Abstract—A bottom-trawl survey with a stratified-random sampling design has been used to inform stock assessments for commercially important species in the Gulf of Alaska since 1984. A new stratified sampling design was evaluated to determine whether its use could improve the precision and accuracy of abundance estimates. In this proposed approach to defining strata, historical survey data are used to generate what we refer to as *information scores* (ISes). We compared the traditional stratification scheme with the new method and 2 other sampling designs, using both a design-based estimator and a model-based estimator with each design, to determine if the existing approach is optimal. Statistical robustness, measured in terms of coefficient of variation, bias, and root mean square error, was compared among 7 scenarios with different combinations of estimators and sampling designs by using simulation with a spatiotemporal generalized linear mixed model conditioned on historical observations of catch per unit of effort of 3 species. The combination of the design-based estimator with the IS-based stratification scheme was the best scenario across all performance metrics for all species. This scenario consistently had the lowest variance and smallest total error, and it was generally unbiased. In contrast, the pairing of the model-based estimator with this sampling design was by far the worst-performing scenario. The performance of the existing approach was average.

Reducing uncertainty in survey abundance estimates by considering alternative designs and estimators: a case study with 3 species in the Gulf of Alaska

Paul G. von Szalay (contact author)

Stan Kotwicki

Lewis A. K. Barnett

Lou J. Rugolo

Kotaro Ono

Email address for contact author: paul.von.szalay@noaa.gov

Alaska Fisheries Science Center
National Marine Fisheries Service, NOAA
7600 Sand Point Way NE
Seattle, Washington 98115

Surveys for estimating abundance of natural resources, and the precision of abundance estimates based on survey data, are limited in scope by a number of logistical constraints, including but not limited to funding, staffing, time, and access to appropriate and sufficient sampling tools and platforms. The precision of abundance estimates, including those of biomass, generally increases with sample size, which is limited by available resources, assuming the availability of a species to sampling remains constant. In practice, the scope of a survey effort, as indicated by the number of stations where sampling occurs, is typically known before the sampling design is generated and is dictated by logistical constraints. The goal of a sampling design, therefore, is to optimize the geographical distribution of the predetermined number of survey stations to achieve the highest possible precision for survey data products. A potential problem for survey continuity arises when circumstances compel a reduction or increase in sampling effort, especially after a survey time series is well established.

In such an event, it may be appropriate to review the existing sampling design of the survey for optimality and to consider changes to the design while assuring continuity of the time series. The long-established and relatively large-scale multispecies bottom-trawl survey that has been conducted in the Gulf of Alaska (GOA) since 1984 can serve as an illustrative example of how these issues can be addressed.

The NOAA Alaska Fisheries Science Center has conducted a bottom-trawl survey in summer in the GOA biennially since 1999 (and triennially between 1984 and 1999), using a stratified-random sampling design to assess the distribution and abundance of groundfish for the purposes of fisheries management (von Szalay and Raring, 2018). Traditionally, the survey effort consisted of 3 vessels sampling approximately 820 stations in 59 different strata. Prior to 1990, the focal survey areas in a given year were adapted to meet specific scientific or management needs, and as a result the effort levels in the different strata could be highly variable from one survey year

to another. This practice was discontinued in 1990, when the sampling design of the survey was standardized to adhere to a modified Neyman allocation scheme. A variety of factors in recent years, such as funding limitations and inability to acquire a sufficient number of suitable charter vessels, have resulted in the need to reduce the survey effort in some years (von Szalay, 2015). Other fisheries resource management agencies, both domestic and international, similarly have had to or are preparing to contend with potential unavoidable reductions in survey effort for a variety of reasons, including vessel breakdowns, lack of enough survey vessels or staff due to the COVID-19 pandemic, bad weather, and failure to obtain sampling permits for all parts of the survey area (Peel et al., 2013; ICES, 2020).

The reduction of survey effort has amplified the need not only to ensure the use of the optimal sampling design but also to seek the most robust and precise abundance estimator possible that would perform well in years with fewer data. This issue can be addressed by comparing the statistical robustness of the existing approach with that of alternative scenarios based on different combinations of estimators and sampling designs, a valuable exercise even if survey effort has not been reduced. Until recently, it has been difficult to make direct comparisons between different estimators and sampling designs because of a lack of an appropriate simulation framework. However, Kotwicki and Ono (2019) developed a framework for simulating distributions of fish population density with a spatiotemporal model conditioned on historical catch and environmental survey data for a number of commercially important species in the GOA. We used this operating model to simulate the "true" distribution of abundance and the results of simulated sampling with different designs.

When the existing stratified-random sampling design was first devised in the early 1980s, no historical catch or environmental data were available to refine the stratification scheme. Now that a long-running time series is well established, it is possible to use these data to explore alternative stratification schemes and sampling designs, which may potentially reduce the uncertainty in abundance estimates by yielding more accurate and precise estimates of biomass and its associated variance. In recent studies, optimization-based methods of stratifying finite populations (Hidiroglou and Kozak, 2018; Oyafuso et al., 2021, 2022) have been proven to be superior to the traditional stratification scheme used for the bottom-trawl survey in the GOA. In our study, we evaluated the performance of 3 different traditional sampling designs, which are among the most commonly used in the world and have already been implemented by the Alaska Fisheries Science Center: simple random, stratified random, and systematic.

For these evaluations, we used a proposed new technique for determining stratum boundaries in the GOA, one in which estimates of abundance and its variability from historical survey data are used to derive an empirical metric referred to herein as an *information score* (IS). This technique, which ensures that more stations are allocated to strata with higher abundance and variance in abundance

based on historical survey data, fine-tunes the stratification process by decreasing the variance within strata while increasing the variance among strata. This approach contrasts with the traditional stratification scheme whose performance may be compromised by including an ad hoc management area and depth bounds to define strata. Management area boundaries are unlikely to have any bearing on fish density, and arbitrary depth boundaries in regular 100-m intervals are unlikely to coincide with breaks in fish density distributions.

There has recently been a rapid development of methods for estimating population density by using spatial model-based estimators to calculate biomass or abundance (e.g., Thorson and Barnett, 2017; Anderson et al.[1]). Design-based estimators, such as the one developed by Wakabayashi et al. (1985), are simpler and faster, but spatial model-based estimators, such as vector autoregressive spatiotemporal (VAST) models (Thorson, 2019), may be more accurate or precise under some scenarios because they allow for spatial and temporal correlation and covariate effects. In addition, model-based designs can be both efficient and flexible in allowing uneven sampling due to survey logistics and in providing a general framework to answer specific design questions (Peel et al., 2013).
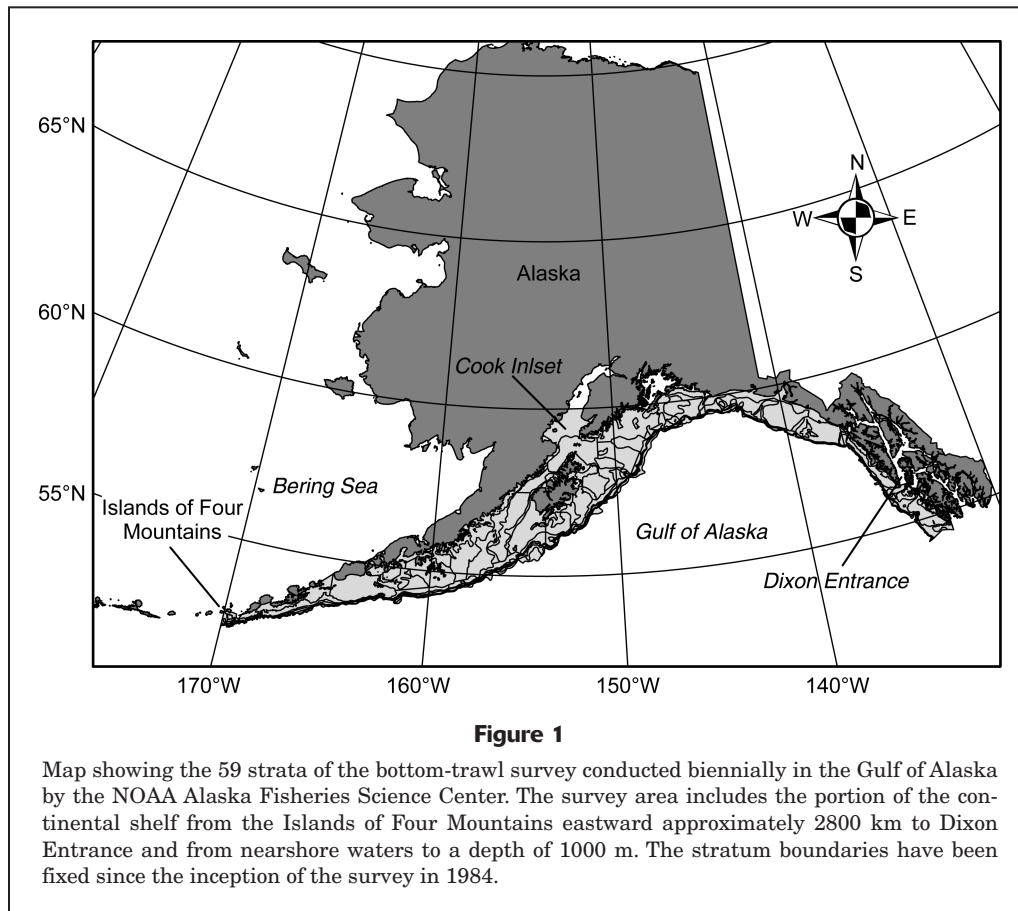
The objective of this study was to compare the statistical robustness of 7 scenarios with different combinations of sampling designs and estimators in order to identify whether an approach improves the accuracy and precision of estimates of biomass, which we used as a measure of abundance. As a case study, we evaluated the statistical robustness of 7 scenarios by simulating the distribution and abundance of 3 species in the GOA: Pacific cod (*Gadus macrocephalus*), Pacific ocean perch (*Sebastes alutus*), and arrowtooth flounder (*Atheresthes stomias*). These species were chosen on the basis of their ecological or commercial importance as well as the diversity of their spatial distribution patterns. In 6 scenarios, a random sampling design was paired with a design-based or model-based estimator, and in 1 scenario a systematic sampling design was paired with a model-based estimator.

