# Incorporating vertical distribution in index standardization accounts for spatiotemporal availability to acoustic and bottom trawl gear for semi-pelagic species

**Cole C. Monnahan[1,2*], James T. Thorson[1], Stan Kotwicki[1], Nathan Lauffenburger[1], James N. Ianelli[1], Andre E. Punt[2]**

[1] Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, National Marine Fisheries Service 7600 Sand Point Way NE, Seattle, WA 98115, USA.

[2] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA, 98195, USA

[*] Corresponding author's email: cole.monnahan@noaa.gov

**Keywords:** index standardization; spatiotemporal model; walleye pollock (*Gadus chalcogrammus*); VAST; vertical gear availability

## Abstract

Abundance indices from scientific surveys are key stock assessment inputs, but when the availability of fish varies in space and time, the estimated indices and associated uncertainties do not accurately reflect changes in population abundance. For example, indices for many semi-pelagic species rely on acoustic and bottom trawl gear that differ in water column coverage, and so spatiotemporal trends in fish vertical distribution affect the availability of fish to each gear type. The gears together cover the whole water column, and so in principle can be used to estimate more accurate combined indices of the whole population. Here, we extend previous methods and develop a vertically-integrated index which accounts for spatiotemporal correlation and works with data unbalanced spatially or unpaired from distinct surveys. Using eastern Bering Sea walleye pollock (*Gadus chalcogrammus*) as an example, we identified clear spatial and temporal patterns in vertical distribution and gear availability from 2007-2018. Estimated acoustic annual vertical availability ranged from 0.339 to 0.888 among years, and from 0.588 to 0.911 for the bottom trawl survey. Our results highlight the importance of accounting for the spatiotemporal and vertical distribution of semi-pelagic fish to estimate more accurate indices, and provide important context for gear availability.

# Introduction

Information about fish distribution in space and time is valuable both for understanding diverse ecological processes and for guiding applied fisheries management decisions. One important applied case is in quantifying how the relative biomass of a fish stock varies over time, known as an index of biomass or abundance. These indices are typically derived from catch and effort data after controlling for external factors (index standardization; Maunder and Punt, 2004). Resulting indices then inform stock assessments either by direct application or within statistical population dynamics models to provide fisheries management advice (Hilborn and Walters, 1992), and so the accuracy and precision of indices is important to provide reliable fisheries management advice. The accuracy of indices can vary based on changes in the catchability coefficient, a parameter typically used to link indices to modelled abundances caused by changes in survey gear efficiency and fish availability (i.e., the fraction of the stock available to the gear). Scientific surveys of fish stocks use standardized sampling and data collection protocols to minimize changes in the catchability coefficient (Gunderson, 1993). Despite this, fish availability may still vary in time and space and adversely affect the index trends and accuracy of uncertainty estimates (e.g., Kotwicki *et al.*, 2018; Kotwicki and Ono, 2019). Two important examples of changing availability to a survey are when the population moves outside of the spatial extent of the survey (spatial availability), or if fish are present but only partially susceptible to detection by the sampling method (gear availability). Of particular concern is when availability is inconsistent among years, because that appears as a change in abundance which can have negative consequences on the management of a stock (Hilborn and Walters, 1992). Consequently, explicitly accounting for variation in availability should improve the accuracy of indices for stock assessments and the management advice they provide.

Vertical availability has been a longstanding concern for semi-pelagic species because there are vertical regions unavailable to gears used for sampling (Godø and Wespestad, 1993; Michalsen et al., 1996). Bottom trawls miss pelagic fish above the effective fishing height, while acoustic gear misses demersal fish which cannot be detected acoustically (Dead zone; Fig. 1; Kotwicki *et al.*, 2013). Consequently, as the vertical distribution changes (e.g., a population-level shift off bottom, or localized shifts caused by dynamic environmental conditions), the proportion of fish available to each gear type will also vary in space and time (e.g., Michalsen *et al.*, 1996; Kotwicki *et al.*, 2015). Since neither gear can enumerate the entire population in the presence of variation in vertical distribution, previous studies have recognized the need to combine estimates from acoustic and bottom trawl surveys as a way to provide more accurate abundance indices (e.g., Ona *et al.*, 1991; Godø and Wespestad, 1993; Aglen, 1996; Everson *et al.*, 1996). Some studies investigated whether acoustic observations at and between trawl locations could reduce variance (e.g., Beare *et al.*, 2004; Bouleau *et al.*, 2004; Hjellvik *et al.*, 2007). In contrast, Kotwicki *et al.* (2018) predicted gear overlap as a function of environmental covariates using only acoustic data collected at trawling locations (i.e., paired data). These studies relied on a single survey using both gears, and none directly estimated spatially-correlated vertical density (and thus availability) which we consider limitations in many situations.

Accounting for spatial autocorrelation is important because it has several advantages over conventional post-stratification of design-based estimators of survey data. This includes improved precision with little change in bias and the ability to extract spatial statistics such as range shifts or concentration that provide useful contextual and ecological information (Thorson *et al.*, 2015b). Spatiotemporal index standardization methods are increasingly used in a variety of situations (e.g., table 1 of Thorson, 2019), and are available as stand-alone analyses (e.g., Kai *et al.*, 2017; Monnahan and Stewart, 2018) or within generic software platforms such as the vector autoregressive state space modeling platform (VAST; Thorson and Barnett, 2017; Thorson, 2019). Another important advantage of spatial modeling is the capability to mitigate potential bias arising from spatially unbalanced sampling. This is particularly advantageous when combining gears because it means the data from the two gear types do not need to sample at the same places in space and time but are sufficiently similar in seasonal timing that they sample the same spatiotemporal patterns. This may occur if e.g., one gear is unavailable at some locations, or if there are distinct acoustic and bottom trawl surveys with different sampling designs and protocols and may cover different but overlapping spatial footprints. In the extreme, one gear type may be missing for one or more entire years due to budget limitations or planned survey reductions, or unexpected cancellations (O'Leary *et al.*, 2020). Spatial models thus provide both improved estimators and the flexibility to use a wider variety of data beyond spatially balanced, paired data, effectively expanding the potential applications to a wider set of stocks and regions.

Despite popularity and advantages of spatial models, no previous analyses estimated the vertical distribution of the study species using such methods. We hypothesize that extending previous analyses of vertical distribution (Kotwicki *et al.*, 2013, 2018) using spatiotemporal index standardization methods will account for changes in gear availability and provide more accurate indices, i.e., those that are more likely to be proportional to true abundance. For instance, consider the vertical distribution of walleye pollock (*Gadus chalcogrammus*; hereafter pollock) in the eastern Bering Sea (EBS), one of the largest and most valuable commercial fisheries in the US with $1.38 billion USD in wholesale value in 2018 (table 7 of Ianelli *et al.*, 2019). The vertical distribution of pollock is affected by abiotic sources such as water temperature, current velocity, bottom depth, light conditions, sediment size, and biotic sources including size-structure and prey availability (Kotwicki *et al.*, 2013, 2015). The pollock stock assessment uses design-based indices based on two distinct surveys as independent data sources due to a lack of methodology to estimate a reliable combined index. Currently, estimates of time-varying bottom trawl catchability are used to account for variation in vertical availability (Ianelli *et al.*, 2019). Estimating a combined index is not unique to pollock or the Bering Sea region, but extends to other gadoids and semi-pelagic species such as Argentinian hake (Álvarez-Colombo *et al.*, 2014), various species in the Barents Sea (Aglen, 1996; Jakobsen *et al.*, 1997; Ono *et al.*, 2018), cod and haddock in the Northeast Atlantic (e.g., Godø and Wespestad, 1993; Michalsen *et al.*, 1996), among others. Thus, a modeling framework capable of combining acoustic and trawl data, while accounting for spatial dependence, would be valuable for global stock assessment and management.

Here we present a method to explicitly estimate the vertical distribution of fish density in discrete depth layers and how it changes in space and time using two disparate data sources. We develop and describe a new vertically-integrated 'combined' spatiotemporal index standardization method that can be fitted to paired or unpaired acoustic and bottom trawl data with missing years and unbalanced spatial sampling designs, and accounts for spatial autocorrelation, covariate effects, and gear efficiency (catchability). We use pollock survey data from 2007 to 2018 as a case study, estimating the vertical gear availability (percentage fish available) of each gear type in space, and aggregated total vertical density across space to construct an abundance index. Finally, we use simulation to test the statistical properties of the combined index. This allows us to evaluate biases in stock assessment when changes in fish vertical distribution occur but are ignored in their evaluation. We show the extent that this new method presented here can improve an index when fish are vertically distributed with annual variation in that distribution.

# Methods

In this section we provide an overview of the analysis, discuss the data requirements, and provide the statistical details of the combined index. We then describe a fit to pollock as a case study and do a series of model simulations to test the properties of this index, including if we mistakenly assume that there is no variation in vertical availability of fish.

## Modeling framework overview and assumptions

Conceptually, we assume that fish density $d$ in a given year varies in three-dimensional space as a function of spatially-correlated physical and biological processes. This three-dimensional space includes latitude, longitude, and depth (vertical dimension above seafloor). Throughout this work, the water column is divided into three vertical layers referenced from the seafloor. We use data from two survey platforms to estimate this three-dimensional density: one survey that uses bottom trawls to catch fish on and near the sea floor, and the second using acoustic echosounders that detect fish in mid-water but miss near-bottom fish (Fig. 1a). The acoustic observations can be post-processed into arbitrary vertical layers, to be treated separately. The whole water column (with the exception of a near-surface acoustic blind zone) is thus sampled by at least one survey and this allows the density $d$ to be estimated for each depth layer (i.e., vertically).

The depth layers are defined by two heights off bottom $h_1$ (the lower limit of the acoustic gear) and $h_2$ (the upper limit of the bottom trawl gear). These define three depth layers: 1) from sea bottom to $h_1$ which is only sampled by the trawl, 2) $h_1$ to $h_2$ which is sampled by both gears (the overlap layer), and 3) above $h_2$ sampled only by the acoustic gear (Fig. 1a). To illustrate, assume the true vertically-integrated density (in kg/km$^2$) of fish is 500 below $h_1$, 250 between $h_1$ and $h_2$, and 40 above $h_2$, for a total (entire water column) vertical density of 500+250+40=790. The fish available to the acoustic gear is then 290/790 ≈ 0.37 and to the bottom trawl is 750/790 ≈ 0.95. Note that the sum of vertical availabilities exceeds one because of the overlap in sampling. Now, after sampling with error, we have two acoustic observations (expected values 250 and 40 since they sample from the overlap layer and layer above $h_2$) and one bottom trawl observation with an expected value of 750 (the two layers below $h_2$; Fig. 1). Our goal is to infer the density in each depth layer while accounting for spatial correlation and other factors. From these estimates, we can integrate over areas and depth layers to get total biomass, vertical availability, and other quantities of interest and investigate how they vary in space and time.

