sampling was discontinued in another area of the Big Bend (St. Marks; Table 1; Figure 6) as a cost-saving measure and to reduce redundancy in the survey. Information for nearly all species of interest was being obtained from the other two Big Bend sampling areas, which had trends similar to those in St. Marks, suggesting some redundancy in the sampling design (unpublished data).

With even more reductions in funding for the work, critical assessments focused on further streamlining the survey while still being able to provide statistically powerful indices of abundance for stock assessments and managers. Therefore, all available WI data were used to compare catches between the two gear types and to conduct simulations to estimate the statistical power of each type of gear to detect changes in abundance for seven species (Schrandt et al., 2021b). The study concluded that the 6.1-m otter trawl was more efficient than the haul seine in collecting many of the reef species and that other data were similar between the two gear types. Furthermore, a modest increase in sample size of the otter trawl would achieve statistical power to track changes in abundance (Schrandt et al., 2021b). After much consideration, WI haul seine sampling was discontinued in all estuaries and the sample size for trawls was increased in 2019.

## 4.4 Implications

The FIM program's WI survey added a new habitat to a long-term survey design that addressed an evolving stock assessment need (i.e., less variable, more powerful indices of juvenile abundance); the habitat addition capitalized on a new grant funding award and complemented the long-term survey. Subsequent modifications to the WI survey design were necessary to meet stock assessment needs as funding sources changed and overall funding was reduced. The complementary WI survey was able to be seamlessly added to estuarine sampling efforts after habitat and location reconnaissance and gear-testing for a new deployment technique had been completed. The WI survey has expanded the estuaries and habitats sampled under the long-term estuarine survey. A comparison of fish communities sampled via haul seines in the long-term survey and those sampled in the WI survey documented differences in fish communities between the seagrass habitat along shorelines and the shoal seagrass habitat, away from the shorelines (De Angelo et al., 2014), indicating that the haul seines in the WI survey were providing information that was not being obtained with the FIM standard estuarine inshore survey. Furthermore, they noted that shoal habitats had greater densities of several estuarine-dependent reef fish species, like gag, gray snapper, *Lutjanus griseus*, and lane snapper, *L. synagris*.

In 2019, results from additional gear comparisons and power simulations (Schrandt et al., 2021b) led to discontinuing WI haul seine sampling, which was the largest and potentially most impactful change made to the survey. Although this change resulted in the end of a time series for haul seine data in polyhaline seagrass beds along Florida's Gulf coast, it resulted in the continuance and enhancement of the program's trawl time series. The routine analysis of data and the geographic extent of the long-term FIM program were critical to the implementation of this

**TABLE 1** Summary of annual 6.1-m otter trawl effort (number of net hauls) in the five estuaries sampled by the West Florida Shelf Inshore (WI) survey along the Gulf coast of Florida, USA.

| Year | St. Andrew Bay | Apalachicola Bay | Big Bend region | | | | Tampa Bay | Charlotte Harbor | Yearly total |
|---|---|---|---|---|---|---|---|---|---|
| | | | St. Marks | Econfina | Keaton Beach | Steinhatchee | | | |
| 2008 | 42 | 56 | 70 | 70 | 68 | 68 | 90 | 70 | 534 |
| 2009 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2010 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2011 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2012 | 39 | 56 | 70 | 70 | | 70 | 70 | 56 | 431 |
| 2013 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2014 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2015 | 42 | 56 | 70 | 70 | | 70 | 70 | 56 | 434 |
| 2016 | 36 | 48 | 60 | 60 | | 60 | 60 | 48 | 372 |
| 2017 | 36 | 48 | 60 | 60 | | 60 | 60 | 48 | 372 |
| 2018 | 30 | 48 | 60 | 60 | | 60 | 60 | 48 | 366 |
| 2019 | 72 | 96 | | 90 | | 90 | 144 | 120 | 612 |
| 2020 | | 48 | | 45 | | 45 | 72 | 60 | 270 |
| 2021 | 72 | 96 | | 90 | | 90 | 144 | 119 | 611 |

Sampling was intentionally reduced in 2009 and 2016. In 2019, trawl sampling effort was increased to improve statistical power of indices of juvenile reef fish abundance (see Schrandt et al., 2021b for details of the change in effort). In 2020, because of COVID-19, effort was reduced to bimonthly sampling in Apalachicola Bay, Tampa Bay, and Charlotte Harbor and suspended for St. Andrew Bay.

survey design. Without routine analysis of data from the long-term FIM program survey, supporting documentation for the awarded grant would not have existed. Additionally, the cost of the grant would have been excessive had the program not been able to leverage established surveys. All amendments were considered exhaustively, and all available data were used to inform the FIM program's decisions, with minimal data loss and prioritizing the continuity of new survey data with the earlier data. Ultimately, the decisions to focus the survey over space and time and to streamline to a single gear type allowed the FIM program to provide better indices of juvenile abundance. For example, the polyhaline seagrass survey has reduced variability and coefficients of variation for catch-per-unit-effort for gray snapper (Flaherty-Walia et al., 2015). The WI survey has also improved abundance estimates for young-of-the-year gag and reduced coefficients of variation for young-of-the-year gag when WI data were combined with data from the standard long-term estuarine survey (Switzer et al., 2015). Furthermore, 6.1-m otter trawl abundance estimates for seven reef fish species of interest were greater for the WI survey, which sampled deep polyhaline seagrass beds, than for the long-term monitoring survey, which had not sampled such habitat (Figure 7). Finally, the WI survey has greater statistical power to discern trends in time-series of abundance data for juvenile reef fishes (Schrandt et al., 2021b).

Statistically powerful indices of juvenile abundance can help fishery managers make better informed decisions in forecasting and managing Gulf of Mexico reef fish stocks and allow for the opportunity to assess connectivity between the juvenile fish populations using estuarine habitats and their adult counterparts using offshore reef habitats. In addition to data on juvenile reef fish provided by the WI survey, the FIM program has expanded to conduct offshore sampling in both reef (Keenan et al., 2022; Switzer

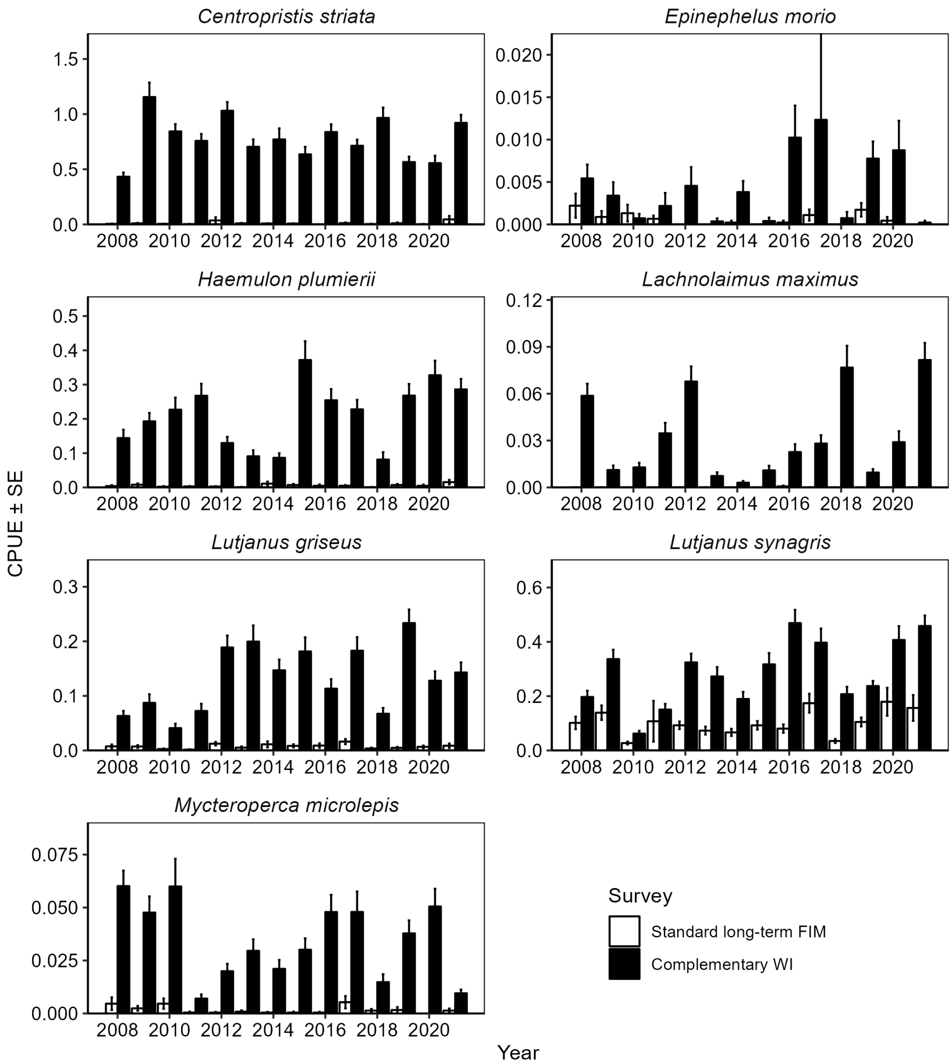



FIGURE 7
Mean catch-per-unit-effort (CPUE; fish per 100 m$^2$) ± SE for estuarine-dependent reef fishes collected in 6.1-m otter trawls in the fishery-independent long-term monitoring survey and the complementary West Florida Shelf Inshore (WI) survey. The data presented here include catch data for five estuaries (St. Andrew Bay, Apalachicola Bay, the Big Bend region, Tampa Bay, and Charlotte Harbor) from June through November for each year from 2008 through 2021. All five estuaries are sampled under the WI survey, but only Apalachicola Bay, Tampa Bay, and Charlotte Harbor are sampled under the long-term monitoring survey.

et al., 2023) and nonreef environments throughout the eastern Gulf (Matheson et al., 2017; Christiansen et al., 2022b). Combined, data from these comprehensive surveys of reef fish populations—juveniles through adults—have already proven critical to assessing managed reef fish populations (see SEDAR (2018) for an example for gray snapper).

# 5 Case study 3: Spotted seatrout in the Western Florida Panhandle

## 5.1 Rationale

Spotted seatrout is an estuarine-dependent, economically important species throughout its U.S. range, which encompasses the Atlantic and Gulf coasts from Cape Cod, MA, southward through Texas (Tabb, 1966; Brown-Peterson and Thomas, 1988; Bortone, 2003). In Florida, in the mid-1990s, net limitation regulations and declines in catch rates led the spotted seatrout fishery, traditionally both commercial and recreational, to shift to almost exclusively recreational, with approximately 98% of the state harvest coming from the recreational fishery (Murphy et al., 2011; Addis et al., 2017). The FWC has managed the spotted seatrout fishery with bag, gear, and slot limits across management regions that have changed over time. From 2012 to 2019, four regions were used: Northwest (northern Gulf coast), Southwest (southern Gulf coast), Southeast (southern Atlantic coast), and Northeast (northern Atlantic coast). Ongoing, standard long-term FIM sampling surveys existed in each of these regions. In 2020, FWC configured five management regions (Figure 8), with input from stakeholders, that aligned with evidence of genetic breaks in Florida (Seyoum et al., 2018). One of these newly created management regions, Western Panhandle, did not have a standard FIM survey in place, resulting in data limitations in abundance estimates. The FIM survey design for this region needed to address several spotted seatrout metrics: data for age-and-growth determination, and abundance data for three size classes of spotted seatrout (young-of-the-year, pre-fishery, and fishery). Additionally, the survey design would need to address other species, managed and unmanaged, and provide data for ecosystem-management assessments (e.g., Schrandt and MacDonald, 2020; Schrandt et al., 2021a; zu Ermgassen et al., 2021), as are standard for all FIM program surveys. Ideally, the implemented sampling design would collect data consistent with data collected in other estuaries sampled by the FIM program. A final, important consideration was that less than 20% of the annual funding needed to establish a traditional FIM survey in just one estuary was available for surveys in this new region, which includes six main estuaries (St. Joseph Bay, St. Andrew Bay, Choctawhatchee Bay, Santa Rosa Sound, Pensacola Bay, and Perdido Bay). The main objective of the Western Panhandle survey was to address the data limitations by providing data for abundance and age-and-growth information for multiple size and age classes of spotted seatrout to inform the state stock assessment and management decisions.

## 5.2 Evaluation phase

Data from two nontraditional FIM surveys were available from the new spotted seatrout management region: 1) a short-term,

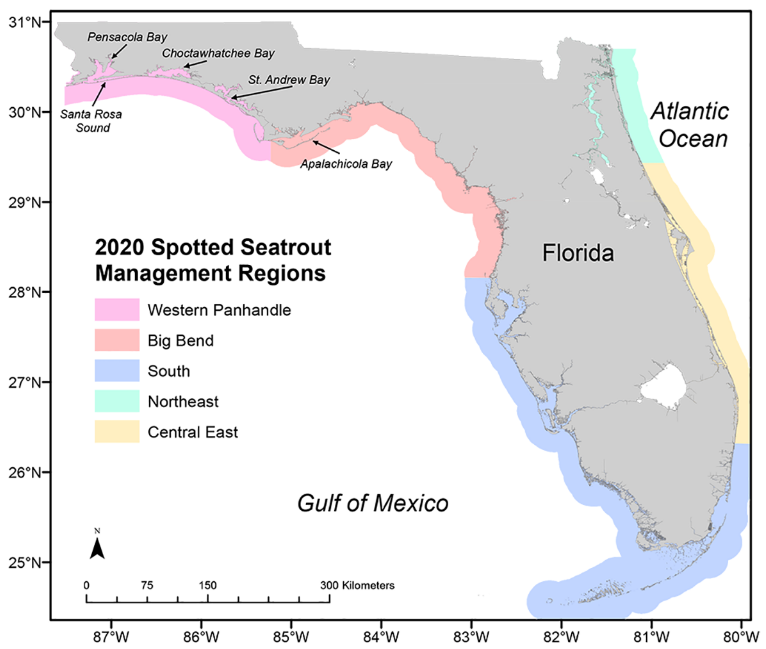



FIGURE 8
Map of five fishery management regions for spotted seatrout, *Cynoscion nebulosus*, established in 2020. Apalachicola Bay, one of the Fisheries-Independent Monitoring (FIM) program's long-term monitoring estuaries, is the FIM laboratory closest to the Western Panhandle management region established in 2020. The FIM program began sampling efforts in the Western Panhandle to support data collection for the new management region.

seasonal (June–December) reconnaissance survey from 2017–2019 in four western Panhandle estuaries (St. Andrew Bay, Choctawhatchee Bay, Santa Rosa Sound, and Pensacola Bay; Figure 8); and 2) a relatively long-term, seasonal (June–November) West Florida Shelf Inshore survey (WI; see section 4) from 2008 to the present day, that targets polyhaline seagrass beds in St. Andrew Bay. There were also standard long-term FIM survey and WI survey data available from Apalachicola Bay, an estuary immediately east of the Western Panhandle region (Figure 8). Data from those two surveys were examined to determine a suitable, cost-effective sampling design to address data needs for stock assessment and ecosystem management purposes in the Western Panhandle. These analyses included examining monthly spotted seatrout length-frequency and abundance between gear types, survey designs, and estuaries.

Monthly length-frequency distributions for spotted seatrout were compared between the two gear types (21.3-m seines and 6.1-m otter trawls) that targeted age-0 (<100 mm SL) spotted seatrout in two sampling areas (Western Panhandle and Apalachicola Bay) to discern differences between gear types and regions, identify monthly differences, and assist in assessing appropriate months in which to survey the population. Spotted seatrout <100 mm SL were collected with both gear types between June and November (Figure 9) and there were no visually identifiable differences in the range of sizes sampled between gear types or areas.

Monthly abundance plots were prepared for each estuary for the two gear types that target age-0 (<100 mm SL) spotted seatrout (21.3-m seines and 6.1-m otter trawls) to compare trends between gear types and estuaries, determine what estuaries should be sampled, and assist in assessing appropriate survey months. Monthly abundance trends for the seines in Apalachicola Bay were similar to the monthly abundance trends for the otter trawls in St. Andrew and Apalachicola bays (Figure 10), with low abundance in June, peak abundance in August, and then a decline to very low abundance by November. Reconnaissance sampling with 21.3-m seines in the Western Panhandle region demonstrated little variation between months, and monthly trends were not in agreement with the seine or trawl data from Apalachicola Bay, or with the otter trawl data from St.

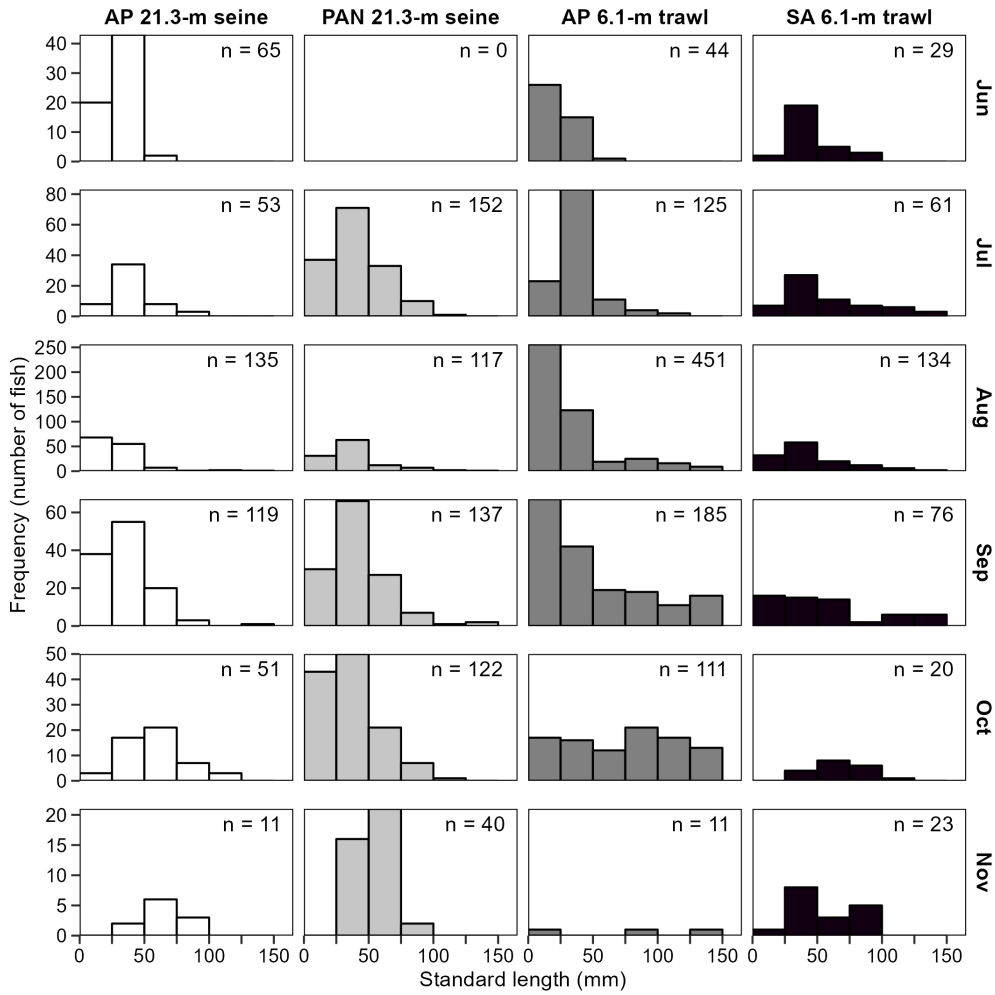



FIGURE 9
Size distribution of spotted seatrout, *Cynoscion nebulosus*, collected from 21.3-m seines from Apalachicola Bay (AP) and the Western Panhandle region (PAN) during the reconnaissance survey and from 6.1-m otter trawls during the West Florida Shelf Inshore (WI) survey. Data from 2017–2019 were analyzed. Fish >150 mm SL were collected but the x-axis has been truncated to emphasize sizes that represent the majority of the catch (age-0).

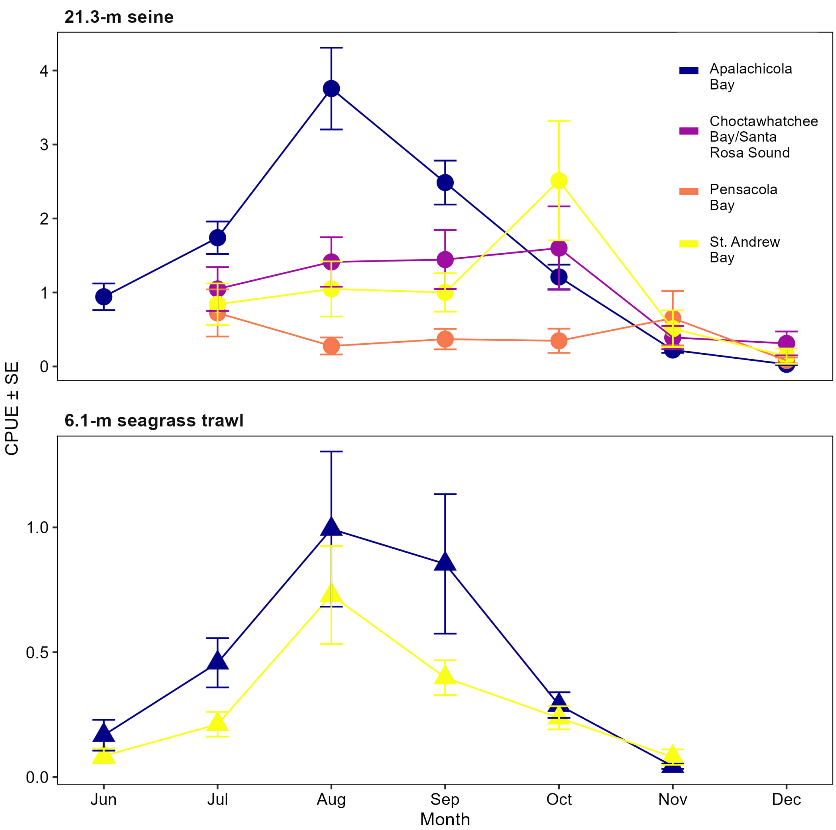



FIGURE 10
Monthly mean catch-per-unit-effort (CPUE) ± SE for all sizes of spotted seatrout, *Cynoscion nebulosus*, collected by the FIM program sampling surveys. CPUE for the 21.3-m seine data is expressed as number of fish per haul and represents long-term FIM sampling in Apalachicola Bay (2001–2019) and reconnaissance sampling from 2017 to 2019 in the Western Panhandle (St. Andrew Bay, Choctawhatchee Bay/Santa Rosa Sound, and Pensacola Bay). CPUE for the 6.1-m otter trawls is expressed as number of fish per 100 $m^2$ and represents the West Florida Shelf Inshore (WI) survey conducted from June through November (2008–2019) in Apalachicola and St. Andrew bays.

Andrew Bay. Although the seines in the Western Panhandle reconnaissance survey included seagrass bed habitats like the other surveys, the sample size within seagrass beds might not have been large enough to allow monthly trends to be discerned, which may explain this lack of agreement between surveys.

The assessment of pre-fishery (100–325 mm SL) and fishery (326–415 mm SL) sized spotted seatrout in the Western Panhandle was necessary for stock assessments. Available data collected before 2020 consisted of limited fishery-dependent data for a few metrics (age, growth, sex proportions, and maturity), but no abundance data. The Western Panhandle region had never been sampled with the 183-m haul seine, but length-frequency and abundance data from Apalachicola Bay were available and were assessed to refine the survey design for this gear type. Pre-fishery and fishery sized spotted seatrout were collected regularly (Figure 11) and were most abundant in 183-m haul seine samples in Apalachicola Bay between July and February.

## 5.3 Survey implementation

In June 2020, a seasonal survey with monthly sampling was initiated to provide spotted seatrout data for the newly created Western Panhandle region. Sampling in two estuaries (Choctawhatchee Bay and Santa Rosa Sound) would be combined with the WI survey in St. Andrew Bay to assess spotted seatrout in the region. Pensacola Bay, a Western Panhandle region estuary sampled during the reconnaissance survey, was not included in the survey design because funding was inadequate. The survey design used two gear types: 6.1-m otter trawls (all three estuaries) and 183-m haul seines (Choctawhatchee Bay and Santa Rosa Sound, only).

The 6.1-m otter trawl was chosen to sample young-of-the-year spotted seatrout, rather than the 21.3-m seine that is used in estuaries with standard long-term FIM surveys. The WI survey trawls collected the same size animals with similar monthly trends as the 21.3-m seine in Apalachicola Bay and these trends mirrored trends in other estuaries sampled by the FIM program (Kupschus, 2003), whereas the monthly abundances from the 21.3-m seine reconnaissance sampling in the Western Panhandle varied little between months (Figure 10). By using the WI survey with 6.1-m trawls, the St. Andrew Bay WI survey data immediately provided a 13-year (2008–2020) data set for spotted seatrout in this newly established region. Using the WI survey protocol in the Western Panhandle had the additional benefit of extending the FIM program's WI survey into two additional estuaries (Choctawhatchee Bay and Santa Rosa Sound). Although abundance of spotted seatrout from trawls in June and November was less than that from trawls from July through October, those

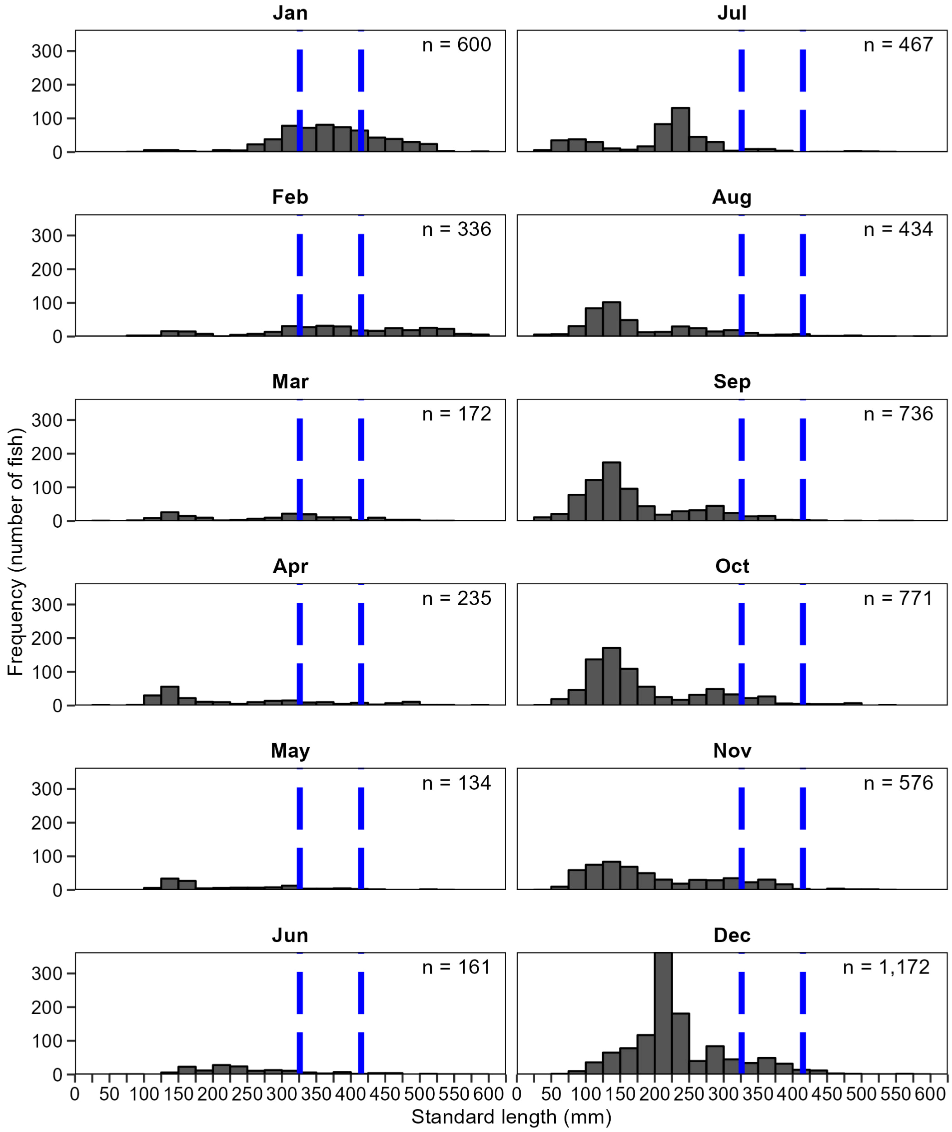



FIGURE 11
Size distribution of spotted seatrout, *Cynoscion nebulosus*, collected by month with 183-m haul seines, 2001–2019 in Apalachicola Bay. Fish to the left of the blue vertical bar are age-0 (<100 mm SL) and pre-fishery (100–326 mm SL). Fish between 326 mm and 415 mm SL are fish within the fishery, and fish larger than 415 mm SL are post-fishery fish. Age-0 and post-fishery spotted seatrout are not efficiently collected by this gear.

months were retained to ensure full comparability with WI surveys, which sample from June through November, in other estuaries.

The need to initiate a 183-m haul seine survey in the Western Panhandle region (Choctawhatchee Bay and Santa Rosa Sound) was identified as soon as the new management region was designated. The 183-m haul seine is the only standard FIM gear type that addresses two critical size classes of spotted seatrout (pre-fishery and fishery). Spotted seatrout catch data from Apalachicola Bay were used to establish the sampling period, June–December (Figure 12). The almost complete overlap in months with the 6.1-m otter trawl survey was economically fortuitous. Sampling with two gear types during fewer multiday events is less expensive than conducting more single-gear-type multiday events.

All gear deployments in the Western Panhandle region follow standard FIM sampling protocols and all fish and selected invertebrate species captured are identified to the lowest possible taxonomic level, counted, and a random sample of at least 10 individuals is measured. Although the data for this region are not year-round as for the estuaries that have been monitored long-term, following the same sampling protocol allows comparisons of data for estuaries in the region and statewide for months sampled by all surveys (June–November). This ensures that the data are available and appropriate for assessment of other taxa and for ecosystem management-type assessments.