Materials and methods

Survey characteristics

The GOA forms the northeastern border of the Pacific Ocean and has complex bathymetric features ranging from jagged, mountainous pinnacles to flat, muddy areas (von Szalay and Raring, 2018). These features provide a variety of habitats, creating a complex ecosystem. The standard survey area of the biennial bottom-trawl survey in the GOA is approximately 320,000 km^2 and includes

[1] Anderson, S. C., E. J. Ward, P. A. English, and L. A. K. Barnett. Unpubl. manuscript. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. bioRxiv 2022.03.24.485545. [Available from website.]

Figure 1

Map showing the 59 strata of the bottom-trawl survey conducted biennially in the Gulf of Alaska by the NOAA Alaska Fisheries Science Center. The survey area includes the portion of the continental shelf from the Islands of Four Mountains eastward approximately 2800 km to Dixon Entrance and from nearshore waters to a depth of 1000 m. The stratum boundaries have been fixed since the inception of the survey in 1984.

the portion of the continental shelf from the Islands of Four Mountains, in the Aleutian Islands, eastward to Dixon Entrance, near the border of Canada and the United States, and from nearshore waters to a depth of 1000 m (Fig. 1). The continental shelf, which makes up 82% of the survey area, is bordered by the continental slope, a region that is approximately 20 km in width, but only the portion of the slope at depths between 200 and 1000 m, is sampled in the standard surveys. The total survey area has been effectively decreased to 308,415 km^2 in recent surveys as reductions in survey effort have resulted in the deepest strata (700–1000 m) not being sampled.

In the bottom-trawl surveys conducted in the GOA by the Resource Assessment and Conservation Engineering Division of the Alaska Fisheries Science Center, approximately 820 locations within 59 strata traditionally have been sampled with a stratified-random sampling design. Abundance estimates are generated by using the estimator developed by Wakabayashi et al. (1985). Henceforth, this combination of sampling design and estimator will be referred to as the *TRS*.

The same standard trawl gear has been used since the beginning of the survey, a Poly Nor'Eastern 4-seam bottom trawl with 24.2-m roller gear constructed with 36-cm rubber bobbins separated by 10-cm rubber disks (Stauffer, 2004). Surveys start in the western end of

the survey area and proceed eastward. Tow duration is approximately 15 min at 1.54 m/s (3 kt). The catch per unit of effort (CPUE) is estimated by using the area-swept method (e.g., Alverson and Pereyra, 1969), which defines the effort as the product of the distance fished and the average distance between wing tips (for details, see von Szalay and Raring, 2018).

A grid with a resolution of 5 km superimposed on the survey area is used for sampling design. Each 5-by-5-km cell in the grid is a potential survey station, but some potential stations are excluded from the pool of stations eligible for survey selection because they are untrawlable as a result of rough bottom or other factors. Furthermore, because they are truncated by the coastline and the deepest edge of the sampling domain, grid cells along the boundaries of the survey area have an area smaller than a standard 25-km^2 grid cell. Only grid cells large enough to accommodate a complete tow path (those with an area >5 km^2) are eligible for sampling.

Simulated distributions of catch per unit of effort

Simulated CPUE distributions of Pacific cod, Pacific ocean perch, and arrowtooth flounder were created by fitting a spatiotemporal delta model to historical GOA survey data from 1996 through 2015, with the package

R-INLA (vers. 18.07.12; available from website, accessed October 2018; Rue et al., 2009) in R (vers. 3.6.1; R Core Team, 2019); this package accounts for both environmental covariates and spatiotemporal dependency in catches (Kotwicki and Ono, 2019) (Fig. 2). The delta model has 2 components: one that models the species occurrence and another that models positive CPUE.

Species occurrence at locations s during year t, $\pi_t(s)$, was modeled by using a binomial generalized linear mixed model with the logit link function (see Lindgren et al., 2011):

$$\text{logit}\left(\pi_t\left(s\right)\right) = X_t\left(s\right)b + \omega_t\left(s\right), \text{where} \qquad (1)$$

$$\omega_t\left(s\right) \sim N\left(\rho_1\omega_{t-1}\left(s\right), \Sigma_1\right)$$

and $X_t(s)$ = the matrix of covariates at locations s during year t;

   b = the vector of regression coefficients;
   ω = the spatiotemporal variation that follows a first-order autoregressive process;
   N = normal distribution;
   ρ_1 = the degree of autoregression in encounter probability between successive years; and
   Σ_1 = the spatial covariance modeled as a Mátern function with smoothness of 1.

The covariates used in this study were log(*depth*), $(\log(depth))^2$, *bottom temperature*, *bottom temperature* squared, *surface temperature*, *surface temperature* squared, and a fixed year effect (Kotwicki et al., 2005, 2015).

The non-zero species density at a set of locations s during year t, $\mu_t(s)$, was modeled by using a lognormal distribution:

$$\log\left(\mu_t\left(s\right)\right) = Z_t\left(s\right)a + \delta_t\left(s\right), \text{where} \qquad (2)$$

$$\delta_t\left(s\right) \sim N\left(\rho_2\delta_{t-1}\left(s\right), \Sigma_2\right)$$

and $Z_t(s)$ = a matrix of covariates at locations s during year t;

   a = the vector of regression coefficients;
   δ_t = the spatial field for year t assumed to follow an autoregressive 1 process;
   ρ_2 = autocorrelation of the autoregressive 1 process in which the current value is based on the immediately preceding value; and
   Σ_2 = the spatial covariance modeled as a Mátern function with smoothness of 2.

Annual distributions of simulated CPUE within the survey area were predicted over a nominal grid with a
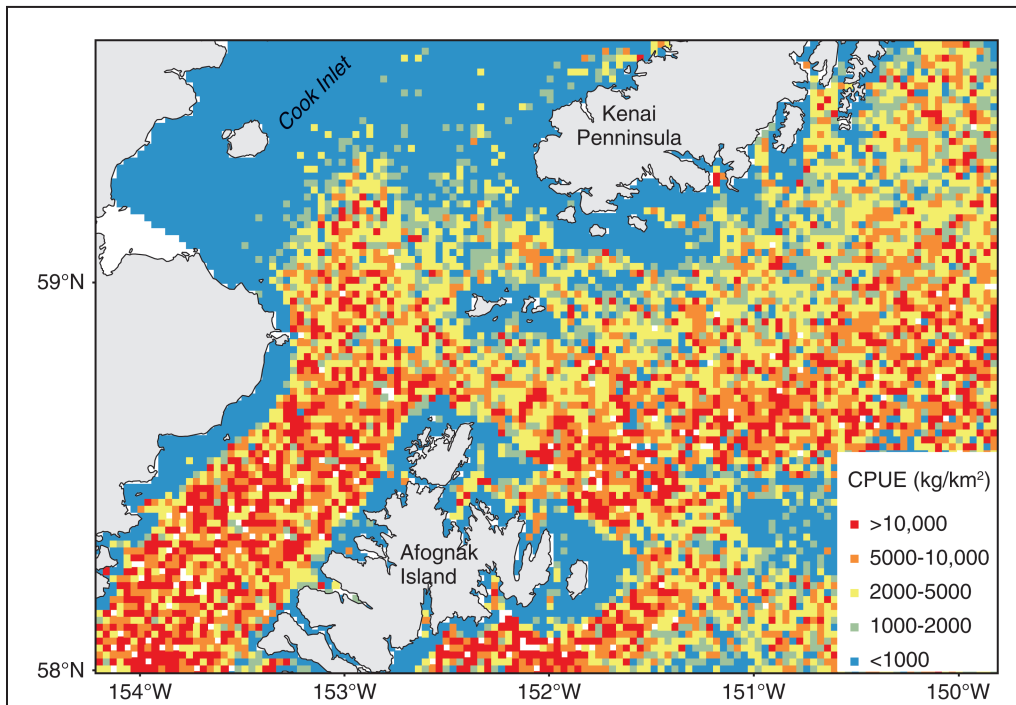
Figure 2

Simulated distribution of catch per unit of effort (CPUE) of arrowtooth flounder (*Atheresthes stomias*) in 2015 at the entrance to Cook Inlet in southcentral Alaska. The simulations were created by fitting a spatiotemporal delta model that accounts for both environmental covariates and spatiotemporal dependency in catches to historical data from the bottom-trawl survey in the Gulf of Alaska. Each pixel on the map is 2 km^2 and represents the CPUE realized when a station is sampled at that location in the simulated surveys conducted in this study.

2-km resolution on the basis of a single but random sample from the posterior predictive distribution of the models while accounting for the sampling process. In other words, simulated CPUE was calculated as the product of sampled fish occurrence and sampled density. Sampled fish occurrence was estimated by using a Bernoulli trial with the probability determined by the parameter samples taken from the posterior predictive distribution (in practice, a random Markov chain Monte Carlo [MCMC] run was chosen and parameter values were taken from it). Sampled density was estimated by using a sample from a Gaussian distribution on the link scale with the mean and variance derived from parameter values taken from a randomly chosen MCMC run (the sample value was exponentiated by using the natural exponential function to convert it to real space).

With this approach, predictions accounted for all sources of uncertainty included in the model and created a noisier CPUE distribution that was more reflective of "true" patterns than the mean MCMC prediction. Environmental covariate values were determined by kriging with the semivariogram model that best fit the historical survey data (Kotwicki and Ono, 2019), as implemented by the function autofitVariogram in the R package automap (vers. 1.0-16; Hiemstra et al., 2009). All geographic coordinates were converted into an Albers projection in order to preserve distances prior to analysis.

Simulation of surveys

The CPUE distributions defined over the grid with the finer resolution of 2 km were superimposed on the grid with a 5-km resolution for the surveys in the GOA before the simulated surveys were generated. Each station sampled during surveys in the GOA (i.e., each 25-km^2 grid cell), therefore, contained an average of approximately 4–7 of the 2-km^2 grid cells with density data. A variety of scenarios with different combinations of estimators and sampling designs were then implemented to generate simulated surveys over the modeled CPUE distributions for each of the 10 survey years between 1996 and 2015 and for each of 3 commercially important species: Pacific cod, Pacific ocean perch, and arrowtooth flounder.

In addition to the traditional stratified-random sampling design and the proposed stratification sampling design based on an IS, we considered 2 additional sampling designs, simple random sampling and systematic sampling, in conjunction with both the model-based estimator and the design-based estimator, for a total of 8 potential scenarios. The scenario in which the design-based estimator is paired with the systematic sampling design was excluded from this analysis because of its unsuitability for the survey area in the GOA. Because the survey area is characterized by relatively wide strata on the continental shelf but extremely narrow strata along the continental slope, stations are too far apart to adequately cover the slope area. The remaining 7 scenarios consisted of the use of the design-based estimator with 3 different sampling designs—simple random sampling

(SRS), stratified-random sampling based on ISes (WIS) (described later in the "Stratified-random sampling based on information scores" section), and TRS—and the use of the model-based estimator in the R package VAST (vers. 3.3.0; Thorson, 2019) with 4 different sampling designs—systematic sampling (VSY), simple random sampling (VRS), stratified-random sampling based on ISes (VIS), and traditional stratified-random sampling (VTRS).