## Acoustic and bottom trawl surveys

We provide details of the data collection and processing in the supplementary materials, and only provide key summaries here. The total spatial extent of our study, the eastern Bering Sea (EBS), is defined as the spatial extent of the bottom trawl data (Fig. 2), to maintain consistency with previous studies and the current stock assessment (Ianelli *et al.*, 2019). Acoustic data used in this analysis were available annually from 2007-2010 and biennially from 2010-2018 along transects that are spaced at 20 nmi apart (Fig. 2) vertically between 0.5 m off bottom to 16 m from the surface. Some acoustic transects extended beyond the extent of the EBS as defined here and were filtered out. These acoustic data were processed to be referenced from the sea floor instead of the sea surface as is commonly done, and were aggregated into groups of 20 consecutive 0.5 nmi intervals, resulting in $n$=3,830 total observations, to reduce the computational burden and result in roughly the same number of observations between gears. Annual bottom trawl (BT) surveys have been conducted in the EBS since 1982, but for our study we used data from 2007-2018 to match available data from the acoustic surveys, resulting in $n$=4,511 observations (Table 1, Fig. 2).

It is important to highlight that the observations comprising these two data sets were not paired nor coordinated in the collection process. They are from different vessels at different times using different spatial sampling protocols, and these differences in spatial and temporal coverage could affect our analysis. The acoustic survey has typically not sampled the northeastern region of the EBS because it was assumed that there were negligible densities of pollock off-bottom (> 0.5 m) nearer to the coast (Fig. 2). However, new evidence suggests that this assumption may be incorrect in recent years (see Fig. 8 of Levine and De Robertis 2019). Our case study could be affected by this assumption of negligible densities near the coast because there is a large region of missing acoustic data with presumably low densities in earlier years but potentially higher (but still relatively small) densities in recent years. In a preliminary analysis, we explored two approaches to address this issue. First, we used the data in its original form with "missing" acoustic data nearer to the coast (the model then extrapolated densities into this domain). As a second option, we extended the acoustic transects to cover the inner domain assuming they would have observed mostly zeroes and small positive observations. These "inflated" observations (Fig. 2) were then appended to the observed data. We proceed with the latter solution because it more closely represents how the gear types are currently analyzed for use in the stock assessment, and the former led to implausibly high estimates in the unsampled region (Supplementary Material). We acknowledge this as a key assumption and discuss it further below.

Differences in the timing of the two surveys were evaluated by calculating the time difference between the closest bottom trawl for each acoustic observation. We found clear spatial patterns in the time differences (Fig. S1), with some observations being up to a month apart. Both surveys only sample during daytime to avoid effects of diurnal vertical migrations by pollock (defined as 30 minutes after sunrise and 30 minutes before sunset for the bottom trawl and between 0600-2400 hours for the acoustic survey per their protocol; Stauffer, 2004; Honkalehto *et al.*, 2018). We proceeded with the analysis under the assumption that our model predicts average density during summer months of a given year. However, we recognize that the spatiotemporal differences in design between the two surveys is not ideal and examine model residuals for negative effects of this. Another consequence of the spatiotemporal mismatch is that bottom depth (bottom trawl averaged 81.1 m and acoustic 98.3 m) was the only covariate available for use.

## Model structure

We define depth layer $c=1$ between bottom-0.5 m, $c=2$ between 0.5-16 m off bottom (the effective fishing height), and $c=3$ between 16 m off bottom and 16 m below the surface, with a corresponding total vertical density (in kg/km$^2$) $d=d(1)+d(2)+d(3)$, where $d(c)$ is expected density in depth layer $c$. Then the acoustic gear samples from $d(2)$ (data set referred to as "AT2") and $d(3)$ (data set "AT3"), both of which are separated in post-processing as described above, while the bottom trawl (BT) samples from $d(1)+d(2)$ combined, with no ability to post-process data to *c=1* and *c=2* separately. Note that both gears sample from *c=2*, the overlap layer, and this is a key structural feature of our model.

Then we assume a widely-used "delta-model" to represent the observations (Aitchison, 1955; Maunder and Punt, 2004), where the two key processes are the expected values of encounter probability (*p*) and positive observations (*r*), and are separate processes. We use the Poisson-link reformulation of the delta model (Thorson, 2017), where *n* represents the "group density" and *w* is the "biomass per group." Given *n* and *w*, *p* and *r* are calculated for depth layer *c* as

$$p(c) = \text{encounter rate} = 1 - \exp(-\mathrm{n(c)}) \in (0,1) \tag{1}$$

$$r(c) = \text{positive observation} = \frac{n(c)w(c)}{p(c)} \in (0, \infty), \tag{2}$$

where the phrase 'positive observation' represents either positive catches or backscatter depending on the data type. The expected value of biomass density in layer *c* is then $d(c) = p(c)r(c) = n(c)w(c)$. These expected values represent averages over the time period in which sampling occurs, and then are used in calculating the likelihood of the observed data.

## Deriving the combined likelihoods

We compare model expectations of density to the observations (*b*) using the delta-likelihood formulation:

$$f_{B(c)}(b) = \begin{cases} 1 - p(c) & b = 0 \\ lognormal(b|r(c), \sigma_M) & b > 0 \end{cases}, \tag{3}$$

where we assume $g$ is a log-normal distribution for the three data sets (AT2, AT3, BT) with estimated standard deviation of the positive observations, $\sigma_M$, for the two gears (i.e., $\sigma_{BT}$ and $\sigma_{AT}$). The likelihoods for AT2 are $P(b=0)=1-p(2)$ for zero observations and $f_{B(c)}(b)=lognormal\big(b;\ \log\big(r(2)\big),\sigma_{AT}\big)$ for positive observations (e.g., table 1 of Thorson, 2019), and those for AT3 are the same except that *c=3*.

Likelihood equations for the bottom trawl must account for the fact that it samples from layers *c*=1 and *c*=2 jointly so we derive them here. A non-encounter in the BT (*b=0*) means that no fish were encountered in either layer. If we assume the bottom trawl gear samples from the first two layers independently, conditioned on the model estimates, then the probability of not encountering any fish in a trawl sample is:

$$P(b=0)=1-p(BT)=\big(1-p(1)\big)\big(1-p(2)\big)=\exp(-n(1)-n(2)), \tag{4}$$

where *p(BT)* is our notation signifying the joint sampling of layers 1 and 2 with a bottom trawl.

When *b>0*, we assume the expected catch is approximated by the sum of expected catches *r* across the first two layers:

$$f_{B(c)}(b)=\mathrm{lognormal}\big(b;\ \log\big(r(BT)\big),\sigma_{BT}\big), \tag{5}$$

where

$$r(BT)=\frac{n(1)w(1)+n(2)w(2)}{1-\exp\big(-n(1)-n(2)\big)}. \tag{6}$$

Note that $d(BT)=p(BT)r(BT)=n(1)w(1)+n(2)w(2)=d(1)+d(2)$, such that the expected bottom trawl observations are the sum of the density in the first and second vertical layers, as desired.

Equations (4)-(6) comprise the key methodological development in our vertically-integrated model, and were derived assuming the Poisson-link delta-model. Similar derivations could be done for a conventional delta-model but were not explored here. The derivation required an assumption of statistical independence for sampling of the two layers by the bottom trawl (i.e., that measurement errors are independent for those two layers conditional upon estimated density in each layer). In the discussion section, we hypothesize why statistical independence might be true and outline possible approaches to test it. The derivation also assumed that the sum of the expected positive catches in the first two depth layers approximates the bottom trawl data. As shown above, our approximation guarantees that the expected density is the same as the true combined distribution. However, there is no guarantee that other properties such as variance or higher statistical moments will match. We use Monte Carlo sampling to demonstrate that for the estimated sampling properties of our case study this approximation holds well (Fig. S2), and encourage this test for other applications.

## Model predictors

Next we define how the model expectation of *n* and *w* are determined, from which *p* and *r* can be calculated from eqns. (1-2). We assume fish density varies continuously in space, and changes annually in response to spatially-dynamic environmental and biological processes. Thus, we include spatial, temporal, spatiotemporal, and covariate effects in the context of modeling spatial data. We assumed linear effects of explanatory variables on the log of the expected value *w*, as follows:

$$\log w(i) = \log w(c_i, s_i, t_i, x_i)$$
$$= \beta_w(c_i, t_i) + \sum_{f=1}^{3} L_{\omega_w}(c_i, f)\omega_w(s_i, f) + \sum_{f=1}^{3} L_{\varepsilon_w}(c_i, f)\varepsilon_w(s_i, f, t_i) + \gamma_w(c_i)X_w(x_i) + \lambda_w Q(i), \tag{7}$$

where $c_i$, $s_i$, $t_i$, and $x_i$ are the depth layer, spatial cell, year, and covariate for observation *i,* respectively. Each layer and year has an intercept term, $\beta_w(c,t)$, which we assumed followed a Gaussian random walk temporal smoother, necessitated by the years without acoustic data (Fig. 2). Specifically, $\beta_w(c,1) \sim N\left(\mu_{\beta_w}(c), \sigma_{\beta_w}^2(c)\right)$ for *t=1* and

$\beta_w(c,t) \sim N\left(\beta_w(c,t-1), \sigma^2_{\beta_w}(c)\right)$ for $t$>1, where the mean $\mu_{\beta_w}(c)$ and variance $\sigma^2_{\beta_w}(c)$ are fixed effects and the $\beta_w$ terms are random effects, with separate estimates for each *c.* An autoregressive process could be employed instead, but is not presented for the case study here. The spatial and spatiotemporal random effects, ω and ε, are Gaussian Markov random fields used by the SPDE method (Lindgren *et al.*, 2011), which is a computationally efficient approach for approximating Gaussian continuous spatial processes (Illian *et al.*, 2012; Thorson *et al.*, 2015b). We used 400 spatial knots for the approximation (spatial resolution). We only present a fully specified spatiotemporal model (i.e., ω and ε for both linear predictors) because configurations with less spatial complexity fit poorly based on model selection with PSIS-LOO (Vehtari *et al.*, 2017) and estimated indices (Supplementary material). We also included the same Gaussian random walk temporal smoothing structure on the spatiotemporal effects, $\varepsilon_w(s,c,t)$, as specified above for the annual intercepts above (see eqn. 2.10 of Thorson 2019). Note that this formulation implies a spatially correlated surface of log *w* for each of the three depth layers, and that values at each point in space are correlated across years for each layer. The correlation in log *w* implies a spatiotemporal correlation in biomass density in each layer, given *n* and as detailed above. How log *w* is correlated among the depth layers at a given point in space depends on how the model is configured as follows.