## 5.4 Implications

This case study represents a situation in which the FIM program was able to implement a monitoring survey in

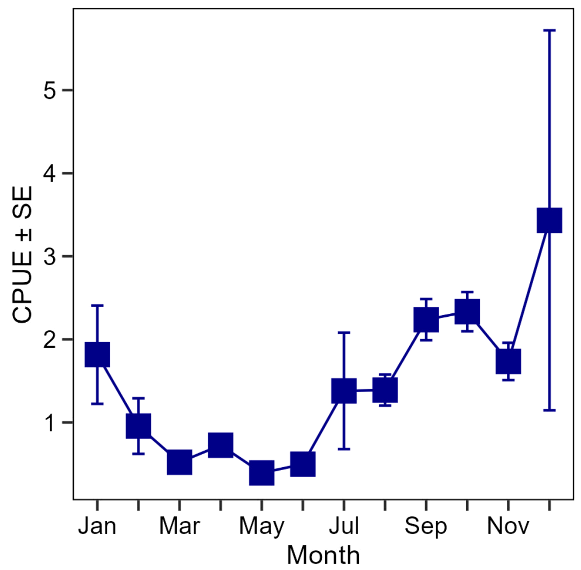



FIGURE 12
Monthly mean catch-per-unit-effort (CPUE; number of fish per haul) ± SE for all sizes of spotted seatrout, *Cynoscion nebulosus*, collected in Apalachicola Bay, 2001–2019, by the FIM program's long-term 183-m haul seine survey.

unsampled estuaries to provide timely and accurate data for state fisheries managers after an existing management region was split into two. The location of the split created a new region (Western Panhandle) without long-term fishery survey data. Although the establishment of a full FIM survey in the Western Panhandle region would have been preferred, adequate funding to initiate a full survey was not available or appropriated. Funding to provide the necessary data had to be reallocated from a federal Sportfish Restoration grant that had a slight surplus in 2020 ($150k). Continued funding was maintained by trimming other surveys; slight reductions were made to the juvenile snook survey (case study 1) and the standard FIM trawling survey (not case study 2, WI) was reduced by half. Standard FIM trawls mostly collect data on ecologically important taxa although some loss of data on managed species (especially blue crab, *Callinectes sapidus* and pink shrimp, *Farfantepenaeus duorarum*) occurred. The available funds, about 33% of what it currently costs to maintain a FIM field laboratory, shaped the final survey design, limiting it to seasonal sampling with just two of the three standard FIM program sampling gear types. By retaining trawls rather than one type of seine in the Western Panhandle, the FIM program was able to increase the geographic scope of data available for assessing juvenile reef fish (case study 2). And despite funding limitations, the survey immediately provided data for management of spotted seatrout in the Western Panhandle. The Western Panhandle management zone was specific to spotted seatrout; adequate data on other managed species (e.g., red drum and sheepshead, *Archosargus probatocephalus*) and ecologically important species, that would have been provided by a full FIM survey, were not necessary from this zone at the time. As Florida's fishery managers continue to consider management zone changes for various species, however, re-evaluation of the adequacy of the FIM program survey design within management geographic boundaries and fiscal stagnation is an ongoing challenge and necessity. The cost to establish even this limited sampling survey in the Western Panhandle region would have far exceeded the available funding without leveraging the resources from an established long-term FIM program in an adjacent estuary, Apalachicola Bay.

The FIM program had data to develop a survey design for the Western Panhandle: reconnaissance data from this region, long-term traditional FIM data from Apalachicola Bay (an estuary immediately east of the Western Panhandle), and 13 years of WI survey data from an estuary within the new region. Without these data, at least a year of reconnaissance would have been necessary for developing a survey design and four years would have been necessary before enough data would have been available to incorporate into the state stock assessment for spotted seatrout. The 183-m haul seine would have been included in the new survey whether data had been available or not; the necessary pre-fishery and fishery data could only have been obtained with the 183-m haul seine. Having long-term data available for the haul seine from an adjacent estuary was beneficial in cost-effectively determining when sampling was done. The selection of the 6.1-m otter trawl with the WI survey design to collect young-of-the-year spotted seatrout was not an obvious choice, as this gear type has not been used to assess spotted seatrout in any other FIM program estuary. After data had been analyzed and the alternatives discussed with the FWC stock assessment analysts, it was determined that the trawl was the best alternative. Maintaining standard FIM protocols in the survey for this new region further ensures the utility of these data in assessing stock of other species, comparisons among estuaries, and ecosystem management analyses. The data being developed by this survey meet the needs for spotted seatrout stock assessment in this region. The Western Panhandle sampling design, which lacked a full suite of FIM gear types and called for only limited seasonal sampling, addressed only some of the fishery data needs for this region. The design and implementation of the Western Panhandle survey relied on analysis of previously collected data and the geographically dispersed laboratories of the long-term FIM program, without which adapting FIM surveys to address this critical state management need would have been much more difficult and time-consuming. It is hoped that at some point, a full, traditional FIM survey can be established in the Western Panhandle region to support management of additional species.

## 6 Discussion

The three case studies discussed herein document only a handful of the changes the FIM program has had to implement to its long-term monitoring survey design in response to emerging management needs and funding fluctuations. In the first case study, the FIM program was able to provide improved data (abundance estimates and age-length keys) on juvenile snook for the state stock assessments by modifying its existing long-term monitoring survey to include areas and habitats historically under sampled in the long-term monitoring survey. The complementary WI survey in the second case study improved available data and statistical power to detect changes in abundance of juvenile reef fish species to inform management and expanded FIM ecosystem sampling into areas and habitats not previously addressed by the program. The third case study showcased how the FIM program was able to quickly adapt to develop and implement an appropriate survey design to address data limitations caused by the creation of a new management region for spotted seatrout. In each case, the FIM program had to revisit sampling design to adjust for realities of funding. Using the data already collected by the FIM program, the program was able to adapt existing monitoring plans, rather than initiate entirely new surveys, to maximize effectiveness and efficiency. All survey modifications were essential to address management needs and each modification considered the balance between monitoring costs and meeting the programmatic mission (Caughlan and Oakley, 2001; Strayer and Smith, 2003), to provide timely data and analysis for fisheries management for the conservation and protection of Florida's fisheries. The FIM program will continue to capitalize on its flexibility to address changing management needs in the future (e.g., snook, spotted seatrout).

Although dedicated surveys may often be the most effective approach for addressing emerging data needs for a particular species, the FIM program has generally focused on implementing design modifications into the existing long-term survey design as opposed to creating new surveys. For example, rather than creating independent surveys for juvenile snook or spotted seatrout, the FIM program leveraged the existing dispersed geographical model with field laboratories throughout the state to inform survey designs and implement enhancements to the long-term monitoring survey (e.g., additional areas, habitats, seasonal survey effort). Although species-specific surveys could have been implemented, enhancing the existing multispecies long-term monitoring survey to incorporate habitats and areas important to the species of interest better addressed the statewide FIM program mission. Multispecies surveys are also beneficial because they are more cost-effective for long-term monitoring programs and provide data for species that may not currently be assessed, but eventually may be (e.g., Gulf flounder, *Paralichthys albigutta*; sheepshead). Cost-effectiveness though, needs to be balanced with the collection of statistically powerful data (Nieman et al., 2021). Managers and decision makers need to be able to reliably understand fluctuations in species abundance over time (Wauchope et al., 2019) and depending on the species, decades of data may be needed to reliably detect trends in abundance (White, 2019). The FIM program has considered this as well when implementing survey design changes, as evident by the statistical power simulations conducted for the WI survey prior to amending the design in 2019. Furthermore, all modified and new designs follow statewide standardized procedures to collect biological and environmental data to produce an ecological dataset that is comparable among estuaries.

Shortfalls in funding will always restrict the scope of a long-term monitoring program. To address funding issues, the FIM program has frequently had to consider reducing staff, closing field labs, and sample reductions. Admittedly, a large portion of the FIM program budget is dedicated to data collection and design optimization, and staff workloads are ever-expanding as data collection increases. Throughout the program's history, it has had to critically evaluate processes to streamline data collection and improve data management, reporting, and analyses. Proactively recognizing and budgeting for these needs in the future will help the FIM program better consider all costs, beyond data collection, associated with long-term ecological monitoring (Caughlan and Oakley, 2001).

The FIM program has a strong (up to 50% of the budget) reliance on grant funding to conduct long-term fishery independent monitoring and produce robust ecological datasets. The program has been fortunate to apply for and receive grant funding to implement new or complementary surveys to address emerging needs and specific questions that cannot be directly addressed through the long-term monitoring survey design. Although these grants represent a temporary increase in funding, they also create the need for additional balancing. Before funding is received there is an additional workload required for researching various funding opportunities and writing proposals within the bounds of the research foci. In response to the grant award, a long-term monitoring program needs to address specific questions and provide specific deliverables for grant requirements, within the context of the broader program. Furthermore, additional grants do not always equate to additional staff, so the program must further maximize efficiency, while also managing what can be widely fluctuating budgeting as awards start and end. A dependence upon grant funding also creates a challenge as grants frequently require additional data (e.g., GIS shapefiles/layers, video camera data, acoustic data) and analyses (e.g., multi-metric indices) be incorporated into the existing database and workflow. A final consideration with grant funding is that when the funding ends, the program may be faced with evaluating and modifying survey designs to maintain critical data needed for fishery management decisions despite a reduction in funding.

## 7 Lessons learned

Since the FIM program's inception, it has had to balance any changes in the long-term survey design with maintaining critical long-term time series, addressing management needs, funding, and staff changes. For all three case studies highlighted here, as well as other programmatic amendments over the years, the FIM program has learned to improvise to meet changing needs but also to make sure that data resulting from sampling reductions, modifications, and expansions, are compatible to other aspects of FIM sampling throughout Florida. This consistency among data is critical to maintaining long-term time series and allowing for statewide data comparisons. Before a survey is modified, programs should review the knowledge on the species and ecosystems of interest, summarize and analyze data from earlier surveys, leverage data from adjacent systems as best as possible, conduct reconnaissance trips with various gear configurations to inform design and gear selection, and test different gear types based on the previous work so that the data can be used to evaluate survey design. We suggest that survey changes examine multiple options to best determine which option provides the necessary data, minimizes any compromising of data quality, and ideally, provides ecosystem-level data rather than single-species data, so that program data will have broader application.

One key application of FIM data requiring careful consideration of long-term consistency is the generation of indices of relative abundance for stock assessment. A key assumption of fishery independent indices is that changes in relative abundance through time represent actual changes in the stock being assessed, and not changes in survey design; therefore, it is important to consider what analytical approaches can be applied to maintain long-term time series prior to considering change in survey design. There are various ways to account for or address changes in survey design, although the approaches used are likely to vary with the species of interest or the particular assessment models being fit. Most commonly, analysts account for changes in survey design through standardization techniques (e.g., generalized linear models, generalized additive models (e.g., Switzer et al., 2012; Bacheler et al., 2022) that can adjust for resultant changes while retaining the maximum amount of data possible. For example, if sample size is increased and samples are collected in a new geographic location, or spatial sampling zone, within an existing sampled estuary (e.g., Section 3: Juvenile common snook), the stock assessment model can account for the change by including zone as a factor in the model. Similarly, if additional samples are collected within an existing sampled zone, as opposed to a new zone, the increase in sample size refines the model by reducing variation in estimates. When sampling designs are changed to the point where previous data may not fully align with the new design, analysts may consider either truncating the time series or developing a split index. Alternately, one could censure the data to resolve potential temporal or spatial mismatch. For example, if a time series is reduced from monthly to bimonthly sampling, one could remove the months that are no longer sampled from the previous data to yield one continuous time series of commonly sampled months. It is also important to consider any potential changes to length composition data. For example, if sampling is no longer conducted during periods corresponding to peak juvenile recruitment, or gear modifications are implemented that alter the size selectivity of a particular sampling gear, the assessment model would need to be adjusted (e.g., Christiansen et al., 2022a). In each scenario, the fundamental action while using the data in subsequent analyses is exploration of the data to determine how to account for changes in survey design. Ideally, there is also an open dialogue between those collecting the data and those using the data in various analyses, like the FIM program and the stock assessment analysts, so that they can work collaboratively to address survey design changes. To that end, the FIM program has invested significant effort in examining available data to make decisions that keep the long-term data as consistent and comparable as possible.

Key factors in the success of the statewide FIM program and its ability to adapt to funding shifts and management decisions are 1) its mission-focused programmatic goals, 2) geographically dispersed laboratories, 3) standardized protocols, 4) ongoing critical analysis of the data, 5) grant award success, and 6) high-level data management. The mission and objectives established for the program 30 years ago allow survey design changes to be considered within the realm of the programmatic mission. The geographically dispersed laboratories throughout the state ensure that standardized survey data exist for sites near a new area or for any new management-critical species. It also allows the FIM program to easily conduct reconnaissance sampling—and even long-term sampling—in estuaries adjacent to long-term survey sites at a fraction of the cost of establishing a full survey in such an estuary. The standardized sampling protocols ensure that sampling completed anywhere in the state can be readily implemented in a new estuary and that the data allow comparison between regions. Ongoing analysis of data provides insights that help identify deficiencies in stock assessments and survey design modifications that address management needs. The FIM program's ability to achieve success in its grant proposals has been instrumental for pilot studies, field reconnaissance, and implementation of survey design modifications. Finally, high-level data management is critical to the FIM program's success in ensuring that the data are available for analyses and other uses. Ultimately, these key characteristics of the FIM program have made possible successful data sharing, products, and publications (www.myfwc.com/research/saltwater/fim/), which are vital to the program's grant award success.

Often, the greatest expenses in the budget of a monitoring program are related to data collection, so it is not surprising that the first considerations of monitoring costs focused on optimizing sampling design (Caughlan and Oakley, 2001). But the focus on data collection and design optimization often leads to the neglect of other critical aspects, such as training, quality assurance, reporting, scientific oversight, and, fundamentally, data management (Caughlan and Oakley, 2001). Data management is fundamental to any long-term monitoring program (Burns et al., 2018), especially since one characteristic of long-term monitoring programs is the continued collection and availability of consistent data over time. Ensuring that data are available for analysis requires use of detailed and comprehensive procedures to manage the data (Fancy and Bennetts, 2012). Data management must be one of the critical components of a long-term monitoring program's budget (Caughlan and Oakley, 2001). For example, successful long-term monitoring programs for U.S. national parks use 25–30% of the monitoring budget for data management, assessment, and reporting (Graber et al., 1993; Mulder et al., 1999). A practical way to budget for data management and integrate it into long-term ecological monitoring programs is to adopt standard and comprehensive procedures for data management (Sutter et al., 2015), including metadata, database design overview, data verification and editing procedures, archival procedures, data summaries, reporting schedules and formats, and describing potential analyses (Oakley et al., 2003). The three survey design modifications highlighted herein could be folded into the FIM database structure with only minor changes to the database back-end and front-end because the database was designed to be structured, yet flexible. Collection under FIM data and survey protocols ensured that data were readily available for analyses for stock assessments and fishery managers. This is not always the case, however, as grant awards often collect streams of data (e.g., acoustic tags) that are new to the FIM database, something that must be accounted for in budgeting. Another component of data management that is key to data integrity, accessibility, and use of data is clearly assigned roles for database management and front-end software development (Oakley et al., 2003; Sutter et al., 2015). Though the FIM program does not have a dedicated database manager, it has maintained long-term database needs through scientific staff with this skill set. For the long term, however, this is inadequate. High-level data management is essential to continued accessibility of FIM data and data products for fishery management analyses, ecosystem analyses, and program changes, such as the three case studies presented here. And the FIM program is still growing, and database management is an increasingly critical program area that needs to be directly budgeted in future grant proposals (~25%).

## Author contributions

MS, GO, and TS originally discussed and planned the review, which was further developed in collaboration with TM, RP, BW, DB, and DG. MS, TM, RP, BW, DB, and DG wrote the manuscript. GO designed and formatted the maps and MS formatted the figures. All authors were involved in at least one of the three case studies and contributed to manuscript text, figures, and revisions. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by funding from the State of Florida saltwater recreational fishing licenses and snook permits, from the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service (grant numbers NA08NMF4720645 and NA09NMF4330152), by funding from the U.S. Department of the Interior, U.S. Fish and Wildlife Service, Federal Aid for Sportfish Restoration (grant numbers F14AF00328, F15AF01222, F16AF00898, F17AF00932, F18AF00281, F18AF00665, F19AF00105, F19AF00210, and F20AF00167), and by funding from the National Fish and Wildlife Foundation Gulf Environmental Benefit Fund (grants numbers FL 40624, FL 45766, FL 50347, FL 54269, and FL 58101).

## Acknowledgments

We thank the staff of the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute's Fisheries Independent Monitoring (FIM) program for their continued dedication to field sampling and data collection. A special thanks to R.H. McMichael Jr. for having the vision for and dedication to the development of the FIM program; this rigorous statewide program would not have been possible without him.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Author disclaimer

The statements, findings, views, conclusions, and recommendations contained in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of the Interior or the U.S. Department of Commerce and should not be interpreted as representing the opinions or policies of the U.S. government. Mention of trade names or commercial products does not constitute their endorsement by the U.S. government.

## References

Adams, D. H., Tremain, D. M., Paperno, R., and Sonne, C. (2019). Florida lagoon at risk of ecosystem collapse. *Science* 365, 991–992. doi: 10.1126/science.aaz0175

Addis, D. S., Allen, S., Herdter, E., Mahmoudi, B., Muller, R. G., Munyandorero, J., et al. (2017). *Florida's inshore and nearshore species: 2017 status and trends report*. In House Report 2017-005. (St. Petersburg, FL: Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute).

Bacheler, N. M., Gregalis, K. C., Gillum, Z. D., Pickett, E. P., Schobernd, C. M., Schoebernd, Z. H., et al. (2022). Using stationary video data to infer relative abundance and distribution of four *Seriola* species along the southeast United States Atlantic coast. *Fish. Res.* 249, 106238. doi: 10.1016/j.fishres.2022.106238

Bortone, S. A. (2003). *Biology of the spotted seatrout* (Boca Raton, FL, USA: CRC Press).

Brame, A. B., McIvor, C. C., Peebles, E. B., and Hollander, D. J. (2014). Site fidelity and condition metrics suggest sequential habitat use by juvenile common snook. *Mar. Ecol. Prog. Ser.* 509, 255–269. doi: 10.3354/meps10902

Brown-Peterson, N. J., and Thomas, P. (1988). Differing reproductive life histories between temperate and subtropical groups of *Cynoscion nebulosus*. *Cont. Mar. Sci.* 30, 71–78.

Bryan, M. D., and McCarthy, K. (2015). *Standardized catch rates for red grouper from the United States Gulf of Mexico vertical line and longline fisheries*. SEDAR42-AW-02.

SouthEast Data, Assessment, and Review (North Charleston, South Carolina: SouthEast Data, Assessment, and Review).

Burns, E. L., Tennant, P., Dickman, C. R., Gillespie, G., Green, P. T., Hoffmann, A., et al. (2018). Making monitoring work: Insights and lessons from Australia's long term ecological research network. *Austral. Zool.* 39, 755–768. doi: 10.7882/AZ.2018.030

Casey, J. P., Poulakis, G. R., and Stevens, P. W. (2007). Habitat use by juvenile Gag, *Mycteroperca microlepis* (Pisces: Serranidae), in subtropical Charlotte Harbor, Florida (USA). *Gulf Carib. Res.* 19, 1–9. doi: 10.18785/gcr.1901.01

Caughlan, L., and Oakley, K. L. (2001). Cost considerations for long-term ecological monitoring. *Ecol. Ind.* 1, 123–134. doi: 10.1016/S1470-160X(01)00015-2

Christiansen, H. M., Solomon, J. J., Switzer, T. S., and Brodie, R. B. (2022a). Assessing the size selectivity of capture gears for reef fishes using paired stereo-bated remote underwater video. *Fish. Res.* 249, 106234. doi: 10.1016/j.fishres.2022.106234

Christiansen, H. M., Tyler-Jedlund, A. J., Switzer, T. S., Keenan, S. F., and Stahl, S. D. (2022b). Potential implications of excluding live-bottom habitat from a bottom-trawl survey in the eastern Gulf of Mexico. *N. Am. J. Fish. Man.* 42, 893–905. doi: 10.1002/nafm.10775

Coull, B. C. (1985). The use of long-term biological data to generate testable hypotheses. *Estuaries* 8, 84–92. doi: 10.2307/1351859

De Angelo, J. A., Stevens, P. W., Blewett, D. A., and Switzer, T. S. (2014). Fish assemblages of shoal-and shoreline-associated seagrass beds in eastern Gulf of Mexico estuaries. *Trans. Am. Fish. Soc* 143, 1037–1048. doi: 10.1080/00028487.2014.911209

Fancy, S. G., and Bennetts, R. E. (2012). "Institutionalizing an effective long-term monitoring program in the US National Park Service," in *Design and analysis of long-term ecological monitoring studies.* Eds. R. A. Gitzen, J. J. Millspaugh, A. B. Cooper and D. S. Licht (Cambridge, MA, USA: Cambridge University Press), 481–497.

Flaherty-Walia, K. E., Switzer, T. S., Winner, B. L., Tyler-Jedlund, J. A., and Keenan, S. F. (2015). Improved ability to characterize recruitment of gray snapper in three Florida estuaries along the Gulf of Mexico through targeted sampling of polyhaline seagrass beds. *Trans. Am. Fish. Soc* 144, 911–926. doi: 10.1080/00028487.2015.1054516

Gillanders, B. M., Able, K. W., Brown, J. A., Eggleston, D. B., and Sheridan, P. F. (2003). Evidence of connectivity between juvenile and adult habitats for mobile marine fauna: an important component of nurseries. *Mar. Ecol. Prog. Ser.* 247, 281–295. doi: 10.3354/meps247281

Gilmore, R. G., Donahoe, C. J., and Cooke, D. W. (1983). Observations on the distribution and biology of the common snook, *Centropomus undecimalis* (Bloch). *Fla. Sci.* 46, 313–336. Available at: https://www.jstor.org/stable/24320342.

Goodyear, C. P. (1985). Relationship between reported commercial landings and abundance of young striped bass in Chesapeake Bay, Maryland. *Trans. Am. Fish. Soc* 114, 92–96. doi: 10.1577/1548-8659(1985)114<92:RBRCLA>2.0.CO;2

Graber, D. M., Haultain, S. A., and Fessenden, J. E. (1993). "Conducting a biological survey: a case study from Sequoia and Kings Canyon national parks," in *Proceedings of the Fourth Conference on Research in California's National Parks*, vol. 9 . Eds. S. D. Veirs Jr., T. J. Stohlgren and C. Schonewald-Cox (Orick, California: U.S. Department of the Interior National Park Service Transactions and Proceedings Series), 17–35.

Hansen, J. A., Carpenter, S. R., Gaeta, J. W., Hennessy, J. M., and Vander Zander, M. J. (2015). Predicting walleye recruitment as a tool for prioritizing management actions. *Can. J. Fish. Aquat. Sci.* 72, 661–672. doi: 10.1139/cjfas-2014-0513

Keenan, S., Switzer, T., Knapp, A., Weather, E., and Davis, J. (2022). Spatial dynamics of the quantity and diversity of natural and artificial hard bottom habitats in the eastern Gulf of Mexico. *Cont. Shelf Res.* 233, 104633. doi: 10.1016/jcsr.2021.104633

Koenig, C. C., and Coleman, F. C. (1998). Absolute abundance and survival of juvenile gags in sea grass beds of the northeastern Gulf of Mexico. *Trans. Am. Fish. Soc* 127, 44–55. doi: 10.1577/1548-8659(1998)127<0044:AAASOJ>2.0.CO;2

Kupschus, S. (2003). Development and evaluation of statistical habitat suitability models: an example based on juvenile spotted seatrout. *Cynoscion nebulosus. Mar. Ecol. Prog. Ser.* 265, 197–212. doi: 10.3354/meps265197

Lefcheck, J. S., Hughes, B. B., Johnson, A. J., Pfirrmann, B. W., Rasher, D. B., Smyth, A. R., et al. (2019). Are coastal habitats important nurseries? A meta-analysis. *Conserv. Lett.* 12 (4), e12645. doi: 10.1111/conl.12645

Marshall, A. R. (1958). *A survey of the snook fishery of Florida, with studies of the biology of the principal species, Centropomus undecimalis (Bloch)* (Miami, FL, USA: Florida State Board of Conservation, Marine Research Laboratory). Technical Series no. 22.

Matheson, R. E. Jr., Flaherty-Walia, K. E., Switzer, T. S., and McMichael, R. H. Jr (2017). The importance of time of day in structuring demersal ichthyofaunal assemblages on the West Florida Shelf. *Bull. Mar. Sci.* 93 (2), 407–437. doi: 10.5343/bms.2016.1047

McMichael, R. H. Jr., Peters, K. M., and Parsons, G. R. (1989). Early life history of the snook, *Centropomus undecimalis*, in Tampa Bay, Florida. *N.E. Gulf Sci.* 10, 113–125. doi: 10.18785/negs.1002.05

Mulder, B. S., Noon, B. R., Spies, T. A., Raphael, M. G., Palmer, C. J., Olsen, A. R., et al. (1999). *The strategy and design of the effectiveness monitoring program for the Northwest Forest Plan.* PNW-GTR-437 (Portland, OR, USA: USDA Forest Service, Pacific Northwest Station).

Munyandorero, J., Trotter, A. A., Stevens, P. W., and Muller, R. G. (2020). *The 2020 stock assessment of common snook, Centropomus undecimalis.* In House Report 2020-004 (St. Petersburg, FL, USA: Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute), 147.

Murphy, M. D., Chagaris, D., and Addis, D. (2011). *An assessment of the status of spotted seatrout in Florida waters through 2009.* In-House Report 2011-002. (St. Petersburg, FL, USA: Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute).

Myers, R. A., and Cadigan, N. G. (1993). Is juvenile mortality in marine demersal fish variable? *Can. J. Fish. Aq. Sci.* 50, 1591–1598. doi: 10.1139/f93-180

Nagelkerken, I., Kleijnen, S., Klop, T., van den Brand, R. A. C. J., de la Morinière, E. C., and van der Velde, G. (2001). Dependence of Caribbean reef fishes on mangroves and seagrass beds as nursery habitats: a comparison of fish faunas between bays with and without mangroves/seagrass beds. *Mar. Ecol. Prog. Ser.* 214, 225–235. doi: 10.3354/meps214225

Nagelkerken, I., Roberts, C. V., van der Velde, G., Dorenbosch, M., van Riel, M. C., de la Morinière, E. C., et al. (2002). How important are mangroves and seagrass beds for coral-reef fish? The nursery hypothesis tested on an island scale. *Mar. Ecol. Prog. Ser.* 244, 299–305. doi: 10.3354/meps244299

Nieman, C. L., Pendleton, R. M., Kenney, G. H., and Solomon, C. T. (2021). Evaluation and optimization of a long-term fish monitoring program in the Hudson River. *Ecol. Indicat.* 133, 108344. doi: 10.1016/j.ecolind.2021.108344

Oakley, K. L., Thomas, L. P., and Fancy, S. G. (2003). Guidelines for long-term monitoring protocols. *Wildl. Soc Bull.* 31 (4), 1000–1003. Available at: https://www.jstor.org/stable/3784444.

Peters, K. M., Matheson, R. E. Jr., and Taylor, R. G. (1998). Reproduction and early life history of the common snook, *Centropomus undecimalis* (Bloch), in Florida. *Bull. Mar. Sci.* 62, 509–529.

Purtlebaugh, C. H., Martin, C. W., and Allen, M. S. (2020). Poleward expansion of common snook *Centropomus undecimalis* in the northeastern Gulf of Mexico and future research needs. *PloS One* 15 (6), e0234083. doi: 10.1371/journal.pone.0234083

Rivas, L. R. (1986). Systematic review of the perciform fishes of the genus *Centropomus*. *Copeia* 1986 (3), 579–611. doi: 10.2307/1444940

Schrandt, M. N., and MacDonald, T. C. (2020). Long-term stability of the faunal community of a subtropical estuary: evaluating disturbances in the context of interannual variability. *Estuaries Coasts* 43, 347–359. doi: 10.1007/s12237-019-00684-1

Schrandt, M. N., MacDonald, T. C., Sherwood, E. T., and Beck, M. W. (2021a). A multimetric nekton index for monitoring, managing and communicating ecosystem health status in an urbanized Gulf of Mexico estuary. *Ecol. Ind.* 123, 1–10. doi: 10.1016/j.ecolind.2020.107310

Schrandt, M. N., Shea, C. P., Kurth, B. N., and Switzer, T. S. (2021b). Amending survey design to improve statistical inferences: monitoring recruitment of juvenile reef fish in the eastern Gulf of Mexico. *Fish. Res.* 241, 106015. doi: 10.1016/j.fishres.2021.106015

Schrandt, M. N., Switzer, T. S., Stafford, C. J., Flaherty-Walia, K. E., Paperno, R., and Matheson, R. E. Jr. (2018). Similar habitats, different communities: fish and large invertebrate assemblages in eastern Gulf of Mexico polyhaline seagrasses relate more to estuary morphology than latitude. *Estuar. Coast. Shelf. Sci.* 213, 217–229. doi: 10.1016/j.ecss.2018.08.022

SEDAR (SouthEast Data, Assessment, and Review) (2018). *SEDAR 51 – Gulf of Mexico gray snapper stock assessment report* (North Charleston, SC, USA: SEDAR).

Seyoum, S., McBride, R. S., Tringali, M. D., Villanova, V. L., Puchutulegui, C., Gray, S., et al. (2018). Genetic population structure of the spotted seatrout (*Cynoscion nebulosus*): simultaneous examination of the mtDNA control region and microsatellite marker results. *Bull. Mar. Sci.* 94, 47–71. doi: 10.5343/bms.2017.1060

Smith, M. W., Goethel, D., Rios, A., and Isley, J. (2015). *Standardized catch rate indices for Gulf of Mexico gray triggerfish (Balistes capriscus) landed during 1993-2013 by the commercial handline fishery*. SEDAR43-WP-05. SouthEast Data, Assessment, and Review (North Charleston, South Carolina).

Stevens, P. W., Blewett, D. A., Boucek, R. E., Rehage, J. S., Winner, B. L., Young, J. M., et al. (2016). Resilience of a tropical sport fish population to a severe cold event varies across five estuaries in southern Florida. *Ecosphere* 7 (8), e01400. doi: 10.1002/ecs2.1400

Stevens, P. W., Paperno, R., Beal, J. L., MacDonald, T. C., Miller, H. N., Klarmann, P. A., et al. (2022). Identification of fish habitat hotspots for use in prioritizing conservation and restoration projects in coastal rivers. *Environ. Biol. Fishes* 106, 221–235. doi: 10.1007/s10641-022-01226-8

Strayer, D. L., and Smith, D. R. (2003). A guide to sampling freshwater mussel populations. *Am. Fish. Soc Mon.* 8, 1–103.

Sutter, R. D., Wainscott, S. B., Boetsch, J. R., Palmer, C. J., and Rugg, D. J. (2015). Practical guidance for integrating data management into long-term ecological monitoring projects. *Wildl. Soc Bull.* 39 (3), 451–463. doi: 10.1002/wsb.548

Switzer, T. S., Keenan, S. F., Stevens, P. W., McMichael, R. H. Jr., and MacDonald, T. C. (2015). Incorporating ecology into survey design: monitoring the recruitment of age-0 gags in the eastern Gulf of Mexico. *N. Am. J. Fish. Man.* 35, 1132–1143. doi: 10.1080/02755947.2015.1082517

Switzer, T. S., Keenan, S. F., Thompson, K. A., Shea, C. P., Knapp, A. R., Campbell, M. D., et al. (2023). Integrating assemblage structure and habitat mapping data into the design of a multispecies reef fish survey. *Mar. Coast. Fish.: Dynam. Manage. Ecosys. Sci.* 15 (4), e10245. doi: 10.1002/mcf2.10245

Switzer, T. S., MacDonald, T. C., McMichael, R. H. Jr., and Keenan, S. F. (2012). Recruitment of juvenile gags in the eastern Gulf of Mexico and factors contributing to observed spatial and temporal patterns of estuarine occupancy. *Trans. Am. Fish. Soc* 141, 707–719. doi: 10.1080/00028487.2012.675913

Tabb, D. C. (1966). The estuary as a habitat for spotted seatrout, *Cynoscion nebulosus*. *Am. Fish. Soc Spec. Publ.* 3, 58–67.

Taylor, G. R., Wittington, J. A., and Grier, H. J. (2000). Age growth, maturation and protandric sex reversal in the common snook *Centropomus undecimalis*, from the east and west coasts of South Florida. *Fish. Bull.* 98, 612–624.