We simulated 100 replicate surveys, each consisting of 820 stations, for each of the 10 years in which surveys had been conducted, resulting in 82,000 simulated stations per survey year for each scenario. All analyses were conducted in R, vers. 3.6.1 (R Core Team, 2019).

Design-based approaches

Traditional stratified-random sampling In the stratified-random sampling design used since the inception of the bottom-trawl survey of the GOA in 1984, the survey area is divided into 59 strata defined by water depth, bottom terrain (e.g., shelf, gully, and slope), and statistical management areas (von Szalay and Raring, 2018). Following this traditional design, we allocated stations among the strata for species k by using a modified Neyman strategy for optimal allocation (Cochran, 1977):

$$n_{hk} = \frac{\dfrac{nN_h s_{hk} a_h}{\sqrt{c_h}}}{\sum_h \dfrac{N_h s_{hk} a_h}{\sqrt{c_h}}}, \qquad (3)$$

where n_{hk} = the sample size for species k in stratum h;
n = the total sample size;
N_h = the population size for stratum h;
s_{hk} = the standard deviation of species k in stratum h;
a_h = the area of stratum h; and
c_h = the cost to conduct a trawl tow in stratum h.

Catch rates, stratum variances, and stratum areas from the surveys conducted in 1990–2015 were used to allocate sampling effort among strata for each previous survey year and for each of the 50 principal GOA groundfish species that the historical survey was designed to sample. The estimated time to perform a trawl tow in a given stratum was used as a cost variable because tows in deeper strata have a greater probability of unacceptable gear performance and take longer to complete. A mean sample size was calculated across years for each principal species, and the sample size for each species was then weighted by each species' commercial value, which was defined as the product of the mean biomass and its ex-vessel value in 2015. The calculated sample sizes, representing the numbers of trawl tows allocated to the various strata, were rounded to whole numbers while ensuring that each stratum was allocated at least 2 samples.

The allocated stations within each stratum were randomly selected without replacement from the 2-km-resolution grid described earlier. Sample mean ($\bar{x}_{hk}$)

and standard deviation (S_{hk}) for species k were calculated by stratum and were used to generate the survey mean ($\overline{X}_k$) and variance ($S_{\overline{x}k}^2$) for a stratified-random sampling design according to the following equations:

$$\overline{X}_k = \frac{1}{N} \sum\nolimits_{h=1}^{20} N_h \overline{x}_{hk}, \text{ and} \qquad (4)$$

$$S_{\overline{x}k}^2 = \sum\nolimits_{h=1}^{20} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_{hk}}{N_h}\right)\left(\frac{s_{hk}^2}{n_{hk}}\right). \qquad (5)$$

Stratified-random sampling based on information scores As an alternative method of stratification to the traditional scheme, we propose the IS approach to stratified-random sampling. This novel approach uses spatial and temporal information of abundance and temporal information of variation in abundance to derive an IS for each sampling unit. A high IS should identify areas where abundance tends to be high, areas where the abundance is highly variable over time, or a combination of these characteristics. Sampling in such areas is likely to provide most information about population abundance, composition, and trends in abundance, all of which are critical for use in stock assessment models and for acquisition of other types of information about sampled populations and temporal trends in population parameters. Information scores could be used to stratify the survey area and provide a basis for prioritizing areas with high fish density and high variance in fish density to ensure that more stations are allocated to strata with higher expected abundance and variability in abundance. Given this logic, we propose an IS that is calculated for each of the 15,718 eligible 5-by-5-km grid cells in the survey area as follows:

$$IS_{ik} = \overline{\widehat{CPUE}}_{ik}^2 + \frac{\text{Var}(\widehat{CPUE}_{ik})}{n}, \qquad (6)$$

where $\overline{\widehat{CPUE}}_{ik}$ = the mean predicted CPUE of species k in grid cell i over the 10 survey years between 1996 and 2015.

The $\widehat{CPUE}_{ik}$ of grid cell i (5-by-5 km) for a given year was calculated as the mean prediction of the simulated densities of all 2-by-2-km grid cells within it. By squaring the mean predicted CPUE, we ensure that the 2 terms have the same units and equal weight.

To determine how to stratify on the basis of discretizing the ISes into bins of similar values, we first examined how the scores were distributed. We found the distribution of the 15,718 ISes to be highly positively skewed for all species (Pearson's coefficient of skewness ranged between 8.7 for arrowtooth flounder and 43.5 for Pacific ocean perch; Suppl. Fig. 1).

These values were sorted from low to high, and evenly spaced breakpoints were inserted to generate 20 strata, each with 786 stations. The choice to use 20 strata resulted from observations about the scores. These observations

were that approximately 5% of the ISes were equal to zero, and a similar proportion of the scores on the opposite end of the distribution were of the same order of magnitude, resulting in 2 strata that contained the extreme values on both ends of the distribution with other strata having intermediate values. In a sensitivity analysis, we found no indication of substantial differences in results when either 19 or 21 strata were used. Furthermore, 20 strata is a reasonable number for the survey in the GOA and is also the number of strata being considered for a revised sampling design for this survey.

The number of survey stations allocated to the different strata was weighted by the mean IS of each stratum. First, we used the Neyman algorithm to determine the sample allocation (Lavrakas, 2008):

$$n_h = \frac{n(N_h s_h)}{\sum\nolimits_{h=1}^{20} N_h s_h}. \qquad (7)$$

Then, to weight the allocation among strata by using the mean IS of each stratum, the weights were calculated as follows:

$$w_h = \frac{\sqrt{\overline{IS}_h}}{\sum\nolimits_{h=1}^{20} \sqrt{\overline{IS}_h}}, \qquad (8)$$

where w_h = the weight of stratum h, and

$\overline{IS}_h$ = the mean information score for stratum h.

In this equation, substituting $\sqrt{\overline{IS}_h}$ for s_h in Equation 7 and algebraically simplifying the new equation (N_h is a constant and cancels out) results in n_h being equal to nw_h, which is the sample size of stratum h weighted by its information score.

Some stations originally assigned to the strata with the highest ISes were reallocated to strata receiving fewer than 2 stations to make variance calculations possible for all strata. This procedure for station allocation among strata was repeated for each species.

The allocated stations within each stratum were randomly selected without replacement from the grid with a 5-km resolution for the survey in the GOA. A random grid cell from the 2-km-resolution grid was then sampled to represent the CPUE of the entire station. Sample mean and standard deviation were calculated by stratum and were used to generate the survey mean and variance for a stratified-random sampling design similar to the procedure used for the traditional stratified-random approach described in the previous section.

Model-based approaches

The model-based estimator was a spatiotemporal delta model consisting of a binomial model for the probability of encounter and a user-selected model for positive catch rates, as implemented in the VAST package (Thorson, 2019). In a report of their recent work, Thorson et al. (2021) noted that the choice of distributional assumptions can

have a substantial effect on the scale of the biomass index estimated from such spatiotemporal models; therefore, we tested a range of assumptions to determine which provided the best fit and accuracy. We considered 2 different observation error distributions for positive catch rates, lognormal and gamma, and whether or not to correct values for retransformation bias (Thorson and Kristensen, 2016).

The hurdle models included *year* as a fixed effect, a spatial random field for each model component (encounter and positive catch rate), and independent spatiotemporal random fields for the positive catch rate component (Suppl. Table). The spatiotemporal term for encounter probability was not estimated because of the lack of consistent convergence among replicate simulated surveys. A model resolution of 500 knots was used, and bilinear interpolation was implemented to extrapolate from knot locations in order to generate predictions at each location in the same 2-km-resolution grid used in the operating model. Anisotropy was estimated. The models for each scenario differed only in terms of the observation error model used for positive catch rates (gamma versus lognormal distribution) and by whether retransformation bias was addressed (i.e., with or without epsilon bias correction) (Thorson and Kristensen, 2016). All models were run with the VAST package (Thorson et al., 2015) and the tools available in FishStatsUtils (vers. 2.6.0; available from website, accessed October 2019).

A customized 2-km-resolution extrapolation grid was created as a subset of the standard GOA survey grid (depths <1000 m), which excluded all grid cells known to be untrawlable. This custom grid was used to ensure that the VAST estimation model only integrated densities from trawlable areas or areas with an unknown trawlability status in order to make the estimates of biomass and its variance from the model-based approaches comparable to those from the design-based approaches. The results of the 4 different combinations of settings, between the 2 observation models for positive catch rates and whether or not the retransformation bias was addressed, for each scenario and species were compared with respect to the performance metrics described later in the "Performance metrics" section. The combination of settings that resulted in the lowest relative root-mean-square error (rRMSE) for a given scenario and species was selected for final comparison with the other scenarios (Table 1).

Simple random sampling Simulated CPUE values from the operating model were sampled by using simple random sampling. These values were then used as input to the VAST estimation model, along with the associated data on location and year for CPUE observations.

Systematic sampling A rectangular grid encompassing the entire survey area in the GOA was created by identifying the extreme latitudes and longitudes spanning the survey area. An iterative procedure was used to determine that the length of the square grid cells that would result in approximately 820 uniformly spaced stations falling within the survey area was approximately 20 km.

To generate multiple iterations of systematic surveys, the origin of the survey grid was altered between iterations by randomly sampling a set of coordinates from within the grid cell in the extreme southeast corner of the survey grid. This 20-km^2 survey grid was, in turn, superimposed on the nominal 2-km-resolution grid of simulated fish densities. The 2-km^2 density grid cells within each of the 820 uniformly spaced grid cells were identified, and the density associated with the 2-km^2 grid cell closest to the center of the survey grid cell was used to represent the density of the entire survey grid cell.

Stratified-random sampling based on information scores Simulated CPUE estimates from the operating model were sampled by using an IS-based stratified-random sampling design. These values were then used as input to the VAST estimation model, along with the associated data on location and year for CPUE observations.

Traditional stratified-random sampling Based on the same principles described for the traditional stratified-random sampling in the "Design-based approaches" section, the station allocation by stratum scheme replicated the sampling design used for the bottom-trawl survey in the GOA in 2007 that consisted of 825 stations allocated to the 59 survey strata. However, rather than sampling directly from the 2-km-resolution CPUE grid, the stations were sampled from the grid with a 5-km resolution used for the survey in the GOA, and the CPUE associated with each station was based on the mean of the 2-km^2 CPUE grid cells within it.