The sums across factors (*f*) is the spatial factor analysis approach where the estimated loadings matrices $L_{\omega_w}$ and $L_{\varepsilon_w}$ relate to the correlations among depth layers at the same points in space (Thorson *et al.*, 2015a). Instead of directly estimating covariances, triangular loadings matrices (denoted *L*) representing the Cholesky decomposition of the covariance matrix are estimated as fixed effects. They can be fully parameterized, use a low-rank approximation by specifying fewer than three factors, or assumed to be diagonal (so layers are independent). Initial attempts to estimate loadings matrices in the case study with three factors had severe convergence issues (Supplementary Material), so we switched to diagonal loadings matrices. This does not preclude correlation among layers, but means the correlation is not directly estimated. In our analysis the largest effect of this is that the information from the bottom trawl cannot be used to infer fish density in the top depth layer, so only the spatiotemporal smoother informs that layer in years without acoustic data.

We included a single catchability parameter $\lambda_w$ to allow differences between gear types, coded as *Q=0* for acoustic and *Q=1* for bottom trawl observations because previous studies suggest potential differences in gear efficiency and thus catchability (Kotwicki *et al.*, 2013). Configured this way, the quantity $\mathrm{e}^{\lambda_w}$ represents the ratio of expected observations (trawl to acoustic), all else being equal. We also explored the effect of alternative structures for catchability, including a time-varying structure and constant catchability in the first linear predictor *n* (Supplementary Material). We included normalized log depth (due to the skewness of depth) as a covariate, $\gamma_w(c)$. Other covariates were unavailable (see above) although they could be incorporated if available.

The structure for *n* is identical, except that it excludes the catchability parameter (i.e., $\lambda_n = 0$) and is thus left off for brevity. Equations 1-7 define model expectations and data likelihoods, so now we turn to model fitting. In addition to this combined model, we use above model description but fit it separately to the acoustic and bottom trawl data, and additionally compare our predictions to design-based estimates of the same data sets (Conner and Lauth, R. R., 2017; Honkalehto *et al.*, 2018).

## Model fitting

We fit our model in the software framework VAST (Thorson and Barnett, 2017; Thorson, 2019), after adding the ability to combine likelihoods (eqns. 4-6), available with release number 1.6.0 (see Supplementary Materials for details and reproducible example). VAST is written in the modeling framework Template Model Builder (TMB; Kristensen *et al.*, 2016) and these models are typically fit with maximum marginal likelihood estimation. But in the combined model of pollock data the Laplace approximation to the marginal likelihood (Skaug and Fournier, 2006) had numerical issues and TMB consistently crashed during optimization (i.e., it could not find the maximum likelihood estimates regardless of initial values). The issue was clearly related to the likelihood for the bottom trawl data (eqn. 6), but we were unable to fully diagnose and avoid the core cause. Instead, we switched to Bayesian inference using the R package *tmbstan* (Monnahan and Kristensen, 2018), which provides an interface for TMB models to the Bayesian platform *Stan* (Carpenter *et al.*, 2017), through the R package *rstan* (R Core Team, 2018; Stan Development Team, 2018). Stan implements the no-U-turn sampler which is an efficient MCMC algorithm for drawing posterior samples from large, complex hierarchical models (Hoffman and Gelman, 2014; Monnahan *et al.*, 2017). Bayesian integration with MCMC provided an alternative algorithm for inference that worked better than maximum marginal likelihood estimation, but required explicit priors (see below).

As configured for the case study, the combined model has 32,520 random effects (intercepts, spatial and spatiotemporal effects for each depth layer) and 33 fixed effects, 18 of which were hypervariance parameters (Table 2). The spatial correlation and anisotropic parameters (*logkappa* and *ln_H_input* in VAST) proved to be inestimable in preliminary runs given their poor mixing relative to the other parameters. Consequently, we set these parameters to mean values from the preliminary runs, and tested the sensitivity to this assumption (Supplementary Materials).

We used informative priors for some parameters (Figs. S3a,b). The prior for the catchability parameter $\lambda_w \sim N(0,0.15)$ was based on expert knowledge that the catchability of gear types should not be very different. This is similar to the “bias ratio” from Kotwicki *et al.* (2018), but we did not use results from that study to inform the prior so that we could instead compare results as a way to corroborate our approach. Priors on the effects of standardized depth by layer were broad and normally distributed: $\gamma_n(c), \gamma_w(c) \sim N(0,5)$. For the random walk temporal structure, we set priors on the initial intercept to give approximately uniform encounter probability and the log of positive observations between roughly 1 and 13, which we found reasonable based on expert knowledge of the system and other species. For the remaining parameters, we used implicit uniform priors. We ran six chains of 800 iterations, each initialized from diffuse values and using the first 300 iterations as warmup. We increased the target acceptance rate to 0.85 to eliminate divergences and set the maximum tree depth to 17. As typical and recommended for *Stan* analyses, we use posterior medians and credible intervals to quantify estimates and uncertainty, ensured sufficient estimated effective samples for all parameters (at least 800), the potential scale reduction statistic diagnostic $\hat{R} < 1.02$ for all parameters, and no divergent NUTS transitions (Gelman *et al.*, 2014; Stan Development Team, 2017). We also used posterior predictive distributions, where observed data are compared to data simulated given the posterior draws, to validate the model (Gelman *et al.*, 2014; Conn *et al.*, 2018).

## Simulation study

We used a simulation experiment to check the statistical properties of the combined method, and demonstrate potential inaccuracies in biomass indices. The model used to generate the pseudo data closely reflected the structure of our fitted case study model, but omitted spatiotemporal variation and had lower hypervariances and observation errors for computational expediency. We generated unbiased random samples from the 'assumed truth' using the two-step sampling process described above (eqns. 1-6), and fit the combined model. We also made "independent" estimates where data from the acoustic and bottom trawl surveys were fit separately mimicking the standard application of these data for assessment purposes. We specified changes in the "true" index by manipulating the annual effects ($\beta_{n,}(c,t)$ and $\beta_{w,}(c,t)$) for the three depth layers to produce a vertical distribution that had a downward trend for depth layer <0.5 m, a constant trend for the overlap, and an increasing trend for >16 m. Such trends increase availability to the acoustic survey and decrease the availability for the bottom trawl. We computed relative error of the estimated log-index to the log total biomass as a performance metric and examined estimation bias. The simplified structure of the simulation testing allowed use maximum marginal likelihood instead of Bayesian integration, which also allowed us to estimate the geostatistical parameters. We used a maximum gradient of 0.01 as a cutoff for optimizer convergence (i.e., the largest absolute derivative of the marginal likelihood with respect to the parameters), as deviations from zero indicate lack of convergence. Further details of the simulation are given in the Supplementary Materials.

# Results

## Case study on walleye pollock

The fit passed MCMC convergence diagnostics and the posterior predictive distributions showed no systematic patterns in space (Fig. S4a-c) nor against the time difference between the surveys (Fig. S5). Most marginal posteriors were different from the prior, indicating meaningful information in the data to update them (Table 2; Figs. S3a,b). The only exceptions were the $\mu_{\beta_n}$ parameters representing the annual process for the *n* component, which had broad, but informative priors. Of the estimated variance terms, only the terms for the annual intercepts had any meaningful probability mass around zero. All three depth layers had meaningful total (spatial + spatiotemporal) variance, with the spatial component representing about 67.1-82.2% of total variance by depth layer (Table S1). The catchability parameter $\lambda_w$ was estimated to be 0.17 (-0.01 to 0.34) which after exponentiation represents 1.19 (0.99-1.42) times higher catch for bottom trawl vs. acoustic survey in the overlapping depth layer for a given place and time. The median posterior for the covariate effects on depth were positive for both the *n* and *w* predictors for all depth layers, suggesting increasing encounter probability and positive observations with increasing depth. The only exception was the effect $\gamma_n(1)$, which was centered at zero indicating no effect (Table 2).

Early in the time series, fish were concentrated in the northeast corner of the EBS with few fish in any depth layer closer to inshore, but by 2018 fish were more evenly distributed over the study region (Fig. 3). We also found subareas of consistent vertical availability across all years. For example, in the southwest corner of the EBS the bottom trawl availability is low, whereas in the southeast area (which is shallower) the availability of pollock to the acoustic survey is lower (Fig. 4). The spatial patterns in availability by gear type in other areas varied among years.

Each of the three depth layers, summed across the whole EBS, contained approximately equal biomass on average, but this varied by year (Table 3; Fig. 5a). In general, there was a decrease in the proportion of fish <0.5 m off bottom (Fig. 5a), leading the availability in the acoustic survey to increase over time, and the opposite for the bottom trawl (Fig. 5b). Vertical gear availability (% fish available; Table 3) for the acoustic survey ranged from a low of 33.9% in 2008 to a high of 88.8% in 2017, although 2017 was a year without acoustic data so that estimate was more uncertain. The bottom trawl availability ranged from a low of 58.8% in 2016 to a high of 91.1% in 2009. The uncertainty around estimated quantities was notably higher in the years without acoustic data where the model interpolated densities above 16 m using the random walk temporal process (Fig. 5b-d).

The combined index is the total biomass calculated by summing vertical depth layers and across space, and ranged from a low of 15.2 log metric tons in 2009 to a high of 17.3 in 2015 (Table 3, Fig. 5c). The gear-specific log-indices closely matched the trend and uncertainty of design-based estimates (Fig. S6) with a few exceptions. For instance the downward trend in the acoustic index between 2007 and 2009 was steeper compared to the design-based counterpart. The uncertainty in years with acoustic data was similar between the combined model and independent estimates, but for years without data the combined model had a truncated lower end of uncertainty. The bottom trawl indices were similar except a smaller uncertainty in the design-based estimates (Fig. S6).