Thompson, K. A., Switzer, T. S., Christman, M. C., Keenan, S. F., Gardner, C. L., Overly, K. E., et al. (2022). A novel habitat-based approach for combining indices of abundance from multiple fishery-independent video surveys. *Fish. Res.* 247, 106178. doi: 10.1016/j.fishres.2021.106178

Trotter, A. A., Blewett, D. A., Taylor, R. G., and Stevens, P. W. (2012). Migrations of common snook from a tidal river with implications for skipped spawning. *Trans. Am. Fish. Soc* 141, 1016–1025. doi: 10.1080/00028487.2012.675903

Ultang, O. (1977). *Methods of measuring stock abundance other than by use of commercial catch and effort data* (Rome: FAO). FAO Fisheries Technical Paper No. 176.

Wauchope, H. S., Amano, T., Sutherland, W. J., and Johnston, A. (2019). When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods Ecol. Evol.* 10 (12), 2067–2078. doi: 10.1111/2041-210X.13302

White, E. R. (2019). Minimum time required to detect population trends: the need for long-term monitoring programs. *BioScience* 69 (1), 40–46. doi: 10.1093/biosci/biy144

Wilson, J. K., Stevens, P. K., Blewett, D. A., Boucek, R., and Adams, A. J. (2022). A new approach to define an economically important fish as an umbrella flagship species to enhance collaborative stakeholder–management agency habitat conservation. *Environ. Biol. Fishes* 106 (2), 237–254. doi: 10.1007/s10641-022-01214-y

Winner, B. L., Blewett, D. A., McMichael, R. H. Jr., and Guenther, C. B. (2010). Relative abundance and distribution of common snook along shoreline habitats of Florida estuaries. *Trans. Am. Fish. Soc* 139, 015079. doi: 10.1577/T08-215.1

Wolfe, D. A., Champ, M. A., Flemer, D. A., and Mearns, A. J. (1987). Long-term biological data sets: their role in research, monitoring, and management of estuarine and coastal marine systems. *Estuaries* 10, 181–193. doi: 10.2307/1351847

Young, J. M., Yeiser, B. G., and Whittington, J. A. (2014). Spatiotemporal dynamics of spawning aggregations of common snook on the east coast of Florida. *Mar. Ecol. Prog. Ser.* 505, 227–240. doi: 10.3354/meps10774

Zieman, J. C., and Zieman, R. T. (1989). *The ecology of the seagrass meadows of the west coast of Florida: a community profile*. US ( Department of the Interior, Fish and Wildlife Service, Research and Development) 85 (7).

zu Ermgassen, P. S. E., DeAngelis, B., Gair, J. R., zu Ermgassen, S., Baker, R., Daniels, A., et al. (2021). Estimating and applying fish and invertebrate density and production enhancement from seagrass, salt marsh edge, and oyster reef nursery habitats in the Gulf of Mexico. *Estuaries Coasts* 44, 1588–1603. doi: 10.1007/s12237-021-00935-0

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 03 January 2025
DOI 10.3389/fmars.2024.1426756

Check for updates

OPEN ACCESS

EDITED BY
Fabio Campanella,
National Research Council (CNR), Italy

REVIEWED BY
Bradley P. Harris,
Alaska Pacific University, United States
Walter Bubley,
South Carolina Department of Natural Resources, United States

*CORRESPONDENCE
Jill M. Hendon
Jill.hendon@usm.edu

†These authors contributed equally to this work and share first authorship

RECEIVED 02 May 2024
ACCEPTED 06 December 2024
PUBLISHED 03 January 2025

CITATION
Hendon JM, Hoffmayer ER, Pollack AG, Mareska J, Martinez-Andrade F, Rester J, Switzer TS and Zuckerman ZC (2025) Impacts of survey design on a Gulf of Mexico bottom longline survey and the transition to a unified, stratified - random design.
*Front. Mar. Sci.* 11:1426756.
doi: 10.3389/fmars.2024.1426756



# Impacts of survey design on a Gulf of Mexico bottom longline survey and the transition to a unified, stratified - random design

Jill M. Hendon[1*†], Eric R. Hoffmayer[2†], Adam G. Pollack[2†], John Mareska[3], Fernando Martinez-Andrade[4], Jeff Rester[5], Theodore S. Switzer[6] and Zachary C. Zuckerman[7]

[1]Center for Fisheries Research and Development, The University of Southern Mississippi, Ocean Springs, MS, United States, [2]Mississippi Laboratories, National Marine Fisheries Service, Southeast Fisheries Science Center, Pascagoula, MS, United States, [3]Marine Resources Division, Alabama Department of Conservation and Natural Resources, Dauphin Island, AL, United States, [4]Coastal Fisheries Division, Texas Parks and Wildlife Department, Corpus Christi, TX, United States, [5]Gulf States Marine Fisheries Commission, Ocean Springs, MS, United States, [6]Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, St. Petersburg, FL, United States, [7]Louisiana Department of Wildlife and Fisheries, Grand Isle, LA, United States

The Southeast Area Monitoring and Assessment Program (SEAMAP) Bottom Longline (BLL) survey was established to provide a nearshore complement to the offshore National Marine Fisheries Service (NMFS) BLL survey. SEAMAP state partners (i.e., Texas, Louisiana, Mississippi, and Alabama) used identical gear and sampling protocol to NMFS; however, temporal window, sampling universe, sampling frequency, and station selection were determined independently by each state based on available resources and capabilities. Although each state collected high quality data, the lack of a unified design complicated the efforts to combine state partner data to develop an index of abundance for stock assessment purposes. To improve the value of the survey and prioritize the quality and utility of the resultant data, the SEAMAP BLL survey was modified to implement a unified design that included consistency in spatial coverage and sampling frequency, and proportional allocation of sampling effort. Data from the early (2008-2014) and modified (2015-2021) SEAMAP BLL surveys, and from the modified SEAMAP and NMFS surveys were compared to determine the effects of this unified design on data precision. Overall catch composition slightly differed between the early and modified SEAMAP BLL surveys; however, taxa with declined abundance under the modified SEAMAP BLL survey were adequately sampled by the complementary NMFS BLL survey. Size composition was compared for three managed species. The size composition of Atlantic Sharpnose Sharks and Blacktip Sharks differed significantly between the modified SEAMAP BLL survey and the NMFS BLL survey, indicating that the modified survey is indeed providing complementary data. Further, implementing the modified design reduced the coefficient of variation for the indices of abundance for both Blacktip Sharks and Red Drum. The evolution of this survey highlights the benefit of unifying survey designs that build upon existing efforts to enhance the utility of survey data for multiple applications.

KEYWORDS

**fishery-independent survey, coefficient of variation, multispecies, standardization, stock assessment, Atlantic Sharpnose Shark, Blacktip Shark, Red Drum**

## Introduction

All survey methodologies attempt to understand a population by assessing a representative sample (Cadima et al., 2005), and all methodologies have inherent biases that can impact the data collected. These biases should be identified and controlled for, to the greatest extent practicable, when designing the survey so that their impact can be accounted for in multi-model assessments. Best survey designs encompass the full spatial scope of the target species and conduct sampling during the full temporal period of the target species' presence. Any time these frameworks are limited or are sampled with unequal effort, the full population may be misrepresented in the sample (Hansen et al., 2007). For this reason, standardized survey methodologies with a spatially balanced sampling design are essential for producing improved accuracy and precision of key population dynamic metrics (Cheng et al., 2024).

Fish populations are not evenly distributed within a system and their range may change based on environmental parameters, habitat availability, food availability, reproductive and ontogenetic phases, and/or pressures and stressors, thereby violating the statistical assumption of independence (Pennington and Strømme, 1998; Perry et al., 2005; Blanchard et al., 2008; Nye et al., 2009). Fishers are also dynamic, altering fishing methods, gear types, fishing time/location/effort, or target species to maximize harvest (Simpfendorfer et al., 2002) under a constantly evolving management framework. Due to the complexities of these dynamics, fishery dependent data may not accurately track the status of managed fish populations (de Mutsert et al., 2008; Pennington and Godø, 1995; Pennington and Strømme, 1998). Therefore, fishery independent surveys are invaluable data for assessing stocks (Pennington and Strømme, 1998; Wilberg et al., 2010).

Fishery independent surveys monitor diversity and abundance within an area (Xu et al., 2015) as limited by the selectivity and catchability of the gear being used and the design of the underlying sampling frame (Gunderson, 1993; Rago, 2005; Miller et al., 2007; Liu et al., 2009). These data complement other data inputs (e.g., fishery dependent data, life history data) in stock assessment models to discern population biomass and trends. Fishery independent data are often the most statistically robust assessment inputs, as their standardization, continuity, and random stratified designs result in relative abundance estimates with comparably high precision and low uncertainty (Pennington and Strømme, 1998; Miller et al., 2007). These surveys are of greatest utility when they cover the full spatial distribution of the stock being assessed (Walters, 2003; Wilberg et al., 2010; Gunderson, 1993) and are designed with sufficient statistical power so that as fish populations/distributions shift, changes in relative abundance can be detected (Pennington, 1985; Kimura and Somerton, 2006; Wang et al., 2018; Wilberg et al., 2010; Grace et al., 2012). As depth is a known driver of fish distribution, multispecies surveys typically implement a depth stratified random sampling effort to survey designs (Hansen et al., 2007). This ensures that effort is partitioned among heterogeneous strata based on standardized allocation criteria (Raj, 1968; Cochran, 1977; Gunderson, 1993; Smith et al., 2011; Richards et al., 2016; Ault et al., 2018).

Although valuable, fishery independent surveys are expensive and funding can often restrict the temporal or spatial coverage of the surveys (Dennis and Plagányi, 2015; Howard et al., 2023). To address these limitations, multiple sources of data are often integrated to attempt to capture the full range of species being assessed, especially for highly migratory species such as sharks (Simpfendorfer et al., 2002). The multi-sourced data must be standardized to ensure that the time series data are compatible (e.g., Maunder and Punt, 2004; Francis, 2011; Grüss et al., 2019). With this concept in mind, fishery independent surveys are often modeled after established well designed surveys to allow for effective comparisons and/or combinations and result in a more comprehensive and representative assessment. These standardized datasets can further allow for a more accurate assessment of population dynamic changes as related to broad issues such as climate change, management actions, and migratory pathway dynamics (Bonar et al., 2009). Survey compilations, such as the one published by Grüss et al. (2018) for the northern Gulf of Mexico (GOM), can help in determining what surveys are active and how to best leverage existing resources without duplicating efforts.

The Southeast Area Monitoring and Assessment Program (SEAMAP) is a federal/state/university collaboration that focuses on collecting and disseminating fisheries independent data. Gulf of

**Abbreviations:** BLL, bottom longline; CPUE, catch per unit effort; CV, coefficient of variation; FMP, federal management plan; GOM, Gulf of Mexico; NMFS, National Marine Fisheries Service; SEAMAP, Southeast Area Monitoring and Assessment Program; SEDAR, Southeast Data and Assessment Review.

Mexico SEAMAP partners recognized that the existing National Marine Fisheries Service Bottom Longline (NMFS BLL) survey, which targets Atlantic shark species, only sampled depths greater than 9 m in the northern GOM due to vessel limitations. As there are concerns that many shark stocks in the region are in decline (Stone et al., 1998), and many shark species are common to the coastal region, the nearshore waters (waters <9 m) were a notable gap in the survey. Further, as the NMFS BLL survey data have proven to be informative to the assessment of ten stocks in the GOM (e.g., SEDAR 29, 2012; SEDAR 34, 2013; SEDAR 54, 2017; SEDAR 77, 2024), the SEAMAP program hoped to fill this data gap in the coastal waters. Therefore, a complementary fishery independent BLL survey was initiated in waters off Texas, Louisiana, Mississippi, and Alabama (hereafter SEAMAP BLL survey). The intended goal of the SEAMAP BLL survey was to conduct a fisheries independent, gear standardized, survey in GOM coastal waters that generated data useful for fisheries assessment and management.

While exploring the utility of SEAMAP BLL data for use stock assessments, the Southeast Data Assessment and Review (SEDAR) identified several limitations. Although the sampling protocol was standardized, the variation in other survey design parameters made the datasets challenging to combine. The Texas data was spatially disjunct from the other state sampling universes, resulting in concern that the combined data would not reflect the same population trends as the other states (SEDAR 34, 2013) (Figure 1B). Louisiana sampled depths deeper than other states, resulting in spatial overlap with the NMFS BLL Survey (Figure 1C). Finally, the higher sampling intensity within Mississippi and Alabama waters (Figure 1D) could disproportionately drive population-level trends by artificially lowering coefficients of variation (CV) for target species. Ultimately, various *post hoc* weighting mechanisms and complex analyses were needed to generate an index of abundance (Hoffmayer et al., 2013a). Weighting, however, can be subjective and could lead to unintended consequences, especially when the parameter estimates are conflicting (Francis, 2011; Thorson et al., 2017). Further *post hoc* weighting is not always able to account for survey design shortfalls (Gunderson, 1993).

The SEAMAP partners made plans to solve these complications with a modified survey design (Christman, personal communication) and hopefully improve index of abundance precision and lower variance. Accordingly, the state partners integrated survey efforts under a unified, spatially balanced survey design. This sampling design has the advantages that 1) stratification improves the precision of parameter estimates by subdividing a heterogeneous population into relatively homogeneous strata and effectively partitioning population variance (Smith et al., 2011; Richards et al., 2016; Ault et al., 2018), and 2) assures that sampling effort is appropriately assigned to all strata.

This paper follows the management and evolution of the regional, multi-partner, fishery independent SEAMAP BLL survey. Data from the original survey design, hereafter referred to as “early” (2008-2014), was compared to data generated from the GOM-wide unified design, hereafter referred to as “modified” (2015-2021), as well as to the NMFS BLL survey. Herein, we explore the impacts of an independent (early) versus unified (modified) survey design (i.e., spatial scope, effort allocation, frequency) on survey statistics and estimates in a Gulf of Mexico bottom longline survey.

## Methods

### Independent state designs: early BLL design (2008-2014)

The SEAMAP BLL Survey began in 2008 when Mississippi and Texas started sampling their respective coastal waters and the waters off eastern Louisiana (Chandeleur Sound). Alabama joined

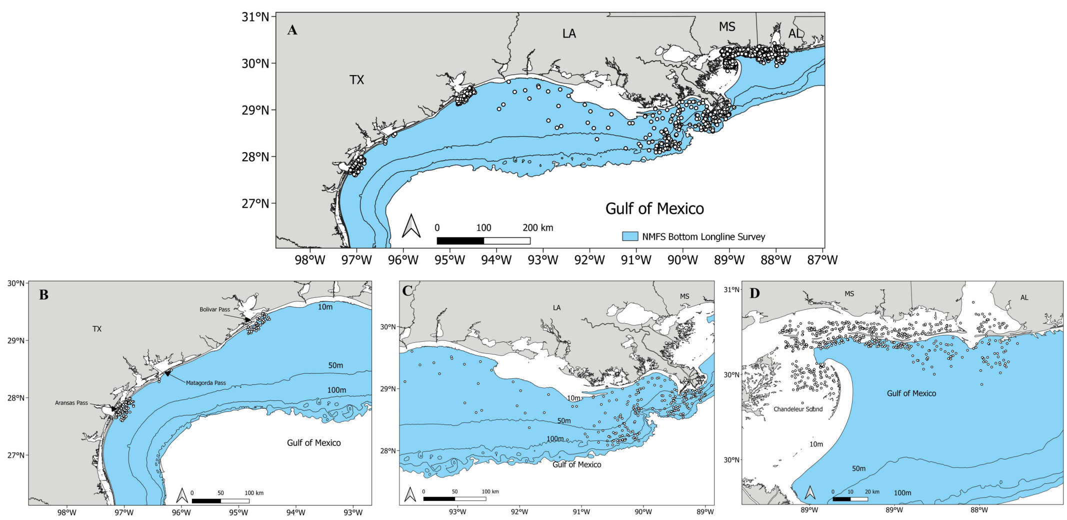



FIGURE 1
Maps of the sampling effort for the early Southeast Area Monitoring and Assessment Program Bottom Longline Survey in the northern Gulf of Mexico for stations conducted by **(A)** all SEAMAP state partners, **(B)** Texas (TX), **(C)** Louisiana (LA), and **(D)** Mississippi (MS) and Alabama (AL), during the early sampling period (2008-2014). White dots represent a sampling location. The blue region represents the universe covered by the National Marine Fisheries Service Bottom Longline Survey.

10.3389/fmars.2024.1426756

the survey efforts in 2010 and Louisiana began in 2011. All partners used sampling methodology identical to that of the NMFS BLL Survey, which involved the deployment of a 1.85 km main longline, weighted to the bottom at the beginning, middle, and endpoints, outfitted with one hundred 15/0 circle hooks baited with Atlantic Mackerel, *Scomber scombrus*, on 3.66 m gangions, fished for 60 minutes (Grace and Henwood, 1997; Driggers et al., 2008). Catch data (e.g., species, length, weight, sex, etc.) and environmental data (e.g., surface, middle, and bottom temperature, salinity, dissolved oxygen) were collected at each site. Although fishing protocol and sampling window (March through October) were standardized, each state partner independently determined the size of the sampling universe, the number of random stations conducted, the frequency of sampling, and the depth range sampled based on logistical and funding limitations (Table 1; Figure 1).

## Transition to a unified design: modified BLL design (2015-2021)

The first change implemented was a revision of the depth boundaries for the SEAMAP BLL Survey. It was known that the NMFS BLL survey implements a 9 m minimum depth limit due to the draft limitations of their vessels, and the BLL gear requires a minimum of 3 m for effective fishing. Therefore, SEAMAP set the modified design depth range as 3 – 10 m to ensure inclusion of unsampled waters while minimizing overlap with the NMFS BLL survey (9 - 366 m).

The second modification to the SEAMAP BLL survey was a spatial expansion of the survey universe and redistribution of sampling effort. In the modified design, the entire 3 - 10 m coastal contour from the Mexico-Texas border to just east of the Alabama-Florida border was eligible in the universe. This was a much larger sampling area and, therefore, required a station allocation protocol that would result in sufficient statistical power in the resultant data. Accordingly, the sampling effort was allocated among NMFS statistical reporting zones (zones 10 – 21) based on the proportion of the total universe 3 – 10 m depth contour present in each zone (Figure 2). Since the 3 – 10 m depth stratum is smaller in some zones relative to others, each statistical zone was allocated a minimum of two sampling stations to ensure that a measure of variability could be estimated. The random stratified design with proportional allocation ensured that the heterogeneity of the universe would be captured *a priori* through the station selection process.

The final change to the SEAMAP BLL survey defined the temporal sampling window. In the early SEAMAP design, the partners conducted the work from March to October with either seasonal or monthly sampling efforts when sharks and Red Drum were prevalent in coastal waters. The NMFS BLL survey, however, is only conducted from August through September due to ship availability. It was therefore decided that the nearshore effort would involve a seasonal sampling strategy where sampling effort was allocated among three sampling seasons: Spring (April-May), Summer (June-July), and Fall (August-September). This ensured that, 1) there was consistency in the sampling frequency by maintaining the majority of the original sampling period in the new design, and 2) the Fall survey period would directly correspond to the NMFS offshore survey (and the SEAMAP data could be truncated to that, if necessary).

Under the modified SEAMAP design, fifty-five stations were randomly selected per season throughout the statistical zones (10 – 21) in the 3 – 10 m depth stratum. The largest sampling area occurred off the Louisiana coast, accounting for 74.7% of the sampling universe (10,300 km$^2$), followed by Texas 12.7% (1,742 km$^2$), Mississippi 7.5% (1,040 km$^2$), and Alabama 5.1% (700 km$^2$). The proportional allocation resulted in 37 stations sampled off Louisiana, 10 off Texas, 5 off Mississippi, and 3 off Alabama during each season, totaling 165 stations completed each year. Each SEAMAP state partner was primarily responsible for conducting efforts off their respective coastlines, although sampling often extended into neighboring states when logistically efficient.

## Comparison between early and modified SEAMAP BLL survey designs

During the early (2008 – 2014) SEAMAP period, states joined the survey in different years and expanded and/or shifted their universe annually as their survey capability matured. To allow for comparison of these spatial scopes, the sampled area for each state partner was calculated by year and a mean and standard error were determined.

Statistical analyses were conducted to compare depth data between the early (2008 – 2014) and modified (2015 – 2021) SEAMAP designs to quantify how the design change impacted the fish assemblages sampled. The mean depth sampled was compared between states and by survey design using a two-way analysis of variance (ANOVA; Zar, 2010).

**TABLE 1** Comparison of the early Southeast Area Monitoring and Assessment Program Bottom Longline Survey (2008-2014) by Gulf of Mexico SEAMAP partner.

| SEAMAP Partner | Year Started | Mean Universe Size (km$^2$) | SE | Mean No. of Stations | SE | Depth Range (m) | Sampling Frequency |
|---|---|---|---|---|---|---|---|
| AL | 2010 | 2,405.4 | 0.0 | 31.0 | 4.3 | 2 – 26.8 | Monthly (Mar-Oct) |
| MS | 2008 | 1,073.9 | 91.7 | 62.0 | 5.3 | 2.5 – 16.4 | Monthly (Mar-Oct) |
| LA | 2011 | 47,157.4 | 16,652.3 | 54.5 | 16.8 | 3 – 332.2 | Seasonally |
| TX | 2008 | 1,453.6 | 212.7 | 19.5 | 2.0 | 1.7 – 25.0 | Seasonally |

SEAMAP Partner: state conducting the sampling, Alabama (AL), Mississippi (MS), Louisiana (LA), and Texas (TX); Year Started: initial year; Universe (km$^2$): mean sampling universe size from initial year through 2014; No. of Stations: mean number of stations per year from initial year through 2014; Depth Range (m): depth range of sampled stations; and Sampling Frequency: monthly or seasonal. As state protocols changed over time, mean area and mean number of stations are reported with standard error (SE).

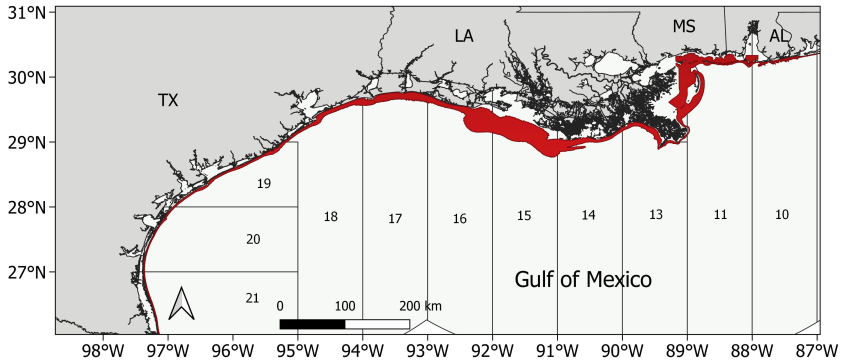



FIGURE 2
Map of the National Marine Fisheries Service Statistical Reporting Zones (labeled 10-21) in the northern Gulf of Mexico used to proportionally allocate sampling effort for the Southeast Area Monitoring and Assessment Botom Longline Survey. Effort was allocated based on the proportion of the total 3 - 10 m depth contour (red) present in each zone.

A multivariate analysis compared the catch composition of the early and modified survey design periods. Catch per unit effort (CPUE) for all sets with positive catch were 4th-root-transformed to down-weight the contribution of highly abundant taxa. Next, an analysis of similarity (ANOSIM) was conducted on the Bray-Curtis similarity matrix (Clarke and Warwick, 2001) to compare catch composition between the early and modified SEAMAP designs. ANOSIM produces an R statistic where values of 0 indicate that groups are not distinct from the entire dataset, while values of 1 indicate that groups of samples are completely distinct; the p-value indicates the significance of this statistic. Further, a similarity percentage analysis (SIMPER) was conducted to identify the species that discriminate between the compared surveys (Warwick et al., 1988). The SIMPER procedure compares the average abundances per design and examines the contribution of each species to the average Bray–Curtis dissimilarity. Both ANOSIM and SIMPER non-parametric statistical tests were carried out using PRIMER v.7 (Clarke and Gorley, 2015). Since there was a significant depth effect when comparing the early to the modified SEAMAP designs, the deeper stations (>15 m) within the early design were removed when comparing catch composition and length data. In addition, since Louisiana did not fully participate in the SEAMAP BLL survey until 2012, the years 2012 - 2014 were used to represent the early design period, and the years 2015 - 2017 were used to represent the modified design period. Catch composition was also compared between the modified SEAMAP design and the NMFS BLL survey for the years 2015 - 2021 using ANOSIM and SIMPER analyses. Since the NMFS survey only occurred during the months of August and September, the modified SEAMAP data was further truncated to this time period.

Three important species under federal management plans (FMP), Atlantic Sharpnose Shark, *Rhizoprionodon terraenovae*, Blacktip Shark, *Carcharhinus limbatus*, and Red Drum, *Sciaenops ocellatus*, were investigated for potential changes in abundance and length frequency. Catch per unit effort was compared between the early and modified SEAMAP designs and between the modified SEAMAP and NMFS BLL surveys (2015-2021) using a t-test (Zar, 2010). Mean length (shark by sex: Fork Length, FL; teleost: Total Length, TL) and length distributions were compared for both early and modified SEAMAP designs using t-tests and Kolmogorov-Smirnov (K-S) tests, respectively (Zar, 2010). Mean length and length distributions were compared for the modified SEAMAP design and the NMFS BLL survey for the August-September 2015 - 2021 temporal scope, using t-tests and K-S tests, respectively.

Further, to determine if the changes to the SEAMAP modified design improved the relative abundance estimates for the three species, a delta-lognormal model was used to generate relative abundance indices for each assessed FMP species. An index of relative abundance was built using a combined dataset consisting of the early (2008 – 2014) and the modified (2015 – 2021) SEAMAP data following the method outlined in Ingram et al. (2017), as this is a common method used in stock assessments in the southeast region. The early data were truncated to only include stations with depths from 3 - 15 m as this range matched the modified SEAMAP design dataset. Factors that were included in the initial model run included depth (m), bottom dissolved oxygen (mg/l), bottom salinity (psu), bottom temperature (°C), SEAMAP partner (Texas, Louisiana, Mississippi, and Alabama), month (March-October), and year. The submodels of the delta-lognormal model were built using a backward selection procedure based on type III analyses with an inclusion level of significance of $\alpha = 0.05$ (Lo et al., 1992). Coefficients of variance around the annual abundance estimates were compared (Lo et al., 1992) as a performance metric to assess the effect of the modified survey design.

## Results

A total of 1,866 BLL sets were completed by the SEAMAP BLL survey from 2008 to 2021, with 920 completed under the early SEAMAP design (2008 – 2014) and 946 completed under the modified SEAMAP design (2015 – 2021; Figures 1 and 3). With the early SEAMAP design, 53% of the stations spatially overlapped the NMFS BLL sampling universe, while 19.2% overlapped both spatially and temporally (August-September) (Figure 1). However, with the modified SEAMAP design, only 20% of the stations

10.3389/fmars.2024.1426756

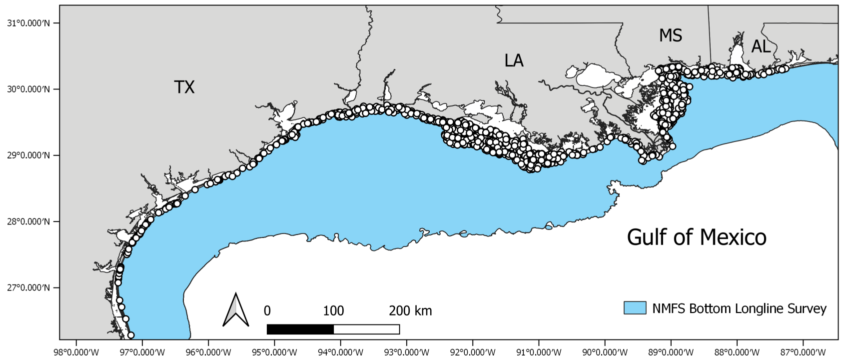



FIGURE 3
Map of the modified Southeast Area Monitoring and Assessment Program Botom Longline Survey in the northern Gulf of Mexico for the sampling period 2015-2021. White circles represent sampling locations conducted by SEAMAP partners. The blue region represents the universe covered by the National Marine Fisheries Service Botom Longline Survey.

overlapped spatially with the NMFS sampling universe (Figure 3), of which only 6.9% overlapped both spatially and temporally (August-September).

When comparing the two SEAMAP BLL survey designs, the biggest difference observed was in the depths sampled. There was a significant difference in station depth across state partners ($F_{3,1847}$ = 343.7, $p < 0.001$) and between the early and modified sampling designs ($F_{1,1849}$ = 272.5, $p < 0.001$). Stations sampled under the early SEAMAP design were significantly deeper (mean: 26.4 ± 0.7 m, range: 1.7 - 332.2 m) compared to those sampled under the modified SEAMAP design (mean: 7.0 ± 0.9 m, range: 1.4 – 15.1 m). These differences were most pronounced off Louisiana where station depths between the early (mean: 74.0 ± 1.4 m) and modified (mean: 6.5 ± 0.9 m) designs were significantly different from each other ($p < 0.05$).

From 2008 to 2021, 34,589 organisms were caught on the SEAMAP BLL survey with 15,944 caught during the early design and 18,645 caught during the modified design. During the early design, 73.0% of the total catch consisted of: Atlantic Sharpnose Sharks (46.9%), Blacktip Sharks (12.2%), Gafftopsail Catfish, *Bagre marinus* (8.6%), and Red Drum (5.4%). With the modified survey design, the top species present were similar, but their frequency of occurrence was substantially different, with 82.6% of the catch made up of Gafftopsail Catfish (34.1%), Blacktip Sharks (24.6%), Atlantic Sharpnose Sharks (15.2%), Red Drum (8.7%), and Bull Sharks, *C. leucas* (5.5%). Six species were caught in relatively high numbers with the early survey design that were caught in lower numbers (n < 30) or were absent in the modified design, including: Smoothhounds, *Mustelus* spp. (early: n = 680; modified n = 0), Red Snapper, *Lutjanus campechanus* (early: n = 584; modified: n = 0), King Snake Eels, *Ophichthus rex* (early: n = 369; modified n = 1), Tiger Sharks, *Galeocerdo cuvier* (early: n = 103; modified: n = 20), Scalloped Hammerheads, *Sphyrna lewini* (early: n = 94; modified: n = 9), and Sandbar Sharks, *C. plumbeus* (early: n = 81; modified: n = 30).

Catch composition was slightly different between the early (station n = 269) and modified (station n = 433) SEAMAP BLL survey designs (R = 0.044, p = 0.001). The SIMPER analysis revealed that Atlantic Sharpnose Shark (17.3%), Blacktip Shark (14.6%), Gafftopsail Catfish (13.2%), Red Drum (10.1%), Bull Shark (9.6%), and Southern Stingray (7.0%) accounted for the majority of the dissimilarity (71.7%) in species composition between the survey designs (Table 2). Catch composition was also found to be significantly different between the modified SEAMAP (station n = 303) and the NMFS BLL (station n = 444) surveys (R = 0.263, p = 0.001). The SIMPER analysis revealed that Blacktip Sharks (12.6%), Atlantic Sharpnose Sharks (11.9%), Gafftopsail Catfish (10.2%), Red Snapper (9.0%), Bull Sharks (7.2%), Blacknose Sharks (5.1%), Southern Stingrays (4.3%), Spinner Sharks (4.2%), Red Drum (3.9%), and Hardhead Catfish (3.6%) accounted for the majority of the difference (72.0%) in composition (Table 3).