Performance metrics

Three performance metrics were used to evaluate the relative performance of the 7 scenarios that were used to simulate abundance: coefficient of variation (CV), relative bias, and rRMSE. Each metric was applied to estimates of biomass and its associated variance and was calculated separately for each year and species. The true biomass was calculated as the product of the trawlable area of the survey area in the GOA and the arithmetic mean of simulated CPUE (with observation error) from all 65,863 trawlable grid cells. The true variance was calculated for each year as the variance of the replicate simulated biomass estimates:

$$\sigma_T^2 = \frac{\sum_S (B_S - B_T)^2}{N},\qquad(9)$$

where σ_T^2 = a "true" variance of the abundance index based on 100 simulated surveys;

N = the total number of simulated surveys (i.e., 100);

B_S = the estimated biomass realized in simulation survey *s*; and

B_T = the "true" biomass estimated from simulated density grids.

Bias of both biomass and its variance was defined as the mean of the deviations between the respective value of each simulation and the true value. The relative bias (*RB*)

Table 1

Relative root-mean-square errors (rRMSEs) for 7 scenarios with different combinations of estimators and sampling designs used to estimate biomass and its variance for arrowtooth flounder (*Atheresthes stomias*) (ATF), Pacific cod (*Gadus macrocephalus*) (COD), and Pacific ocean perch (*Sebastes alutus*) (POP) in the Gulf of Alaska. The observation error model for positive catch rates and the approach to bias correction are indicated for the scenarios in which the model-based estimator was used. The model-based estimator was the vector autoregressive spatiotemporal (VAST) model in the R package VAST. The design-based estimator was developed by Wakabayashi et al. (1985). One of the sampling designs was a stratified-random design based on information scores. The scenarios with the best overall performance for each species have the lowest values for biomass and its variance, which are indicated with asterisks. Analysis of scenarios was based on data from the bottom-trawl surveys conducted in the Gulf of Alaska during 1996–2015 by the NOAA Alaska Fisheries Science Center. The abbreviations in parentheses represent the 7 scenarios.

Species	Estimator and sampling design (scenario)	Observation error model	Bias correction	rRMSE Biomass	rRMSE Variance
Model-based approaches					
ATF	VAST systematic (VSY)	delta-gamma	Off	0.16	>1.00
	VAST simple random (VRS)	delta-lognormal	Off	0.05	0.30
	VAST information score (VIS)	delta-gamma	On	0.24	0.25
	VAST stratified random (VTRS)	delta-gamma	On	0.04*	0.18*
COD	VAST systematic (VSY)	delta-lognormal	On	0.07	0.69
	VAST simple random (VRS)	delta-lognormal	On	0.10	0.18
	VAST information score (VIS)	delta-gamma	On	0.13	0.15*
	VAST stratified random (VTRS)	delta-gamma	On	0.05*	0.15*
POP	VAST systematic (VSY)	delta-gamma	On	0.17	0.98
	VAST simple random (VRS)	delta-lognormal	On	0.14	0.31
	VAST information score (VIS)	delta-gamma	Off	0.40	0.18*
	VAST stratified random (VTRS)	delta-gamma	On	0.12*	0.28
Design-based approaches					
ATF	Simple random (SRS)	–	–	0.07	0.19
	Information score (WIS)	–	–	0.02*	0.05*
	Stratified random (TRS)	–	–	0.08	0.28
COD	Simple random (SRS)	–	–	0.06	0.15
	Information score (WIS)	–	–	0.03*	0.09*
	Stratified random (TRS)	–	–	0.06	0.19
POP	Simple random (SRS)	–	–	0.17	0.39
	Information score (WIS)	–	–	0.03*	0.10*
	Stratified random (TRS)	–	–	0.16	0.40

of these estimates was calculated within each year and species as follows:

$$RB = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\theta_i - \theta_{True}\right)}{\theta_{True}}, \qquad (10)$$

where θ_i = the value (biomass or variance) of the ith simulation replicate, and

θ_{True} = the true value (biomass or variance), over 100 simulation replicates.

The CV was defined as follows:

$$CV = \frac{\sqrt{Var\left(\theta\right)}}{\theta_{True}}, \text{ where} \qquad (11)$$

$$Var(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\theta_i - \theta_{mean}\right)^2,$$

and $Var(\theta)$ = the mean of the square deviations between the value (biomass or variance) of each simulation and the mean biomass or variance.

The mean square error was defined as the sum of the bias squared and variance, and the rRMSE was calculated as follows:

$$rRMSE = \frac{\sqrt{Bias^2 + Var}}{\theta_{True}}, \qquad (12)$$

where $Bias$ = the absolute bias as defined by the numerator in Equation 10.

Results

Of the 7 scenarios with different combinations of estimators and sampling designs that were considered in this
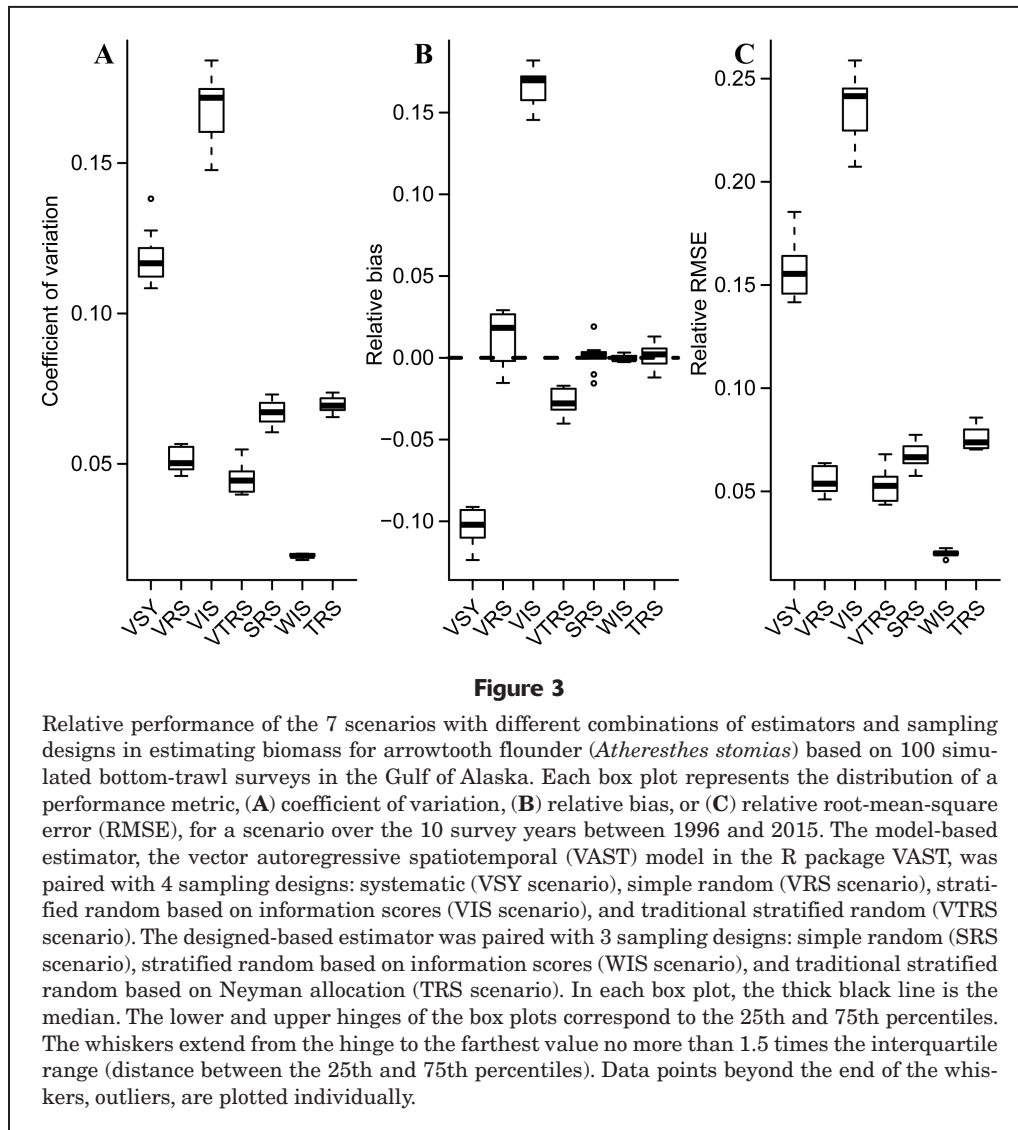
study, the best scenario for estimating biomass and its variance across all performance metrics and all species was the design-based estimator coupled with the stratified-random sampling design in which strata were defined by ISes. This IS scenario consistently had the lowest CV and rRMSE and was generally unbiased (Table 1, Figs. 3–8). In contrast, the scenario that paired the model-based estimator of biomass with this sampling design (VIS scenario) was by far the worst-performing scenario by all 3 performance metrics for all species. It had the highest CV and rRMSE and was the most biased for all species. However, for variance, the VIS scenario performed considerably better, with no bias for 2 of the species and with an rRMSE similar to or smaller than that of the other model-based variance scenarios.

The performance of the combination of the design-based estimator with the traditional stratified-random sampling design (TRS scenario) was generally average for estimates of biomass and its variance among the 7 scenarios. The performance of this scenario was most similar to that of the scenario that pairs the design-based estimator with a simple random sampling design (Figs. 3–8).

The performance of scenarios differed among species. Of the 7 scenarios for estimating biomass, 5 scenarios performed relatively well for arrowtooth flounder (VRS, VTRS, SRS, WIS, and TRS). A sixth scenario (VSY) performed well for Pacific cod in terms of rRMSE (rRMSE <0.10). However, only the WIS had an rRMSE below 0.10 for Pacific ocean perch (Table 1). All of the scenarios with the designed-based estimator of biomass were consistently unbiased (mean relative bias <0.005 for all species), whereas the scenarios with the model-based estimator were generally negatively biased, except the VRS for arrowtooth flounder and the VIS across all species (Figs. 3–5).