## Simulation

Our simulated data produced log-indices for the two gear types with distinct patterns (Fig. 6a,b). When fitting with maximum marginal likelihood, the median maximum gradients were all less than 1E-07 for the different model types, well below the cutoff for convergence. However, the combined model had several replicates with maximum gradients larger than 100. Despite this, the percentage of replicates failing our convergence criterion of 0.01 were roughly the same at 4.21%, 4.74%, and 5.79% for the acoustic, bottom trawl, and combined models respectively.

After filtering out replicates that failed to converge, relative errors of the log-index for the self-test cases (Fig. 6c) were all small, generally less than 5%. They were unbiased for the combined model, but showed some bias for the acoustic and a slight trend for the bottom trawl. Only the combined model accurately estimated the total biomass, reflected as a vertical availability of one (Fig. 6d). Both the acoustic and bottom trawl had significant fractions of the biomass unavailable and this trend varied over time as the vertical distribution changed.

# Discussion

We developed and applied a flexible approach to estimate semi-pelagic fish density divided into three vertical layers, covering the water column from sea floor to near surface. Because the method estimates spatiotemporal variation, it is not restricted to paired data with a consistent sampling design. Importantly, this means it can be fitted to data from distinct surveys providing unpaired data, with different spatial extents, sampling protocols, and spatial or temporal gaps (e.g., missing years). This occurs with dedicated acoustic and bottom trawl surveys, or failure of a single gear on a paired gear survey, and this flexibility in data used to fit the model increases the number of real-world data sets for which these methods could be applied relative to methods that require paired data. We found large variation in the availability of pollock to both gear types (among years and spatially), suggesting that neither data type sufficiently characterizes a consistent portion or the entire population. In contrast, our combined index directly accounts for the total vertical population, regardless of how vertical availability to gears changes in space and time. Thus, this developed approach mitigates index inaccuracies compared to either survey in isolation. In addition, the combined index accounts for variability in vertical availability and consequently provides more accurate uncertainty estimates, another important property required for stock assessments (Kotwicki *et al.*, 2018; Kotwicki and Ono, 2019). We argue that our combined method represents an important new tool for analyzing vertically overlapping survey data for semi-pelagic stocks, and note that important insights can be gleaned from existing data without the cost of additional survey effort (as demonstrated by the pollock case study). Our method could be applied to other semi-pelagic species with similar data, and we provide a working example based on freely available software tools as a starting point (Supplementary Materials).

The key statistical development of this paper is the derivation of approximate likelihoods for the joint sampling by bottom trawl gear over two depth layers (eqns. 4-6), and their implementation in the open-source statistical software VAST (Thorson, 2019). This derivation relies on two important assumptions. First, we assumed the measurement process between the two depth layers is independent, conditional upon the latent state (random effects). We hypothesize that fish behavior (e.g., net response, microhabitat selection) occurs at finer spatial scales than the course spatial processes that represent average density as estimated by the model (i.e., the spatial resolution). Second, we assumed the expectation of the positive observation component of the sampling process can be approximated by summing the expected positive observation of the two overlapping layers. The resulting distribution of joint sampling has the correct mean, but differs in the other statistical moments. We saw no evidence in the model residuals that the independence assumption was violated (Figs. S4a-c), nor that the approximation was inaccurate (Fig. S2). Furthermore, our predicted indices closely matched the trend and uncertainty in design-based methods and spatiotemporal models fitted to each survey data separately (Fig. S6). Despite this, it is important to highlight these statistical assumptions and further research investigating their effects would be valuable. We suggest exploring this through simulation or explicitly modeling the correlation of the data among the depth layers.

We encountered statistical problems fitting the model to the pollock data. We used Bayesian inference because maximum marginal likelihood estimation via the Laplace approximation, the standard VAST approach, was unreliable. Since VAST is not parameterized to optimize MCMC sampling for Bayesian inference, we had to rescale some parameters to improve MCMC convergence (Supplementary Materials). We also had to simplify the loadings matrices which control how depth layers are correlated, such that each layer was independent, to avoid a multi-modal posterior, which also causes issues with maximum likelihood. Finally, we had to assume the geostatistical parameters (related to decorrelation range and anisotropy) were known because of the inability to estimate them using available statistical software. Fortunately, the result was good convergence and a general insensitivity of the resulting index to these assumptions (Supplementary Materials). Unfortunately, these specialized modifications to VAST make it more difficult to apply our method on other case studies, and omitting estimation of the geostatistical properties within the model could be an issue in other contexts. Many of these challenges could be alleviated if maximum likelihood estimation were viable. Our simulation study demonstrated that it can be viable and reliable for simple models, at least when the data are consistent with the model structure and assumptions. However, simulation configurations with spatiotemporal effects had similar estimation issues as our case study (Supplementary Material), suggesting our method may not be compatible with maximum likelihood in general. Bayesian integration in VAST could be improved with rescaling and reparameterizing to be more commensurate with integration instead of optimization, which could also lead to benefits in other settings. In particular, we recommend testing alternative parameterizations for the spatial and spatiotemporal effects, which affect performance in hierarchical models, and the parameterization of the geostatistical range and anisotropy.

The pollock case study also has some important caveats. First, data came from two surveys on separate vessels, which sampled in different spatial footprints (Fig. 2) and at different times (up to a month apart; Fig. S1). We ignored seasonal (within year) differences in population density between surveys, and recognize that analyzing density for two surveys that occur several weeks apart could be problematic due to complex environmental dynamics driving fish behavior and distribution. However, we found no evidence that this led to difficulties in fitting the pollock data (Fig. S5). We also added acoustic data from the eastern EBS, which typically has low densities off-bottom and was not sampled by the acoustic survey (Fig. 2). Spatial patterns in the region were highly dependent on this assumption, but the combined index was generally insensitive to the approach taken because of the relatively low biomass. Second, the only covariate included in this study was depth because data on other potentially important factors known to affect pollock availability (e.g., light attenuation, Kotwicki *et al.*, 2015) were unavailable for the entire time series. We expect that including factors such as fish length, sediment size, water temperature and light intensity could lead to an improved model fit. Despite these issues, our results are corroborated by a similar estimate of efficiency between gears (1.18, 0.99-1.41) using a different set of acoustic data (0.96, 0.49-1.42; Kotwicki *et al.*, 2013), and similar results using design-based estimators (Fig. S6). Ultimately, it may be possible to resolve both of these data issues using acoustic data collected on the bottom trawl survey (e.g., Honkalehto *et al.*, 2011; Kotwicki *et al.*, 2013), although it is beyond the scope of this analysis. We used acoustic data distant from trawl sites (via separate surveys), and similarly found that the precision of the trawl index was relatively unchanged (Fig. S6), reflecting the general finding of other studies which used inter-site transects (e.g., Hjellvik *et al.*, 2007). As noted by von Szalay *et al.* (2007), this could be caused by a poor correlation due to vertical distributions that varies spatiotemporally, an observation we confirmed in this study (Fig. 3). Similar to Kotwicki *et al.* (2018) we modeled the overlap and thus directly relied on an accurate effective fishing height resulting from fish response to the gear. We used 16 m in our case study (based on Kotwicki *et al.*, 2013), but a sensitivity run with 3 m demonstrated similar index trends, albeit a different scale (Supplementary material). This work would benefit from further refinements to fish diving behavior and effective fishing heights based on experimental studies for pollock, but also application to other data sets such as the Barents Sea cod or haddock for general validation of the method.

Pollock appeared to exhibit a general decreasing trend of availability to the bottom trawl survey over time. We initially hypothesized that this could be explained by a general decrease in average age within the population with time. For e.g., younger pollock are typically more pelagic than older ages, and thus are less available to the bottom trawl survey and more available to the acoustic trawl survey. During the study period, the proportion (by mass) of 2- and 3-year-old pollock estimated to be in the population ranged from a low of 15% to a high of 52% (as computed from tables in Ianelli *et al.*, 2019). However, counter to our expectations, the trends in bottom trawl pollock availability and proportion of 2-3 year-olds in the population were not negatively correlated over time. This lack of correspondence in trends could be due to differences in the (horizontal) spatial effort of the two surveys or to spatial patterns that were obscured when aggregated across space. Our study framework also assumes that movement into and out of the survey area are negligible. For example, multiple years of acoustic-trawl survey extension into Russia have observed a mean of only 7% of the entire shelf-wide pollock (n=9, 1-22%; Honkalehto and McCarthy, 2015) present on the Russian side of the U.S.-Russia maritime boundary. If the negligible movement assumption were violated, it could partially explain the lack of a relationship in trends. There are also various potential biological drivers such as reaction to changing pollock density or shifts in predatory or prey species, and oceanographic ones like changes in distribution of light or temperature. Further exploration of potential environmental and biological drivers of vertical and horizontal distribution shifts would be worthwhile to improve understanding and application in the stock assessment.

One application of our results within a stock assessment model would be to allow inclusion of annual availability estimates. This would better inform the model about how an index may change due to vertical shifts (in contrast to standard practice, which assumes a constant value for the catchability coefficient for this index). For example, if the estimate of annual acoustic availability decreased then it would not errantly attribute such changes to changes in stock biomass. Alternatively, each index could be adjusted externally based on estimated availability, but this may complicate how the uncertainty of the index should be specified. Another extension of our work would be to model size- or age classes from both surveys within the layers considered for the combined index. This would be computationally more intensive but could resolve some observed patterns in different age structures throughout the water column.

Although our focus here was on improving indices used in stock assessment models, our method could provide valuable insights in other applications as well. For example, much effort has been dedicated to understanding distribution shifts and how they are likely to be affected by future climate change (e.g., Perry *et al.*, 2005) but typically draw conclusions from a single survey source (e.g., Thorson *et al.*, 2017). If such surveys have inaccuracies that vary over time, these types of analyses may lead to inappropriate conclusions. If vertical shifts manifest as changes in density, then estimates of distributional quantities such as center of gravity would also be incorrect. In general, estimates from our method could also help improve ecological studies which use spatial abundance such as food web studies (e.g., Aydin and Mueter, 2007) or design of surveys (e.g., Overholtz *et al.*, 2006). Thus, improved estimates of the distribution of semi-pelagic species would improve ecological understanding in addition to the benefits to fisheries management.