There were differences in CPUE and length frequencies for the three FMP species examined when comparing the early and modified SEAMAP designs (Figure 4). Atlantic Sharpnose Shark catch rates were significantly higher with the early (CPUE: 6.6 ± 0.4 sharks/100 hook hrs, n = 275) compared to the modified (CPUE: 3.0 ± 0.2 sharks/100 hook hrs, n = 444) design (t = 9.3, $p < 0.001$). Female Atlantic Sharpnose Shark mean size was significantly larger during the early (mean: 587.5 ± 5.9 mm FL, range: 265 – 965 mm FL, n = 466) compared to the modified (mean: 534.9 ± 11.3 mm FL, range: 265 – 975 mm FL, n = 152; t = 4.2, $p < 0.001$; Figure 4A) SEAMAP design, and the length distributions were significantly different between the two designs (D = 0.209, $p < 0.001$). Male Atlantic Sharpnose Shark mean size was significantly smaller in the early (mean: 671.9 ± 1.7 mm FL, range: 255 – 931 mm FL, n = 3,081) compared to the modified (mean: 724.7 ± 2.2 mm FL, range: 273 – 890 mm FL, n = 2,086) SEAMAP design (t = 19.5, $p < 0.001$; Figure 4A), and the length distributions were significantly different between the early and modified designs (D = 0.365, $p < 0.001$).

Blacktip Shark catch rates were significantly higher with the modified (CPUE: 4.8 ± 0.2 sharks/100 hook hrs, n = 444) compared to the early SEAMAP design (CPUE: 2.4 ± 0.2 shark/100 hook hrs, n = 275; t = 8.0, $p < 0.001$). Female Blacktip Shark mean size was significantly larger with the modified (mean: 1,079.4 ± 4.7 mm FL, range: 331 – 1,574 mm FL, n = 2,484) compared to the early (mean: 927.4 ± 9.6 mm FL, range: 446 – 1,590 mm FL, n = 623) SEAMAP

TABLE 2 Catch composition similarity percentage analysis (SIMPER) results for early and modified Southeast Area Monitoring and Assessment Program (SEAMAP) botom longline survey designs.

| Species | Early Average Abundance | Modified Average Abundance | Average Dissimilarity | Diss/SD | Contribution (%) | Cumulative Contribution (%) |
|---|---|---|---|---|---|---|
| Atlantic Sharpnose Shark | 1.24 | 0.68 | 12.03 | 1.17 | 17.26 | 17.26 |
| Blacktip Shark | 0.88 | 0.99 | 10.19 | 1.08 | 14.62 | 31.88 |
| Gafftopsail Catfish | 0.44 | 0.69 | 9.16 | 0.85 | 13.15 | 45.03 |
| Red Drum | 0.41 | 0.49 | 7.06 | 0.84 | 10.13 | 55.16 |
| Bull Shark | 0.3 | 0.58 | 6.7 | 0.93 | 9.61 | 64.77 |
| Southern Stingray | 0.2 | 0.35 | 4.86 | 0.69 | 6.98 | 71.74 |

Average Abundance; Average Dissimilarity: average Bray-Curtis dissimilarity between the two surveys; Diss/SD: average dissimilarity divided by the standard deviation; Contribution %: percent contribution to the total average dissimilarity; Cumulative Contribution %: percent cumulative contribution to the total within-group dissimilarity.

design ($t = 14.3$, $p < 0.001$; Figure 4B), and the length distributions were significantly different ($D = 0.367$, $p < 0.001$). Similarly, male Blacktip Shark mean size was significantly larger for the modified (mean: 985.8 ± 5.3 mm FL, range: 355 – 1,524 mm FL, n = 1,512) compared to the early (mean: 873.6 ± 8.0 mm FL, range: 433 – 1,441 mm FL, n = 536, Figure 4B) SEAMAP design ($t = 11.1$, $p < 0.001$), and the length distributions were significantly different ($D = 0.345$, $p < 0.001$).

Red Drum catch rates were significantly higher with the modified (CPUE: 1.7 ± 0.1 fish/100 hook hrs, n = 444) compared to the early (CPUE: 1.2 ± 0.1 fish/100 hook hrs, n = 275; $t = 2.5$, $p < 0.0134$) SEAMAP design, and significantly larger fish were caught with the modified (n = 1,478, mean: 945.3 ± 1.5 mm TL, range: 492 – 1,180 mm TL) compared to the early design (n = 618, mean: 934.4 ± 2.3 mm TL, range: 669 – 1,090 mm TL; $t = 4.0$, $p < 0.001$; Figure 4C). In addition, the distribution of lengths was significantly different ($D = 0.0846$, $p = 0.004$).

When the length frequency data from the modified SEAMAP and NMFS survey datasets were compared for the FMP species, some interesting differences were observed (Figure 5). Fewer Atlantic Sharpnose Sharks were caught on the modified SEAMAP survey (n = 485) than the NMFS survey (n = 2,247). This was especially true for female Atlantic Sharpnose Sharks, where sharks with significantly smaller mean size were caught with the modified SEAMAP design (mean: 469.7 ± 14.3 mm FL, range: 340 - 672 mm FL, n = 42) as compared to the NMFS BLL (mean: 781.9 ± 2.4 mm FL, range: 389 – 990 mm FL, n = 1,177) survey ($t = 23.643$, $p < 0.001$; Figure 5A). Length distributions were also significantly different ($D = 0.904$, $p < 0.001$). Similarly, male Atlantic Sharpnose Shark mean size was significantly smaller for the modified SEAMAP BLL survey (mean: 681.6 ± 5.6 mm FL, range: 356 – 850 mm FL, n = 373) compared to the NMFS BLL survey (mean: 764.6 ± 2.4 mm FL, range: 310 – 958 mm FL, n = 1,069, Figure 5A) survey ($t = 16.007$, $p < 0.001$), and the length distributions were also significantly different ($D = 0.503$, $p < 0.001$).

TABLE 3 Catch composition similarity percentage analysis (SIMPER) results for modified Southeast Area Monitoring and Assessment Program (SEAMAP) and National Marine Fisheries Service (NMFS) botom longline survey designs.

| Species | SEAMAP Average Abundance | NMFS Average Abundance | Average Dissimilarity | Diss/SD | Contribution (%) | Cumulative Contribution (%) |
|---|---|---|---|---|---|---|
| Blacktip Shark | 0.89 | 0.32 | 11.14 | 1.07 | 12.58 | 12.58 |
| Atlantic Sharpnose Shark | 0.39 | 0.83 | 10.54 | 0.97 | 11.9 | 24.48 |
| Gafftopsail Catfish | 0.76 | 0.16 | 9.03 | 0.83 | 10.2 | 34.68 |
| Red Snapper | 0 | 0.67 | 7.94 | 0.78 | 8.97 | 43.64 |
| Bull Shark | 0.53 | 0.14 | 6.37 | 0.84 | 7.2 | 50.84 |
| Blacknose Shark | 0.06 | 0.32 | 4.54 | 0.55 | 5.12 | 55.97 |
| Southern Stingray | 0.32 | 0.02 | 3.82 | 0.56 | 4.32 | 60.28 |
| Spinner Shark | 0.23 | 0.13 | 3.74 | 0.54 | 4.22 | 64.51 |
| Red Drum | 0.31 | 0.04 | 3.47 | 0.51 | 3.92 | 68.43 |

Average Abundance; Average Dissimilarity: average Bray-Curtis dissimilarity between the two surveys; Diss/SD: average dissimilarity divided by the standard deviation; Contribution %: percent contribution to the total average dissimilarity; Cumulative Contribution %: percent cumulative contribution to the total within-group dissimilarity.

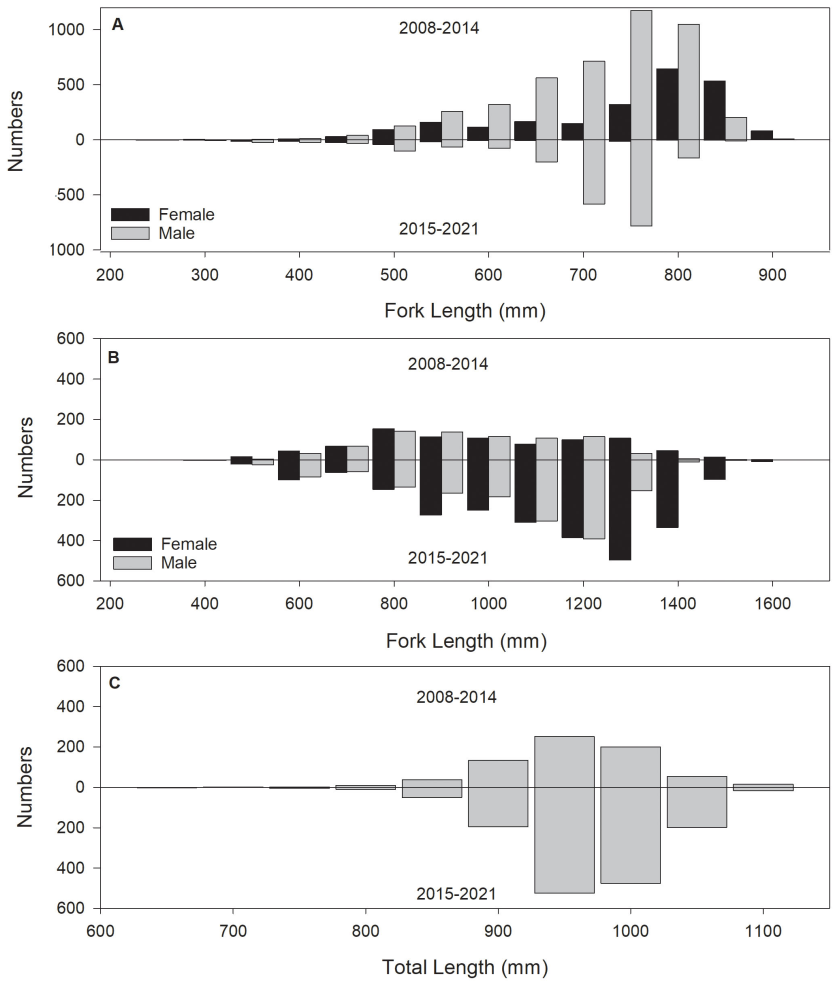



FIGURE 4
Length frequency distribution for **(A)** female (black) and male (gray) Atlantic Sharpnose Shark, *Rhizoprionodon terraenovae*, **(B)** female (black) and male (gray) Blacktip Shark, *Carcharhinus limbatus*, and **(C)** Red Drum, *Sciaenops ocellatus*, from the early (2008-2014; top) and modified (2015-2021; bottom) Southeast Area Monitoring and Assessment Program Bottom Longline Survey design conducted in the northern Gulf of Mexico.

More Blacktip Sharks were caught during the modified SEAMAP BLL survey (n = 1,073) compared to the NMFS BLL survey (n = 590). Female Blacktip Shark mean size was significantly larger with the NMFS BLL survey (mean: 1,166.9 ± 9.8 mm FL, range: 593-1,530 mm FL, n = 349) compared to the modified SEAMAP BLL survey (mean: 1,043.8 ± 9.6 mm FL, range: 389 – 1,341 mm FL, n = 848) survey (t = 8.263, p < 0.001; Figure 5B), and the length distributions were significantly different (D = 0.214, p < 0.001). Similarly, male Blacktip Shark mean size was significantly larger for the NMFS BLL survey (mean: 1036.6 ± 9.7 mm FL, range: 645 – 1,346 mm FL, n = 189) compared to the modified SEAMAP BLL survey (mean: 944.5 ± 12.5 mm FL, range: 355 – 1,298 mm FL, n = 356, Figure 5B) survey (t = 5.101, p < 0.001), and the length distributions were significantly different (D = 0.263, p < 0.001).

Red Drum catch was relatively low within the NMFS BLL survey (n = 77) compared to the modified SEAMAP BLL survey (n = 400). There was no significant difference in the mean size of Red Drum for fish caught in the modified SEAMAP (mean: 932.5 ± 3.0 mm TL, range: 492 – 1,120 mm TL, n = 400) as compared to the NMFS (mean: 933 ± 2.2 mm TL, range: 669 – 1,090 mm TL, n = 705) survey (t = 0.203, p = 0.839). In addition, the distribution of lengths was not significantly different (D = 0.111, p = 0.409; Figure 5C).

Standardized relative abundance indices were generated for Atlantic Sharpnose Sharks, Blacktip Sharks, and Red Drum (Figure 6). For Atlantic Sharpnose Sharks, the final model retained year, SEAMAP partner, month, and bottom temperature, salinity and dissolved oxygen in the binomial submodel, whereas the lognormal submodel retained year, SEAMAP partner, month, depth, and bottom temperature and salinity as significant factors (Supplementary Table 1). For Blacktip Sharks, the final model retained the same factors in the binomial and lognormal submodels, which included year, SEAMAP partner, month, and bottom temperature and salinity as significant factors (Supplementary Table 2). For Red Drum, the final model retained year, SEAMAP partner, month, and bottom salinity and dissolved oxygen in the binomial submodel, and the lognormal submodel retained year, SEAMAP partner, depth, and bottom temperature as significant factors (Supplementary Table 3). The CVs generated from the modified SEAMAP BLL survey dataset were slightly higher for Atlantic Sharpnose Sharks than those generated from the early design (Figure 6). There was an increasing trend in CV

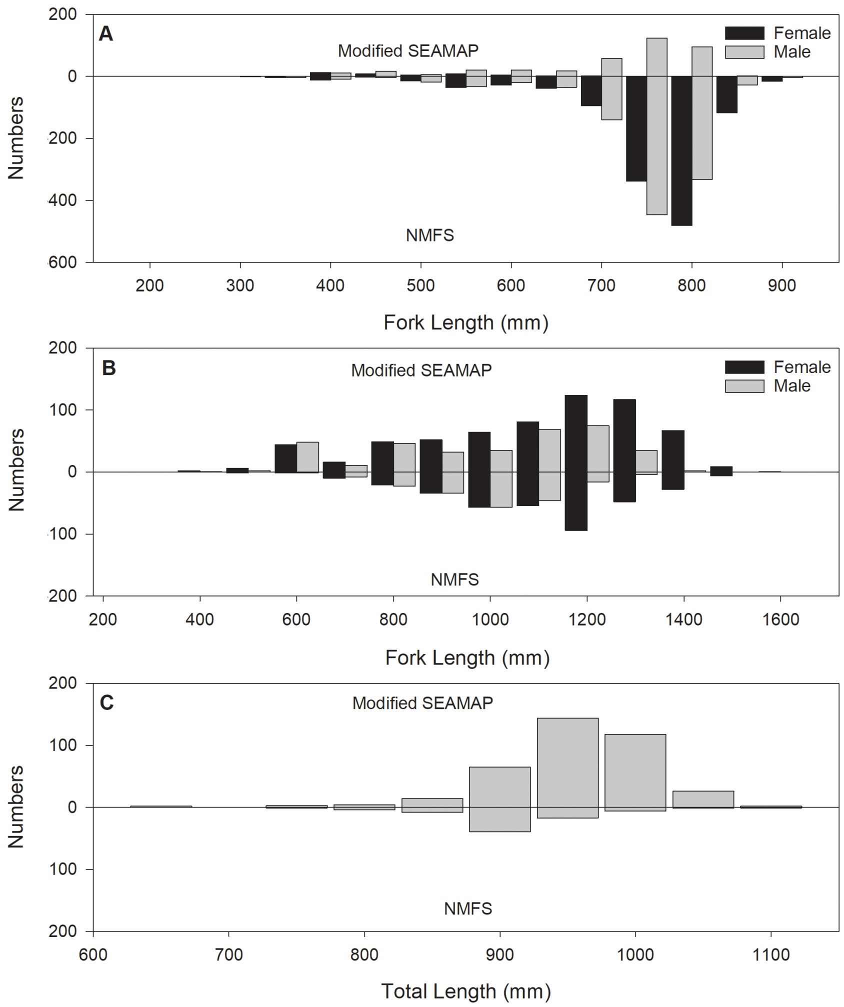



FIGURE 5
Length frequency comparison for **(A)** Atlantic Sharpnose Shark, *Rhizoprionodon terraenovae*, **(B)** Blacktip Shark, *Carcharhinus limbatus*, and **(C)** Red Drum, *Sciaenops ocellatus*, caught in the modified Southeast Area Monitoring and Assessment Program Bottom Longline Survey during August and September and the National Marine Fisheries Service Bottom Longline Survey from 2015-2021.

from 2013 -2021, corresponding with a decrease in total numbers and frequency of occurrence for this species. The CVs generated from the modified SEAMAP BLL design for Blacktip Sharks slightly improved and consistently remained low (Figure 6). Red Drum showed the most substantial CV reduction with the modified SEAMAP design (Figure 6).

## Discussion

Under the modified design, the restriction of the depth stratum allowed survey resources to be redistributed to prioritize the goal of evenly distributing station effort throughout the 3-10 m depth stratum. The survey was then able to fully complement half of the spatial scope (Mexico-Texas border to the east of the Alabama-Florida border) of the NMFS BLL survey with minimal overlap and fill a data gap for the target species. The proportional allocation of stations by depth and the distribution based on NMFS statistical reporting zones ensured a spatially balanced sampling design. Finally, the seasonal timing of the survey ensured that seasonal movement patterns of target species could be captured (Grace et al., 2012). This exclusion of the winter season further maximized survey resources as it is known that sharks move out of the coastal region during this time (Springer, 1940; Parsons and Hoffmayer, 2005; Peterson and Grubbs, 2024).

Most state partners were conducting the early SEAMAP BLL efforts within the boundary of the modified stratum simply due to logistics and resource limitations. Sampling conducted off Louisiana was the only state that showed a significant change in mean depth sampled between the early and modified design. This state had the capability of sampling deeper offshore waters, however, these data were one of the primary sources of complication during the SEDAR data combination. As 53% of early design stations conducted by Louisiana overlapped spatially with the NMFS BLL sampling universe, the most parsimonious solution was to restrict depth and increase coverage in the unified stratum. This depth restriction was a major factor in many of the effects seen in the survey comparisons.

The most common species encountered in both the early and modified SEAMAP surveys were relatively similar, with Atlantic

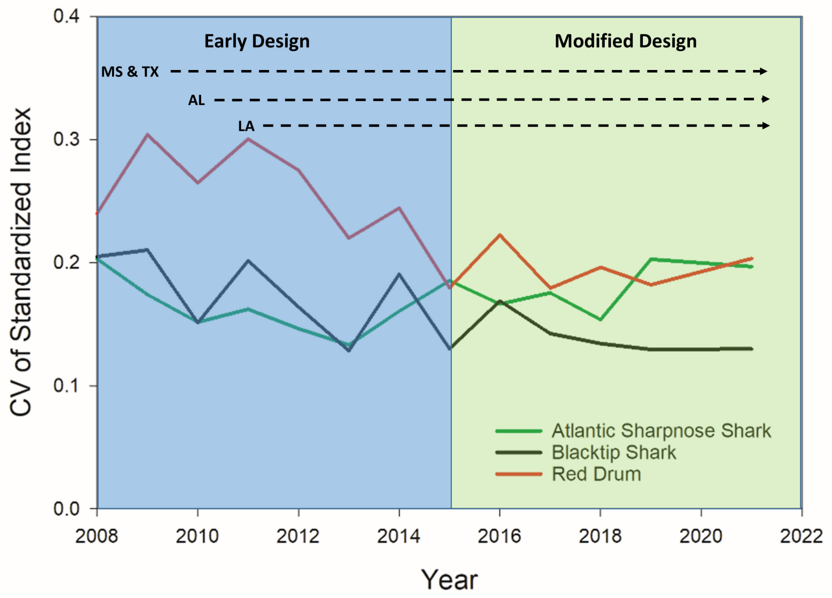



FIGURE 6
Coefficients of variation (CV) by year from standardized catch rates of Atlantic Sharpnose Shark, *Rhizoprionodon terraenovae* (green), Blacktip Shark, *Carcharhinus limbatus* (black), and Red Drum, *Sciaenops ocellatus* (red) using the delta-lognormal modeling method for the Southeast Area Monitoring and Assessment Program Bottom Longline survey early (blue region) and modified designs (light green region). Time of state partner participation in the survey is shown by the labeled dashed lines (MS, Mississippi; TX, Texas; AL, Alabama; LA, Louisiana).

Sharpnose Sharks, Blacktip Sharks, Gafftopsail Catfish, and Red Drum being predominant. There were six species, however, that were caught in noticeably lower numbers in the modified survey. This shift was likely a result of the depth restriction in the modified design. Smoothhounds, Red Snapper, and King Snake Eel are known deeper water species (McEachran and Fechhelm, 2005; Castro, 2011), while Tiger Sharks, Scalloped Hammerheads, and Sandbar Sharks are large coastal shark species that often partition to deeper habitats than the small coastal shark species (Castro, 2011). Several of these species are federally managed, including Red Snapper, Smoothhounds, Scalloped Hammerheads, and Sandbar Sharks, and their loss in the composition was concerning; however, as all are caught in relatively high numbers on the NMFS BLL survey, their population signature is still being captured. The ANOSIM also detected a significant difference in the composition between the early and modified designs; however, the Global R statistic confirmed that the effect is negligible due to the composition overlap. The dissimilarity was driven by Atlantic Sharpnose Sharks, Blacktip Sharks, Gafftopsail Catfish, and Red Drum, which were the common species in both designs. This small effect was likely due to the exclusion of the deeper Louisiana stations in the early design from the analysis, where the mature female Atlantic Sharpnose Sharks were caught in higher numbers.

When comparing the modified SEAMAP and NMFS survey, the ANOSIM detected a significant difference in the composition; although the Global R statistic inferred the effect was moderate due to some species overlap. The dissimilarity was driven by Blacktip Sharks, Atlantic Sharpnose Sharks, Gafftopsail Catfish, and Red Snapper. In this case, Red Snapper was the only species not present in the modified design and likely had the greatest effect on the R value.

Atlantic Sharpnose Shark showed the greatest shift between the early and modified designs. With the modified design, there was a substantial decrease in the overall catch rate, as well as a shift from a mixed catch of males and females to mainly males. Female Atlantic Sharpnose Sharks are known to remain in deeper offshore waters once mature (size at maturity: 620 mm FL; Hoffmayer et al., 2013b), and mature males return to coastal areas of the GOM each spring (Parsons and Hoffmayer, 2005; Hoffmayer et al., 2006), therefore this shift in the size composition is not surprising as the depth strata for the modified design was restricted to 3 -10 m. Comparison of the Atlantic Sharpnose Shark catch between the modified SEAMAP design and NMFS BLL survey corroborates this as the NMFS survey captures the offshore female Atlantic Sharpnose Shark signature while the modified SEAMAP survey captures the portion of adult males that are moving inshore.

This species was the only assessed species that showed an increase in CV with the modified design indicating a reduction in the frequency of occurrence for this species. This may be due to a gear bias selecting against the smallest Atlantic Sharpnose Shark size classes. Hoffmayer et al. (2013b) showed that Atlantic Sharpnose Sharks pup at approximately 350 mm resulting in a neonate shark with a mouth gape that could not consume a 15/0 hook with a large piece of bait. The length frequency histograms show that Atlantic Sharpnose Sharks are not caught in larger numbers until the 500 mm size class. There may also be a concerning trend occurring for this species that our analysis is capturing. The modified SEAMAP design showed a reduction in catch, which was surprising considering that the Atlantic Sharpnose Shark typically had the highest catch rates. However, looking at trends in catch across multiple GOM surveys since 2013, there has been a consistent decline in Atlantic Sharpnose shark catch (NMFS, unpub. data). Therefore, the increase in CV and the decrease in catch may highlight a true population-level shift in abundance.

Blacktip Sharks exhibited an increase in frequency of occurrence and relative abundance with the modified SEAMAP BLL design compared to the early design. The number of Blacktip Sharks caught was more than double, and included a similar mix of males and females, and length frequencies. This is not surprising since the survey's new depth boundary encompassed the species' known depth range (Compagno, 1984; Castro, 2011). The low catch numbers in the early design were likely driven by the imbalance of sampling effort across the different state partners. It was found that the modified SEAMAP BLL survey catches a larger number as well as a broader size range of Blacktip Sharks than the NMFS BLL survey and will ultimately become a better indicator for population abundance. Although the CVs for the Blacktip Shark's relative abundance index only slightly improved with the modified design, the increased catch that this survey will have as compared to the NMFS BLL survey will improve the precision for this species.

Red Drum exhibited an increased frequency of occurrence between the early and modified SEAMAP BLL survey designs. There were more than twice as many Red Drum caught across the adult size range with the modified design resulting in greatly improved CVs. As Red Drum is a coastal, shallow water species that prefers sand and mud bottoms in the GOM (McEachran and Fechhelm, 2005), these changes were likely driven by the restricted sampling depth with the modified design. Red Drum is a highly sought after recreational and commercial species that was recently assessed in the GOM in a data limited stock assessment (SEDAR 49, 2016) with only a single fishery independent dataset from a limited spatial area eligible for evaluation (SEDAR 49, 2016). It was recommended to either expand current inshore surveys that catch Red Drum or develop a new survey to characterize the relative abundance and size composition of Red Drum across the GOM (SEDAR 49, 2016). The modified SEAMAP BLL survey, especially with the recent expansion into Florida waters (2024), shows promise for generating a future index of abundance for Red Drum.

Multi-species surveys that target highly migratory species such as sharks will continually be faced with concern over the spatial and temporal universe covered by the survey, as catchability can be continually changing and influenced by many factors such as environmental, management, or biological processes (Simpfendorfer et al., 2002; Wilberg et al., 2010). Surveys are always designed with the best intentions for gathering data on the target species; however, the decision to change the design of a long-term survey should never be taken lightly. Instead, routine evaluation of the survey data to ensure that it continues to meet management needs is suggested (Bonar et al., 2009; Wang et al., 2018; Vecchio et al., 2023; Cheng et al., 2024).

The survey modification steps discussed herein are typical of the maturation process for a multi-agency survey. The transition to a spatially balanced, unified sampling design and the reduction of spatial overlap with concurrent surveys improved target species' relative abundance index CVs and reduced the potential redundancy in information provided by the respective BLL surveys. Although catch of some species were lost in the design change (i.e., Red Snapper, Smoothhound, Sandbar Shark; all still present in the NMFS survey), the change has resulted in higher catches of several FMP teleost and elasmobranch species that are highly important to commercial and recreational fisheries, including Blacktip Sharks, Bull Sharks, and Red Drum. Further, length frequencies for Atlantic Sharpnose Sharks and Blacktip Sharks showed that the modified SEAMAP BLL survey data was reflecting a different portion of the population than the NMFS BLL Survey.

With the addition of Florida to the modified SEAMAP BLL Survey in 2024, the SEAMAP BLL Survey will cover the entire spatial scope of the 3 – 10 m depth zone in the northern GOM. This consistent effort across the entire GOM basin will further improve the precision and utility of the data collected. Ultimately, these two surveys (SEAMAP and NMFS BLL) will represent the most comprehensive, standardized BLL dataset in the Gulf of Mexico.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.gsmfc.org/seamap-sis.

## Ethics statement

This study was conducted in accordance with the local legislation and institutional requirements. The study was approved by The University of Southern Mississippi Institutional Animal Care and Use Commitee: 9031202, 11092217, 18010502, 18010502.1.

## Author contributions

JH: Writing – original draft, Writing – review & editing. EH: Writing – original draft, Writing – review & editing. AP: Writing – original draft, Writing – review & editing. JM: Writing – review & editing. FM-A: Writing – review & editing. JR: Writing – review & editing. TS: Writing – review & editing. ZZ: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this survey was allocated by Congress through NOAA to the GOM SEAMAP partners. Funds are obligated annually via multi-year cooperative agreements. This paper focuses on the SEAMAP allocation that was designated to the Gulf states for the years 2008-2021 and only reports on the bottom longline survey efforts.

## Acknowledgments

The authors would like to thank R. Waller of The University of Southern Mississippi, as he was instrumental in directing the initial funding for the SEAMAP BLL Survey, the Southeast Area Monitoring and Assessment Program for funding, Gulf States Marine Fisheries Commission for coordination, and Southeast Data Assessment and Review for data consideration, M. C. Christman for survey analysis, and the staff of the Center for Fisheries Research and Development at The University of Southern Mississippi.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1426756/full#supplementary-material

## References

Ault, J. S., Smith, S. G., Richards, B. L., Yau, A. J., Langseth, B. J., O'Malley, J. M., et al. (2018). Towards fishery-independent biomass estimation for Hawaiian Islands deepwater snappers. *Fisheries Res.* 208, 321–328. doi: 10.1016/j.fishres.2018.08.012

Blanchard, J. L., Maxwell, D. L., and Jennings, S. (2008). Power of monitoring surveys to detect abundance trends in depleted populations: the effects of density-dependent habitat use, patchiness, and climate change. *ICES J. Mar. Sci.* 65, 111–120. doi: 10.1093/icesjms/fsm182

Bonar, S. A., Contreras-Balderas, S., and Iles, A. C. (2009). "An introduction to standardized sampling," in *Standard Methods for Sampling North American Freshwater Fishes*. Eds. S. A. Bonar, W. A. Hubert and D. W. Willis (American Fisheries Society, Bethesda, MD).

Cadima, E. L., Caramelo, A. M., Afonso-Dias, M., Conte de Barros, P., Tandstad, M. O., and de Leiva-Moreno, J. I. (2005). *Sampling Methods Applied to Fisheries Science: A Manual. FAO Fisheries Technical Paper No. 434* (Rome: FAO), 99.

Castro, J. I. (2011). *The sharks of North America* (New York: Oxford University Press), 640.

Cheng, W., Zhang, C., Ji, Y., Xue, Y., Ren, Y., and Xu, B. (2024). Performance evaluation of spatially balanced sampling designs in fishery independent surveys. *Fisheries Res.* 270, 106879. doi: 10.1016/j.fishres.2023.106879

Clarke, K. R., and Gorley, R. N. (2015). *Primer v. 7: User Manual and Tutorial* (Plymouth: PRIMER-E).

Clarke, K. R., and Warwick, R. M. (2001). *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation. 2nd ed* (Plymouth, UK: PRIMER-E Ltd), 203.

Cochran, W. G. (1977). *Sampling Techniques. 3rd ed* (New York: John Wiley & Sons).

Compagno, L. J. V. (1984). *Sharks of the World: An Annotated and Illustrated Catalogue of Shark Species Known to Date. Vol. 4: Hexanchiformes to Lamniformes* (Rome, Italy: Food and Agriculture Organization of the United Nations).

de Mutsert, K., Cowan, J. H., Essington, T. E., and Hilborn, R. (2008). Reanalysis of Gulf of Mexico fisheries data: Landings can be misleading in assessments of fisheries and fisheries ecosystems. *Proc. Natl. Acad. Sci.* 105, 2740–2744. doi: 10.1073/pnas.0704354105

Dennis, D., and Plagányi, É. (2015). Cost benefit of fishery-independent surveys: are they worth the money? *Mar. Policy* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

Driggers, W.B.III., Ingram, G. W.Jr., Grace, M. A., Gledhill, C. T., Henwood, T. A., Horton, C. N., et al. (2008). Pupping areas and mortality rates of young tiger sharks *Galeocerdo cuvier* in the western North Atlantic Ocean. *Aquat. Biol.* 2, 161–170. doi: 10.3354/ab00045

Francis, R. I. C. C. (2011). Data weighting in statistical fisheries stock assessment models. *Can. J. Fisheries Aquat. Sci.* 68, 1124–1138. doi: 10.1139/f2011-165F

Grace, M., and Henwood, T. (1997). Assessment of the distribution and abundance of coastal sharks in the U.S. Gulf of Mexico and Eastern Seaboard 1995 and 1996. *Mar. Fisheries Review* 59, 23–32.