Of the 7 scenarios for estimating variance in biomass, only the WIS consistently performed well as measured by rRMSE (rRMSE: 0.06–0.10). Although most of the other scenarios were relatively unbiased for Pacific cod and

Figure 3

Relative performance of the 7 scenarios with different combinations of estimators and sampling designs in estimating biomass for arrowtooth flounder (*Atheresthes stomias*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.
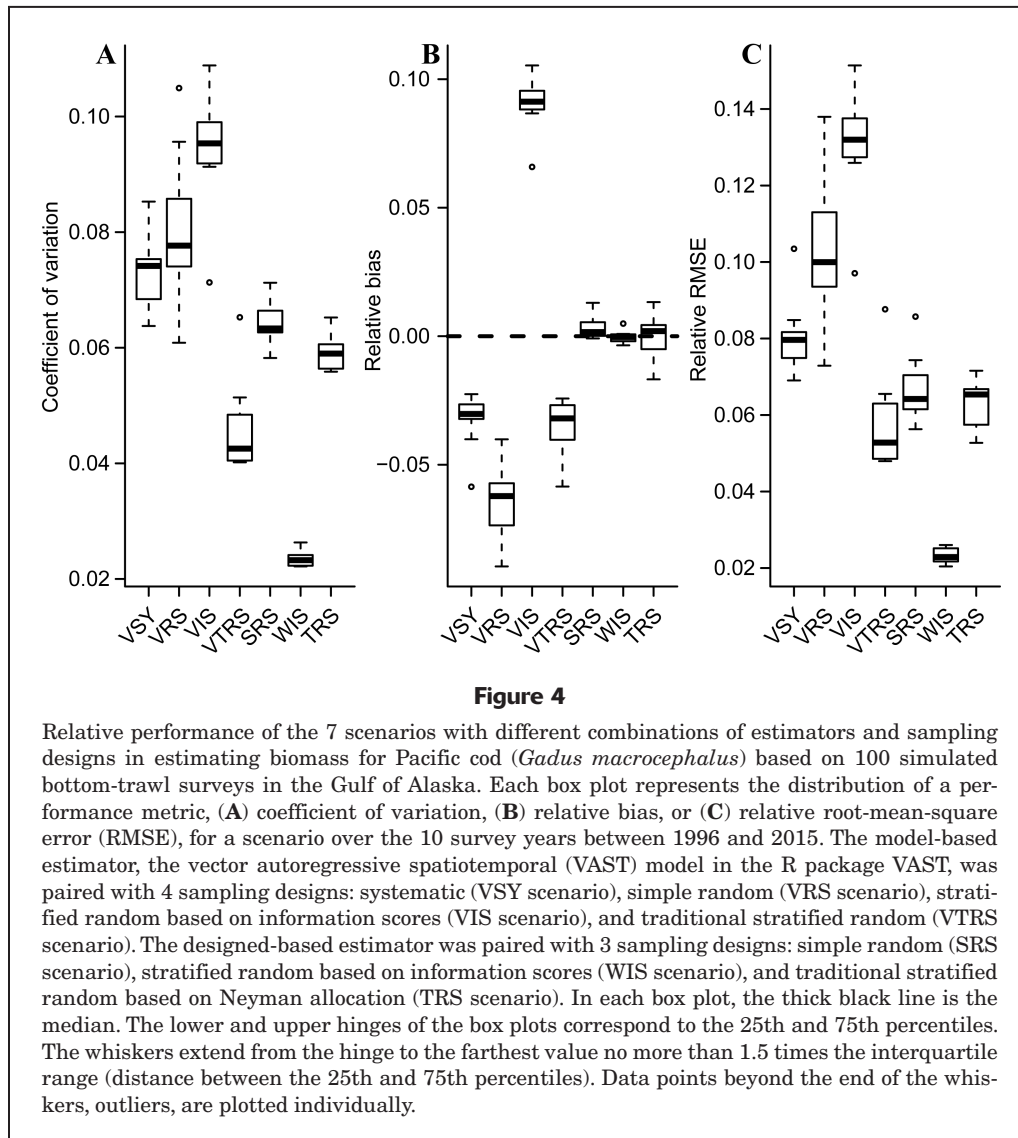
Pacific ocean perch (with VSY being the only exception), they produced relatively large variances, which resulted in relatively large rRMSEs (Figs. 6–8).
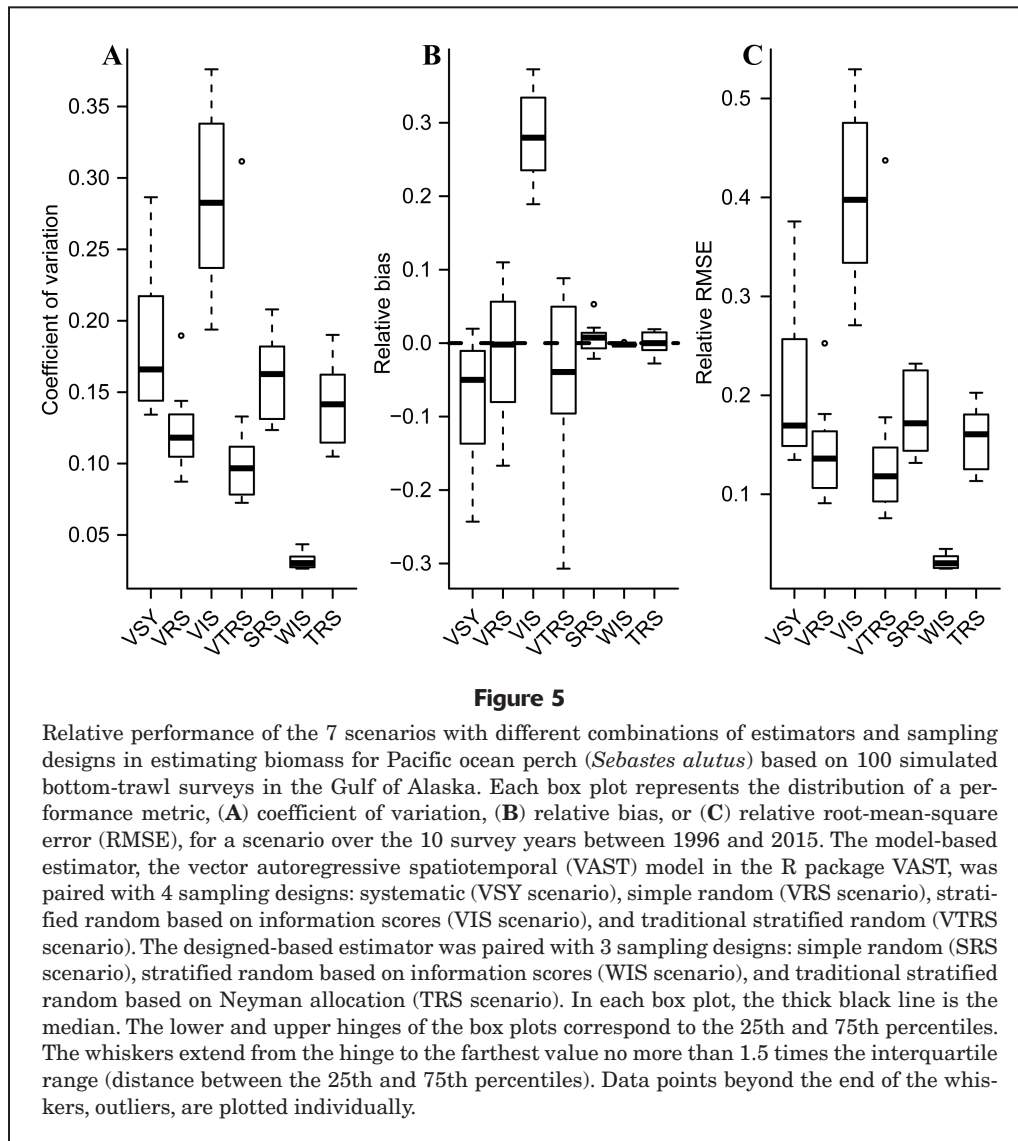
The VTRS scenario performed better in estimating biomass than 2 of the design-based methods (TRS and SRS), as well as all of the other model-based scenarios in terms of CV and rRMSE, but this scenario was somewhat negatively biased (mean relative bias ranged between −0.02 and −0.05). The only scenario with the model-based estimator of biomass that had a comparable rRMSE was the one with the simple random sampling design (VRS scenario), but only for arrowtooth flounder and Pacific ocean perch (Figs. 3–5). The performance of the VTRS in estimating variance in biomass, on the other hand, was generally similar to that of the other methods (except for the poorly performing VSY and the best performing WIS) across all performance metrics (Figs. 6–8).

The VSY scenario for estimating the variance of biomass had extreme values of all performance metrics

for Pacific cod and Pacific ocean perch (CV: 0.64–0.78; mean relative bias: from −0.28 to −0.60; rRMSE: 0.7–1.0; Figs. 7–8) but even more extreme values for arrowtooth flounder (CV and rRMSE >1.0). The VSY scenario for estimating biomass variance for arrowtooth flounder, therefore, was excluded from the boxplots in Figure 6 because it was an extreme outlier, which if included, would have made comparisons of the other scenarios difficult. The poor performance of the VSY scenario for estimating variance was due to the highly skewed variance distributions for all species, with ranges spanning as much as 5 orders of magnitude for arrowtooth flounder (Suppl. Fig. 2).

Apart from the differences in the relative performance of the various scenarios for estimating biomass and variance, differences in the performance of each scenario among species are notable. All of the scenarios performed considerably worse in estimating biomass for Pacific ocean perch than for the other 2 species in terms of rRMSE. The difference in performance across species was small for the

Figure 4

Relative performance of the 7 scenarios with different combinations of estimators and sampling designs in estimating biomass for Pacific cod (*Gadus macrocephalus*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.

Figure 5

Relative performance of the 7 scenarios with different combinations of estimators and sampling designs in estimating biomass for Pacific ocean perch (*Sebastes alutus*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.

top-performing WIS scenario (an rRMSE of 0.03 for Pacific ocean perch versus values of 0.02 and 0.025 for arrowtooth flounder and Pacific cod, respectively). However, the difference was much higher for all of the other scenarios for estimating biomass, with absolute differences in rRMSE between Pacific ocean perch and the other 2 species ranging from 0.01 to 0.16 for arrowtooth flounder and from 0.04 to 0.27 for Pacific cod (Table1). In contrast, all but one of the scenarios for estimating variance of biomass performed best for Pacific cod, with absolute differences in the rRMSE between Pacific cod and the other 2 species ranging from 0.03 to >0.31 for arrowtooth flounder and from 0.01 to 0.29 for Pacific ocean perch. The single exception was the WIS scenario, which had a smaller rRMSE for arrowtooth flounder (0.05 for arrowtooth flounder versus a value of 0.09 for Pacific cod) but a larger rRMSE for Pacific ocean perch (0.10). The differences in relative performance of the scenarios between Pacific ocean perch and arrowtooth flounder were mixed.