# Supplementary material

The following supplementary material is available at ICESJMS online. Supplemental tables and figures with extra results, further details and sensitivity tests for the effective fishing height, estimation of spatial factors, configuration of catchability parameters, the geospatial assumptions, and spatial configuration. A reproducible example is also provided and links to an online repository. Finally, we give more details of the simulation study and how we setup VAST for Bayesian integration with the package tmbstan.

# Data Availability Statement

The data underlying this article are available in the following repository:

https://github.com/Cole-Monnahan-NOAA/stpollock.

Acknowledgements

This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA15OAR4320063, Contribution No. 2019-1025. We thank Kasper Kristensen and Paul Conn for helpful feedback on technical aspects of the study, and Kelli Johnson, Pete Hulson, Rebecca Thomas, Sam Urmy, Taina Honkalehto, Cecilia O'Leary, Lewis Barnett, Alex De Robertis, Nicholas Bez, Olav Rune Godø, and two anonymous reviewers for helpful feedback on an earlier draft.

# Tables

**Table 1**. Annual number of observations of pollock for the acoustic and bottom trawl surveys, including inflated zeroes which extend the acoustic spatial extent into a region where the bottom trawl samples but we assumed no fish would have been detected by acoustics (see Fig. 2). The acoustic observations were binned (grouped spatially and averaged). Note the years in which there was no acoustic survey conducted.

| Year | Total Binned Acoustic | Acoustic Inflated (%) | Bottom Trawl |
|---|---|---|---|
| 2007 | 469 | 48 (10.2) | 376 |
| 2008 | 470 | 42 (8.9) | 375 |
| 2009 | 455 | 43 (9.5) | 376 |
| 2010 | 454 | 43 (9.5) | 376 |
| 2011 | 0 | 0 | 376 |
| 2012 | 473 | 48 (10.1) | 376 |
| 2013 | 0 | 0 | 376 |
| 2014 | 477 | 48 (10.1) | 376 |
| 2015 | 0 | 0 | 376 |
| 2016 | 509 | 41 (8.1) | 376 |
| 2017 | 0 | 0 | 376 |
| 2018 | 523 | 41 (7.8) | 376 |

**Table 2**. Parameter estimates for fixed effects from the combined model. Posterior median and 95% credible intervals are given by depth layers and for the first ($n$) and second ($w$) linear components (see eqn. 7). The last three parameters are not indexed by layers and instead apply to the gear types. The parameters for anisotropy and decorrelation range are not estimated but instead fixed (see Supplementary material). 'SD' is standard deviation. The spatial and spatiotemporal SD represent the SD of the estimated spatial fields and are hierarchical variance terms (see Table S1 as well). Likewise, the SD of the random walk process is a hierarchical variance controlling the amount of smoothing of annual intercepts.

| **VAST Name** | **Symbol** | **Description** | **< 0.5 m** | **0.5-16 m** | **>16 m** | **Component** |
|---|---|---|---|---|---|---|
| gamma1_ctp[1] | $\gamma_n$ | Depth effect | -0.02 (-0.47–0.43) | -0.20 (-0.39–-0.01) | 0.22 (0.05–0.39) | Group density |
| L_omega1_z[1] | $L_{\omega_n}$ | Spatial SD | 1.58 (0.91–2.31) | 1.88 (1.46–2.36) | 2.28 (1.78–2.78) | Group density |
| L_epsilon1_z[1] | $L_{\varepsilon_n}$ | Spatiotemporal SD | 0.59 (0.26–0.96) | 0.58 (0.41–0.78) | 1.31 (1.12–1.54) | Group density |
| L_beta1_z[1] | $\sigma^2_{\beta_n}$ | SD of random walk temporal process | 0.16 (0.01–0.61) | 0.15 (0.01–0.54) | 0.40 (0.02–1.47) | Group density |
| Beta_mean1_c[1] | $\mu_{\beta_n}$ | Mean of random walk temporal process | -0.62 (-2.39–1.16) | 0.23 (-1.71–2.11) | -1.12 (-3.38–0.97) | Group density |
| gamma2_ctp[1] | $\gamma_w$ | Depth effect | 1.77 (1.20–2.40) | 0.39 (0.16–0.62) | -0.32 (-0.54–-0.10) | Biomass per group |
| L_omega2_z[1] | $L_{\omega_w}$ | Spatial SD | 2.32 (1.71–2.90) | 2.32 (1.88–2.77) | 0.65 (0.05–1.29) | Biomass per group |
| L_epsilon2_z[1] | $L_{\varepsilon_w}$ | Spatiotemporal SD | 1.17 (0.94–1.41) | 1.34 (1.20–1.50) | 1.01 (0.84–1.19) | Biomass per group |
| L_beta2_z[1] | $\sigma^2_{\beta_w}$ | SD of random walk temporal process | 0.39 (0.02–1.10) | 0.33 (0.02–1.08) | 0.56 (0.04–1.64) | Biomass per group |
| Beta_mean2_c[1] | $\mu_{\beta_w}$ | Mean of random walk temporal process | 6.85 (4.26–9.29) | 4.37 (1.77–7.01) | 5.31 (3.38–7.26) | Biomass per group |
| logSigmaM[1] | $\log \sigma_{BT}$ | SD of BT observation error | 0.44 (0.42–0.47) | -- | -- | Bottom trawl |
| logSigmaM[2] | $\log \sigma_{AT}$ | SD of AT observation error | 0.48 (0.46–0.50) | -- | -- | Acoustic |

| | | | | | | |
|---|---|---|---|---|---|---|
| lambda2_k | $\lambda_w$ | Catchability effect | 0.17 (-0.01–0.34) | -- | -- | Biomass per group |

**Table 3.** Annual results from the combined model after integrating across space. <0.5m, 0.5-16m, and >16m are the depth layers used in the model, which respectively are the acoustic dead zone, the overlap, and the bottom trawl blind zone. Quantities are median and 95% credible intervals in parentheses for total density, others left off for clarity. Emphasized rows are those years without any acoustic data, leading to higher uncertainty.

| | Log-density (metric tons/km$^2$) | | | | Proportion biomass by strata | | | Vertical availability by gear | |
|---|---|---|---|---|---|---|---|---|---|
| **Year** | **Total** | **<0.5 m** | **0.5-16 m** | **>16 m** | **<0.5 m** | **0.5-16 m** | **>16 m** | **Acoustic** | **Bottom Trawl** |
| 2007 | 16.18 (15.87-16.63) | 15.51 | 14.85 | 14.65 | 0.51 | 0.27 | 0.22 | 0.49 | 0.78 |
| 2008 | 15.94 (15.60-16.39) | 15.52 | 14.32 | 13.97 | 0.66 | 0.20 | 0.14 | 0.34 | 0.86 |
| 2009 | 15.21 (14.89-15.61) | 14.56 | 14.24 | 12.79 | 0.53 | 0.38 | 0.09 | 0.48 | 0.91 |
| 2010 | 16.36 (16.05-16.73) | 15.72 | 14.30 | 15.25 | 0.53 | 0.13 | 0.34 | 0.46 | 0.67 |
| **2011** | **16.44 (15.93-18.00)** | **15.41** | **15.19** | **15.32** | **0.35** | **0.28** | **0.37** | **0.65** | **0.66** |
| 2012 | 16.30 (16.08-16.55) | 15.31 | 15.04 | 15.21 | 0.37 | 0.29 | 0.34 | 0.63 | 0.66 |
| **2013** | **16.72 (16.18-18.07)** | **15.60** | **15.47** | **15.64** | **0.33** | **0.29** | **0.38** | **0.69** | **0.65** |
| 2014 | 17.01 (16.77-17.32) | 15.99 | 16.07 | 15.58 | 0.37 | 0.39 | 0.25 | 0.64 | 0.76 |
| **2015** | **17.28 (16.80-18.54)** | **16.10** | **16.19** | **16.15** | **0.31** | **0.33** | **0.36** | **0.69** | **0.67** |
| 2016 | 16.94 (16.77-17.19) | 15.43 | 15.93 | 16.05 | 0.23 | 0.36 | 0.41 | 0.78 | 0.59 |
| **2017** | **16.95 (16.42-18.46)** | **14.78** | **16.16** | **16.05** | **0.12** | **0.45** | **0.43** | **0.89** | **0.59** |
| 2018 | 16.31 (16.12-16.70) | 14.88 | 15.26 | 15.36 | 0.26 | 0.35 | 0.39 | 0.76 | 0.61 |

# Figures

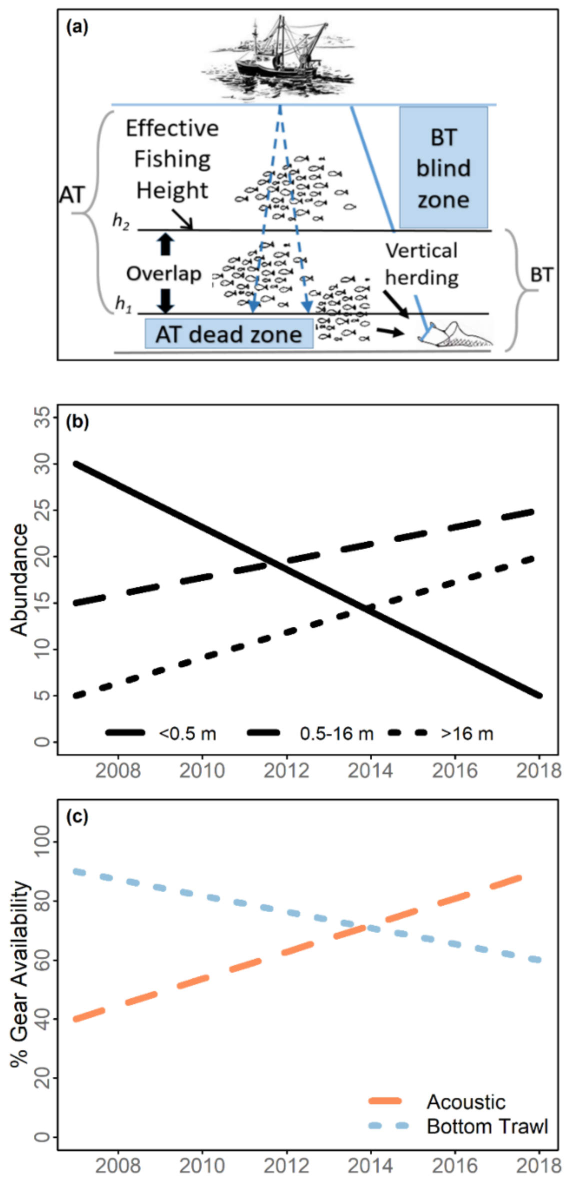


**Figure 1**. Conceptual issue with gears with acoustic dead and blind zones and temporal trends in vertical distribution for a semi-pelagic fish. (a) Schematic of gear types showing acoustic (AT) sampling directly under the vessel, the vertical herding to create a larger effective height than physical fishing height for the bottom trawl (BT), which is behind the vessel; and the three depth layers (horizontal lines, defined by $h_1$ and $h_2$ as measured relative to sea bottom), vertical blind and dead zones

(regions of unavailability), and the overlap where both gears sample; recreated from Kotwicki et al. (2018) with permission. The acoustic blind zone near the surface is left off for visual clarity. (b) A simulated example, where the abundance in the three depth layers (measured from bottom; <0.5 m is the AT dead zone, 0.5-16 m is the overlap, and >16 m the BT blind zone) exhibit distinct annual trends. (c) The percent of fish available to each gear type relative to the total (sum of all three depth layers). Note that in a given year the sum of the gears' availability is not 100% because of the overlap layer sampled by both.