Grace, M. A., Ingram, W., Driggers, W.B.III., and Henwood, T. (2012). Assessing adequate sampling levels with time-series resampling of fishery-independent bottom longline surveys of the U.S. Gulf of Mexico. *Gulf Mexico Sci.* 30. doi: 10.18785/goms.3001.06

Grüss, A., Perryman, H. A., Babcock, E. A., Sagarese, S. R., Thorson, J. T., Ainsworth, C. H., et al. (2018). Monitoring programs of the U.S. Gulf of Mexico: inventory, development and use of a large monitoring database to map fish and invertebrate spatial distributions. *Rev. Fish Biol. Fisheries* 28, 667–691. doi: 10.1007/s11160-018-9525-2

Grüss, A., Walter, J. F.III, Babcock, E. A., Forrestal, F. C., Thorson, J. T., Lauretta, M. V., et al. (2019). Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. *Fisheries Res.* 213, 75–93. doi: 10.1016/j.fishres.2019.01.008

Gunderson, D. R. (1993). *Surveys of Fisheries Resources* (New York, NY: Wiley), 248.

Hansen, M. J., Beard, T. D.Jr., and Hayes, D. B. (2007). "Sampling and Experimental Design," in *Analysis and Interpretation of Freshwater Fisheries Data*. Eds. C. S. Guy and M. L. Brown (American Fisheries Society, Bethesda, MD), 51–120.

Hoffmayer, E. R., Driggers, W.B.III., Jones, L. M., Hendon, J. M., and Sulikowski, J. A. (2013b). Variability in the reproductive biology of the Atlantic Sharpnose Shark in the Gulf of Mexico. *Mar. Coast. Fisheries: Dynamics Management Ecosystem Sci.* 5, 139–151. doi: 10.1080/19425120.2013.783518

Hoffmayer, E. R., Parsons, G. R., and Horton, J. (2006). Seasonal and interannual variation in the energetic condition of adult male Atlantic Sharpnose Shark *Rhizoprionodon terraenovae* in the northern Gulf of Mexico. *J. Fish Biol.* 68, 645–653. doi: 10.1111/j.0022-1112.2006.00942.x

Hoffmayer, E., Pollack, A., Hendon, J., Drymon, M., and Grace, M. (2013a). *Standardized catch rates of Atlantic sharpnose sharks (*Rhizoprionodon terraenovae*) collected during bottom longline surveys in Mississippi, Louisiana, Alabama, and Texas coastal waters 2004-2011. SEDAR34-WP- 11* (North Charleston: SEDAR).

Howard, R. A., Ciannelli, L., Wakefield, W. W., and Haltuch, M. A. (2023). Comparing fishery-independent and fishery-dependent data for analysis of the distributions of Oregon shelf groundfishes. *Fisheries Res.* 258, 106553. doi: 10.1016/j.fishres.2022.106553

Ingram, G. W., Alvarez-Berastegui, D., Reglero, P., Balbín, R., García, A., and Alemany, F. (2017). Incorporation of habitat information in the development of indices of larval Bluefin Tuna (*Thunnus thynnus*) in the Western Mediterranean Sea, (2001–2005 and 2012–2013). *Deep Sea Res. Part II: Topical Stud. Oceanogr.* 140, 203–211. doi: 10.1016/j.dsr2.2017.03.012

Kimura, D. K., and Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fishes. *Rev. Fisheries Sci. Aquaculture* 14, 245–283. doi: 10.1080/10641260600621761

Liu, Y., Chen, Y., and Cheng, J. (2009). A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. *ICES J. Mar. Sci.* 66, 1873–1882. doi: 10.1093/icesjms/fsp157

Lo, N. C. H., Jacobson, L. D., and Squire, J. L. (1992). Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fisheries Aquat. Sci.* 49, 2515–2526. doi: 10.1139/f92-278

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Res.* 70, 141–159. doi: 10.1016/j.fishres.2004.08.002

McEachran, J. D., and Fechhelm, J. D. (2005). *Fishes of the Gulf of Mexico: Volume 2 Scorpaeniformes to Tetraodontiformes* (Austin: University of Texas Press).

Miller, T. J., Skalski, J. R., and Ianelli, J. N. (2007). Optimizing a stratified sampling design when faced with multiple objectives. *ICES J. Mar. Sci.* 64, 97–109. doi: 10.1093/icesjms/fsl013

Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Mar. Ecol. Prog. Series* 393, 111–129. doi: 10.3354/meps08220

Parsons, G. R., and Hoffmayer, E. R. (2005). Seasonal changes in the distribution and relative abundance of the Atlantic Sharpnose Shark *Rhizoprionodon terraenovae* in the north central Gulf of Mexico. *Copeia* 2005 (4), 913–919. doi: 10.1643/0045-8511(2005)005[0914:SCITDA]2

Pennington, M. (1985). Estimating the relative abundance of fish from a series of trawl surveys. *Biometrics* 41, 197–202. doi: 10.2307/2530654

Pennington, M., and Godø, O. R. (1995). Measuring the effect of changes in catchability on the variance of marine survey abundance indices. *Fisheries Res.* 23, 301–310. doi: 10.1016/0165-7836(94)00345-W

Pennington, M., and Strømme, T. (1998). Surveys as a research tool for managing dynamic stocks. *Fisheries Res.* 37, 97–106. doi: 10.1016/S0165-7836(98)00129-5

Perry, A. L., Low, P. J., Ellis, J. E., and Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science* 308, 1912–1915. doi: 10.1126/science.1111322

Peterson, C. T., and Grubbs, R. D. (2024). Temporal community structure and seasonal climatic migration of coastal sharks and large teleost fishes in the northeast Gulf of Mexico. *Can. J. Fisheries Aquat. Sci.* 80, 1335–1350. doi: 10.1139/cjfas-2022-0124

Rago, P. J. (2005). "Fishery independent sampling: survey techniques and data analyses," in *Management Techniques for Elasmobranch Fisheries*. Eds. J. A. Musick and R. Bonfil (Food and Agriculture Organization of the United Nations, Rome, Italy), 201–215.

Raj, D. (1968). *Sampling Theory* (New Delhi: TATA McGraw-Hill Publishing).

Richards, B. L., Smith, J. R., Smith, S. G., Ault, J. S., DiNardo, G. T., Kobayashi, D., et al. (2016). Design and implementation of a bottomfish fishery-independent survey in the main Hawaiian Islands. *U.S. Department of Commerce, National Oceanic and Atmospheric Administration, NOAA Technical Memorandum NOAA-TM-NMFS-PIFSC-53*. 54 p. Available at: hhtps://repository.library.noaa.gov/view/noaa/10761.

SEDAR 29. (2012). *SEDAR 29 stock assessment report HMS gulf of Mexico blacktip shark* (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/sedar-29-final-stock-assessment-report-gulf-of-Mexico-blacktip-shark/ (Accessed April 10, 2024).

SEDAR 34. (2013). *SEDAR 34 stock assessment report HMS atlantic sharpnose shark* (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/sedar-34-final-stock-assessment-report-atlantic-sharpnose-shark/ (Accessed April 10, 2024).

SEDAR 49. (2016). *SEDAR 49 stock assessment report gulf of Mexico data-limited species* (North Charleston, SC: SEDAR), 618. Available at: https://sedarweb.org/documents/sedar-49-final-stock-assessment-report-gulf-of-Mexico-data-limited-species/ (Accessed April 10, 2024).

SEDAR 54. (2017). *SEDAR 54 stock assessment report HMS sandbar shark* (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/sedar-54-final-assessment-report-hms-sandbar-shark/.

SEDAR 77. (2024). *SEDAR 77 stock assessment report HMS hammerhead sharks* (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/sedar-77-hms-hammerhead-sharks-final-stock-assessment-report/ (Accessed April 10, 2024).

Simpfendorfer, C. A., Hueter, R. E., Bergman, U., and Connett, S. M. H. (2002). Results of a fishery-independent survey for pelagic sharks in the western North Atlantic 1977-1994. *Fisheries Res.* 55, 175–192. doi: 10.1016/S0165-7836(01)00288-0

Smith, S. G., Ault, J. S., Bohnsack, J. A., Harper, D. E., Luo, J., and McClellank, D. B. (2011). Multispecies survey design for assessing reef-fish stocks, spatially explicit management performance, and ecosystem condition. *Fisheries Res.* 109, 25–41. doi: 10.1016/j.fishres.2011.01.012

Springer, S. (1940). The sex ratio and seasonal distribution of some Florida sharks. *Copeia* 1940, 188–194. doi: 10.2307/1437982

Stone, R. B., Bailey, C. M., McLaughlin, S. A., Mace, P. A., and Schulze, M. B. (1998). Federal management of US Atlantic shark fisheries. *Fisheries Res.* 39, 215–221. doi: 10.1016/S0165-7836(98)00185-4

Thorson, J. T., Johnson, K. F., Methot, R. D., and Taylor, I. G. (2017). Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinominal distribution. *Fisheries Res.* 192, 84–93. doi: 10.1016/j.fishres.2016.06.005

Vecchio, J. L., Bubley, W. J., and Smart, T. I. (2023). Increased fishery-independent sampling effort results in improved population estimates for multiple target species. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1192739

Walters, C. J. (2003). Folly and fantasy in the analysis of spatial catch data. *Can. J. Fisheries Aquat. Sci.* 60, 1433–1436. doi: 10.1139/f03-152

Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. (2018). Evaluation of alternative stratifications for a stratified random fishery independent survey. *Fisheries Res.* 207, 150–159. doi: 10.1016/j.fishres.2018.06.019

Warwick, R. M., Carr, M. R., Clarke, K. R., Gee, J. M., and Green, R. H. (1988). A mesocosm experiment on the effects of hydrocarbon and copper pollution on a sublittoral soft-sediment meiobenthic community. *Mar. Ecol. Prog. Series* 46, 181–191. doi: 10.3354/meps046181

Wilberg, M. J., Thorson, J. T., Linton, B. C., and Berkson, J. (2010). Incorporating time-varying catchability into population dynamic stock assessment models. *Rev. Fisheries Sci.* 18, 7–24. doi: 10.1080/10641260903294647

Xu, B., Zhang, C., Xue, Y., Ren, Y., and Chen, Y. (2015). Optimization of sampling effort for a fishery-independent survey with multiple goals. *Environ. Monit. Assess.* 187, 252. doi: 10.1007/s10661-015-4483-9

Zar, J. H. (2010). *Biostatistical Analysis. 5th Edition* (Upper Saddle River, New Jersey: Prentice Hall).

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 04 February 2025
DOI 10.3389/fmars.2025.1425362

Check for updates

OPEN ACCESS

EDITED BY
Fabio Campanella,
National Research Council (CNR), Italy

REVIEWED BY
Michela Martinelli,
National Research Council (CNR), Italy
Walter Bubley,
South Carolina Department of Natural Resources, United States

*CORRESPONDENCE
Adam G. Pollack
adam.pollack@noaa.gov

RECEIVED 29 April 2024
ACCEPTED 15 January 2025
PUBLISHED 04 February 2025

CITATION
Pollack AG, Hoffmayer ER, Switzer TS, Hanisko DS, Hendon JM, Mareska J, Martinez-Andrade F, Rester J, Zuckerman ZC and Pellegrin G (2025) Assessing survey design changes of long-term fishery-independent groundfish trawl surveys in the Gulf of Mexico. *Front. Mar. Sci.* 12:1425362. doi: 10.3389/fmars.2025.1425362



# Assessing survey design changes of long-term fishery-independent groundfish trawl surveys in the Gulf of Mexico

Adam G. Pollack[1]*, Eric R. Hoffmayer[1], Theodore S. Switzer[2], David S. Hanisko[1], Jill M. Hendon[3], John Mareska[4], Fernando Martinez-Andrade[5], Jeff Rester[6], Zachary C. Zuckerman[7] and Gilmore Pellegrin[8]

[1]Mississippi Laboratories, Southeast Fisheries Science Center, National Marine Fisheries Service, Pascagoula, MS, United States, [2]Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, St. Petersburg, FL, United States, [3]Center for Fisheries Research and Development, The University of Southern Mississippi, Ocean Springs, MS, United States, [4]Marine Resources Division, Alabama Department of Conservation and Natural Resources, Dauphin Island, AL, United States, [5]Coastal Fisheries Division, Texas Parks and Wildlife Department, Corpus Christi, TX, United States, [6]Gulf States Marine Fisheries Commission, Ocean Springs, MS, United States, [7]Fisheries Research Laboratory, Louisiana Department of Wildlife and Fisheries, Grand Isle, LA, United States, [8]Retired, Pascagoula, MS, United States

Long-term fishery-independent surveys provide a wealth of information on fisheries stocks that inform stock assessments. One of the strengths of these surveys is that the design and methods are consistent through time. However, maintaining an unchanged time series can pose several potential issues as management needs change resulting in the need to alter either the survey design or its spatial extent. In the United States Gulf of Mexico, bottom trawl surveys targeting groundfish and shrimp (hereafter, groundfish surveys) have been conducted since the 1950s, with standardized surveys beginning in 1972. The resulting data can provide a great deal of information on commercially and recreationally important species. However, many of the alterations to the survey design have been buried in gray literature or otherwise poorly noted. The history of these surveys is discussed, along with the rationale behind these changes and the impacts they had on stock assessments in the region. Starting in 1981, the Southeast Area Monitoring and Assessment Program began groundfish surveys, continuing the time series. Over time, the survey's spatial extent and design have been altered to meet changing management needs. Delta-lognormal models are used to draw inferences concerning the effects of the survey design change on the relative abundance and their associated coefficients of variation for several commercially and recreationally important species. The expansion of the surveys across the Gulf of Mexico is examined in relation to stock assessments. Overall, the design changes and spatial expansion have been beneficial from a stock assessment standpoint, resulting in an increase in the number of indices used for single-species stock assessments and the utility of survey data in support of ecosystem modeling efforts. Finally, a discussion around the lesson learned (i.e., the critical need for overlap of survey designs), emphasizing the potential impacts of these changes on the overall time series concerning stock assessments, is presented.

KEYWORDS

**design change, survey expansion, SEAMAP, groundfish, shrimp, trawl survey, Gulf of Mexico**

# 1 Introduction

Under increasing pressure from commercial and recreational fisheries, timely and accurate assessment and management are essential to maintaining sustainable fish stocks. In the United States (U.S.), the National Marine Fisheries Service (NMFS) is responsible "for the stewardship of the nation's living marine resources and their habitat" (NOAA Fisheries, 2022a). Part of this stewardship involves the management of 492 stocks or stock complexes, of which 48 are currently overfished (the population size is too small) and 24 are subject to overfishing (the annual catch rate is too high; NOAA Fisheries, 2023). The Magnuson-Stevens Fishery Conservation and Management Act requires that once stocks are identified as overfished, a rebuilding plan must be put in place to immediately end overfishing and rebuild the stock to sustainable levels (MSFCMA, 2007). Effective management of these stocks requires information on an array of metrics, including annual landings and discards, abundance, size and age composition, reproduction, and mortality among other factors (Punt, 2023).

Fishery managers rely on outputs from stock assessments to not only assess stock status, but also to project future fisheries productivity and guide the establishment of sustainable harvest regulations. The stock assessment process requires both fishery-dependent and fishery-independent data to develop a statistical model that accurately estimates overall population dynamics while also capturing sources of uncertainty (Punt, 2023). The complexity of a stock assessment model is heavily dependent on the quantity and quality of data available for a particular stock. Regardless of the assessment model used, time series (indices) of abundance and size/age composition are essential. Although there are various statistical approaches that can be used to generate these indices, model-based approaches are most commonly used to standardize annual estimates of catch (Lo et al., 1992; Maunder and Starr, 2003; Maunder and Punt, 2004; Thorson, 2019; Thompson et al., 2022).

Historically, stock assessments in the U.S. Gulf of Mexico (GOM) routinely utilized catch per unit effort (CPUE) generated from the analysis of fishery-dependent data. However, fishery-dependent data can be influenced by outside factors (e.g., market prices, changes in regulations, etc.) not directly related to population abundance (de Mutsert et al., 2008). In recent years, increasingly complex and restrictive management regulations have reduced the utility of fishery-dependent data to track abundance trends. Therefore, fishery-independent data from scientific surveys have become essential to the accurate assessment and management of fish stocks (SEDAR, 2015). Fishery-independent indices are especially valuable because they utilize data collected following a statistically rigorous survey design that typically encompasses multiple habitats across broad spatial and temporal scales (Rago, 2005; Thompson et al., 2022). These surveys also often provide additional data for life history stages for which fishery harvest is prohibited (e.g., recruits, juveniles, and sublegal adults), which may help to forecast the future productivity of the stock. Despite their importance, the availability of fishery-independent data is often limited due to the high cost of conducting these surveys.

Furthermore, although data from standardized and long-term fishery-independent surveys are essential to the assessment and management of fish stocks (Rourke et al., 2022), maintaining consistency through time is often a challenge and requires a concerted effort by management agencies and other institutions (Dennis et al., 2015). Changes to long-term surveys can arise from several potential sources, including changes to sampling vessels or platforms, changing technology, or even expansion or contraction of the survey footprint to align survey efforts with available funding. A fundamental consideration when modifying long-term surveys is how best to implement changes to the survey design while maintaining the consistency of the time series to the greatest extent practicable. Ideally, statistically robust calibration studies would be conducted prior to implementing any significant change to determine what, if any, correction factor is required. Oftentimes, this is not possible due to financial constraints, program logistics (e.g., staffing requirements, ship time, etc.), or the need for large survey changes. In these cases, careful consideration must be given as to whether changes can be accounted for statistically, or whether data collected under the new design should be treated as a new time series (Miller et al., 2010; Latour et al., 2023; Switzer et al., 2023; Schrandt et al., 2024).

In the GOM, bottom trawl surveys targeting groundfish and shrimp (hereafter, groundfish surveys) have been conducted since the 1950s, with standardized surveys beginning in 1972. Throughout this period, the survey design and methods of groundfish surveys in the GOM have evolved to meet changing objectives and management needs. These surveys represent the only source of long-term, fishery-independent data on groundfish populations in the GOM. Given its history and the length of available time series, the groundfish surveys represent an ideal case study to investigate the impacts of multi-decadal survey design changes, the evolution of multiple independent surveys, and their utility in stock assessments for a variety of species (e.g., shrimp, snappers, groupers, sharks, etc.) under fishery management plans. The use of this fishery-independent survey data within stock assessments is important because of the 492 stocks that are managed nationwide, 69 are located in the GOM, of which 16 are experiencing overfishing and nine are classified as overfished (NOAA Fisheries, 2023).

The objectives of this paper are to: (1) describe the history of the groundfish surveys in the GOM, and document the changes to survey design and underlying sampling frame, (2) investigate the impacts of the most recent groundfish survey design change/expansion in 2008/2010 to stock assessments and ecosystem models in the region, and (3) present lessons learned from over seven decades of survey evolution to provide guidance to others as to how changes to survey design can be implemented successfully while minimizing the impacts on the overall time series. In doing so, we synthesize the vast amount of gray literature documenting changes in groundfish survey design and how these changes have been dealt with analytically.

# 2 History of fishery-independent groundfish surveys in the Gulf of Mexico

## 2.1 Early fishery-independent surveys (1950 – 1971)

The Gulf States Marine Fisheries Commission (GSMFC) was established in 1949 by the U.S. Congress "to promote better utilization of the fisheries, marine, shell and anadromous, of the seaboard of the Gulf of Mexico, by the development of a joint program for the promotion and protection of such fisheries and the prevention of the physical waste of the fisheries from any cause" (GSMFC, 2024). In 1950, the GSMFC recommended to the U.S. Fish and Wildlife Service's Bureau of Commercial Fisheries (the precursor federal agency to NMFS) that tuna, sharks, and snappers be the primary focus of exploratory surveys in the GOM, although shrimp were subsequently substituted for sharks (Bullis, 1964; NOAA, 2022b). The main objectives of these early surveys (1950 – 1963) were exploration and gear studies designed to assess what underexploited fishery resources may be available to the commercial fishing industry (Springer and Bullis, 1956; Bullis and Thompson, 1965). Multiple cruises were conducted each year throughout all seasons that used a variety of gear types including, but not limited to, trawls (bottom, deepwater, mid-water), gillnets, pelagic longlines, handlines, trap lift nets, dipnets, and purse seines, and extensive modification of those gears were commonplace (Figure 1; Springer and Bullis, 1956; Bullis and Thompson, 1970).

In the GOM, bottom trawls were the most common gear used during these early surveys, partly due to the importance of the shrimp fishery in the region (Springer and Bullis, 1956). Trawling operations were not standardized, and various trawl gear and configurations were used, including multiple-sized trawls and door combinations with varying tow durations (Bullis, 1964; Bullis and Thompson, 1970). Typically, shrimp trawling coverage was limited to two depth strata: 20 – 75 fm (36.6 – 137.2 m) for brown shrimp *Farfantepenaeus aztecus*, pink shrimp *F. duorarum*, and white shrimp *Litopenaeus setiferus* and 175 – 330 fm (320 – 603.5 m) for royal red shrimp *Pleoticus robustus*, and the work was conducted by a variety of vessels with varying lengths and capabilities (Springer and Bullis, 1956; Bullis and Thompson, 1967).

There was a pronounced shift in the fundamental approach of the program between 1964 and 1971. While still exploratory, surveys became more systematic and focused on producing an inventory of marine organisms in the southeastern U.S. and GOM (Bullis and Thompson, 1967). Trawl surveys were conducted throughout the year, but sampling efforts were inconsistent from year to year. Due to the lack of standardization, data provided during this early period do not represent a reliable fishery-

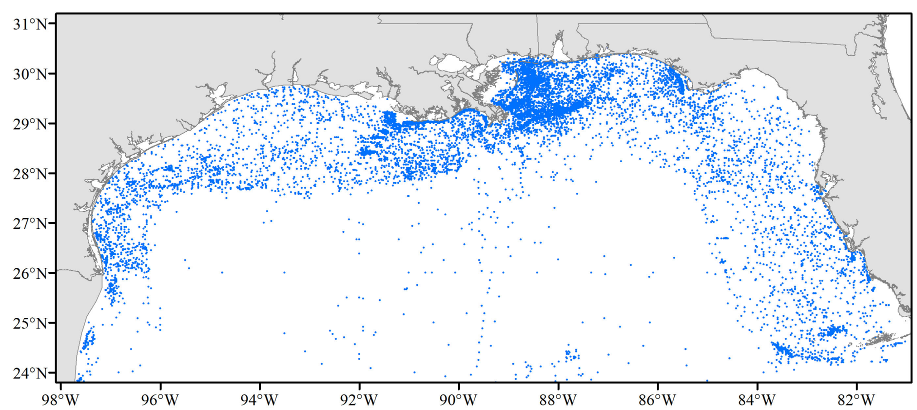



FIGURE 1
Distribution of stations sampled in the United States Gulf of Mexico during fishery-independent surveys conducted by the National Marine Fisheries Service (NMFS) from 1950 – 1971.

independent time series and cannot be used to draw inferences about population structure over time (Nichols, 2004a). However, the trawl surveys did provide valuable information on presence, abundance, and species composition at specific locations and times (Chittenden and Moore, 1977; Darnell et al., 1983; NOAA, 1985; Darnell and Kleypas, 1987); accordingly, records, numbers and weights of species caught in the tows were informative for distributional analyses. In 1970, the U.S. Fish and Wildlife Service's Bureau of Commercial Fisheries became the National Oceanic and Atmospheric Administration's NMFS, with the new task of managing the U.S. fisheries (NOAA Fisheries, 2022b).

## 2.2 A standardized survey to address management needs (1972 – 1980)

In 1972, the commercial groundfish industry raised concerns about declining CPUE in the GOM, particularly for Atlantic croaker *Micropogonias undulatus*, which accounted for roughly 70% of landings at that time (Gutherz, 1977). Discussions were held among commercial fishers and state and federal fisheries managers to address these concerns (Juhl et al., 1973). As a result, the GSMFC Technical Coordinating Committee adopted a measure to develop the Oceanic Resource Surveys and Assessment (ORSA) program. The ORSA program was designed to evaluate the industrial (i.e., non-food fish) and commercial (i.e., food fish) groundfish fishery in the northern GOM, and provide information on availability, abundance, and status of the fisheries resources. Over time, the ORSA program evolved to address all demersal species available to bottom trawls and not just those of commercial importance (NMFS, 1975).

The ORSA program implemented the first resource assessment surveys utilizing a stratified random sampling design and standardized trawl gear configuration (now recognized as the 42 ft (12.8 m) SEAMAP groundfish trawl, GSMFC, 2019) to collect quantitative and qualitative biological and environmental data (NMFS, 1975) during the fall season (primarily October and November). The initial 1972 study area fell between 94.5° N and 85.5° W between depths of 5 and 50 fm (9.1 and 91.4 m) and was divided into a single primary spatial stratum (central) and two (eastern and western) secondary spatial strata (Figure 2; Juhl et al., 1974). The primary and secondary areas were defined on the relative faunal densities of commercial groundfish (e.g., Atlantic croaker, spot *Leiostomus xanthurus*, sand seatrout *Cynoscion arenarius*, and silver seatrout *C. nothus*) based on commercial fishing landings data. The central area was the target of the majority of sampling due to its inclusion of the primary fishing grounds of the commercial groundfish industrial fleet (Roithmayr, 1965), while the secondary strata were only sampled as time permitted (Nichols, 2004a). The minimum depth boundary was based on the operational limit of NOAA Ship *Oregon II*, and the outer depth limit was the expected outer boundary of commercially profitable catches of targeted groundfish and shrimp species (Nichols, 2004a). The defined study area was not inclusive of the full spatial range of the multi-species groundfish stock, but rather an area targeting the greatest densities, as it was considered impossible to study the entire distributional range of desired fishes due to workforce, budgetary, and logistical constraints.

Random site placement within strata was determined by first selecting a random (with replacement) 10 minute (~18.5 km) block of latitude and longitude and then randomly selecting (without replacement) a smaller 2.5 minute (4.6 km) grid of latitude by longitude within the block (i.e., block-grid design; Nichols, 2004a). The total number of blocks, grids and sites for each area was based on the available number of days at sea (Table 1). Up to three 10 minute tows were targeted for sampling within each grid and predominantly towed parallel to depth contours, which was the practice of the commercial fleet (Nichols, 2004a). However, in many cases, fewer tows were conducted, with the actual number of tows often determined by remaining available time at sea for the survey and the amount of catch in the first tow. Three shorter duration tows were employed rather than a single longer tow due to high catch rate variation among tows attributed to the patchy distribution of fauna.

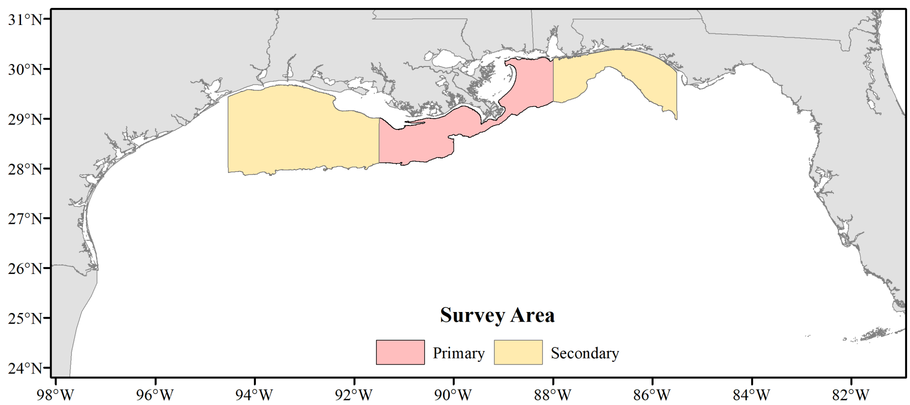



FIGURE 2
Spatial extents of the National Marine Fisheries Service (NMFS) Fall Groundfish Survey (1972 – 1984).

TABLE 1 Summary of the survey design parameters for the groundfish surveys conducted in the Gulf of Mexico from 1972 to present.

| Survey | Timeframe | Sample Design | Sample Allocation | Targeted Stations | Tows per Station | Tow Time (minutes) | Tow Direction | Area Sampled | Depth Sampled (fm) | Area Strata[1] (#) | Depth Strata[2] (#) | Time of Day Strata (#) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fall | 1972-1984 | Block-grid | One per grid | Varied[3] | Up to 3 | 10 | Parallel | Primary (88° W - 91.5° W) Secondary (91.5° W,- 94.5°W, 88° W - 85.5°W) | 5-50 | – | – | 0 - 24 hours[4] |
| Fall | 1985 | Block-grid | One per grid | Varied[3] | 1 | 15 | Perpendicular | 9-17 | 5-100 | – | – | 2 - day/night |
| Fall | 1986 | Block-grid | One per grid | Varied[3] | 1 | 15 | Perpendicular | 9-21 | 5-100 | – | – | 0 - 24 hours[4] |
| Summer | 1982 | Stratified random | One per stratum | 130[5] | Variable[6] | 10-30[7] | Perpendicular | 10-21 | 5-50 | 5 | 26 | 1 - night only |
| Summer | 1983 | Stratified random | One per stratum | 156[5] | Variable[6] | 10-30[7] | Perpendicular | 8-21 | 5-50 | 6 | 26 | 1 - night only |
| Summer | 1984-1985 | Stratified random | One per stratum | 130[5] | Variable[6] | 10-30[7] | Perpendicular | 10-21 | 5-50 | 5 | 26 | 1 - night only |
| Summer | 1986 | Stratified random | One per stratum | 145[5] | Variable[6] | 10-30[7] | Perpendicular | 10-21 | 5-50 | 5 | 29 | 1 - night only |
| Summer/Fall | 1987-1988 | Stratified random | One per stratum | 310[5] | Variable[6] | 10-60[7] | Perpendicular | 10-21 | 5-60 | 5 | 31 | 2 - day/night |
| Summer/Fall | 1989 | Stratified random | One per stratum | 240[5] | Variable[6] | 10-60[7] | Perpendicular | 11-21 | 5-60 | 5 | 24 | 2 - day/night |
| Summer/Fall | 1990-2000 | Stratified random | One per stratum | 230[5] | Variable[6] | 10-60[7] | Perpendicular | 11-21 | 5-60 | 5 | 23 | 2 - day/night |
| Summer/Fall | 2000-2008[8] | Stratified random | One per stratum | 230[5] | Variable[6] | 10-55[7] | Perpendicular | 11-21 | 5-60 | 5 | 23 | 2 - day/night |
| Fall | 2008 | Stratified random | Proportional by area | 350 | 1 | 30 | Random | 11-21 | 5-60 | 10 | 5 | 0 - 24 hours[4] |
| Summer/Fall | 2009 | Stratified random | Proportional by area | 350 | 1 | 30 | Random | 11-21 | 5-60 | 10 | 1 | 0 - 24 hours[4] |
| Summer/Fall | 2010-2012 | Stratified random | Proportional by area | 350/300[9] | 1 | 30 | Random | 1-21 | 5-60 | 20 | 1 | 0 - 24 hours[4] |
| Summer/Fall | 2013-2016 | Stratified random | Proportional by area | 350/300[9] | 1 | 30 | Random | 1-21 | 2-60 | 20 | 2 | 0 - 24 hours[4] |
| Summer/Fall | 2017-present | Stratified random | Proportional by area[10] | 350/300[9] | 1 | 30 | Random | 2-21 | 5-60 | 19 | 2 | 0 - 24 hours[4] |

[1]Full breakdown of area strata can be found in Supplementary Table S1.
[2]Full breakdown of depth strata can be found in Supplementary Table S2.
[3]Targeted number of stations was dependent on the available sea days.
[4]No time of day strata, stations were sampled at arrival, regardless of time of day.
[5]Number represents the total number of strata to be sampled.
[6]Number of tows per station was dependent on covering the depth stratum.
[7]Tow time dependent on the time it took to cover a depth stratum. If maximum tow time was reached during an individual tow, additional tows were made until the depth stratum was completed.
[8]Survey design was changed after the Summer survey in 2008.
[9]Stations counts are representative of Summer and Fall survey respectively.
[10]Area calculated from trawlable area within each NMFS statistical zone (GSMFC, 2019).
Area sampled is representative of the NMFS statistical zones sampled unless otherwise noted.