Discussion

The objective of this study was to assess the performance of 7 scenarios with different combinations of sampling designs and estimators and to compare their statistical robustness. Although we used 3 different performance metrics to evaluate statistical robustness, for conciseness, we highlight herein the best-performing scenarios on the basis of only rRMSE (those with the lowest rRMSEs are considered the best performing), which includes both the bias and variance components teased apart by the other performance metrics.

The results of the simulation analyses indicate that, compared to the scenario that paired the design-based estimator with the existing sampling design (TRS), 4 scenarios performed better for arrowtooth flounder, 3 scenarios performed the same or better for Pacific cod, and 3 scenarios performed better for Pacific ocean perch. Of these scenarios, only the design-based WIS and the model-based VTRS scenarios performed better than the
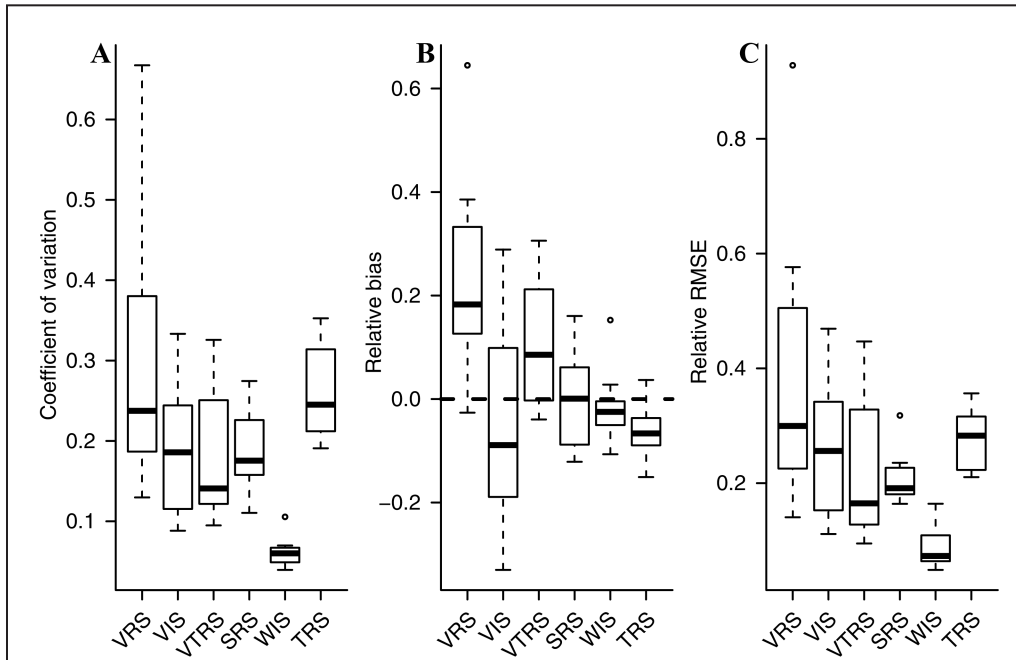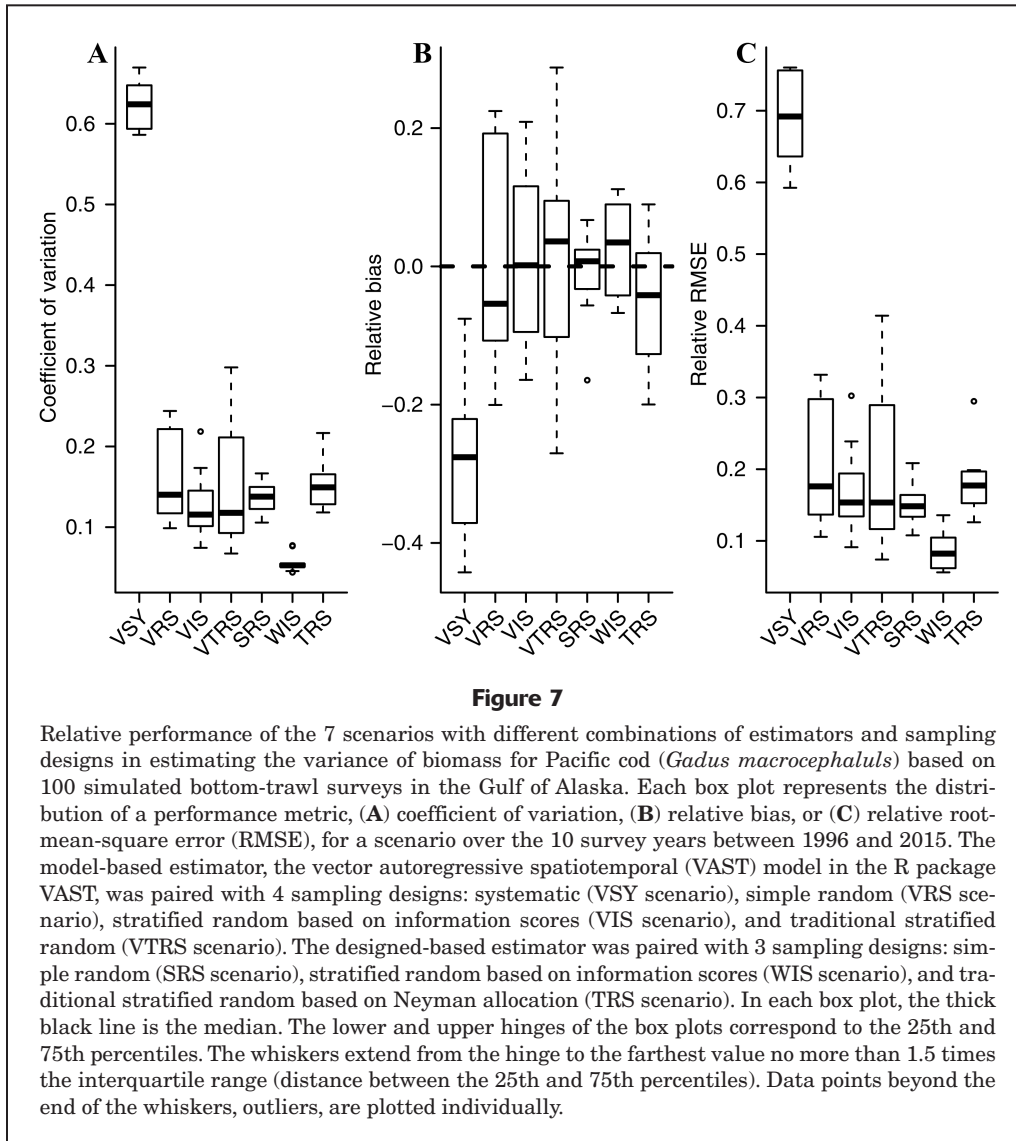
Figure 6

Relative performance of the 7 scenarios with different combinations of estimators and sampling designs in estimating the variance of biomass for arrowtooth flounder (*Atheresthes stomias*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The box plots for the VSY scenario for this species have been omitted for readability of the vertical axis values of the other methods. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.

TRS scenario for all 3 species, with only the WIS scenario consistently performing substantially better than the TRS scenario for estimates of both biomass and its variance for all species (with rRMSEs >50% lower than those for the TRS). The reduction in rRMSE from the WIS scenario to the TRS scenario ranged among species between 58% and 81% for biomass and between 53% and 82% for variance in biomass, with the greatest reduction observed for Pacific ocean perch. It is likely that the traditional sampling design is not as suitable for assessing abundance of Pacific ocean perch because of the patchy distribution of this species in the GOA (Lunsford et al., 2001; Clausen and Fujioka, 2007), a pattern that requires larger sample sizes in areas of high density than in other areas. In summary, our results indicate that the combination of stratified-random sampling with model-based estimation methods is likely to produce the best results.

These results have at least 2 important implications for the redesign of and improvements to the method of abundance estimation for the bottom-trawl survey conducted in the GOA. First, in the short term, estimates of abundance can be improved immediately by using model-based estimators in combination with traditional stratified-random sampling. Second, in the long term, basing stratification on ISes could be an additional change that can improve estimates for species with spatially restricted or patchy distributions, when used in combination with design-based estimators.

Recommendation to develop model-based estimators

Our results indicate clear improvements in estimation for all species when the model-based estimator was used with a stratified-random sampling design. This finding is important because it indicates that some of the losses in