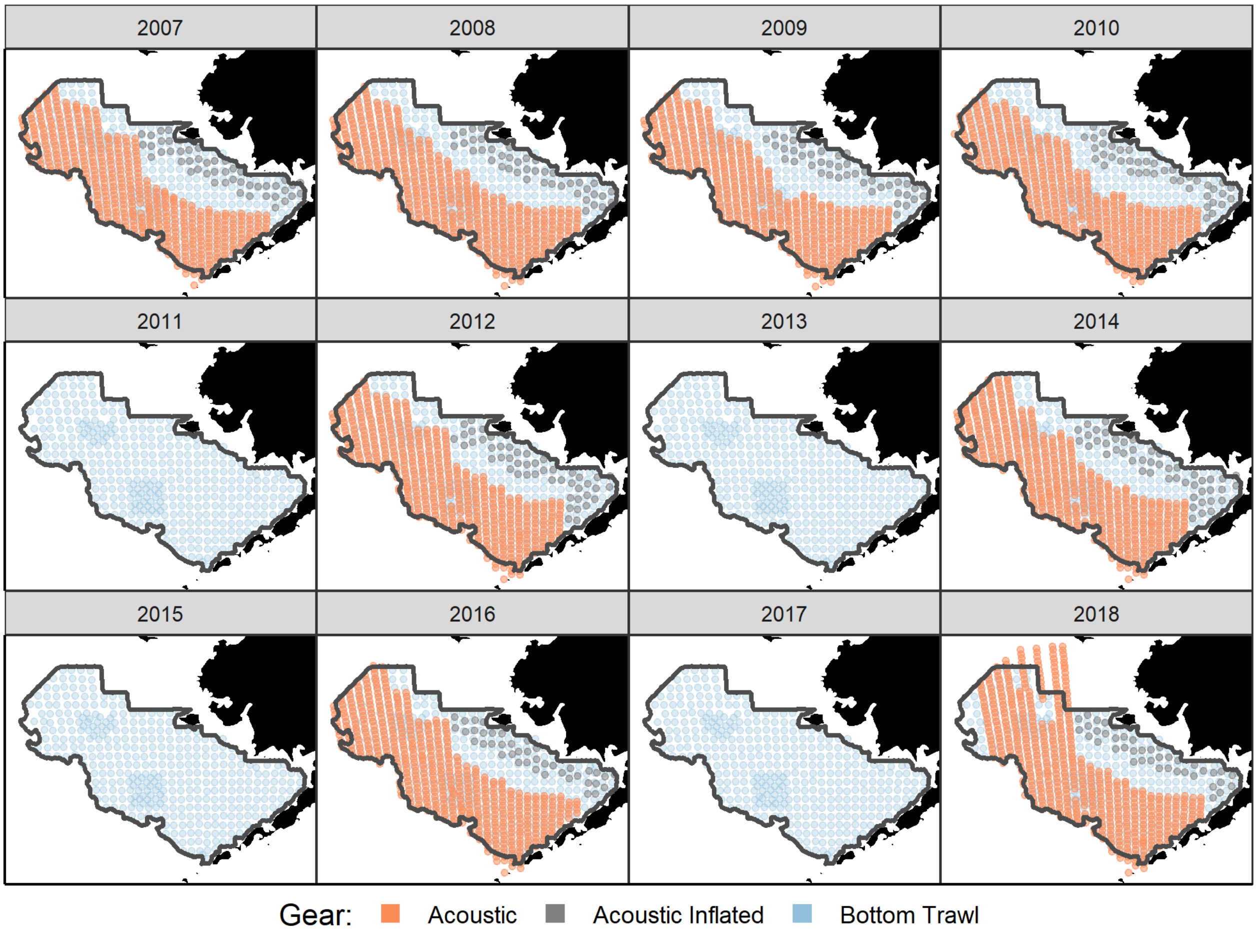


**Figure 2**. Experimental design showing the two surveys that have spatiotemporal sampling patterns. The acoustic survey did not sample in years 2011, 2013, 2015, and 2017, and also never in the southeast portion of the study extent (eastern Bering Sea; black outline), so we inflated it with hypothetical data (gray points, see main text). The bottom trawl points are fixed stations, while the acoustic points are midpoint locations after averaging across 20 acoustic intervals; the black line defines the region where densities are predicted and then summed when calculating an abundance index, despite some acoustic observations being outside this extent.

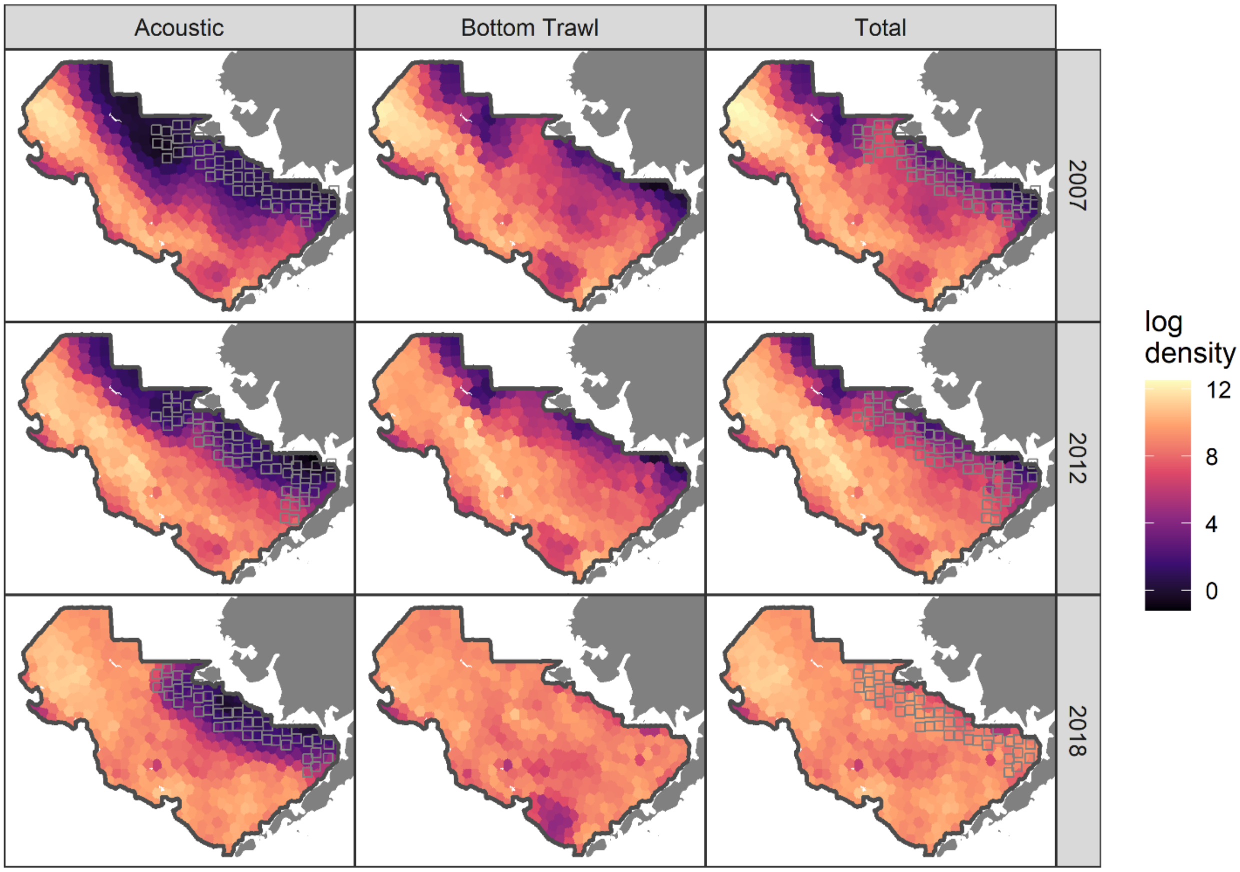


**Figure 3**. Estimated log-density (colors, metric tons/km$^2$) of pollock for three select years (rows) for the combined model. Columns represent the density available to the gear types, which for the acoustic is the sum of density above 0.5 m off bottom, and bottom trawl is the sum of density below 16 m off bottom, while the total is the sum of the entire vertical water column (bottom to surface). The gray squares are the locations of inflated acoustic data which are in a region unsampled by the acoustic survey (see Fig. 2, main text).