## 2.3 Initiation of SEAMAP (1981 – 2008)

The Southeast Area Monitoring and Assessment Program (SEAMAP) was initiated in 1981 with the objective of establishing a collaborative program between state, federal, and academic scientists for the collection, management and dissemination of fishery-independent data in the U.S. GOM (Stuntz et al., 1985). SEAMAP is funded through a series of grants from the NMFS, with the main objective to "provide essential fishery independent data and analyses for evaluating the status of the Nation's fisheries through the SouthEast Data, Assessment and Review (SEDAR) process, while supporting regional fishery management councils and enhanced requirements of the Magnuson-Stevens Reauthorization Act" (SEAMAP, 2024). That same year, NMFS initiated a summer groundfish survey in response to a request to study the effects of the "Texas Closure", an approach to shrimp management implemented following the passage of the Texas Shrimp Conservation Act of 1959. The closure consisted of a 45 to 60 day period during which no trawling was allowed in the Texas Territorial Sea (waters under state jurisdiction) from about mid-May to mid-July. Starting in 1981, by request, NMFS extended this closure to the Exclusive Economic Zone (i.e., waters under federal jurisdiction). The objective of the closure was to allow shrimp to grow larger prior to harvest to reduce the amount of undersized shrimp being discarded while also increasing market prices for harvested shrimp (Nance, 1993; Fuls, 2001).

The following year, SEAMAP established a summer groundfish survey based on results from the Texas Closure study, and sampling protocols developed by the SEAMAP Shrimp and Bottomfish Sampling Gear Workshop (Watson and Bane, 1985). The working group recommended adopting the same 42 ft net used during the ORSA program surveys and similar deployment protocols. Since most of the Texas shrimp landings consist of two nocturnal species, brown and white shrimp, samples were originally collected at night only. Sample sites were randomly allocated to strata based on depth zones and area, with equal sampling across all strata (Stuntz et al., 1985). Twenty-six depth strata were defined between 5 and 50 fm (9.1 and 91.4 m; Table 1; Supplementary Table S1). Area strata were based on NMFS statistical zones (hereafter, statistical zones) which were originally used to report shrimp landings (Kutkuhn, 1962; Supplementary Figure S1). The area covered statistical zones 10 – 21 with each stratum defined by grouping two or three zones (Figure 3A). A full breakdown of the area strata and depth strata can be found in Supplementary Tables S1 and S2, respectively. The groundfish survey was expanded in 1983 to include the spatial area to Apalachicola, Florida (FL) (85° W; statistical zones 8 and 9; Figure 3B). However, this eastward expansion was quickly abandoned due to increased interaction with untrawlable bottom and damage to trawl nets and the survey's spatial extent reverted to the area originally developed in 1982 (Nichols, 2004a; Figure 3A). Specific information about the methods that were used to conduct the tows can be found in Table 1.

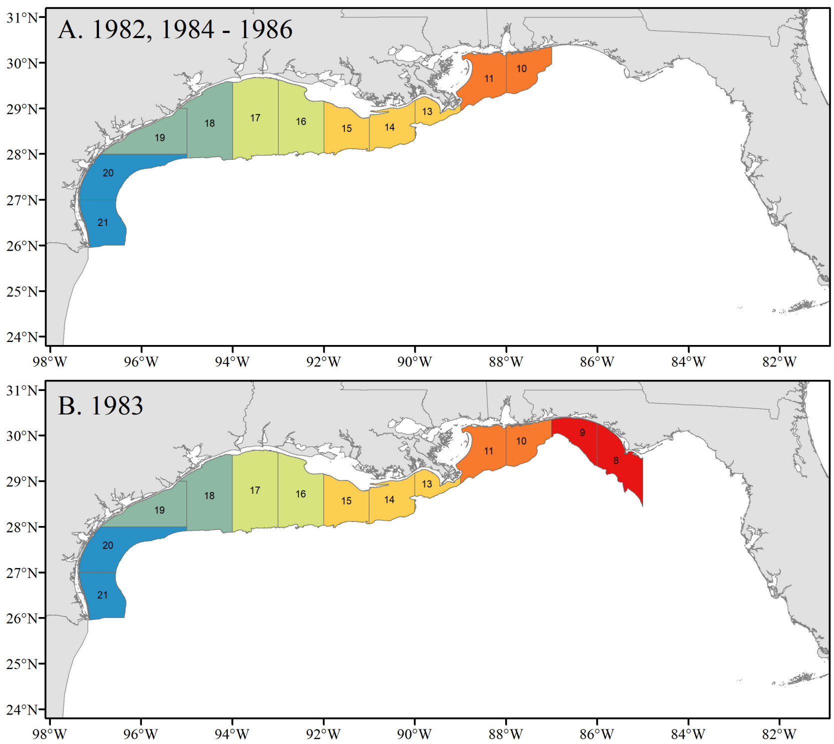



FIGURE 3
Spatial extents of the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer Shrimp/Groundfish Surveys in **(A)**. 1982 and 1984 – 1986 and **(B)**. 1983. Numbered blocks indicate National Marine Fisheries Service (NMFS) statistical zones. Individual colors represent statistical zones that were paired as area strata.

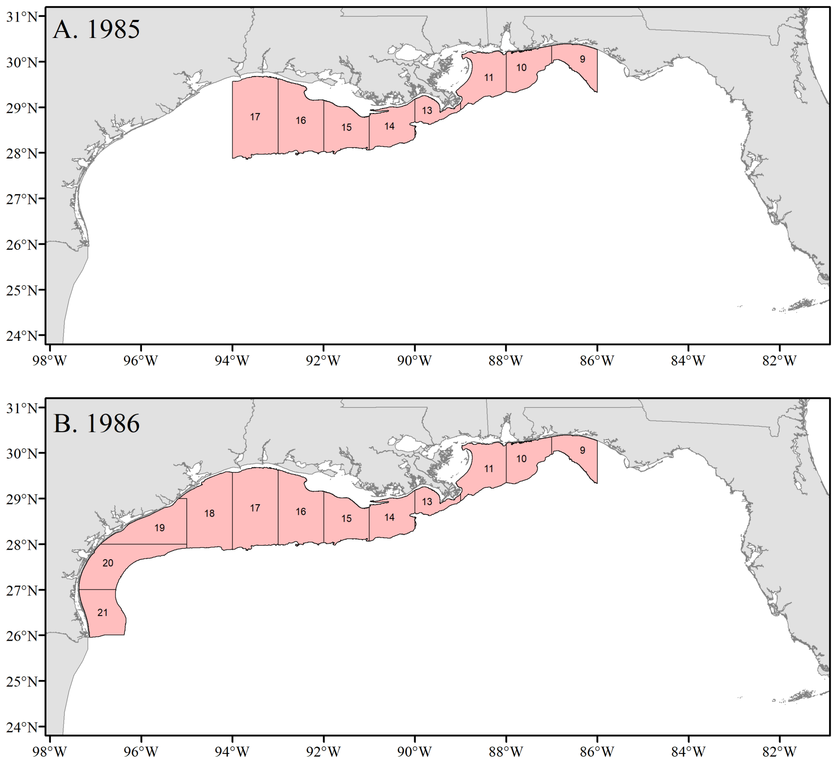



FIGURE 4
Spatial extents of the Southeast Area Monitoring and Assessment Program (SEAMAP) Shrimp/Groundfish Fall Surveys in **(A)** 1985 and **(B)** 1986. Numbered blocks indicate National Marine Fisheries Service statistical zones.

In 1985, SEAMAP initiated the Fall Groundfish Survey incorporating the NMFS Fall Groundfish Survey. The 1985 and 1986 SEAMAP Fall Groundfish Surveys utilized the NMFS block-grid sample allocation, but with sampling within grids limited to a single 15 minute tow in an effort to expand spatial coverage (Nichols, 2004a). The time saved by conducting only a single tow allowed for the spatial expansion of the study area from Rosemary Beach, FL (86° W) to Sabine, Louisiana (LA) (94° W) (Figure 4A), with the maximum depth extended to 100 fm (182.9 m). The survey was further expanded in 1986 to include the spatial area from Sabine, LA (94° W) to the Texas-Mexico border (Sanders et al, 1990a, Figure 4B). The fall groundfish survey expansion served to increase the scope of the survey to better assess the distribution and abundance of many species, particularly brown and white shrimp, throughout the region. It also served to increase the usefulness of the data in stock assessments (e.g., red snapper *Lutjanus campechanus*, SEDAR, 2022). Sampling from 60 to 100 fm (109.7 to 182.9 m) was dropped with the intent to align the fall groundfish survey with the summer groundfish survey design, and partly due to the small segment of the survey area actually covered from 60 to 100 fm (109.7 to 182.9 m, Nichols, 2004a).

In 1987, the summer and fall groundfish surveys adopted the same sampling protocol and spatial extent (Pensacola, FL to Brownsville, Texas (TX), statistical zones 10 – 21), thereby standardizing data collection for the two seasons and providing seasons as another variable of interest (Figure 5A). At this time three changes were introduced to the groundfish survey: (1) a diurnal stratum (day/night) was incorporated into the design, meaning that a full set of daytime and nighttime samples were collected in each combination of area and depth strata, (2) two additional depth strata were added to extend depth coverage to 60 fm (109.7 m), and (3) the maximum tow duration was extended to 60 minutes (Sanders et al., 1990b). Between 1988 and the summer of 2008 only three minor modifications were made to the SEAMAP groundfish survey sampling design: (1) depth strata were collapsed in 1989 and 1990 (Sanders et al., 1991, 1992; Table 1; Supplementary Table S1), (2) statistical zone 10 was dropped from the survey universe in 1989 because of the increased number of obstructions in the area as Alabama expanded its artificial reef permitting area (Nichols, 2004a; Figure 5B), and (3) maximum tow times were limited to 55 minutes in 2000 due to concerns over sea turtle bycatch, as turtle excluder devices are not used in the trawl net.

## 2.4 Survey design change (2008) and expansion to the West Florida Shelf (2010)

In the fall of 2008, the SEAMAP groundfish surveys undertook a series of significant changes that were intended to increase the usefulness and applicability of the data. These changes were implemented to increase survey efficiency by: (1) transitioning to a stratified random sampling design with effort proportional to the spatial area of each statistical zone between 5 and 60 fm (9.1 and 109.7 m; Figure 5C), (2) implementing a standardized 30-minute

10.3389/fmars.2025.1425362

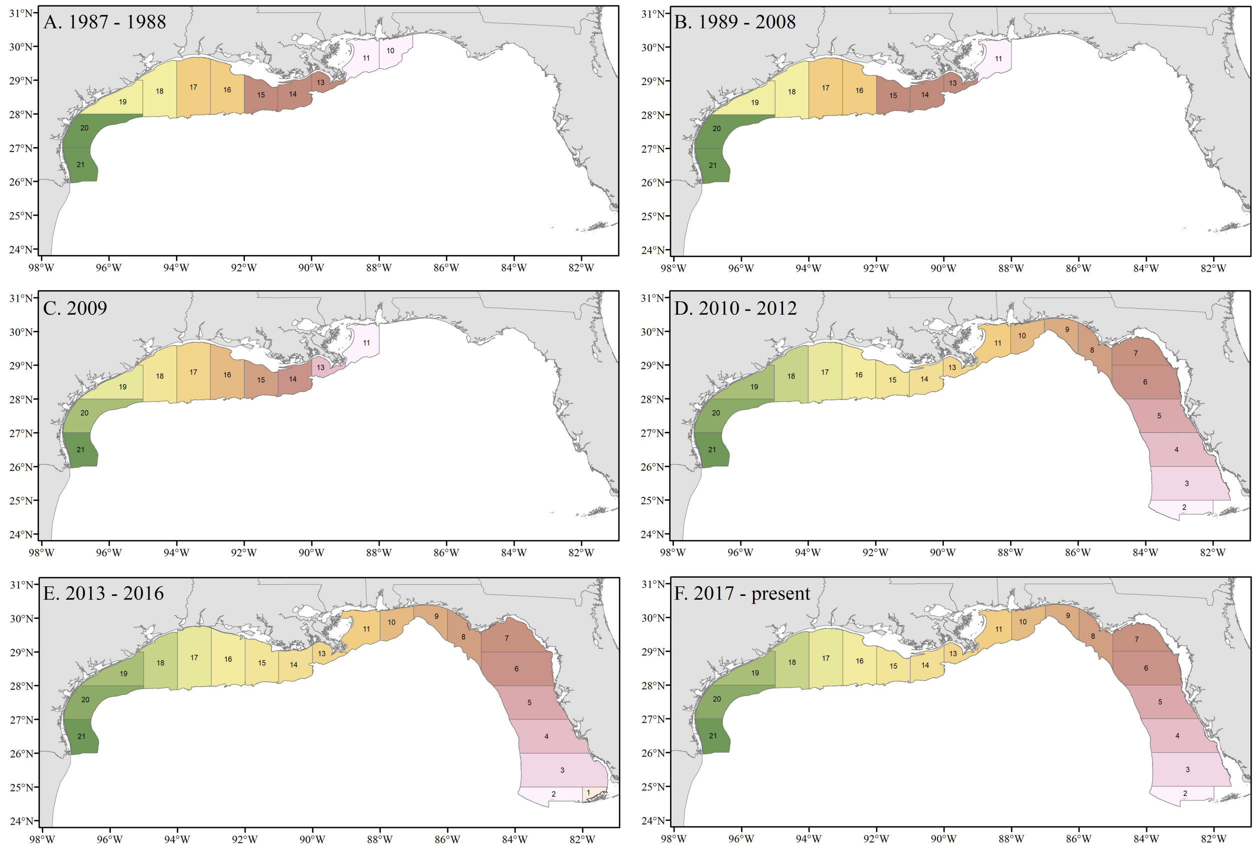



FIGURE 5
Spatial extents of the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer and Fall Shrimp/Groundfish Surveys in **(A)** 1987 – 1988, **(B)** 1989, **(C)** 2009, **(D)** 2010 – 2012, **(E)** 2013 – 2016, and **(F)** 2017 – present. Numbered blocks indicate National Marine Fisheries Service (NMFS) statistical zones. Individual colors in **(A, B)** represent statistical zones that were paired as area strata while those in panels **(C, F)** represent its own strata.

tow time for all samples, and (3) eliminating the day/night stratification (GSMFC, 2009). With the dropping of the day/night stratification, stations could be towed whenever they were reached, as opposed to having to wait for sunrise or sunset under the previous design, which designated when certain stations could be sampled. Under the new survey design, the number of stations sampled was expected to increase, which in turn, would lead to a reduction in the coefficient of variation (CV) in the time series (Ingram, 2008). Several changes to the depth strata were initiated between 2008 and 2013 to better distribute the sampling effort throughout the survey area and increase station density from the standard 230 stations to a target of 350 stations (Table 1).

The official spatial extent of the summer and fall groundfish surveys continued to range from Brownsville, Texas (TX) to Mobile Bay, Alabama (AL) (statistical zones 11 – 21) through 2009 (Figure 5C). Concurrently, the SEAMAP received supplemental funding that allowed the state of Florida to begin experimental groundfish surveys over the West Florida Shelf (WFS) in 2008 and 2009 using similar protocols to those implemented by the summer and fall groundfish surveys (Figure 5D). Based on the success of the experimental sampling, SEAMAP groundfish surveys were expanded GOM-wide in 2010 to include the area from Mobile Bay, AL to Key West, FL (statistical zones 2 to 10). The decision to expand the fishery-independent trawl sampling into the eastern GOM was based on recommendations by the SEDAR Red Snapper Update Assessment Workshop Committee (SEDAR, 2009). The recommendation was derived from the need to obtain essential information concerning age 0 and age 1 red snapper, as well as other managed species occurring in the eastern GOM (e.g., red grouper *Epinephelus morio* and gray snapper *Lutjanus griseus*). The groundfish survey expansion provided gulf-wide coverage, but at the expense of reduced sampling effort within statistical zones 11 through 21 as days at sea remained relatively constant after the expansion. The increased station allocation of 350 stations was now spread across the entire survey area (statistical zones 1 – 21).

After the groundfish survey expansion onto the WFS, there was also an attempt to investigate the composition of catch inside of the 5 fm (9.1 m) line with an additional depth strata added from 2 to 5 fm (3.7 to 9.1 m) in 2013 (Figure 5E). However, this was only possible for statistical zones 2 – 17 because of the depth limitation of the NOAA Ship *Oregon II*, which conducts all sampling in statistical zones 18 – 21. Sampling in waters less than 5 fm (9.1 m) was ended in 2017 due to the lack of gulfwide coverage (Shrimp/Groundfish Work Group, 2017; Figure 5F).

In 2017, there was a refinement to the SEAMAP groundfish surveys area. While the spatial extent of the survey remained

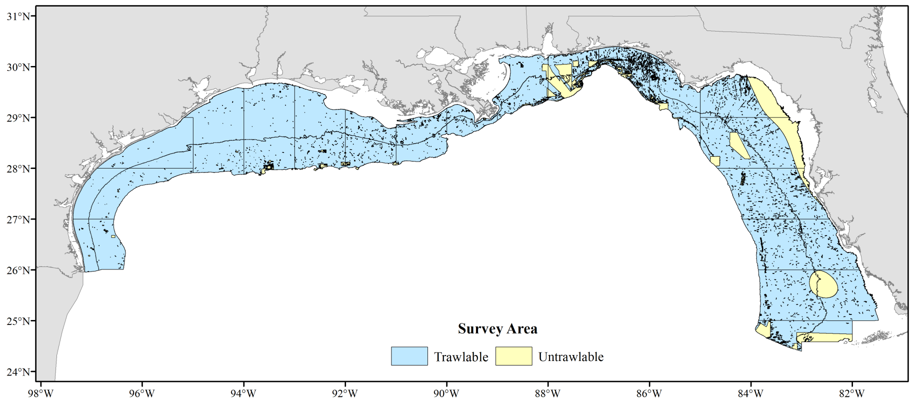



FIGURE 6
Current spatial extent of the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer Shrimp/Groundfish Surveys. Blue shaded areas represent the trawlable area within each National Marine Fisheries Service (NMFS) statistical zone, while the untrawlable areas are shaded in yellow.

unchanged (statistical zones 2 – 21), untrawlable areas within the statistical zones were removed and the area of each statistical zone/ depth strata was recalculated (Figure 6). These changes were made to better reflect the available trawlable habitat within the survey area and to help avoid obstructions, which damage the gear, as well as sensitive live bottom areas (Hanisko et al., 2018). Under the current design, the spatial coverage for the survey ranges from Key West, FL to Brownsville, TX (statistical zones 2 – 21, excluding 12) at depths of 5 to 60 fm (9.1 to 109.1 m), which includes the core area of the initial SEAMAP groundfish surveys design. Random sampling with proportional allocation of stations by area within the statistical and depth strata is used to select the stations, with a target of 350 stations for the summer survey and 300 stations for the fall survey (Table 1).

# 3 Evaluation of groundfish survey design change

## 3.1 Assessing change

One of the main reasons for changing the groundfish survey design was to increase survey efficiency (measured in stations sampled per day), which would lead to an increase in the number of stations sampled during each survey. Using data from 1987 to 2022 from the statistical zones 11 – 21, the average number of stations sampled per day by the NMFS was calculated to determine the survey efficiency of the different sampling designs during the old survey design (1987 – summer 2008), new survey design - expanded sampling years (fall 2008 – 2009) and the new survey design – full GOM sampling (2010–2022).

To examine if the survey design change in 2008/2009 affected the trends of the relative abundance indices and their respective CV of the mean (SE/Mean), delta-lognormal models were fit for a suite of species commonly caught in the survey. The relative abundance index computed by this method is a product of yearly abundance estimates from two distinct generalized linear models: a binomial (logistic) model that describes proportion of positive abundance values (i.e., presence/absence) and a lognormal model, which describes variability in only the nonzero abundance data (i.e., CPUE in number per hour; Lo et al., 1992; Ingram et al., 2017). The submodels of the delta-lognormal model were built using a backward selection procedure based on type 3 analyses with an inclusion level of significance of $\alpha$ = 0.05 (Ingram et al., 2017). Variables that could be included in the submodels were year (1987 – 2022), time of day (day/night), statistical zone (11 – 21), and depth (fm). Data were limited to statistical zones 11 – 21 to evaluate relative abundance trends and impacts to CVs over a common survey area. Species that were analyzed included Atlantic croaker, brown shrimp, gray triggerfish *Balistes capriscus*, longspine porgy *Stenotomus caprinus*, red snapper, and white shrimp. Separate abundance indices were calculated for the SEAMAP Summer and Fall Groundfish Surveys to account for any seasonal changes that may have affected the survey design. All of the abundance indices presented hereafter were scaled to a mean of one to make comparing different magnitudes of catch between species easier. The binomial and lognormal submodels were computed using the GLMMIX macro and MIXED procedure in the SAS software (ver. 9.4, Copyright © 2016 by SAS Institute Inc., Cary, NC, USA.).

## 3.2 Results

The design changes implemented in the fall of 2008 have been beneficial to the SEAMAP groundfish surveys with respect to survey efficiency, primarily due to the elimination of the day/night stratum and the reduction in tow times to 30 minutes. Historically, a complete groundfish survey under the original 1987 to 2008/2009 survey design would have consisted of 230 SEAMAP groundfish stations. Under the new survey design, sampling effort increased to an average of 350 stations (52% increase) throughout the historic survey area (Brownsville, TX to Mobile Bay, AL) for the fall 2008 and both the summer and fall surveys in 2009. Survey efficiency also increased from a mean of 7.7 (SD=0.78) stations sampled per day under the old survey design to 9.9 (SD=1.1) stations sampled per day under the new survey design over the historic area in 2008 and 2009. This increase in sampling across the historic survey area did lead to some reductions

TABLE 2 Comparison of coefficients of variation (CV) and standard deviations (SD) from delta-lognormal models for selected species from the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer Shrimp/Groundfish Survey conducted from 1987 to 2022 in National Marine Fisheries Service (NMFS) statistical zones 11 to 21.

| Species | | Old Survey Design (1987-2008) | | New Survey Design | | |
|---|---|---|---|---|---|---|
| | | | | Increased Sampling (2009) | Expanded Coverage (2010-2022) | |
| Common Name | Scientific Name | Mean CV (SD) | CV Range | CV | Mean CV (SD) | CV Range |
| Red snapper | *Lutjanus campechanus* | 0.1694 (0.0319) | 0.1345-0.2556 | 0.1400 | 0.1511 (0.0159) | 0.1351-0.1953 |
| Atlantic croaker | *Micropogonias undulatus* | 0.2571 (0.0438) | 0.2192-0.3893 | 0.1806 | 0.2559 (0.0250) | 0.2193-0.3152 |
| Longspine porgy | *Stenotomus caprinus* | 0.1521 (0.0139) | 0.1294-0.1936 | 0.1167 | 0.2065 (0.0394) | 0.1632-0.3094 |
| Gray triggerfish | *Balistes capriscus* | 0.3035 (0.0627) | 0.2079-0.4119 | 0.2737 | 0.3544 (0.1299) | 0.2109-0.6290 |
| Brown shrimp | *Farfantepenaeus aztecus* | 0.1327 (0.0095) | 0.1214-0.1538 | 0.0939 | 0.1376 (0.0131) | 0.1206-0.1605 |
| White shrimp | *Litopenaeus setiferus* | 0.3118 (0.0452) | 0.2337-0.3989 | 0.1739 | 0.2785 (0.0360) | 0.2153-0.3373 |

Column headings refer to the sampling design used while conducting those years of the survey, with Increased Sampling referring to an increase in station density across statistical zones 11 to 21 prior to the survey expanding in 2010 and the associated reduction in station density.

in the CVs for all species tested (Tables 2, 3), but was only observed during the periods of high, concentrated sampling efforts (e.g., fall of 2008 and summer and fall of 2009). This decrease in the CVs fulfilled the initial goal of the survey design change.

Ultimately, the decision was made to expand survey coverage into the eastern GOM to better assess the distribution and abundance of species across a wider spatial extent and range of benthic habitats. The gains in efficiency inherent in the new survey design allowed for 7.2 (SD=0.80) stations sampled per day to remain near the 7.7 (SD=0.78) samples per day under the old survey design, all while covering nearly twice the original spatial extent. However, the expansion came at the cost of reduced sampling effort and increased CVs in the indices over the historic survey area. Under the new survey design (GOM-wide target of 350 stations), sampling effort in statistical zones 11 to 21 was reduced to roughly 181 stations, a significant reduction from the old survey design target of 230 stations and the 350 stations averaged during 2008/2009. Given the reduction in sample sizes, abundance indices CVs generated from the summer and fall groundfish surveys for the majority of selected taxa from statistical zones 11 – 21 under the new survey design were in line with those generated under the old survey design (Tables 2, 3).

When examining the relative abundance trends for the selected species, there were some differences between the survey designs.

TABLE 3 Comparison of coefficients of variation (CV) and standard deviations (SD) from delta-lognormal models for selected species from the Southeast Area Monitoring and Assessment Program (SEAMAP) Fall Shrimp/Groundfish Survey conducted from 1987 to 2022.

| Species | | Old Survey Design (1987-2007) | | New Survey Design | | | |
|---|---|---|---|---|---|---|---|
| | | | | Increased Sampling (2008/2009) | | Expanded Coverage (2010-2022) | |
| Common Name | Scientific Name | Mean CV (SD) | CV Range | Mean CV (SD) | CV Range | Mean CV (SD) | CV Range |
| Red snapper | *Lutjanus campechanus* | 0.1180 (0.0161) | 0.0979-0.1752 | 0.0912 (0.0064) | 0.0848-0.0976 | 0.1375 (0.0163) | 0.1163-0.1678 |
| Atlantic croaker | *Micropogonias undulatus* | 0.1425 (0.0123) | 0.1272-0.1799 | 0.1082 (0.0016) | 0.107-0.1093 | 0.1628 (0.0184) | 0.138-0.1992 |
| Longspine porgy | *Stenotomus caprinus* | 0.1498 (0.0248) | 0.1131-0.2026 | 0.1162 (0.0069) | 0.1093-0.1231 | 0.1905 (0.0493) | 0.127-0.3163 |
| Gray triggerfish | *Balistes capriscus* | 0.2021 (0.0807) | 0.1238-0.5233 | 0.1735 (0.0546) | 0.1189-0.2281 | 0.2596 (0.0543) | 0.1826-0.3726 |
| Brown shrimp | *Farfantepenaeus aztecus* | 0.1091 (0.0089) | 0.1002-0.1361 | 0.0865 (0.0007) | 0.086-0.0870 | 0.1268 (0.0166) | 0.1062-0.1617 |
| White shrimp | *Litopenaeus setiferus* | Poor model fit | | | | | |

Column headings refer to the sampling design used while conducting those years of the survey, with Increased Sampling referring to an increase in station density across statistical zones 11 to 21 prior to the survey expanding in 2010 and the associated reduction in station density.

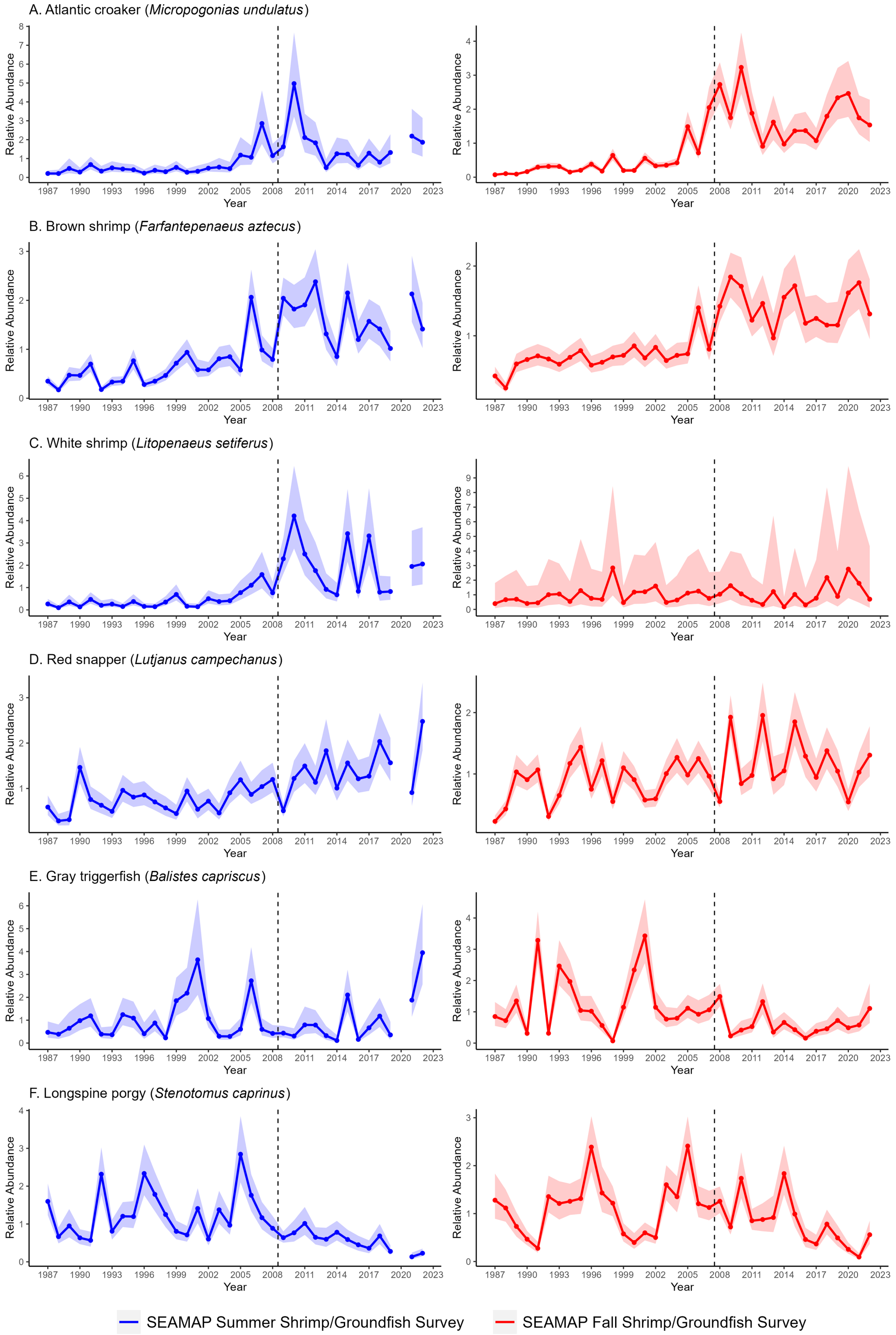



FIGURE 7
**(A-F)** Indices of relative abundance (scaled to a mean of one) for selected species from the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer (blue) and Fall (red) Shrimp/Groundfish surveys conducted from 1987 – 2022 in National Marine Fisheries Service (NMFS) statistical zones 11 – 21. Shaded areas represent the lower and upper 95% confidence intervals. Dashed lines denote implementation of survey design changes.