Figure 7

Relative performance of the 7 scenarios with different combinations of estimators and sampling designs in estimating the variance of biomass for Pacific cod (*Gadus macrocephaluls*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.
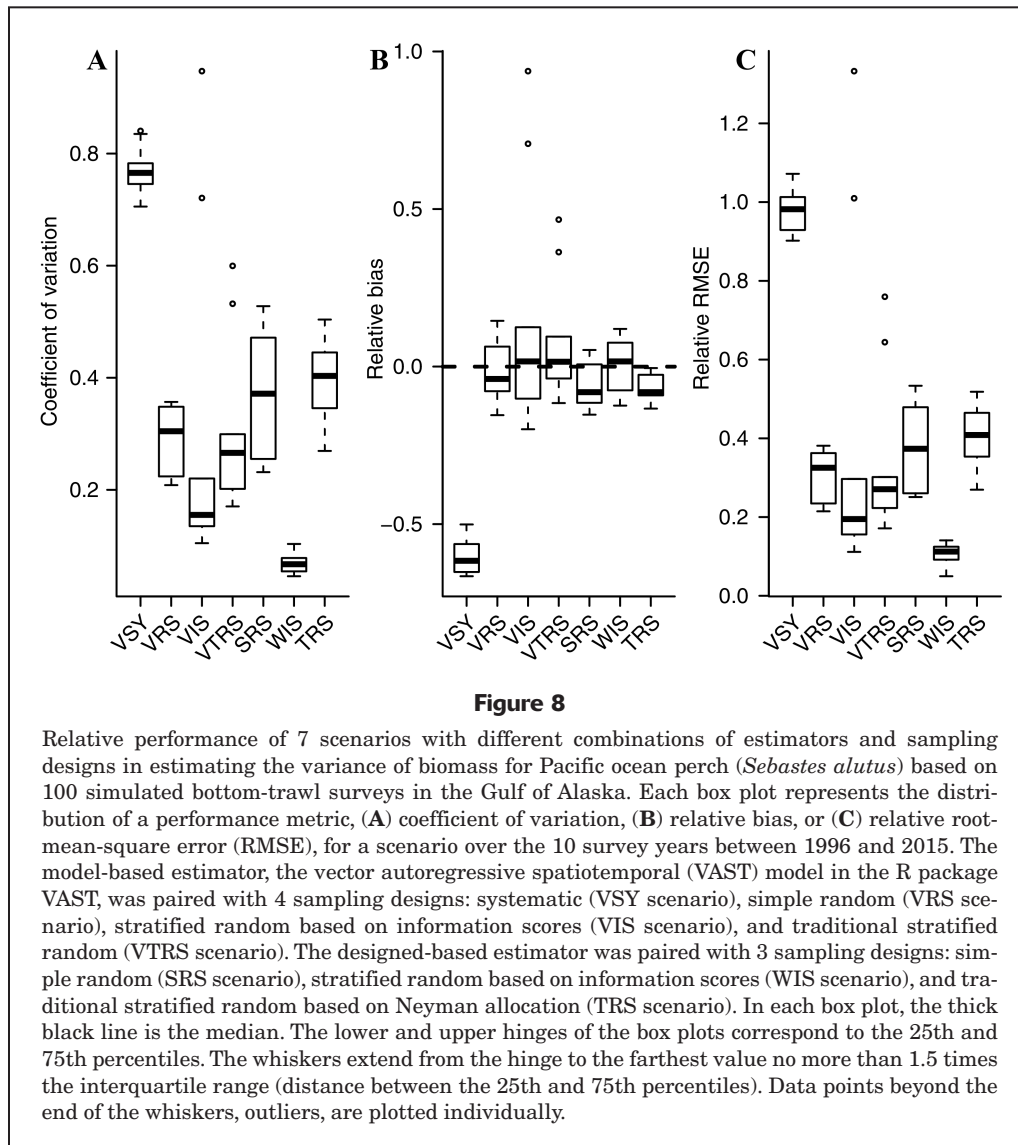
precision expected from design-based estimators in cases with reduced survey effort (Oyafuso et al., 2021) can perhaps be mitigated by the use of spatiotemporal model-based methods, such as those in the VAST package. Similar improvements were also observed when simple random sampling was used with the model-based estimator.

In a related study, in which the effect of sampling density changes on biomass estimates was examined for the same 3 species analyzed here, the model-based estimator when used with the stratified-random sampling design performed better than the design-based estimator for all 3 species at 4 different levels of survey effort (von Szalay et al.[2]). In other studies, the advantages of a spatiotemporal model-based estimator have been demonstrated more broadly. O'Leary et al. (2020) found that the spatial and temporal scope of total biomass for walleye pollock (*G. chalcogrammus*) in the northern Bering Sea was improved in a stock assessment model when a model-based biomass index from multiple surveys was used as model input in combination with the standard design-based estimates. In addition, stock assessment models for dusky rockfish (*S. ciliatus*) and northern rockfish (*S. polyspinis*) in the GOA have incorporated such model-based estimates because they are less noisy than the sampling-design-based CVs (Lunsford et al., 2015; Cunningham et al., 2018).

The improvements we observed with the model-based estimator are likely due to an increase in the effective sample size when data from multiple surveys are used to derive biomass estimates. Thorson and Haltuch (2019) showed that the effective sample size increased by 17% and that the model fit was better, resulting in smaller standard errors of estimated spawning biomass, when a spatiotemporal model-based estimator was used in a stock synthesis assessment

Figure 8

Relative performance of 7 scenarios with different combinations of estimators and sampling designs in estimating the variance of biomass for Pacific ocean perch (*Sebastes alutus*) based on 100 simulated bottom-trawl surveys in the Gulf of Alaska. Each box plot represents the distribution of a performance metric, (**A**) coefficient of variation, (**B**) relative bias, or (**C**) relative root-mean-square error (RMSE), for a scenario over the 10 survey years between 1996 and 2015. The model-based estimator, the vector autoregressive spatiotemporal (VAST) model in the R package VAST, was paired with 4 sampling designs: systematic (VSY scenario), simple random (VRS scenario), stratified random based on information scores (VIS scenario), and traditional stratified random (VTRS scenario). The designed-based estimator was paired with 3 sampling designs: simple random (SRS scenario), stratified random based on information scores (WIS scenario), and traditional stratified random based on Neyman allocation (TRS scenario). In each box plot, the thick black line is the median. The lower and upper hinges of the box plots correspond to the 25th and 75th percentiles. The whiskers extend from the hinge to the farthest value no more than 1.5 times the interquartile range (distance between the 25th and 75th percentiles). Data points beyond the end of the whiskers, outliers, are plotted individually.

model for lingcod (*Ophiodon elongatus*) in the California Current. However, we acknowledge that additional work with more than the 3 species examined in our study is needed before our results can be generalized. It is possible that the VAST estimator does not perform as well for species with a more limited spatial distribution.

Furthermore, both bias and variance tended to be high for the model-based estimator when the sampling design is highly unbalanced. This was the case for all 3 species when the VAST estimation model was paired with the IS-based sampling design (VIS scenario), which concentrated sampling density in areas of high biomass.

Recommendation to develop stratification methods based on information scores

For all 3 species, IS-based stratification in combination with design-based estimation yielded results that were by far better than any other method. The main reason this

method outperformed the traditional design schemes (random and systematic) was that stratification was informed by historical data in the form of ISes, which can be described as estimated expectation of population abundance and its temporal variability. Of course, for this approach to be valid, one must be confident in the predictions of density based on historical data to adequately represent current and perhaps future species distributions. In contrast, the stratum boundaries in traditional random sampling are defined by depth and management area. Using ISes to stratify the survey area resulted in strata with more homogenous variances than the strata in the traditional Neyman allocation. This difference is due to the variance component of the IS, which increases the efficiency of IS-based stratification compared to that of the traditional stratification scheme. Furthermore, in contrast to using Neyman allocation, the use of ISes ensures that samples are representative in areas where density is persistently high and temporal variance is low. In Neyman allocation, strata with high density

but low variance are apportioned a relatively small number of stations. Such allocation is undesirable for fisheries surveys because, in addition to precise estimates of abundance, representative samples for other population measures, such as age and size compositions, are also needed. In the extreme case of the abundance of a stratum being very high but uniform, the Neyman strategy would allocate no samples to that stratum.

In the future, it may be problematic to use historical data to define stratum boundaries under the IS approach if fish distributions change in response to climate change as they are in many regions (Hobday and Evans, 2013; Maureaud et al., 2021; Hollowed et al., 2022). If the shift in distributions is relatively gradual over time, it may still be appropriate to use relatively recent data while excluding older data to define stratum boundaries on the basis of ISes. However, if the shift in distribution patterns is more abrupt, it may not be appropriate to use any sampling design that defines strata on the basis of historical data, as is the case for both the WIS and TRS approaches. In this situation, it may be necessary to conduct a pilot survey prior to the main survey to establish new stratum boundaries.

The finding that the design based on ISes outperformed the traditional sampling design is not surprising given that the stratification was performed individually for each species in the case of the WIS scenario. Therefore, this method would be applicable to only single-species surveys. In contrast, the stratification scheme used for the TRS scenario was first optimized for each species, but because the GOA survey is a multispecies survey, the realized station allocation to each stratum was based on a weighted average of several principal species. This weighting effectively resulted in less than optimal station allocation for each species. The comparison is nevertheless informative because it illustrates how much better the biomass and variance estimates are if we can focus on sampling individual species, given their importance to the fishery. Perhaps it has been a mistake to make so many compromises with the traditional weighting across too many species.

More research is needed to investigate how the relative performance of the WIS scenario compares with the TRS scenario in the context of a multispecies survey, in which the IS-based station allocation scheme would be subjected to the same compromises among several principal species as the traditional stratified-random sampling design. We envision generating multispecies ISes, which are linear combinations of the single-species scores associated with each of the principal species included in the TRS scenrio. Each component of these combinations could be weighted by a number of different factors important to management (e.g., the commercial value of each species, ecological importance of each species, or non-biological elements, such as cultural or political considerations).

Although both the absolute and relative performance of a multispecies version of the WIS would necessarily be lower than the performance of the single-species version, we expect it to perform better than the TRS scenario because the stratum boundaries and station allocation scheme are strictly driven by historical data and

determined by the objective factors of fish density and its variance. Furthermore, it is likely that the IS approach can be improved by optimizing for both the relative weighting of the variance component and the number of strata used. In contrast, in the TRS scenario, strata are defined by ad hoc boundaries and depth of the management area, likely reducing the efficiency of the traditional sampling design. The allocation scheme of the TRS scenario also includes a penalty against strata that are relatively costly to sample in terms of time. This cost penalty may be problematic for estimating biomass and its variance for slope species, such as Pacific ocean perch and sablefish (*Anoplopoma fimbria*), which primarily inhabit these high-cost strata and are also more patchily distributed than species, such as Pacific cod and arrowtooth flounder, that inhabit the continental shelf. As a compromise with the objective to preserve the TRS approach, it may be feasible in the future to introduce a hybrid sampling design in which sampling for some species is done with the traditional design and sampling for other species, such as Pacific ocean perch, may be enhanced by sampling more in areas with high ISes.

The relatively poor performance of the model-based VIS scenario can likely be attributed to the unbalanced sampling design resulting from stratification based on ISes. Under this stratification scheme, in which the distributions of the ISes were highly skewed (Suppl. Fig. 1), the bulk of the stations were assigned to just a few strata while the minimum of 2 required stations were assigned to all of the remaining strata, concentrating the spatial distribution of samples used to extrapolate to the entire survey area. The performance of the VIS in estimating biomass was particularly poor for Pacific ocean perch because this species is primarily confined to a depth range of approximately 150–300 m, which exists only in a relatively narrow band on the upper continental slope of the survey area. Pacific cod and arrowtooth flounder, on the other hand, are more evenly distributed over a larger portion of the survey area (primarily over the relatively wide continental shelf); therefore, the effect of the unbalanced sampling design is less pronounced for them. The IS-based sampling design tailors the station allocation to individual strata, resulting in better performance than the traditional stratification scheme with designed-based estimators because stations are allocated to areas where fish are most likely to be, thereby reducing the variance in biomass estimates while having no effect on bias. However, it does not work well with the model-based estimator, which is sensitive to the highly unbalanced sampling dictated by the IS-based stratification.

Conclusions

The results of our analysis of 6 scenarios with different combinations of sampling designs and estimators, in comparison to the method used for the bottom-trawl survey conducted in the GOA since 1984, not only identify which scenario consistently outperformed the traditional approach but also indicate that the traditional approach of using a stratified-random sampling design with a design-based

estimator is not optimal for any of the 3 species considered in this study. Estimates of abundance can likely be improved immediately by using the model-based estimator in the VAST package in combination with the traditional stratified-random sampling design (VTRS scenario). In the long term, stratification based on the IS proposed in this study, when used with a design-based estimator, could be a useful alternative to the traditional approach in improving estimates of abundance for species, such as Pacific ocean perch, with spatially restrictive or patchy distributions. However, in subsequent work, interspecific tradeoffs in a multispecies context will be required to justify reconsidering the traditional sampling design. Nonetheless, this approach of combining an IS-based sampling design with a design-based Wakabayashi estimator is promising, particularly for single-species surveys or for multispecies surveys in areas with few interspecific differences in distribution.

Resumen

Desde 1984, se han utilizado prospecciones de arrastre de fondo con un diseño de muestreo aleatorio estratificado para la evaluación de poblaciones de especies de importancia comercial en el Golfo de Alaska. Se evaluó un nuevo diseño de muestreo estratificado para determinar si su uso podría mejorar la precisión y exactitud de las estimaciones de abundancia. En el enfoque propuesto para definir los estratos, se utilizaron datos de estudios históricos para generar lo que denominamos puntajes de información (ISes). Para determinar si el enfoque existente es óptimo, se comparó el esquema de estratificación tradicional con el nuevo método y otros 2 diseños de muestreo, utilizando tanto un estimador basado en el diseño como un estimador basado en el modelo con cada diseño. La robustez estadística, medida en términos del coeficiente de variación, sesgo y la raíz del error cuadrático medio, se comparó entre 7 escenarios con diferentes combinaciones de estimadores y diseños de muestreo mediante simulación con un modelo lineal generalizado mixto espaciotemporal condicionado a observaciones históricas de capturas por unidad de esfuerzo de 3 especies. La combinación del estimador basado en el diseño con el esquema de estratificación basado en los ISes fue el mejor escenario en todas las medidas de desempeño para todas las especies. Este escenario consistentemente presentó la varianza más baja y el menor error total, y en general, no estuvo sesgado. En contraste, la comparación del estimador basado en el modelo con este diseño de muestreo fue por mucho, el escenario con peor desempeño. El desempeño del enfoque existente fue promedio.

Acknowledgments

Literature cited

Alverson, D. L., and W. T. Pereyra.
　1969. Demersal fish explorations in the northeastern Pacific Ocean—an evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. J. Fish. Res. Board Can. 26:1985–2001. Crossref

Clausen, D. M., and J. T. Fujioka.
　2007. Variability in trawl survey catches of Pacific ocean perch, shortraker rockfish, and rougheye rockfish in the Gulf of Alaska. *In* Biology, assessment, and management of North Pacific rockfishes. Univ. Alsk. Sea Grant Coll. Program Rep. AK-SG-07-01 (J. Heifetz, J. Dicosimo, A. J. Gharrett, M. S. Love, V. M. O'Connell, and R. D. Stanley, eds.), p. 411–428. Univ. Alsk. Sea Grant Coll. Program, Fairbanks, AK. [Available from website.]

Cochran, W. G.
　1977. Sampling techniques, 3rd ed., 428 p. John Wiley and Sons, New York.

Cunningham, C. J., P.-J. F. Hulson, C. R. Lunsford, and D. H. Hanselman.
　2018. Assessment of the northern rockfish stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska, 87 p. North Pac. Fish. Manag. Counc., Anchorage, AK. [Available from website.]

Hidiroglou, M. A., and M. Kozak.
　2018. Stratification of skewed populations: a comparison of optimisation-based versus approximate methods. Int. Stat. Rev. 86:87–105. Crossref

Hiemstra, P. H., E. J. Pebesma, C. J.W. Twenhöfel, and G. B.M. Heuvelink.
　2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. Comp. Geosci. 35:1711–1721. Crossref

Hobday, A. J., and K. Evans.
　2013. Detecting climate impacts with oceanic fish and fisheries data. Clim. Change 119:49–62. Crossref

Hollowed, A. B., A. C. Haynie, A. J. Hermann, K. K. Holsman, A. E. Punt, and C. S. Szuwalski.
　2022. Implications of climate change on the Bering Sea and other cold water systems. Deep-Sea Res., II 201:105110. Crossref

ICES (International Council for the Exploration of the Sea).
　2020. ICES workshop on unavoidable survey effort reduction (WKUSER). ICES Sci. Rep. vol. 2, issue 72, 92 p. [Available from website.]

Kotwicki, S., and K. Ono.
　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. Crossref

Kotwicki, S., T. W. Buckley, T. Honkalehto, and G. Walters.
　2005. Variation in the distribution of walleye pollock (*Theragra chalcogramma*) with temperature and implications for seasonal migration. Fish. Bull. 103:574–587.

Kotwicki, S., J. K. Horne, A. E. Punt, and J. N. Ianelli.
    2015. Factors affecting the availability of walleye pollock to
        acoustic and bottom trawl survey gear. ICES J. Mar. Sci.
        72:1425–1439. Crossref
Lavrakas, P. J. (ed.).
    2008. Encyclopedia of survey research methods, 2 vols.,
        1072 p. SAGE Publ. Inc., Thousand Oaks, CA.
Lindgren, F., H. Rue, and J. Lindström.
    2011. An explicit link between Gaussian fields and Gaussian
        Markov random fields: the stochastic partial differential
        equation approach. J. R. Stat. Soc., B 73:423–498. Crossref
Lunsford, C., L. Haldorson, J. T. Fujioka, and T. J. Quinn II.
    2001. Distribution patterns and survey design consider-
        ations of Pacific ocean perch (*Sebastes alutus*) in the Gulf
        of Alaska. *In* Spatial processes and management of marine
        populations. Univ. Alsk. Sea Grant Coll. Program Rep.
        AK-SG-01-02 (G. H. Kruse, N. Bez, A. Booth, M. W. Dorn,
        S. Hills, R. N. Lipcius, D. Pelletier, C. Roy, S. J. Smith, and
        D. Witherell, eds.), p. 281–302. Univ. Alsk. Sea Grant Coll.
        Program, Fairbanks, AK. [Available from website.]
Lunsford, C. R., P.-J. F. Hulson, S. K. Shotwell, and D. H. Hanselman.
    2015. Assessment of the dusky rockfish stock in the Gulf
        of Alaska. *In* Stock assessment and fishery evaluation
        report for the groundfish resources of the Gulf of Alaska,
        p. 1013–1102. North Pac. Fish. Manag. Counc., Anchorage,
        AK. [Available from website.]
Maureaud, A. A., R. Frelat, L. Pécuchet, N. Shackell, B. Mérigot,
    M. L. Pinsky, K. Amador, S. C. Anderson, A. Arkhipkin, A. Auber,
    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. Glob. Change Biol. 27:220–236.
        Crossref
O'Leary, C. A., J. T. Thorson, J. N. Ianelli, and S. Kotwicki.
    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. Oceanogr. 29:541–557. Crossref
Oyafuso, Z. S., L. A. K. Barnett, and S. Kotwicki.
    2021. Incorporating spatiotemporal variability in multispe-
        cies survey design optimization addresses trade-offs in
        uncertainty. ICES J. Mar. Sci. 78:1288–1300. Crossref
Oyafuso, Z. S., L. A. K. Barnett, M. C. Siple, and S. Kotwicki.
    2022. A flexible approach to optimizing the Gulf of Alaska
        groundfish bottom trawl survey design for abundance esti-
        mation. NOAA Tech. Memo. NMFS-AFSC-434, 142 p.
Peel, D., M. V. Bravington, N. Kelly, S. N. Wood, and I. Knuckey.
    2013. A model-based approach to designing a fishery-
        independent survey. J. Agric. Biol. Environ. Stat. 18:1–21.
R Core Team.
    2019. R: a language and environment for statistical com-
        puting. R Foundation for Statistical Computing, Vienna,
        Austria. [Available from website, accessed August 2019.]

Rue, H., S. Martino, and N. Chopin.
    2009. Approximate Bayesian inference for latent Gaussian
        models using integrated nested Laplace approximations
        (with discussion). J. R. Stat. Soc., B 71:319–392. Crossref
Stauffer, G.
    2004. NOAA protocols for groundfish bottom trawl surveys
        of the nation's fishery resources. NOAA Tech. Memo.
        NMFS-F/SPO-65, 205 p.
Thorson, J. T.
    2019. Guidance for decisions using the Vector Autoregres-
        sive Spatio-Temporal (VAST) package in stock, ecosystem,
        habitat and climate assessments. Fish. Res. 210:143–161.
        Crossref
Thorson, J. T., and K. Kristensen.
    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.
        Crossref
Thorson, J. T., and L. A. K Barnett.
    2017. Comparing estimates of abundance trends and dis-
        tribution shifts using single- and multispecies models of
        fishes and biogenic habitat. ICES J. Mar. Sci. 75:1311–1321.
        Crossref
Thorson, J. T., and M. A. Haltuch.
    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. Crossref
Thorson, J. T., A. O. Shelton, E. J. Ward, and H. J. Skaug.
    2015. Geostatistical delta-generalized linear mixed models
        improve precision for estimated abundance indices for
        West Coast groundfishes. ICES J. Mar. Sci. 72:1297–1310.
        Crossref
Thorson, J. T., C. J. Cunningham, E. Jorgensen, A. Havron,
    P.-J. F. Hulson, C. C. Monnahan, and P. von Szalay.
    2021. The surprising sensitivity of index scale to delta-model
        assumptions: recommendations for model-based index
        standardization. Fish. Res. 233:105745. Crossref
von Szalay, P. G.
    2015. The impact of sample size reduction on the precision
        of biomass estimates in the Gulf of Alaska. NOAA Tech.
        Memo. NMFS-AFSC-297, 17 p.
von Szalay, P. G., and N. W. Raring.
    2018. Data report: 2017 Gulf of Alaska bottom trawl survey.
        NOAA Tech. Memo. NMFS-AFSC-374, 260 p.
Wakabayashi, K., R. G. Bakkala, and M. S. Alton.
    1985. Methods of the U.S.–Japan demersal trawl surveys.
        In Results of cooperative U.S.–Japan groundfish inves-
        tigations in the Bering Sea during May–August 1979
        (R. G. Bakkala and K. Wakabayashi, eds.), p. 7–29. Int.
        North Pac. Fish. Comm. Bull. 44. [Available from website.]