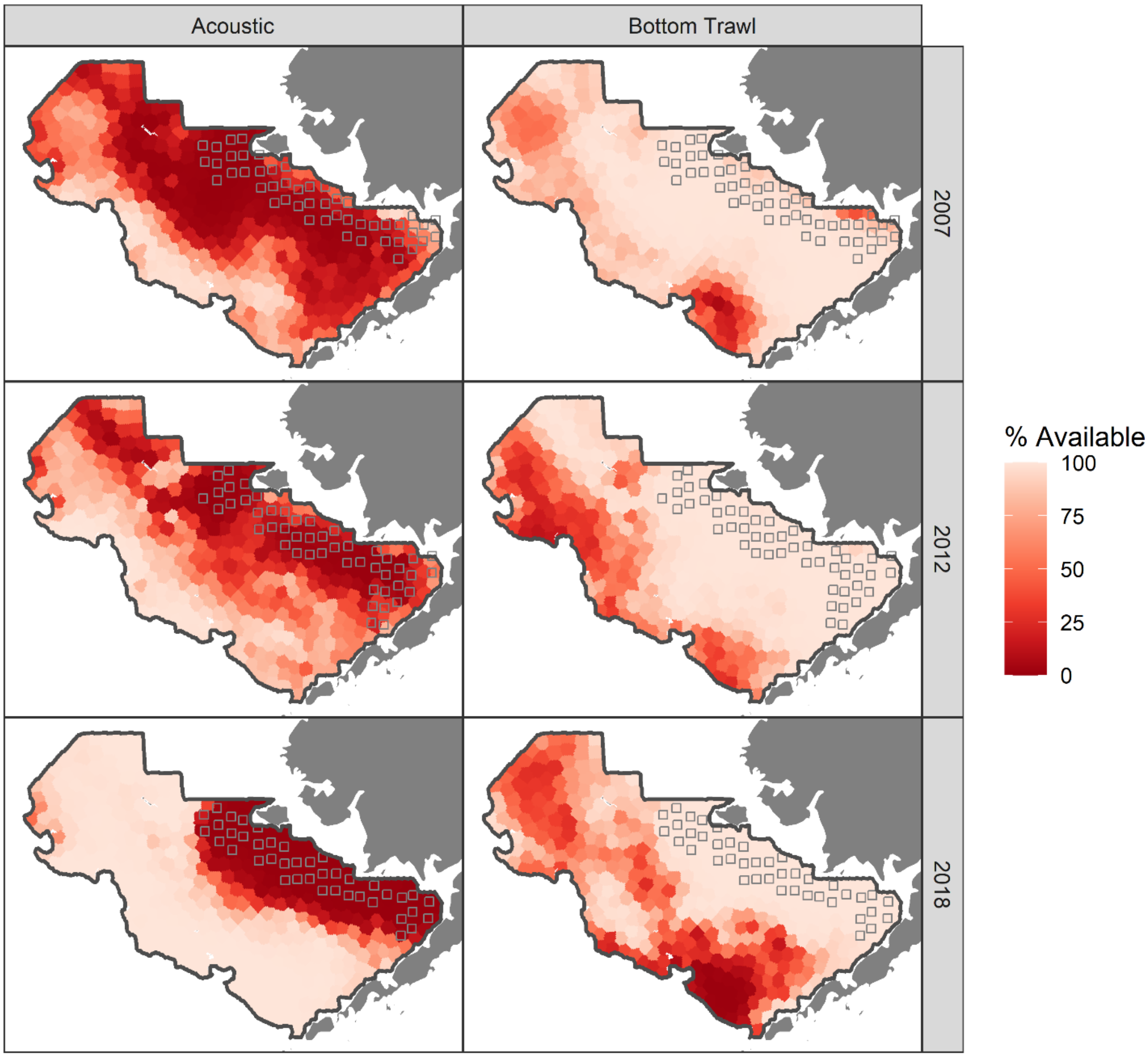


**Figure 4**. Estimated spatial availability (i.e., percentage of pollock available to a gear type at a location) for three select years (rows) for the acoustic and bottom trawl surveys (columns) from the combined model. The gray squares are the locations of inflated acoustic data which are in a region unsampled by the acoustic survey (see Fig. 2 and main text).

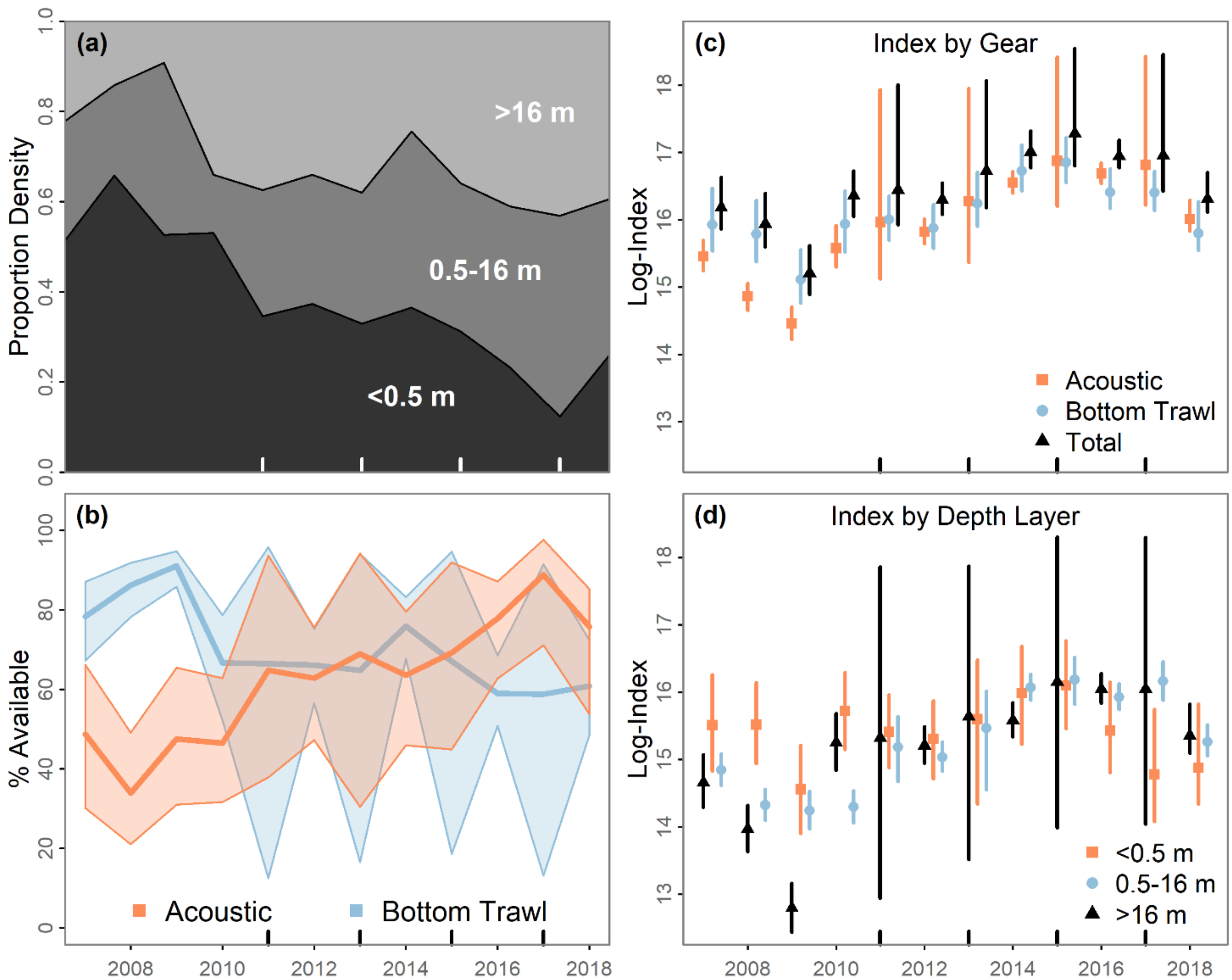


**Figure 5**. Annual results from the combined model fit to pollock, where years without acoustic data are indicated in all panels using *x*-axis dashes and are estimated with a temporal smoother within the model, resulting in higher uncertainty. (a) Vertical distribution (posterior median proportions after integrating across space) of fish density for the three depth layers, with uncertainty left off for visual clarity. (b) Vertical availability by gear type (colors), shown as the median and 95% credible interval (lines and ribbons). Comparison of estimated abundance indices by gear type (c) and depth layer (d), where points are medians and vertical bars are 95% credible intervals.

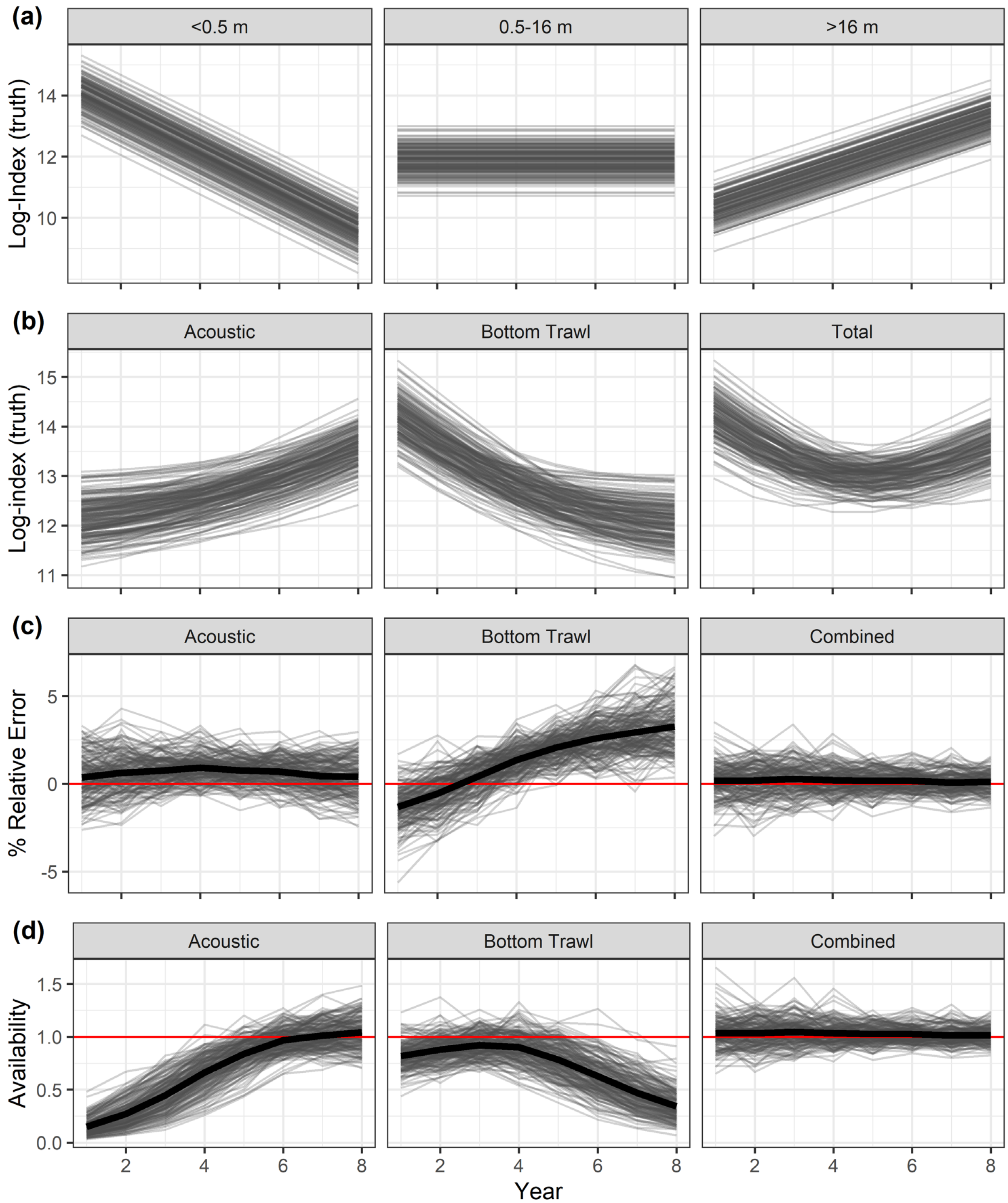


**Figure 6**. Results of simulation study. Individual lines show simulation replicates and think black lines the average across them. (a) the true log-index by depth layer used; (b) the true index available to each gear type; (c) the

estimation bias compared to the truth for each gear type to its own truth (a self-test); and (d) the resulting estimated vertical availability compared to the truth for the whole water column when fitting to the two gear types separately or with the combined model developed here.

# References


Aglen, A. 1996. Impact of fish distribution and species composition on the relationship between acoustic and swept-area estimates of fish density. ICES Journal of Marine Science, 53: 501–505.

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association, 50: 901–908.

Álvarez-Colombo, G. L., Dato, C. V., Machinandiarena, L., Castro-Machado, F., and Betti, P. 2014. Daylight vertical segregation of young-of-the-year Argentine hake *Merluccius hubbsi*: Advances in assessment of juvenile abundance with acoustic methods. Fisheries Research, 160: 85–95.

Aydin, K., and Mueter, F. 2007. The Bering Sea—a dynamic food web perspective. Deep-sea Research Part II: Topical Studies in Oceanography, 54: 2501–2525.

Beare, D., Reid, D., Greig, T., Bez, N., Hjellvik, V., Godø, O. R., Bouleau, M., *et al.* 2004. Positive relationships between bottom trawl and acoustic data. Working paper. ICES. https://imr.brage.unit.no/imr-xmlui/handle/11250/100595 (Accessed 23 July 2020).

Bouleau, M., Bez, N., Reid, D. G., Godo, O. R., and Gerritsen, H. 2004. Testing various geostatistical models to combine bottom trawl catches and acoustic data. ICES CM 2004/R:28: 19.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., *et al.* 2017. Stan: A probabilistic programming language. Journal of Statistical Software, 76: 1–29.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. 2018. A guide to Bayesian model checking for ecologists. Ecological Monographs, 88: 526–542.

Conner, J., and Lauth, R. R. 2017. Results of the 2016 eastern Bering sea continental shelf bottom trawl survey of groundfish and invertebrate resources. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC352,159 p.

Everson, I., Bravington, M., and Goss, C. 1996. A combined acoustic and trawl survey for efficiently estimating fish abundance. Fisheries Research, 26: 75–91.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2014. Bayesian data analysis. Chapman & Hall. 675 p.

Godø, O. R., and Wespestad, V. G. 1993. Monitoring changes in abundance of gadoids with varying availability to trawl and acoustic surveys. ICES Journal of Marine Science, 50: 39–51.

Gunderson, D. R. 1993. Surveys of Fisheries Resources. John Wiley & Sons. 278 pp.

Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media Dordrecht. 570 p.

Hjellvik, V., Tjøstheim, D., and Godø, O. R. 2007. Can the precision of bottom trawl indices be increased by using simultaneously collected acoustic data? The Barents Sea experience. Canadian Journal of Fisheries and Aquatic Sciences, 64: 1390–1402.

Hoffman, M. D., and Gelman, A. 2014. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15: 1593–1623.

Honkalehto, T., Ressler, P. H., Towler, R. H., and Wilson, C. D. 2011. Using acoustic data from fishing vessels to estimate walleye pollock (*Theragra chalcogramma*) abundance in the eastern Bering Sea. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1231–1242.

Honkalehto, T., and McCarthy, A. 2015. Results of the acoustic-trawl survey of walleye pollock (*Gadus chalcogrammus*) on the U.S. and Russian Bering Sea Shelf in June - August 2014 (DY1407). AFSC Processed Rep. 2015-07, 62 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. http://www.afsc.noaa.gov/Publications/ProcRpt/PR2018-03.pdf.

Honkalehto, T., McCarthy, A., and Lauffenburger, N. 2018. Results of the acoustic-trawl survey of walleye pollock (*Gadus chalcogrammus*) on the U.S. Bering Sea shelf in June - August 2016 (DY1608). AFSC Processed

Rep. 2018-03, 78 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115. http://www.afsc.noaa.gov/Publications/ProcRpt/PR2018-03.pdf.

Ianelli, J. N., Fissel, B., Holsman, K., Kotwicki, S., Monnahan, C. C., Siddon, E., Stienessen, S., *et al.* 2019. Assessment of the walleye pollock stock in the Eastern Bering Sea, pp. 51-156. In Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fishery Management Council, Anchorage, AK. Available at https://archive.fisheries.noaa.gov/afsc/refm/stocks/plan_team/2019/EBSPollock.pdf.

Illian, J. B., Sørbye, S. H., and Rue, H. 2012. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). The Annals of Applied Statistics: 1499–1530.

Jakobsen, T., Korsbrekke, K., Mehl, S., and Nakken, O. 1997. Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. Working paper. ICES. https://imr.brage.unit.no/imr-xmlui/handle/11250/100335 (Accessed 23 July 2020).

Kai, M., Thorson, J. T., Piner, K. R., and Maunder, M. N. 2017. Predicting the spatio-temporal distributions of pelagic sharks in the western and central North Pacific. Fisheries Oceanography, 26: 569–582.

Kotwicki, S., De Robertis, A., Ianelli, J. N., Punt, A. E., Horne, J. K., and Jech, J. M. 2013. Combining bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl efficiency parameters for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences, 70: 208–219.

Kotwicki, S., Horne, J. K., Punt, A. E., and Ianelli, J. N. 2015. Factors affecting the availability of walleye pollock to acoustic and bottom trawl survey gear. ICES Journal of Marine Science, 72: 1425–1439.

Kotwicki, S., Ressler, P. H., Ianelli, J. N., Punt, A. E., and Horne, J. K. 2018. Combining data from bottom-trawl and acoustic-trawl surveys to estimate an index of abundance for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences, 75: 60–71.

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 and Fisheries, 20: 760–774.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70: 21.

Levine, M., and De Robertis, A. 2019. Don’t work too hard: Subsampling leads to efficient analysis of large acoustic datasets. Fisheries Research, 219: 105323.

Lindgren, F., Rue, H., and Lindstrom, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B-Statistical Methodology, 73: 423–498.

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

Michalsen, K., Godø, O. R., and Fernö, A. 1996. Diel variation in the catchability of gadoids and its influence on the reliability of abundance indices. ICES Journal of Marine Science, 53: 389–395.

Monnahan, C. C., Thorson, J. T., and Branch, T. A. 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution, 8: 339–348.

Monnahan, C. C., and Kristensen, K. 2018. No-U-turn sampling for fast Bayesian inference in ADMB and TMB: Introducing the adnuts and tmbstan R packages. Plos One, 13: e0197954.

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

O’Leary, C., Thorson, J. T., Ianelli, J., 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.

Ona, E., Pennington, M., and Vølstat, J. H. 1991. Using acoustics to improve the precision of bottom trawl indices of abundance. ICES CM 1991/D:13: 11.

Ono, K., Kotwicki, S., Dingsør, G. E., and Johnsen, E. 2018. Multispecies acoustic dead-zone correction and bias ratio estimates between acoustic and bottom-trawl data. ICES Journal of Marine Science, 75: 361–373.

Overholtz, W. J., Jech, J. M., Michaels, W. L., Jacobson, L. D., and Sullivan, P. J. 2006. Empirical comparisons of survey designs in acoustic surveys of Gulf of Maine-Georges Bank Atlantic herring. Journal of Northwest Atlantic Fisheries Science, 36: 127–144.

Perry, A. L., Low, P. J., Ellis, J. R., and Reynolds, J. D. 2005. Climate change and distribution shifts in marine fishes. Science, 308: 1912–1915.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Skaug, H. J., and Fournier, D. A. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis, 51: 699–709.

Stan Development Team. 2017. Stan modeling language users guide and reference manual, version 2.17.0.

Stan Development Team. 2018. rstan: R interface to Stan. R package version 2.18.2. http://mc-stan.org.

Stauffer, G. D. (compiler). 2004. NOAA protocols for groundfish bottom trawl surveys of the nation's fishery resources, March 16, 2003. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SPO-65, 205 p.

Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., and Kristensen, K. 2015a. Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. Methods in Ecology and Evolution, 6: 627–637.

Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015b. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science, 72: 1297–1310.

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 Journal of Marine Science, 74: 1311–1321.

Thorson, J. T. 2017. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Canadian Journal of Fisheries and Aquatic Sciences, 75: 1369–1382.

Thorson, J. T., Ianelli, J. N., and Kotwicki, S. 2017. The relative influence of temperature and size-structure on fish distribution shifts: A case-study on walleye pollock in the Bering Sea. Fish and Fisheries, 18: 1073–1084.

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

Vehtari, A., Gelman, A., and Gabry, J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27: 1413–1432.

von Szalay, P. G., Somerton, D. A., and Kotwicki, S. 2007. Correlating trawl and acoustic data in the eastern Bering Sea: A first step toward improving biomass estimates of walleye pollock (*Theragra chalcogramma*) and Pacific cod (*Gadus macrocephalus*)? Fisheries Research, 86: 77–83.