However, in most cases, there had already been increases or decreases in the years immediately preceding the survey design change, which makes it difficult to discern if they had an actual effect on the trends. For example, Atlantic croaker (Figure 7A), brown shrimp (Figure 7B), and white shrimp (Figure 7C) showed increases in abundance during the 2006 summer and fall groundfish surveys and increased values overall during the new design period, with the exception of white shrimp during the fall groundfish survey. However, the model fits for white shrimp during the fall groundfish survey were poor, evidenced by the large confidence intervals. Trends in red snapper abundance also increased after the survey design change during the summer groundfish survey, but were more erratic during the fall groundfish survey, which is most likely due to the fact the fall groundfish survey is mainly catching age 0 fish (Figure 7D). Gray triggerfish relative abundance trends did not seem to be affected by the survey design change (Figure 7E). Longspine porgy relative abundance trends have shown decreases since the survey design change in the summer and fall groundfish surveys, although in the fall groundfish survey the decrease occurs several years after the survey design change (Figure 7F). The final model results detailing the variables retained in each submodel and their significance can be found in Supplementary Tables S3 and S4 for the summer and fall surveys, respectively.

# 4 Evaluation of spatial survey expansion

## 4.1 Evaluation methods

One of the main motivations for the eastern expansion of the SEAMAP groundfish surveys was to more accurately capture species diversity in the GOM in support of ongoing ecosystem modeling efforts (Ainsworth et al., 2015; Chagaris et al., 2020) and to better understand the spatial range of commercially and recreationally important species (Grüss et al., 2018; Figure 8A–J). To examine the diversity of species captured across the range of the SEAMAP groundfish surveys, the catch composition from stations sampled during the summer and fall surveys from 2010 to 2022 was examined in relation to their region of capture. In addition, the average number of taxa per station by statistical zone was calculated to see if any differences existed in the catch composition.

From a stock assessment perspective, a review is presented of how the survey data are being utilized in stock assessments in the GOM. In doing so, indices of abundance for lane snapper *L. synagris*, red snapper, wenchman *Pristipomoides aquilonaris*, gray triggerfish, and pink shrimp were compared between the western GOM and the full northern GOM. For lane snapper and pink shrimp, indices of abundance were also calculated separately for the eastern GOM because higher catches rates in the area (Table 4; Figures 8B, I). Indices of abundance were calculated using the delta-lognormal index that was described in section 3.1. Each submodel incorporated the variables year (2010 – 2022), time of day (day/night), statistical zone (western GOM: 11 – 21, eastern GOM: 2 – 10, full GOM: 2 – 21), and depth (fm). All variables were retained in both submodels regardless of significance. This analysis is limited to the SEAMAP Summer Groundfish Survey because of its more complete spatial coverage throughout the time series.

## 4.2 Results

SEAMAP groundfish survey operations between 2010 and 2022 encountered 1866 unique taxa. Taxa that were only found in the western GOM numbered 280, while those only found in the eastern GOM numbered 766, with 820 taxa occurring in both regions. On average, the total number of taxa captured per station does not vary drastically between statistical zones (Figure 9). However, catches in the eastern GOM often exhibited a higher range of taxonomic diversity when compared to stations in the western GOM.

Indices of abundance for important commercial and recreational selected species were divided into three groups: (1) similar relative abundance trends with similar CVs, (2) similar relative abundance trends with improved CVs, and (3) divergent relative abundance trends between the western GOM and full GOM indices. Species in the first group included red snapper and wenchman (Figures 10A, B). Vermilion snapper and gray triggerfish were in the second group with similar relative abundance trends and showed an average reduction in CVs of 49% (6% SD) and 37% (11% SD), respectively (Figures 10C, D). Species with divergent relative abundance trends included lane snapper and pink shrimp (Figures 10E, F). For lane snapper, indices of abundance for the eastern GOM and full GOM showed similar increasing trends with the trends in the western GOM being more divergent, showing a relatively flat trend over the same time period. The CVs for the western GOM were on average 36% (18% SD) higher when compared to the eastern GOM model. Indices of abundance for pink shrimp were divergent for the eastern GOM and western GOM, with the full GOM model roughly falling between them. The CVs for both regional models were higher than those of the full GOM model.

## 4.3 Stock assessment applications

The eastward expansion onto the WFS has resulted in an increase in the number of indices used for single-species stock assessments and increased the utility of the survey data to support ecosystem modeling efforts. The SEAMAP groundfish surveys currently provide abundance indices for 17 single-species stock assessments both across the entire GOM and for the eastern GOM. With the expansion of the SEAMAP groundfish surveys onto the WFS, abundance indices for juvenile vermilion snapper *Rhomboplites aurorubens*, red grouper, gray snapper, lane snapper, and hogfish *Lachnolaimus maximus* can now be calculated because they are now captured during the surveys. For several of these species (e.g., red grouper, gray snapper, hogfish), only data from the eastern GOM is included in the stock assessment, while other species (e.g., red snapper) are assessed across the entire GOM. Prior to the eastward expansion of the survey, any index that was produced for the eastern GOM was severely limited in its spatial coverage due to the survey universe ending at Mobile Bay, AL (Figure 3), even if the species' abundance continued onto the WFS [e.g., red snapper (Figure 8A) and gray triggerfish (Figure 8G)]. However, now indices for pink shrimp, gray triggerfish, red snapper, and Spanish mackerel cover a greater spatial extent of their abundance in the northern GOM. In addition, the improvement in CVs for vermilion snapper and gray triggerfish is indicative of the survey expansion into their preferred habitat. Finally, the divergent trends for lane snapper and pink shrimp show the importance of sampling across a wider range of species' habitat to get a more informative picture of abundance, and sampling within their core habitats.

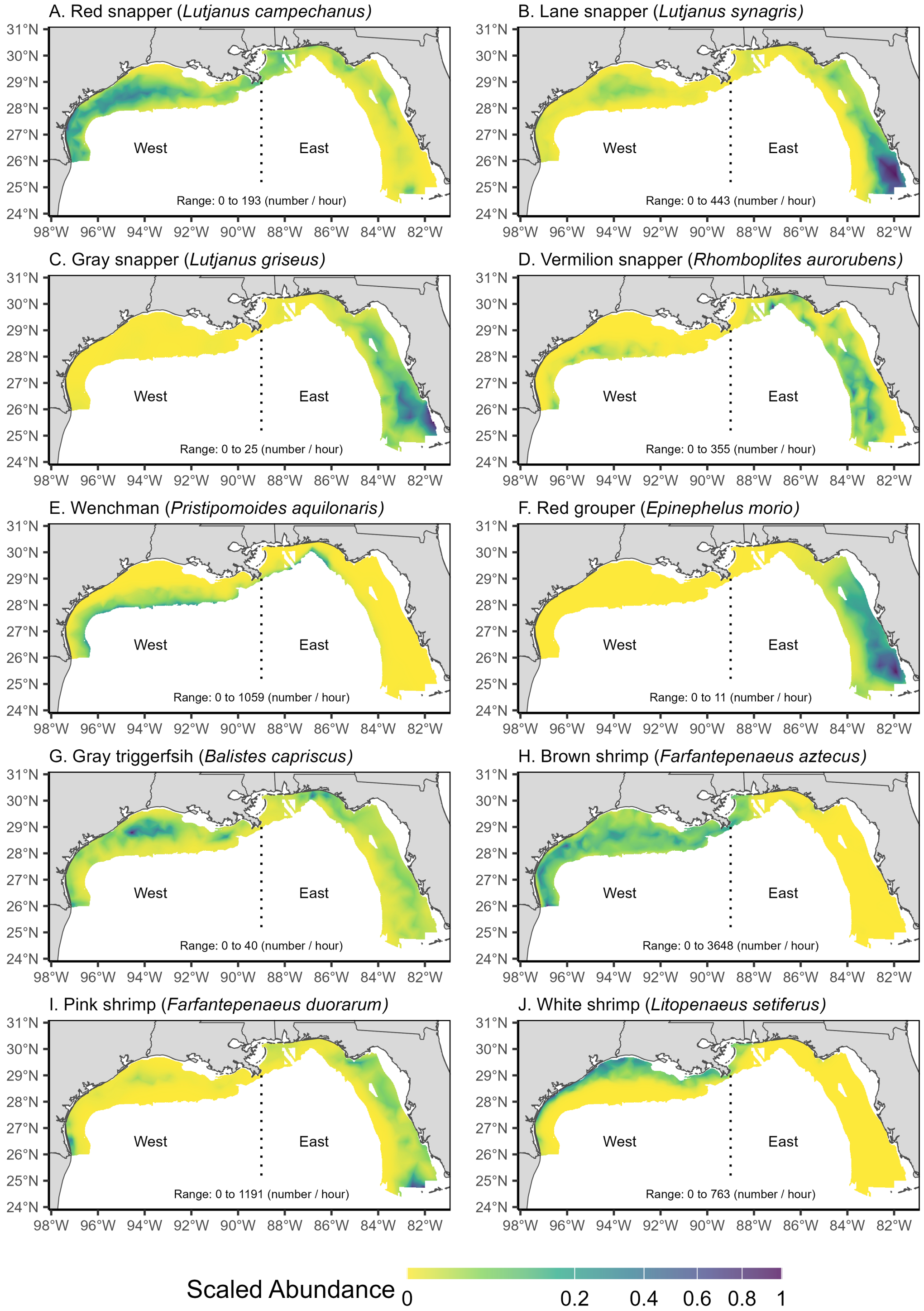



FIGURE 8
**(A-J)** Scaled abundance (0 to 1 with respect to the maximum catch rates in number per hour) of selected species captured during Southeast Area Monitoring and Assessment Program (SEAMAP) Summer and Fall Shrimp/Groundfish Surveys from 2010 – 2022. The range of catch rates are listed for each species in their respective panel.

## 5 Lessons learned

Groundfish surveys conducted by NMFS and SEAMAP have exhibited a pronounced shift in their objectives and design since their inception in the early 1950s. The first significant change came between 1965 and 1971 when NMFS transitioned from exploratory surveys to surveys that attempted to better quantify the assemblages of fish and invertebrates (Bullis and Thompson, 1967). While this was a move in the right direction, the surveys still lacked standardized sampling protocols and/or gear that made

TABLE 4 Number of stations (N), proportion of stations with a positive catch (PPOS), catch per unit effort (CPUE, number per hour), and standard error (SE) for selected species in the historic and expanded coverage from the SEAMAP Summer and Fall Shrimp/Groundfish Surveys (2010-2022) in the Gulf of Mexico (GOM), along with respective proportion of total Gulf of Mexico CPUE.

| Species | | Historic Area (NMFS statistical zone (11–21) | | | | | Expanded Area (NMFS statistical zone (1–10) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Common Name | Scientific Name | N | PPOS | CPUE | SE CPUE | *% Total GOM CPUE* | N | PPOS | CPUE | SE CPUE | *% Total GOM CPUE* |
| Red snapper | *Lutjanus campechanus* | 3792 | 0.57 | 14.34 | 0.51 | 92.28 | 2952 | 0.15 | 1.54 | 0.18 | 7.72 |
| Lane snapper | *Lutjanus synagris* | 3792 | 0.19 | 2.24 | 0.14 | 7.30 | 2952 | 0.43 | 36.49 | 1.98 | 92.70 |
| Gray snapper | *Lutjanus griseus* | 3792 | < 0.01 | 0.01 | 0 | 0.43 | 2952 | 0.20 | 1.71 | 0.12 | 99.57 |
| Vermilion snapper | *Rhomboplites aurorubens* | 3792 | 0.06 | 1.13 | 0.19 | 9.11 | 2952 | 0.31 | 14.47 | 1.98 | 90.89 |
| Wenchman | *Pristipomoides aquilonaris* | 3792 | 0.31 | 15.66 | 0.75 | 86.85 | 2952 | 0.04 | 3.05 | 0.7 | 13.15 |
| Red grouper | *Epinephelus morio* | 3792 | 0 | 0 | 0 | 0 | 2952 | 0.18 | 0.94 | 0.06 | 100.00 |
| Gray triggerfish | *Balistes capriscus* | 3792 | 0.16 | 1.21 | 0.10 | 67.00 | 2952 | 0.16 | 0.77 | 0.12 | 33.00 |
| Brown shrimp | *Farfantepenaeus aztecus* | 3792 | 0.90 | 273.21 | 10.55 | 99.69 | 2952 | 0.03 | 1.09 | 0.25 | 0.31 |
| Pink shrimp | *Farfantepenaeus duorarum* | 3792 | 0.12 | 4.61 | 0.93 | 19.36 | 2952 | 0.29 | 24.69 | 2.42 | 80.64 |
| White shrimp | *Litopenaeus setiferus* | 3792 | 0.30 | 27.78 | 1.93 | 99.68 | 2952 | < 0.01 | 0.12 | 0.07 | 0.32 |

comparisons across surveys difficult, if not impossible. Beginning in 1972, the use of a standardized survey design and gear marked the second significant change, and the first opportunity to attempt to track changes in the offshore fisheries in the northcentral GOM. The continued refinement of sampling methods and the adjustments to the survey designs have continued to help improve the surveys and expand its usefulness across a wider range of species and scientific applications.

Changes in the survey design of a long-term fishery-independent survey have proven to be beneficial but also introduced some uncertainty as to how to accurately account for these changes when using these data in stock assessments. The spatial expansion of the groundfish surveys have led to a better representation of species ranges and distributions across the U.S. GOM (Figure 8). It has also led to more abundance data being available for use in stock assessments, especially for species that primarily occur across the WFS [e.g., red grouper (Figure 8F)]. Increases in survey efficiency (through both survey design changes and technology [e.g., electronic measuring boards, integrated data collection systems, advancement of database entry]) have allowed the northern GOM-wide survey expansion with the minimal addition of days at sea.

Data from the groundfish surveys serve an important role in the stock assessments as it mainly represents size classes of fishes that are often absent from the fishery-dependent data (e.g., landings) and other fishery-independent surveys (e.g., SEAMAP Reef Fish Video Survey (Campbell et al., 2019), NMFS Bottom Longline Survey (Driggers et al., 2008)). The juvenile red snapper can be used to help inform recruitment since based on their lengths, they are representative of the age 0 (fall groundfish survey) and age 1 (summer groundfish survey) year classes and are not present in other standardized fishery-independent surveys across the GOM (Figure 11). Similar trends are seen in the length distributions of red grouper and gray triggerfish across the groundfish and other fishery-independent surveys. Even in cases where the length data distributions overlap with the SEAMAP Reef Fish Video Survey, the groundfish survey consistently catches smaller fish (e.g., gray snapper, lane snapper).

## 5.1 Analytical considerations

From a stock assessment perspective, combining data from different survey designs (i.e., NMFS Fall Groundfish Survey, SEAMAP Fall Groundfish Survey) has been problematic, particularly due to the survey expansions that accompanied them. Historically, two methods have been used when preparing indices of abundance for inclusion in stock assessment using the groundfish survey data. Method 1 restricts the data spatially to maximize the length of the time series over a core area, and method 2 restricts the data temporally to maximize spatial coverage. Both have tradeoffs that need to be considered. By utilizing a spatially restricted full time series, relative abundance trends over a longer time frame can be tracked, in some cases over several decades. However, this method may miss changes in abundance that occur outside of the spatially restricted area. In contrast, the use of the second method would capture the changes in abundance across the entire spatial area, but would be limited in describing the long-term trends seen in the data with the use of method 1.

Nichols (2004a) first addressed the issue of combining data from different survey designs during the stock assessment workshop

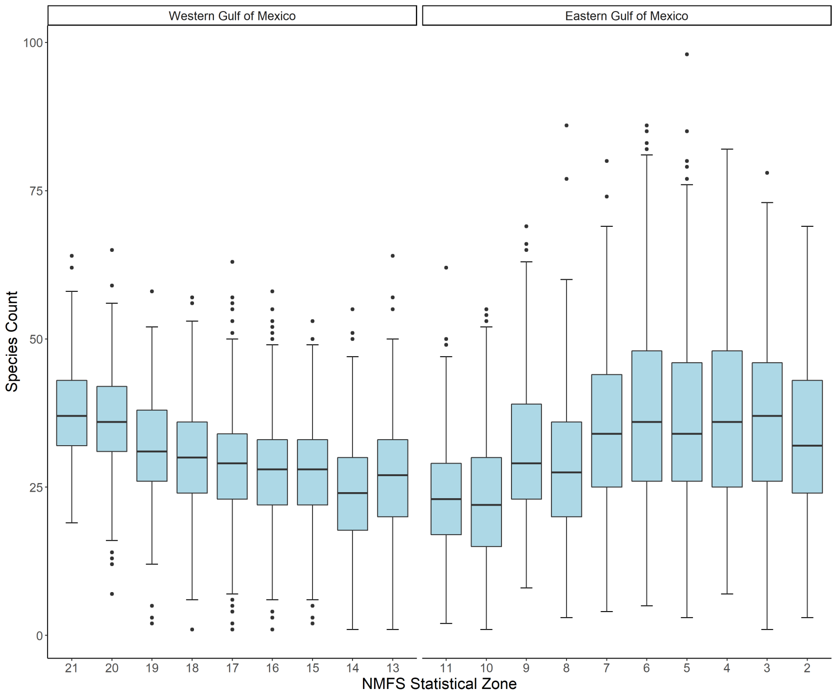



FIGURE 9
Boxplots representing the differences in the number of taxa (Species Count) caught per station by National Marine Fisheries Service (NMFS) statistical zone across the northern Gulf of Mexico from the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer and Fall Shrimp/Groundfish surveys conducted from 2010 – 2022.

(SEDAR, 2005) for red snapper by separating the data by survey design and analyzing annual abundance data as separate time series. Data were also combined as a single continuous time series following the approach of Nichols (2004b), which was ultimately used for the stock assessment model (SEDAR, 2005). Following this precedent, subsequent analysis followed the general recommendation of producing one continuous time series when data were available (e.g., Pollack et al., 2012; Pollack and Ingram, 2013a; SEDAR, 2020). During subsequent stock assessments, a survey design factor was included in the delta-lognormal models to account for any differences in the catches due to the survey design change (Pollack and Ingram, 2013a, 2015). The survey design factor was defined as a class variable with two ("Old Survey Design (1987 – 2008/2009)" and "New Survey Design (2008/2009 – present")) or three ("Early (1972 – 1986)", "Old Survey Design (1987 – 2008/2009, and "New Survey Design (2008/2009 – present")) levels based on the time series being analyzed. Unfortunately, its usefulness was limited due to the lack of overlap between the survey designs.

While the method of combining time series data across survey design changes seemed to be working for several fish species, issues arose with the king mackerel *S. cavalla* index, where initially one index was produced (1972 – present) for the assessment. Further analysis determined there were differing catch rates in the sampled areas (i.e., limited area in the early years vs full western GOM coverage onward from 1987) that raised concerns that one area was driving the index (Pollack and Ingram, 2013a). This resulted in the recommendation that the early part of the time series (1972 – 1986) be removed from the analysis (SEDAR, 2014). A similar issue was encountered during SEDAR 74 for red snapper, where the early parts of the time series for the fall survey were not recommended for use in the final assessment model because of the lack of spatial coverage across the defined subregions (SEDAR, 2022).

Brown and white shrimp also presented an issue with divergence in the scale of catches that was noticed between the survey design periods. As a result, NMFS convened a workgroup to better understand data inputs and assumptions for the GOM penaeid shrimp stock assessment models, current practices for index estimation were evaluated, and best practices were identified. A decision was made to develop split indices for both brown and white shrimp, resulting in early (1987 – 2007/2008) and modern (2008/2009 – present) indices. The reasoning behind this approach was related to possible changes in the catchability of brown and white shrimp resulting from the survey design change, particularly the change in tow direction (Pollack et al., 2021). Following this recommendation from the workgroup, the CPUE indices for red snapper (Pollack and Hanisko, 2022) and Spanish mackerel *Scomberomorus maculatus* (Pollack and Hanisko, 2023) have also been split, echoing the original recommendation put forth by Nichols (2004a).

While the eastward survey expansion in 2008 provided a more complete coverage of several species' distributions (Figure 8), it did present some problems on how to properly account for expanded areas in the abundance indices. In the first years following the expansion, the decision was made to not use any of the data from the expanded area due to the short time frame that the sampling had taken place (Pollack et al., 2012). As more years of data became

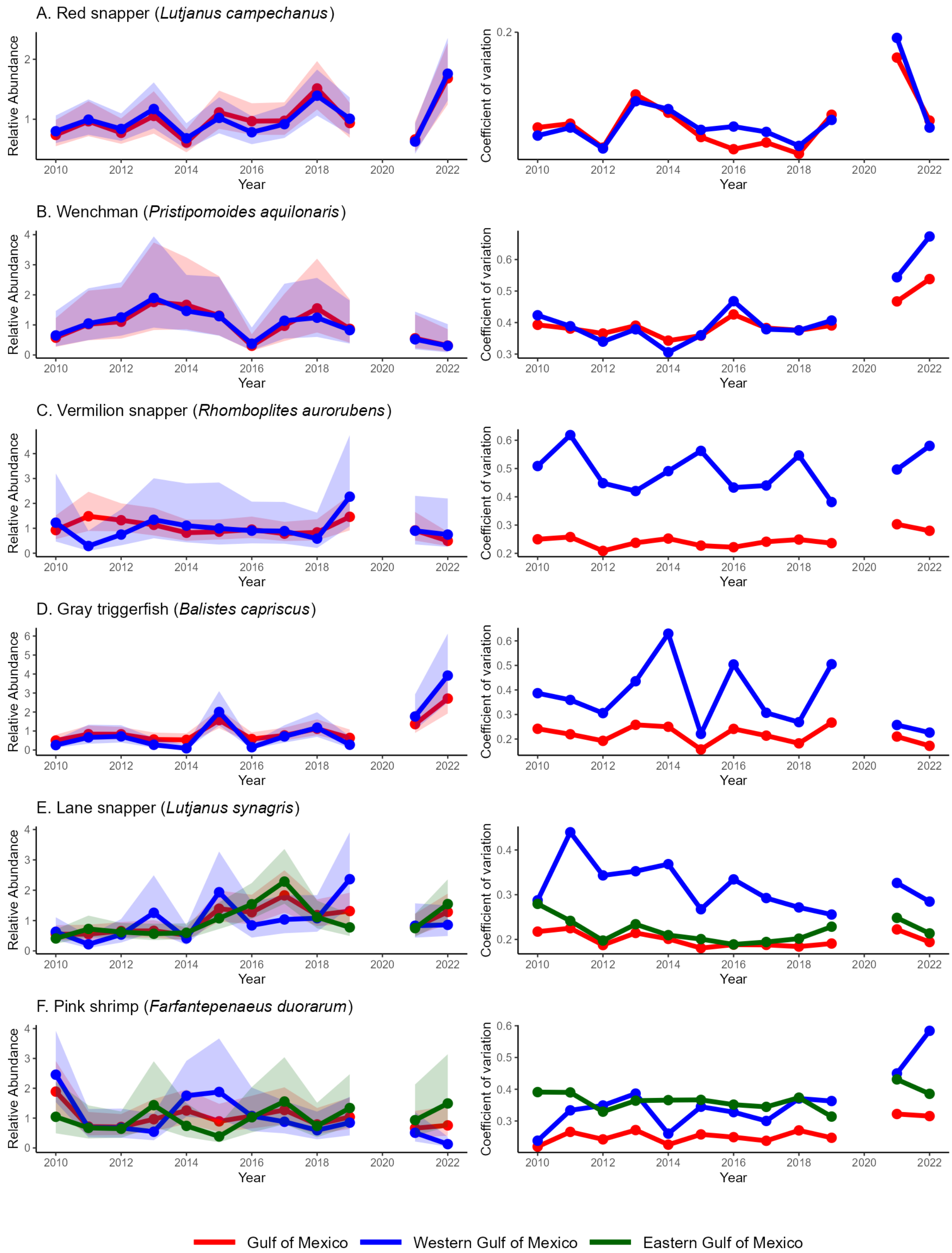



FIGURE 10
**(A-F)** Comparison of indices of relative abundance (scaled to a mean of one) and coefficient of variation for selected species from the Southeast Area Monitoring and Assessment Program (SEAMAP) Summer Shrimp/Groundfish Survey from 2010 – 2022 across the western Gulf of Mexico (GOM), full GOM and eastern GOM. Shaded areas represent the lower and upper 95% confidence intervals. Eastern GOM indices presented only for lane snapper and pink shrimp due to core abundance occurring in that region.

available, data were included using a variable in the delta lognormal model to attempt to account for the expanded area (Pollack and Ingram, 2013b, 2015). There was also an issue of when to begin the time series in the eastern GOM, due to the limited spatial coverage of the surveys in 2008 and 2009. In many cases, as more years of data have become available these early years are dropped from the index in favor of only retaining years with more complete spatial coverage (Pollack and Hanisko, 2022). For many indices in the western GOM, the data prior to 1987 are excluded due to the changes in survey design and spatial coverage (Pollack and Ingram, 2013a, 2015). There are different methods available to deal with some of these issues, such as using an area-weighted approach

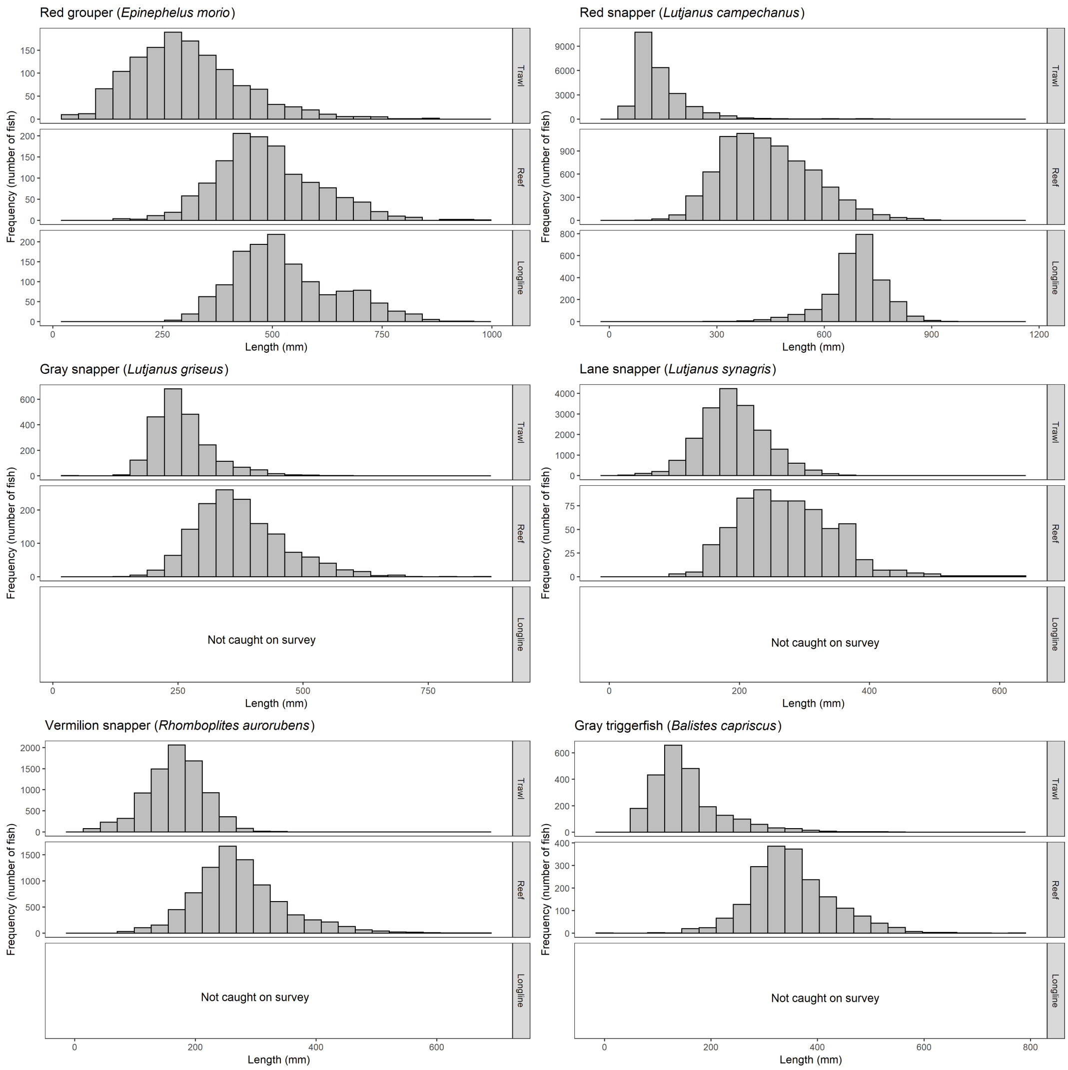



FIGURE 11
Comparison of length distributions for selected species from the Southeast Area Monitoring and Assessment Program (SEAMAP) Shrimp/Groundfish Survey (Trawl), SEAMAP Reef Fish Video Survey (Reef), and National Marine Fisheries Service (NMFS) Southeast Fisheries Science Center (SEFSC) Bottom Longline survey (Longline).

similar to that of Thompson et al. (2022) or a spatiotemporal model (Thorson, 2019), but these options have not been fully explored with the SEAMAP groundfish data.

Another issue becomes how to determine whether changes seen in the groundfish data (e.g., biomass, abundance, species composition) are indicative of survey design changes and/or changes in the commercial groundfish fishery (e.g., regulations, permitting), environmental effects (e.g., hypoxia, river flooding), socioeconomic pressures (e.g., diesel fuel cost, ex-vessel prices), or some combination of them. For example, when examining the annual mean of the total CPUE (kg per km$^2$) of all species within the historic sampling area for the fall groundfish surveys (Figure 2), a case could be made for three distinct trends in the time series (Figure 12). The early part of the time series (1972 – 1986) shows a general decline in annual mean biomass, which mirrors what the commercial groundfish fishery was experiencing, and why the groundfish survey was initiated (Juhl et al., 1973; Gutherz, 1977). The initial years after the survey design change in 1987 saw continued declines in biomass until 1990 when the trend reversed and increased to a peak in 1995. The reversal of the decline in biomass happens to coincide with the implementation of the use of Turtle Excluder Devices in the GOM shrimp fishery (NOAA Fisheries, 2024). During the period of the early SEAMAP groundfish survey (1987 – 2007), the time series was relatively stable and then jumped up in 2008, right after the design change was implemented. The increase in biomass seen during the years of the initial design change (2008 – 2011) appears to be returning to the levels seen during the previous survey design period (1987 – 2007). There were several episodic events, such as a dramatic increase in diesel fuel prices in 2003, Hurricanes Katrina and Rita in 2005, and the Deepwater Horizon Oil Spill in 2010, which had largely negative effects on the shrimp fleet providing a reduction in fishing efforts across the GOM (Posadas and Posadas, 2013; Gallaway et al., 2020).

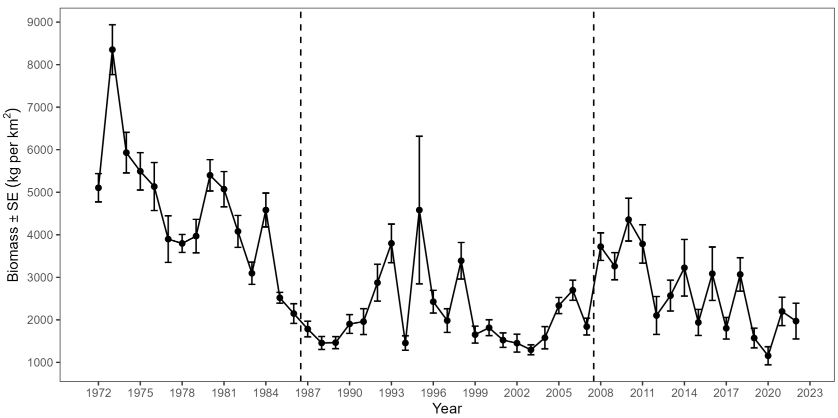



FIGURE 12
Yearly mean biomass (kg per km$^2$) ± standard error (SE) captured in the historic survey area (National Marine Fisheries Service [NMFS] statistical zones 11 – 15) on the NMFS Fall Groundfish Survey and Southeast Area Monitoring and Assessment Program (SEAMAP) Fall Shrimp/Groundfish Survey from 1972 – 2022. Dashed lines denote implementation of survey design changes.

## 5.2 Challenges

There are several challenges to maintaining a long-term fishery-independent monitoring program. Budgetary concerns may be the foremost issue that agencies face when conducting large-scale fishery-independent surveys because these surveys are inherently more expensive than fishery-dependent surveys (Dennis et al., 2015). Often intertwined with budgetary concerns are personnel limitations to conducting the surveys, since they require scientists to staff the surveys and conduct the research. Agency priorities can also present a challenge to maintaining long-term monitoring programs due to the redirection of resources and interest. Ideally, the long-term monitoring project could adapt to these changing priorities, such as survey expansion or design changes that maintain the integrity of the historic data while providing comparable data on the new objectives.

Another potential consequence of changing survey designs are the unanticipated factors that can result in additional changes being made as the survey expands into unfamiliar areas. For the SEAMAP groundfish surveys, the interactions with live bottom (e.g., sponges and soft corals) on the WFS are a particular concern (Hanisko et al., 2018). In this case, as a mitigation measure for avoiding live bottom, trawl paths where more than 50 kg of sponge was collected and trawl paths where any coral was reported in the catch were buffered by 250 m area around the trawl path and removed from the survey universe. However, Christiansen et al. (2022) found that abundance data on certain species such as gray triggerfish and red grouper, of which the latter is found almost exclusively on the WFS, could be impacted by the avoidance of areas with high sponge occurrence. In addition, with the survey expansion, additional expertise was needed to help identify species not previously encountered. Bringing in this expertise, either within an agency or from an outside source, not only ensures the accuracy of the data being collected but can lead to a better understanding of species occurrence and distribution across the sampling area.

## 5.3 Importance of calibration studies

Although survey design improvements have dramatically increased the utility of data provided by the SEAMAP groundfish surveys, the challenges that have been faced in the analysis of groundfish survey data highlight the critical importance of conducting rigorous calibration studies when possible. Initial assessments of the 2008 survey design change indicated that differences between the historical and novel survey designs were minimal (Ingram, 2008). However, the initial assessments of the survey design change were compared between the original SEAMAP design and the new survey design when sampling was much higher in 2008/2009. As additional data were collected after the reduction in sampling effort in the northwestern GOM due to the eastward expansion, questions regarding whether catchability remained constant for some taxa at the lower level of effort have been raised (SEDAR, 2018). Additionally there have also been concerns in regards to alterations to tow direction no longer being consistent across depth strata (SEDAR, 2024). However, there have been no calibration studies conducted between the original SEAMAP design (1987 – summer 2008) and the new survey design (fall 2008 – present) and there are no calibration studies currently being planned.

By conducting calibration studies, one would be able to assess any differences between the survey designs prior to implementing the change while also developing calibration coefficients should any differences be detected (Miller et al., 2010). Nonetheless, calibration studies for large-scale surveys are often not feasible because of (1) their cost, (2) limitation on vessel time, and (3) limitation on personnel time. For example, the calibration study conducted by Miller et al. (2010) used paired tow data from 636 stations to test for differences in catch rates between vessels. While it would be useful to have a similar calibration study done for the SEAMAP groundfish survey, even at a smaller scale, it would be very costly and difficult to implement during the allotted time frames for the

10.3389/fmars.2025.1425362

surveys. However, a calibration study would be one way to fully understand the impacts of the survey design changes for the SEAMAP groundfish survey.

## 6 Conclusion

Groundfish surveys in the northern GOM have evolved over time to be an important source of fishery-independent data for fisheries management. While data from the exploratory surveys conducted in the 1950s and 1960s may not be suitable for tracking abundance over time, they can be used to track species occurrence and distribution throughout the GOM. The SEAMAP groundfish surveys and their predecessor, the NMFS groundfish survey, have and continue to provide data on the abundance, spatial distribution, size distribution, and life history metrics for several commercially and recreationally important species, and in many cases represent data on life stages of fishes not captured in fishery-dependent data. However, changes to the design of a long-term fishery-independent survey can have unforeseen effects on data products and analysis if these changes are not properly accounted for. Documentation and communication is key to data users. Throughout the development of this paper, it was found that while most of the survey design changes are documented, they are primarily reported in gray literature or other obscure resources (e.g., meeting notes or internal memos). While many of these changes have been known to internal data users and have documented in stock assessment reports (e.g., Pollack et al., 2021; Pollack and Hanisko, 2022), this paper serves as the first major step in communicating the changes to SEAMAP groundfish surveys to a wider scientific audience. Finally, changes to the SEAMAP groundfish survey design and expansion of its spatial extent have been beneficial in providing important biological information (e.g., indices of relative abundance) for a greater number of commercially and recreationally important species and spatial regions in the northern GOM to inform stock assessments and fisheries management.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because Research was covered under a federally authorized permit.

## Author contributions

AP: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Visualization, Writing – original draft, Writing – review & editing. EH: Conceptualization, Data curation, Methodology, Resources, Writing – original draft, Writing – review & editing. TS: Conceptualization, Data curation, Methodology, Resources, Visualization, Writing – original draft, Writing – review & editing. DH: Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. JH: Conceptualization, Data curation, Writing – review & editing. JM: Conceptualization, Data curation, Writing – review & editing. FM-A: Conceptualization, Data curation, Writing – review & editing. JR: Conceptualization, Data curation, Writing – review & editing. ZZ: Conceptualization, Data curation, Writing – review & editing. GP: Resources, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding was provided by the National Marine Fisheries Service and the Southeast Area Monitoring and Assessment Project (SEAMAP).

## Acknowledgments

The authors would like to acknowledge and thank all the SEAMAP participants that have contributed their time and expertise since the inception of the program in 1981. Without their dedication and commitment to the mission, none of this data collection would have been possible.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2025.1425362/full#supplementary-material

# References


Ainsworth, C. H., Schirripa, M. J., and Morzaria-Luna, H. N. (2015).An Atlantis Ecosystem Model for the Gulf of Mexico supporting Integrated Ecosystem Assessment. Available online at: https://repository.library.noaa.gov/view/noaa/4875 (Accessed October 24, 2024).

Bullis, H. R. Jr. (1964). *Annual report, exploratory fishing and gear research, Bureau of Commercial Fisheries, Region 2 for fiscal year 1963 ending June 30, 1963* Vol. 193 (Washington, D.C.: U. S. Fish and Wildlife Service. Circular). Available at: https://spo.nmfs.noaa.gov/sites/default/files/legacy-pdfs/CIRC193.pdf (Accessed October 24, 2024).

Bullis, H. R. Jr., and Thompson, J. R. (1965). *Collections by the exploratory Fishing Vessels Oregon, Silver Bay, Combat, and Pelican made during 1956 to 1960 in the southwestern North Atlantic* (Washington, D.C.: U.S. Fish and Wildlife Service. Special Scientific Report-Fisheries No. 510). Available at: https://www.biodiversitylibrary.org/item/39156page/1/mode/1up (Accessed October 24, 2024).

Bullis, H. R. Jr., and Thompson, J. R. (1967). *Progress in exploratory fishing and gear research in region 2, fiscal year 1966* Vol. 265 (Washington, D.C.: U. S. Fish and Wildlife Service. Circular). Available at: https://spo.nmfs.noaa.gov/sites/default/files/legacy-pdfs/CIRC265.pdf (Accessed October 24, 2024).

Bullis, H. R. Jr., and Thompson, J. R. (1970). *Bureau of commercial fisheries exploratory fishing and gear research base, Pascagoula, Mississippi July 1, 1967 to June 30, 1969* Vol. 351 (Washington, D.C.: U.S. Fish and Wildlife Service. Circular). Available at: https://spo.nmfs.noaa.gov/sites/default/files/legacy-pdfs/CIRC351.pdf (Accessed October 24, 2024).

Campbell, M. D., Rademacher, K. R., Noble, B., Salisbury, J., Felts, P., Moser, J., et al. (2019). Status and trends of marbled grouper in the north-central gulf of Mexico. *Mar. Coast. Fish* 11, 114–124. doi: 10.1002/mcf2.10066

Chagaris, D. D., Patterson, W. F. I., and Allen, M. S. (2020). Relative effects of multiple stressors on reef food webs in the northern gulf of Mexico revealed via ecosystem modeling. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00513

Chittenden, M. E. Jr., and Moore, D. (1977). Composition of the ichthyofauna inhabiting the 110-meter contour of the Gulf of Mexico, Mississippi River to the Rio Grande. *Northeast Gulf Sci.* 1, 106–114. doi: 10.18785/negs.0102.06

Christiansen, H. M., Tyler-Jedlund, A. J., Switzer, T. S., Keenan, S. F., and Stahl, S. D. (2022). Potential implications of excluding live-bottom habitat from a bottom-trawl survey in the eastern Gulf of Mexico. *North Am. J. Fish. Manage.* 42, 893–905. doi: 10.1002/nafm.10775

Darnell, R. M., Defenbaugh, R. E., and Moore, D. (1983). *Northwestern Gulf shelf bio-atlas, a study of the distribution of demersal fishes and penaeid shrimp of soft bottoms of the continental shelf from the Rio Grande to the Mississippi River Delta* Vol. 82-04 (U.S. Dept. Interior, Minerals Management Service), xii + 438. Open File Rept.

Darnell, R. M., and Kleypas, J. A. (1987). *Eastern Gulf shelf bio-atlas, a study of the distribution of demersal fishes and penaeid shrimp of soft bottoms of the continental shelf from the Mississippi River Delta to the Florida Keys* (U.S. Dept. Interior, Minerals Management Service), xv + 548. OCS Study MMS 86-0041.

de Mutsert, K., Cowan, J. H., Essington, T. E., and Hilborn, R. (2008). Reanalyses of Gulf of Mexico fisheries data: Landings can be misleading in assessments of fisheries and fisheries ecosystems. *Proc. Natl. Acad. Sci.* 105, 2740–2744. doi: 10.1073/pnas.0704354105

Dennis, D., Plagányi, É., Van Putten, I., Hutton, T., and Pascoe, S. (2015). Cost benefit of fishery-independent surveys: Are they worth the money? *Mar. Policy* 58, 108–115. doi: 10.1016/j.marpol.2015.04.016

Driggers, W. B. III, Ingram, G. W. Jr., Grace, M. A., Gledhill, C. T., Henwood, T. A., Horton, C. N., et al. (2008). Pupping areas and mortality rates of young tiger sharks *Galeocerdo cuvier* in the western North Atlantic Ocean. *Aquat. Biol.* 2, 161–170. doi: 10.3354/ab00045

Fuls, B. E. (2001). *Texas closure dates for the 1993-2000 Texas gulf shrimping seasons* (Austin, TX: Texas Parks and Wildlife Department). Available at: https://tpwd.texas.gov/publications/pwdpubs/media/mds_coastal/Series%202_MDS178.pdf (Accessed October 24, 2024).

Gallaway, B. J., Raborn, S. W., Picariello, L., and Putman, N. F. (2020). Changes in shrimping effort in the gulf of Mexico and the impacts to red snapper. *iScience* 23, 101111. doi: 10.1016/j.isci.2020.101111

Grüss, A., Perryman, H. A., Babcock, E. A., Sagarese, S. R., Thorson, J. T., Ainsworth, C. H., et al. (2018). Monitoring programs of the U.S. Gulf of Mexico: inventory, development and use of a large monitoring database to map fish and invertebrate spatial distributions. *Rev. Fish Biol. Fish.* 28, 667–691. doi: 10.1007/s11160-018-9525-2

GSMFC (Gulf States Marine Fisheries Commission) (2009). *Meeting Minutes from 2008*. Available online at: https://www.gsmfc.org/publications/minutes/2008/2008%20minutes.pdf (Accessed October 24, 2024).

GSMFC (Gulf States Marine Fisheries Commission) (2019). *SEAMAP Operations Manual for Trawl and Plankton Surveys*. Available online at: https://seamapdata.gsmfc.org/trawl/03%20-%20SEAMAP%20Trawling%20Operations%20Manual%203-29-22.pdf (Accessed October 24, 2024).

GSMFC (Gulf States Marine Fisheries Commission) (2024). *Gulf States Marine Fisheries Commission Homepage*. Available online at: https://www.gsmfc.org/ (Accessed October 24, 2024).

Gutherz, E. J. (1977). The northern Gulf of Mexico groundfish fishery, including a brief life history of the croaker (*Micropogon undulatus*). *Proc. Twenty-Ninth Annu. Gulf Caribbean Fish. Instit.* 29, 87–101. Available at: http://hdl.handle.net/1834/28084 (Accessed October 24, 2024).

Hanisko, D. S., Pollack, A. G., Hoffmayer, E. R., and Switzer, T. S. (2018). *Quantifying the impacts to coral and sponge habitats in the eastern Gulf of Mexico during Southeast Area Monitoring and Assessment Program fishery-independent bottom trawl surveys* (Pascagoula, MS: NOAA Technical Memorandum NMFS-SEFSC-723), 17. Available at: https://repository.library.noaa.gov/pdfjs/web/viewer.html?file=https://repository.library.noaa.gov/view/noaa/17865/noaa_17865_DS1.pdf (Accessed October 24, 2024).

Ingram, G. W. (2008). *Evaluation of Summer and Fall SEAMAP Groundfish Surveys in the Western U.S. Gulf of Mexico, (1988-2007) to Assess the Efficacy of Future Changes in Survey Design* (Pascagoula, MS: Internal Memo).

Ingram, G. W., Alvarez-Berastegui, D., Reglero, P., Balbín, R., García, A., and Alemany, F. (2017). Incorporation of habitat information in the development of indices of larval bluefin tuna (*Thunnus thynnus*) in the Western Mediterranean SeSea, (2001–2005 and 2012–2013). *Deep Sea Res. Part II: Topical Stud. Oceanogr.* 140, 203–211. doi: 10.1016/j.dsr2.2017.03.012

Juhl, R. J., Gutherz, E. J., Roithmayr, C. M., and Russel, G. M. (1973). *Southeast groundfish program status report to the Gulf States Marine Fisheries Commission* (Pascagoula, Miss: Mississippi Laboratory), 39567.

Juhl, R. J., Roe, R. B., Gutherz, E. J., Drummond, S. B., Roithmayr, C. M., and Benigno, J. (1974). *Oceanic Resource Surveys and Assessment program status report to the Gulf States Marine Fisheries Commission* (Pascagoula, Miss: Mississippi Laboratory), 39568.

Kutkuhn, J. H. (1962). *Gulf of Mexico commercial shrimp populations-trends and characteristics 1956-59*. Available online at: https://fisherybulletin.nmfs.noaa.gov/sites/default/files/pdf-content/fish-bull/kutkuhn.pdf (Accessed October 24, 2024).

Latour, R. J., Gartland, J., and Bonzek, C. F. (2023). Design and redesign of a bottom trawl survey in Chesapeake Bay, USA. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1217792

Lo, N. C., Jacobson, L. D., and Squire, J. L. (1992). Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish. Aquat. Sci.* 49, 2515–2526. doi: 10.1139/f92-278

Maunder, M. N., and Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* 70, 141–159. doi: 10.1016/S0165-7836(03)00002-X

Maunder, M. N., and Starr, P. J. (2003). Fitting fisheries models to standardised CPUE abundance indices. *Fish. Res.* 63, 43–50. doi: 10.1016/S0165-7836(03)00002-X

Miller, T., Das, C., Politis, P., Miller, A. S., Lucey, S., Legault, C., et al. (2010). *Estimation of Albatross IV to Henry B. Bigelow calibration factors* (Woods Hole, MA: Northeast Fisheries Science Center Reference Document 10-05), 233. Available at: https://repository.library.noaa.gov/view/noaa/3726/noaa_3726_DS1.pdf (Accessed October 24, 2024).

MSFCMA (Magnuson-Stevens Fishery Conservation and Management Act) (2007). *Magnuson-Stevens Fishery Conservation and Management Act (MSFCMA), Public Law 94-265. 2007. As amended by the Magnuson-Stevens Fishery Conservation and Management Reauthorization Act (P.L. 109-479). U.S. Department of Commerce* (Washington, D.C.: National Oceanic and Atmospheric Administration, National Marine Fisheries Service).

Nance, J. M. (1993). *Biological review of the 1992 Texas Closure* (Galveston, TX: NOAA Technical Memorandum, NMFS-SEFSC-325), 30. Available at: https://repository.library.noaa.gov/view/noaa/8985 (Accessed October 24, 2024).

Nichols, S. (2004a). *Derivation of red snapper time series from SEAMAP and groundfish trawl surveys. SEDAR7-DW01* (North Charleston, SC: SEDAR), 28. Available at: https://sedarweb.org/documents/s7dw01-derivation-of-red-snapper-time-series-from-seamap-and-groundfish-trawl-surveys/ (Accessed October 24, 2024).

Nichols, S. (2004b). *Calibration among the separate trawl survey programs to extend the time series for juvenile red snapper indexes. SEDAR7-DW2* (North Charleston, SC: SEDAR), 65. Available at: https://sedarweb.org/documents/s7dw02-calibration-among-the-separate-trawl-survey-programs-to-extend-the-time-series-for-juvenile-snapper-indexes/ (Accessed October 24, 2024).

NMFS (National Marine Fisheries Service) (1975). *Program development plan for resource assessment fiscal year 1977* (Pascagoula, Mississippi: National Marine Fisheries Service, Southeast Fisheries Science Center, Pascagoula Fisheries Laboratory), 39567.

NOAA (National Oceanic and Atmospheric Administration) (1985). *Gulf of Mexico coastal and ocean zones strategic assessment: data atlas* (Rockland, MD: U.S. Dept. Commerce. NOAA, NOS).

NOAA (National Oceanic and Atmospheric Administration) Fisheries (2022a). *Strategic Plan 2022-2025*. Available online at: https://media.fisheries.noaa.gov/2022-12/NOAA-Fisheries-2022-25-StrategicPlan.pdf (Accessed October 24, 2024).

NOAA (National Oceanic and Atmospheric Administration) Fisheries (2022b). *A century of conservation: A brief history of NOAA Fisheries* (NOAA Fisheries). Available at: https://www.fisheries.noaa.gov/new-england-mid-atlantic/about-us/century-conservation-brief-history-noaa-fisheries (Accessed October 24, 2024).

NOAA (National Oceanic and Atmospheric Administration) Fisheries (2023). *Status of stocks 2022: Annual report to Congress on the status of U.S. Fisheries*. Available online at: https://www.fisheries.noaa.gov/s3/2023-04/2022-Status-of-Stocks-RtC-041423-0.pdf (Accessed October 24, 2024).

NOAA (National Oceanic and Atmospheric Administration) Fisheries (2024). *History of Turtle Excluder Devices* (NOAA Fisheries. NOAA). Available at: https://www.fisheries.noaa.gov/southeast/bycatch/history-turtle-excluder-devices (Accessed October 24, 2024).

Pollack, A. G., and Hanisko, D. S. (2022). *Red snapper abundance indices from groundfish surveys in the northern Gulf of Mexico. SEDAR74-DW-30* (North Charleston, SC: SEDAR), 74. Available at: https://sedarweb.org/documents/sedar-74-dw-30-red-snapper-abundance-indices-from-groundfish-surveys-in-the-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., and Hanisko, D. S. (2023). *Spanish mackerel abundance indices from SEAMAP groundfish surveys in the northern Gulf of Mexico. SEDAR81-WP-09* Vol. 28 (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/sedar-81-wp-09-spanish-mackerel-abundance-indices-from-seamap-groundfish-surveys-in-the-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., and Ingram, G. W. Jr. (2013a). *King mackerel abundance indices from SEAMAP groundfish surveys in the northern Gulf of Mexico. SEDAR38-DW-02* (North Charleston, SC: SEDAR), 71. Available at: https://sedarweb.org/documents/s38dw02-king-mackerel-abundance-indices-from-seamap-groundfish-surveys-in-the-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., and Ingram, G. W. Jr. (2013b). *Atlantic sharpnose and bonnethead abundance indices from NMFS Bottom Longline Surveys in the western North Atlantic and northern Gulf of Mexico. SEDAR34-WP-15* (North Charleston, SC: SEDAR), 41. Available at: https://sedarweb.org/documents/s34wp15-atlantic-sharpnose-and-bonnethead-abundance-indices-from-nmfs-bottom-longline-surveys-in-the-western-north-atlantic-and-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., and Ingram, G. W. Jr. (2015). *Gray triggerfish abundance indices from SEAMAP groundfish surveys in the northern Gulf of Mexico. SEDAR43-WP-11* (North Charleston, SC: SEDAR), 30. Available at: https://sedarweb.org/documents/sedar-43-wp11-gray-triggerfish-abundance-indices-from-seamap-groundfish-surveys-in-the-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., Ingram, G. W. Jr., and Foster, D. G. (2012). *Red snapper abundance indices from SEAMAP groundfish surveys in the northern Gulf of Mexico. SEDAR31-DW20* Vol. 236 (North Charleston, SC: SEDAR). Available at: https://sedarweb.org/documents/s31dw20-red-snapper-abundance-indices-from-seamap-groundfish-surveys-in-the-northern-gulf-of-Mexico/ (Accessed October 24, 2024).

Pollack, A. G., Masi, M., Hanisko, D. S., Munch, S., Nance, J., Craig, K., et al. (2021). *SEAMAP Trawl Shrimp Data and Index Estimation Work Group Report* (Pascagoula, MS: A Report to the Gulf of Mexico Marine Fisheries Commission), 118. Available at: https://sedarweb.org/documents/sedar-rd01-seamap-trawl-shrimp-data-and-index-estimation-work-group-report/ (Accessed October 24, 2024).

Posadas, B. C., and Posadas, B. A. Jr. (2013). *Estimation of the Baseline for Assessment of Economic Impacts of the Gulf of Mexico Oil Spill on Mississippi's Commercial Fishing Sector*. Available online at: https://repository.library.noaa.gov/view/noaa/38158 (Accessed October 24, 2024).

Punt, A. E. (2023). Those who fail to learn from history are condemned to repeat it: A perspective on current stock assessment good practices and the consequences of not following them. *Fish. Res.* 261, 106642. doi: 10.1016/j.fishres.2023.106642

Rago, P. J. (2005). "Fishery independent sampling: survey techniques and data analyses," in *Management techniques for elasmobranch fisheries*, vol. 474 . Ed. J. A. Musick (Food & Agriculture Organization, Roma), 201–215. FAO fisheries technical paper.

Roithmayr, C. M. (1965). *Industrial bottomfish fishery of the northern Gulf of Mexico 1959-63* (Washington, D.C.: U. S. Fish and Wildlife Service). Available at: https://www.biodiversitylibrary.org/item/38427page/1/mode/1up. Special Scientific Report-Fisheries, No. 518. (Accessed October 24, 2024).

Rourke, M. L., Fowler, A. M., Hughes, J. M., Broadhurst, M. K., DiBattista, J. D., Fielder, S., et al. (2022). Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. *Environ. DNA* 4, 9–33. doi: 10.1002/edn3.185

Sanders, N., Donaldson, D., and Thompson, P. A. (1990b). *SEAMAP environmental and biological atlas of the Gulf of Mexic* (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 337. Available at: https://www.gsmfc.org/publications/GSMFC%20Number%20022.pdf (Accessed October 24, 2024).

Sanders, N., Donaldson, D., and Thompson, P. A. (1991). *SEAMAP environmental and biological atlas of the Gulf of Mexic* (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 318. Available at: https://www.gsmfc.org/publications/GSMFC%20Number%20025.pdf (Accessed October 24, 2024).

Sanders, N., Donaldson, D., and Thompson, P. A. (1992). *SEAMAP environmental and biological atlas of the Gulf of Mexico 1990* (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 311. Available at: https://www.gsmfc.org/publications/GSMFC%20Number%20027.pdf (Accessed October 24, 2024).

Sanders, N., Van Devender, T., and Thompson, P. A. (1990a). "SEAMAP environmental and biological atlas of the Gulf of Mexico 1986," (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 328. Available at: https://www.gsmfc.org/publications/GSMFC%20Number%20020.pdf (Accessed October 24, 2024).

Schrandt, M. N., MacDonald, T. C., Winner, B. L., Blewett, D. A., Paperno, R., Onorato, G. V., et al. (2024). Sampling design modifications to a fishery-independent monitoring survey balance the maintenance of long-term data with emerging management needs and funding limitations. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1199076

SEAMAP (Southeast Area Monitoring and Assessment Program) (2024). *Gulf States SEAMAP*. Available online at: https://seamapdata.gsmfc.org/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2005). *Stock Assessment Report of SEDAR 7: Gulf of Mexico Red Snapper* (North Charleston, SC: SEDAR), 480. Available at: https://sedarweb.org/documents/sedar-7-sar1-red-snapper-stock-assessment-report/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2009). *Stock Assessment of Red Snapper in the Gulf of Mexico – SEDAR Update Assessment* (North Charleston, SC: SEDAR), 224. Available at: https://sedarweb.org/documents/2009-update-sedar-07-gulf-of-Mexico-red-snapper/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2014). *SEDAR 38 Stock assessment report Gulf of Mexico king mackerel* (North Charleston, SC: SEDAR), 465. Available at: https://sedarweb.org/documents/sedar-38-final-stock-assessment-report-gulf-of-Mexico-king-mackerel/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2015). *SEDAR Procedural Workshop 7: Data Best Practices* (North Charleston SC: SEDAR), 151. Available at: http://sedarweb.org/pw-07 (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2018). *SEDAR 52 Stock Assessment Report Gulf of Mexico Red Snapper* (North Charleston, SC: SEDAR), 434. Available at: https://sedarweb.org/documents/sedar-52-gulf-of-Mexico-red-snapper-final-stock-assessment-report/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2020). *SEDAR 38 Update Gulf of Mexico king mackerel* (North Charleston, SC: SEDAR), 82. Available at: https://sedarweb.org/documents/2019-update-sedar-38-update-assessment-report-gulf-of-Mexico-king-mackerel/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2022). *SEDAR 74 Gulf of Mexico Red Snapper Section III: Data Workshop Report* (North Charleston, SC: SEDAR), 225. Available at: https://sedarweb.org/documents/sedar-74-gulf-of-Mexico-data-workshop-report/ (Accessed October 24, 2024).

SEDAR (Southeast Data, Assessment, and Review) (2024). *SEDAR 87 Gulf of Mexico White, Pink, and Brown Shrimp Section II: Data Workshop Report* (North Charleston, SC: SEDAR), 94. Available at: https://sedarweb.org/documents/sedar-87-gulf-of-Mexico-white-pink-and-brown-shrimp-data-workshop-report/ (Accessed October 24, 2024).

Shrimp/Groundfish Work Group (2017). *Meeting summary for the Shrimp/Groundfish Work Group – Thursday, May 18, 2017* (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 2.

Springer, S., and Bullis, H. R. (1956). *Collections by the Oregon in the Gulf of Mexico: List of crustaceans, mollusks, and fishes identified from collections made by the exploratory Fishing Vessel Oregon in the Gulf of Mexico and adjacent seas 1950 through 1955* (Washington, D.C.: U.S. Department of the Interior, Bureau of Commercial Fisheries). Available at: https://spo.nmfs.noaa.gov/sites/default/files/legacy-pdfs/SSRF196.pdf (Accessed October 24, 2024).

Stuntz, W. E., Bryan, C. E., Savastano, K., Waller, R. S., and Thompson, P. A. (1985). *SEAMAP environmental and biological atlas of the Gulf of Mexic* (Ocean Springs, MS: Gulf States Marine Fisheries Commission), 145. Available at: https://www.gsmfc.org/publications/SEAMAP/SEAMAP%20Environmental%20and%20Biological%20Atlas%201982.PDF (Accessed October 24, 2024).

Switzer, T. S., Keenan, S. F., Thompson, K. A., Shea, C. P., Knapp, A. R., Campbell, M. D., et al. (2023). Integrating assemblage structure and habitat mapping data into the design of a multispecies reef fish survey. *Mar. Coast. Fish* 15, e10245. doi: 10.1002/mcf2.10245

Thompson, K. A., Switzer, T. S., Christman, M. C., Keenan, S. F., Gardner, C. L., Overly, K. E., et al. (2022). A novel habitat-based approach for combining indices of abundance from multiple fishery-independent video surveys. *Fish. Res.* 247, 106178. doi: 10.1016/j.fishres.2021.106178

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fish. Res.* 210, 143–161. doi: 10.1016/j.fishres.2018.10.013

Watson, J. W., and Bane, N. (Eds.) (1985). *Proceedings: SEAMAP Shrimp and Bottomfish Sampling Gear Workshop* Vol. 12 (Ocean Springs, MS: Gulf States Marine Fisheries Commission). Available at: https://www.gsmfc.org/publications/GSMFC%20Number%20012.pdf (Accessed October 24, 2024).

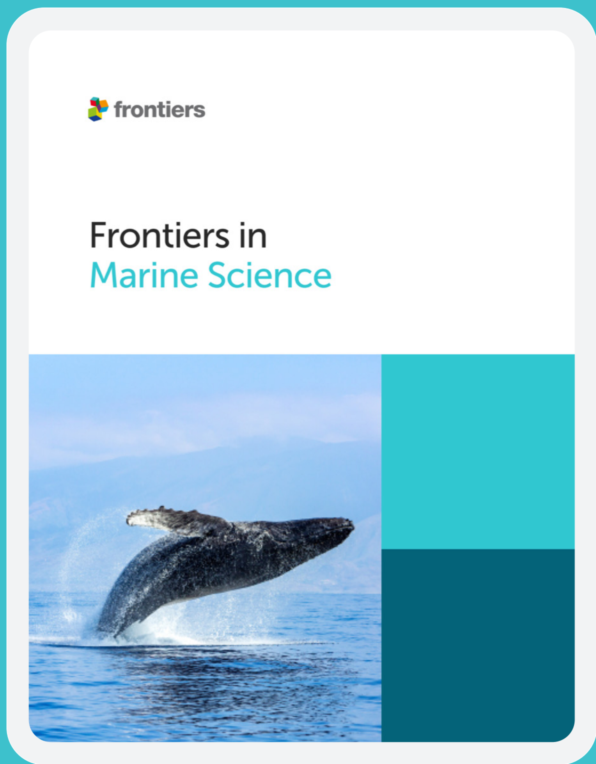
# Frontiers in
# Marine Science

**Explores ocean-based solutions for emerging global challenges**

The third most-cited marine and freshwater biology journal, advancing our understanding of marine systems and addressing global challenges including overfishing, pollution, and climate change.

## Discover the latest Research Topics

See more →




Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact