National Marine
Fisheries Service
NOAA

Fishery Bulletin
established in 1881

Spencer F. Baird
First U.S. Commissioner
of Fisheries and founder
of *Fishery Bulletin*

Abstract—We assessed the effect of survey effort reduction on the accuracy and precision of estimates of abundance for 4 commercially or ecologically important species with differing distributions observed in a bottom-trawl survey conducted in the Gulf of Alaska. Simulations from a spatiotemporal generalized linear mixed model based on historical observations of catch densities were used to evaluate the statistical robustness, measured in terms of coefficient of variation, relative bias, and relative root mean square error, of the abundance estimates and their variances. These metrics were used to compare estimates between the traditional design-based estimator and the alternative estimator, based on a vector autoregressive spatiotemporal model, at 4 different sampling densities, representing 2 historical and 2 theoretical sampling effort levels on either side of the historical range. The recent reduction in the density of survey sampling from 820 to 550 stations had only a modest effect on the performance metrics for both estimators for arrowtooth flounder (*Atheresthes stomias*), Pacific cod (*Gadus macrocephalus*), and Pacific ocean perch (*Sebastes alutus*). However, the effect on the abundance estimates for sablefish (*Anoplopoma fimbria*) was substantial. We attribute this difference in results to the wider depth range utilized by sablefish, which preferentially occupy the relatively under-sampled deep strata (>500 m), and to the truncated survey area at the reduced sampling levels where the deepest strata (>700 m) have been eliminated.

Manuscript submitted 13 March 2024.
Manuscript accepted 16 July 2024.
Fish. Bull. 122:143–161 (2024).
Online publication date: 9 August 2024.
doi: 10.7755/FB.122.4.2

Quantifying the effects of sample size and species distribution on the precision and accuracy of abundance estimates from bottom-trawl surveys in the Gulf of Alaska

Paul G. von Szalay (contact author)[1]
Edward A. Laman[1]
Stan Kotwicki[1]
Lewis A. K. Barnett[1]
Kotaro Ono[2]

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

[1] Resource Assessment and Conservation Engineering Division
Alaska Fisheries Science Center
National Marine Fisheries Service, NOAA
7600 Sand Point Way NE
Seattle, Washington 98115

[2] Institute of Marine Research
P.O. Box 1870 Nordnes
5817 Bergen, Norway

The precision of estimates of abundance based on data from fishery-independent surveys generally increases with sample size, assuming the availability of a species to sampling remains constant. In practice, the number of sampling stations is typically known before a survey design is generated and is dictated by logistical constraints of conducting bottom-trawl surveys, including funding, staffing, time, and the ability to acquire appropriate and sufficient sampling tools and platforms. The goal of a survey design is, therefore, to optimize the geographical distribution of the predetermined number of sampling stations to achieve the highest possible precision and accuracy for survey data products. Furthermore, it is often assumed that sampling effort will remain constant over time when a survey time series is initiated. However, a potential problem for survey continuity arises if circumstances compel alteration of sampling effort because variability in sample size can change the precision of abundance estimates (Cochran, 1977). Large-scale surveys (spanning areas larger than approximately 30,000 km^2), which tend to be relatively long in duration (>1–2 months), are especially prone to unavoidable reductions in survey effort because there is a smaller margin with which to accommodate unforeseen tightening of logistical constraints for them (ICES, 2020).

It is well-known that fisheries resource management agencies, both in the United States and internationally, have had to contend with unavoidable reductions in survey effort for a variety of reasons. These reasons include vessel breakdowns, reduction in survey area due to development of offshore wind farms, failure to acquire a sufficient number of survey vessels or staff, bad weather, and failure to obtain sampling permits in all parts of the survey area, among others (Peel et al., 2013; ICES, 2020). An illustrative example is the change in survey effort made by the NOAA Alaska Fisheries Science Center for the bottom-trawl survey that the organization has conducted during summer in the Gulf of Alaska (GOA) since 1984, using a stratified random survey design to assess the distribution and abundance of groundfish species for fisheries management (von Szalay and Raring, 2018). Traditionally, the survey effort consisted of 3 survey vessels sampling approximately 820 stations from 59 different strata extending from nearshore waters to a depth of 1000 m. However, a variety of factors in recent years, such as funding limitations and difficulties with acquiring 3 suitable charter vessels, have resulted in a reduction of the survey effort to only 2 survey vessels sampling 550 stations from 54 strata extending as deep as only 700 m (von Szalay, 2015). Therefore, the reduction in survey effort also resulted in a reduction in survey area, similar to the consequences of offshore wind farm development and other marine spatial management actions that cause preclusion of sampling (ICES, 2023).

In this study, we assessed the effect of variability in sampling density on estimates of the biomass of species targeted in the GOA bottom-trawl survey. Specifically, we performed a simulation analysis to quantify the effect of changes in sampling density on the accuracy and precision of estimates of biomass, as a measure of abundance, and their associated variances for 4 commercially and ecologically important species that represent diverse geographical distribution patterns: Pacific cod (*Gadus macrocephalus*), Pacific ocean perch (*Sebastes alutus*), arrowtooth flounder (*Atheresthes stomias*), and sablefish (*Anoplopoma fimbria*). The statistical robustness of the biomass estimates was evaluated by using a simulation framework, recently developed by Kotwicki and Ono (2019), in which distributions of fish population density are simulated with a spatiotemporal generalized linear mixed model based on historical catch and environmental survey data. We set up this operating model so that the "true" distribution of abundance and the results of simulated surveys with different sampling densities were represented. Three performance metrics, including coefficient of variance (CV), relative bias, and relative root mean square error (rRMSE), were used to measure the quality of the simulated estimates of biomass and its variance. Using these metrics, we compared estimates from the use of both the traditional design-based biomass estimator (Wakabayashi et al., 1985) and the recently developed estimator based on a vector autoregressive spatiotemporal (VAST) model (Thorson, 2019), across 4 plausible sampling densities, including the 2 levels of sampling effort that have been used in the past and 1 level on each side of the historical range.

Materials and methods

Survey characteristics

The GOA (Fig. 1) forms the northeastern border of the Pacific Ocean and consists of complex bathymetric features ranging from jagged, mountainous pinnacles to flat, muddy areas (Zimmermann et al., 2019). These features provide a variety of habitats in a complex ecosystem. The stratified random sampling design of the standard biennial GOA bottom-trawl survey historically has covered an area of approximately 320,000 km^2 that includes the continental shelf and upper slope from the Islands of Four Mountains east to Dixon Entrance and extends from nearshore waters to a depth of 1000 m (von Szalay and Raring, 2018). In years when survey effort has been unavoidably reduced from the historic 3 charter vessels to 2 vessels, the survey area has been concomitantly reduced to depths shallower than 700 m and to around 30% fewer stations than in those years in which 3 vessels were used. In a bottom-trawl survey with unreduced survey effort, the bulk of survey stations are located on the continental shelf because it accounts for over 80% of the survey area, and the stations on the approximately 20-km-wide continental slope that borders the shelf is sampled only at depths between 200 and 1000 m.

During 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 across 59 strata traditionally have been sampled in years with standard effort (i.e., when 3 vessels were used) and around 550 stations across 54 strata have been sampled in years with reduced effort (when 2 vessels were used). Survey effort in 2001, 2013, and 2017–2023 was reduced because of unavoidable funding limitations as well as challenges in acquisition of 3 qualified charter vessels. Abundance estimates typically have been generated by using the design-based estimator developed by Wakabayashi et al. (1985). The same standard trawl gear has been used for this survey since 1990: 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 begin in the western GOA and proceed east into southeast Alaska. The targeted duration for tows of the trawl net is 15 min at a speed of 1.54 m/s (3 kt). The catch per unit of effort (CPUE), in terms of weight per swept area, typically is estimated by using the area-swept method (e.g., Alverson and Pereyra, 1969), in which the effort is defined as the product of the distance fished and the average distance between wing tips of the trawl gear (von Szalay and Raring, 2018).

A 5-by-5-km grid of sampling units (stations) superimposed on the survey area is used to delineate stations in combination with predetermined stratum designations. Each unique grid cell within a stratum represents a potential survey station. Some stations have been excluded from the pool of stations eligible for survey selection because they are known to be untrawlable with our standard survey trawl gear. Grid cells along the boundaries of the survey area are truncated by the coastline and the deepest edge of the sampling domain and, therefore, have an area smaller than a standard 25-km^2 grid cell.

Following the stratified random sampling design used since the inception of the GOA bottom-trawl survey in 1984, with the survey area divided into 59 strata defined by water depth, bottom terrain (e.g., shelf, gully, and slope),

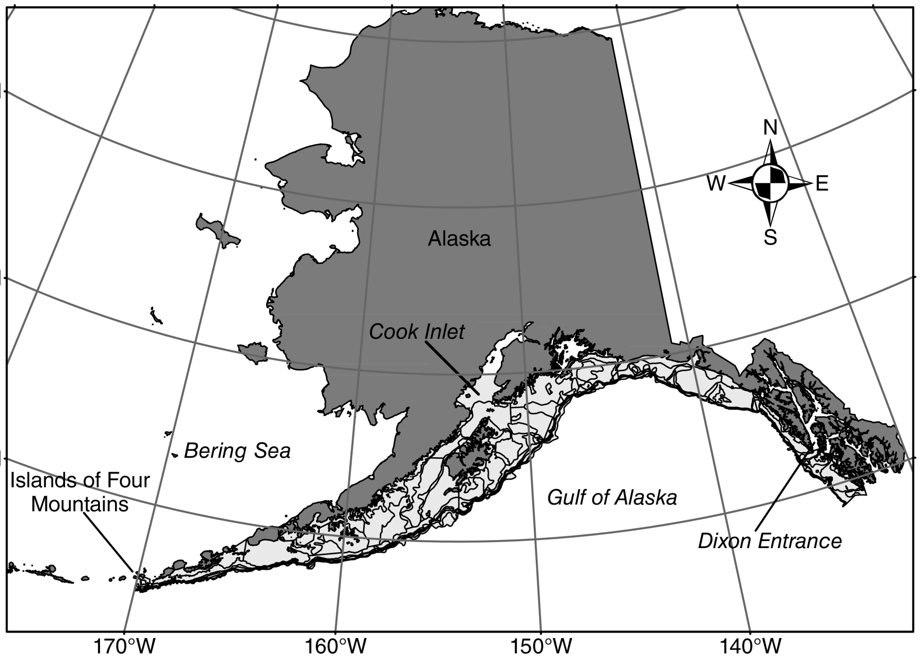

Figure 1

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 total survey area encompasses approximately 320,000 km^2. The stratum boundaries have been fixed since the inception of the survey in 1984.

and statistical management areas (von Szalay and Raring, 2018), we determined the allocation of stations among the strata for species k using a modified Neyman optimal allocation strategy (Cochran, 1977):

$$n_{hk} = \frac{\dfrac{nN_h s_{hk} a_h}{\sqrt{c_h}}}{\sum_h \dfrac{N_h s_{hk} a_h}{\sqrt{c_h}}}, \tag{1}$$

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 tow a trawl in stratum h.

Sampling effort among strata for each of the 10 survey years between 1996 and 2015 and each of 50 principal groundfish species in the GOA was based on stratum variances and their respective areas from the 12 consecutive surveys conducted between 1990 and 2015. A cost variable in terms of the time required to complete a trawl tow in a given stratum was also used to penalize strata that are deeper and for which there is a higher probability of unacceptable gear performance. The mean sample size over the 10 survey years between 1996 and 2015 for each stratum and species was calculated and subsequently weighted by the commercial value of each species to obtain the final allocation assignments. The sample sizes were rounded to whole numbers representing the number of stations allocated to each stratum, with an additional constraint that 2 samples were required for each stratum.

Simulated distributions of catch per unit of effort

Simulated CPUE distributions of Pacific cod, Pacific ocean perch, arrowtooth flounder, and sablefish were created by fitting a spatiotemporal delta model to historical GOA survey data from 1996 to 2015, through the use of the package R-INLA (vers. 18.07.12; available from website, accessed October 2018; Rue et al., 2009) in the statistical computing environment 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 and year t, $\pi_t(s)$, was modeled by using a binomial generalized linear mixed

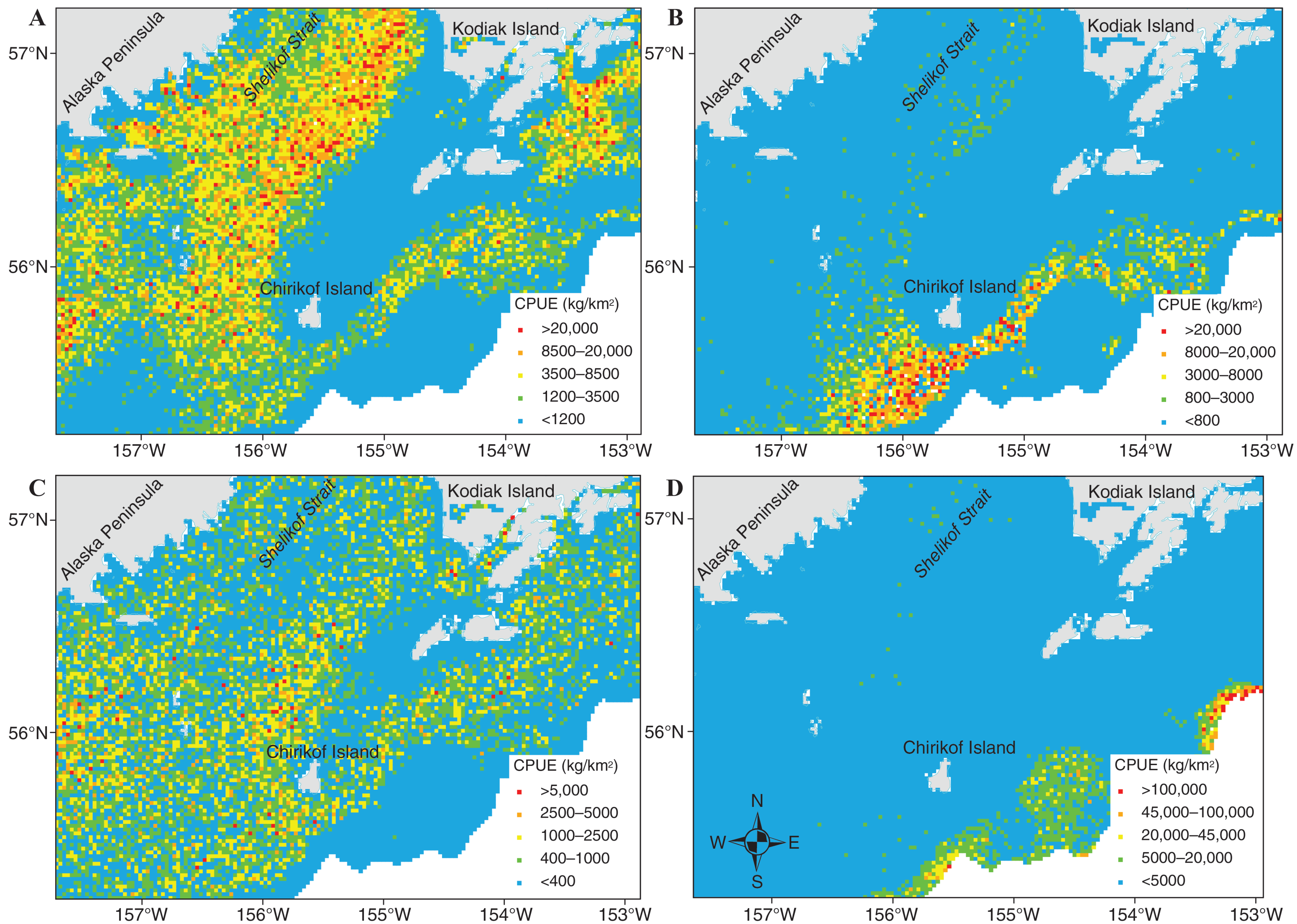

Figure 2

Simulated distributions of catch per unit of effort (CPUE) for (**A**) arrowtooth flounder (*Atheresthes stomias*), (**B**) Pacific ocean perch (*Sebastes alutus*), (**C**) Pacific cod (*Gadus macrocephalus*), and (**D**) sablefish (*Anoplopoma fimbria*) 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 a bottom-trawl survey conducted 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.

model with the logit link function (see Lindgren et al., 2011):

$$\operatorname{logit}\left(\pi_t(s)\right) = X_t(s)b + \omega_t(s), \text{ where} \tag{2}$$

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

and $X_t(s)$ = the matrix of covariates at locations s and year t;
b = the vector of regression coefficients;
ω = the spatiotemporal variation that follows a first-order autoregressive process;
ρ_1 = the degree of autocorrelation 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:

$$\log\left(\mu_t(s)\right) = Z_t(s)a + \delta_t(s), \text{ where} \tag{3}$$

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

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 a first-order autoregressive process;
ρ_2 = autocorrelation of the first-order autoregressive 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 2-km resolution on the basis of a single random sample from the posterior predictive distribution of the models while accounting for the sampling process. 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 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 Markov chain Monte Carlo 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 Markov chain Monte Carlo 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 grid with the finer resolution of 2 km for CPUE distribution was superimposed on the grid with a resolution of 5 km for GOA surveys prior to generation of the simulated surveys. Accordingly, each GOA survey station (i.e., 25-km^2 grid cell) contained between 4 and 7 of the 2-km^2 grid cells with density data. The allocated stations within each stratum were randomly selected without replacement from the 2-km-resolution grid described previously in the "Simulated distributions of catch per unit of effort" section. A total of 100 replicate surveys with the traditional stratified random sampling design were simulated by using the modeled CPUE distributions for each of the 10 survey years between 1996 and 2015. Simulations at 4 different levels of survey effort representing 1, 2, 3, and 4 survey vessels sampling 280, 550, 820, and 1094 stations, respectively, were used as input for 2 different estimators: Wakabayashi's design-based estimator (Wakabayashi et al., 1985) and Thorson's (2019) model-based VAST estimator. All analyses were conducted in R, vers. 3.6.1 (R Core Team, 2019).

The station allocations among strata for the 2-vessel and 3-vessel simulation scenarios were based on the sampling allocation realized during the GOA surveys conducted in 2013 and 2007, respectively. Consequently, the 5 deepest strata (those at depths between 700 and 1000 m) were not sampled in the 1-vessel and 2-vessel simulation scenarios. In contrast, all 59 strata were sampled under the 3-vessel and 4-vessel scenarios. The number of stations allocated by stratum in the 1-vessel scenario was half that in the 2-vessel scenario, subject to rounding and ensuring a minimum of 2 stations per stratum. The station allocation for the 4-vessel scenario was generated by using the Alaska Fisheries Science Center's traditional station allocation program for the GOA survey with an effort level of 1094 stations.

Design-based estimator

The mean CPUE $\left(\bar{x}_{hk}\right)$ and its standard deviation (s_{hk}) of stratum h for species k was calculated and used to generate the survey mean CPUE $\left(\overline{X}_k\right)$ and its variance $\left(S^2_{\bar{x}k}\right)$ for a stratified random sampling design according to Wakabayashi et al. (1985):

$$\overline{X}_k = \frac{1}{N}\sum_{h=1}^{H_r} N_h \bar{x}_{hk}, \text{ and} \tag{4}$$

$$S^2_{x_k} = \sum_{h=1}^{H_t} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right)\left(\frac{s^2_{hk}}{n_h}\right), \quad (5)$$

where N = the total number of 5-by-5-km grid cells in the survey area;
N_h = the number of 5-by-5-km grid cells in stratum h;
H_r = the 55 strata for the surveys with reduced survey effort and 1 or 2 vessels; and
H_t = the 59 strata for the traditional surveys with full survey effort and 3 or 4 vessels.

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:

$$\sigma^2_T = \frac{\Sigma_S \left(B_S - B_T\right)^2}{N}, \quad (6)$$

where σ^2_T = the "true" variance of the abundance index based on 100 simulated surveys;
N = the total number of simulated surveys (100);
Σ_S = the sum over the 100 simulated surveys;
B_S = the estimated biomass realized in simulation survey s; and
B_T = the "true" biomass estimated from simulated density maps.

Model-based estimator

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 R package VAST, vers. 3.3.0 (Thorson, 2019). In a report on 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. A range of assumptions to determine which provided the best fit and accuracy were tested in a previous study (von Szalay et al., 2023), in which alternate observation error distributions for positive catch rates were compared. On the basis of the results of that study (Suppl. Table), we used a delta-gamma distribution for the observation error for positive catch rates and applied the epsilon correction for retransformation bias in the VAST package (Thorson and Kristensen, 2016).

The delta models included *year* as a fixed effect, a spatial random field for each model component (encounter probability and positive catch rate), and independent spatiotemporal random fields for the positive catch rate component. The spatiotemporal term for encounter probability was not estimated because of a 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 by the operating model. Anisotropy was estimated.

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 August 2022).

Performance metrics

Three metrics were used to evaluate the relative performance of the design-based and model-based abundance estimators at the 4 survey effort levels: 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. Bias of 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) of these estimates was calculated within each year and species as follows:

$$RB = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\theta_i - \theta_{True}\right)}{\theta_{True}}, \quad (7)$$

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:

$$CV = \frac{\sqrt{Var(\theta)}}{\theta_{True}}, \text{ where} \quad (8)$$

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

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:

$$rRMSE = \frac{\sqrt{Bias^2 + Var}}{\theta_{True}}, \quad (9)$$

where $Bias$ = the absolute bias as defined by the numerator of Equation 7.

These performance metrics were used as measures of quality of the estimates of biomass and associated variance derived from the simulations. Thresholds were established according to the recommendations by the second Workshop on Unavoidable Survey Effort Reduction (ICES, 2023) to place the values of the performance metrics into 1 of 3 categories:

– desirable and reliable estimate of biomass and variance (metric value <0.20);
– acceptable estimate of biomass and variance (metric value of 0.20–0.40); and
– poor and unreliable estimate of biomass and variance (metric value >0.40).

These 3 categories were, in turn, used to evaluate how estimates of biomass or variance changed with sample size (e.g., from acceptable to desirable or from poor to desirable). For reference, a CV of 0.50 indicates that the bounds of the 95% confidence interval for an estimate of either biomass or variance are ±100%, clearly indicating an unacceptable level of uncertainty.

Results

Realized survey effort reduction from 3 vessels to 2 vessels

Results differed qualitatively between the deep-dwelling sablefish and the other 3 species analyzed, as estimates of the abundance of sablefish were more sensitive to changes in sampling effort as well as to concomitant changes in sampling area. The effect of the reduction in survey effort from 3 vessels (820 stations) to 2 vessels (550 stations) on all 3 performance metrics for biomass estimates was small for both the design-based and model-based estimators for arrowtooth flounder, Pacific cod, and Pacific ocean perch (Figs. 3–5). For these 3 species, all 3 performance metrics remained well within the desirable range for both estimators (<0.20), given that the CV increased only modestly from already low levels (0.060–0.150) to slightly higher levels (0.070–0.160) for the design-based estimator and that the CV was essentially unchanged (change of <0.002) for the model-based estimator. Similarly, the increase in relative bias was modest, as it rose from already low levels (0.000–0.050) to only slightly higher levels (0.003–0.040) for both estimators and all 3 species (Table 1).

For the design-based estimator, the effect of the change in survey effort on the CV and relative bias of the estimates of variance for arrowtooth flounder, Pacific cod, and Pacific ocean perch was generally similar to the effect on those same metrics of the biomass estimates, with only modest increases in both metrics when the survey effort was reduced (Figs. 6–8). The median CVs of the variance estimates from the design-based estimator increased modestly

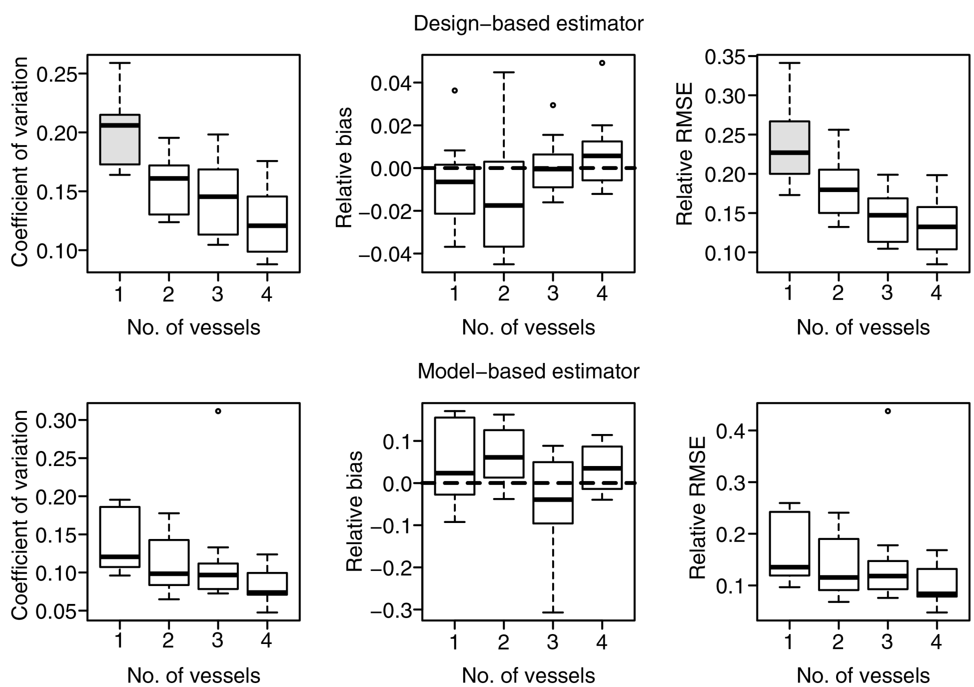

Figure 3

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of biomass of Pacific ocean perch (*Sebastes alutus*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The white and light gray colors of individual boxes indicate that performance metric values are within the ranges at which estimates are classified as desirable (<0.20) and acceptable (0.20–0.40), respectively.

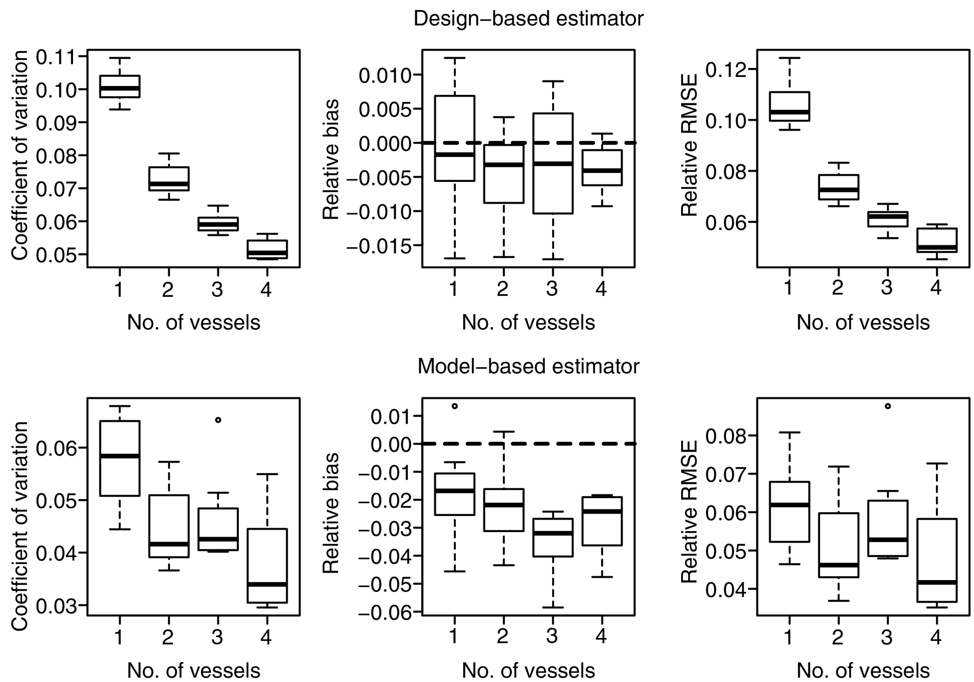

Figure 4

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of biomass of Pacific cod (*Gadus macrocephalus*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. All metric values in this figure are below the threshold at which estimates are classified as desirable (<0.20).

from 0.15 to 0.17 for Pacific cod and from 0.23 to 0.25 for arrowtooth flounder but decreased from 0.45 to 0.39 for Pacific ocean perch, resulting in estimates of variance categorized as either desirable (Pacific cod) or acceptable for all of these species. The relative bias, which was comparatively low (<0.02) for variance estimates for arrowtooth flounder and Pacific cod and was −0.10 for Pacific ocean perch, was unchanged for estimates for Pacific cod and Pacific ocean perch but increased slightly (0.00 to −0.01) for estimates for arrowtooth flounder.

The CVs for variance estimates from the model-based estimator at the 3-vessel effort level were considerably higher for arrowtooth flounder (0.10) and Pacific cod (0.11) than the corresponding CVs of the biomass estimates for these species (0.044 and 0.043, respectively). Only for Pacific ocean perch was the CV of the variance estimate in a lower quality category than that for the biomass estimate with this estimator: the variance estimate was deemed acceptable with a CV of 0.28, and the biomass estimate was categorized as desirable with a CV of 0.100. The CV of variance estimates increased modestly (from 0.28 to 0.32) for Pacific ocean perch, decreased modestly (from 0.11 to 0.09) for Pacific cod, and was unchanged for arrowtooth flounder, but the quality category did not change for any of the species. The relative bias for variance estimates from the model-based estimator was relatively low (0.00–0.12) and generally comparable to the corresponding values for the biomass estimates from this estimator. It decreased to 0.00 from the already low level of 0.03 for Pacific cod, increased from 0.00 to 0.12 for Pacific ocean perch, and was unchanged at 0.09 for arrowtooth flounder (Table 1).

Unlike the comparatively modest effects observed for arrowtooth flounder, Pacific cod, and Pacific ocean perch, the effect on the CV and relative bias of biomass estimates observed for sablefish was unexpectedly high under this survey effort reduction scenario (Fig. 9). Although the CV of the biomass estimate for sablefish was relatively constant between the 2 effort levels for the design-based estimator (0.200 versus 0.210), the relative bias of the biomass estimate changed from a negligible value at the 3-vessel effort level to a large and negative value (−0.550, low enough to categorize the biomass estimate as poor) at the

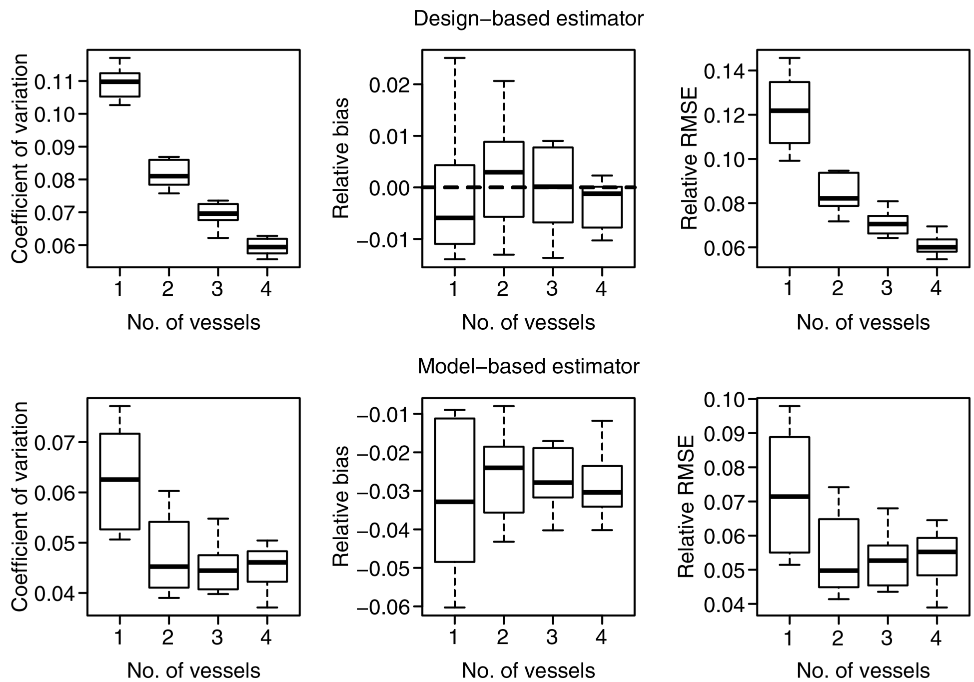

Figure 5

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of biomass of arrowtooth flounder (*Atheresthes stomias*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. All metric values in this figure are below the threshold at which estimates are classified as desirable (<0.20).

2-vessel effort level. For the model-based estimator, the CV (0.600–1.300) and relative bias (0.450–0.750) were high, and meant the estimates were in the poor category, at both survey effort levels (Table 1). Although the effect of the effort reduction on the CV and relative bias of variance estimates was also relatively high for sablefish, the values for these performance metrics already put the variance estimates in the poor category at the 3-vessel effort level and the classification remained there at the 2-vessel effort level for both estimators (Fig. 10, Table 1).

Hypothetical survey effort reduction from 4 vessels to 1 vessel

As expected, the effect on the CV was more pronounced between the 2 hypothetical survey effort levels—1 vessel (280 stations) and 4 vessels (1094 stations)—than it was for the realized survey reduction from 3 vessels to 2 vessels, especially for estimates from the design-based estimator. The median CV of the biomass estimates for arrowtooth flounder, Pacific cod, and Pacific ocean perch from this estimator increased from a range of 0.050–0.120 to a higher and slightly wider range of 0.100–0.210 when the survey effort was reduced from 4 vessels to 1 vessel. However, even in this most extreme scenario for reduction in survey effort, the CV remained in the range that indicates desirable estimates for 2 of the 3 species (arrowtooth flounder and Pacific cod) and only slightly exceeded the range for desirable estimates for 1 species (Pacific ocean perch). The increase in the median CV of biomass estimates from the model-based estimator was even more modest, going from already low values and a relatively narrow range (0.030–0.090) to only slightly higher values and a wider range (0.058–0.120), with all values within the range that indicates desirable estimates at both effort levels. Similar to the values for the 2-vessel and 3-vessel effort levels, the relative bias was small and well within the range for desirable estimates (<0.03) for both estimators and all 3 species at both the 1-vessel and 4-vessel effort levels (Table 1).

The effect on the median CV and relative bias of the variance estimates from the design-based estimator for arrowtooth flounder, Pacific cod, and Pacific ocean perch

Table 1

Comparison of mean values of 3 performance metrics for the traditional design-based estimator and the alternative estimator, based on a vector autoregressive spatiotemporal model, used to estimate biomass and its variance for arrowtooth flounder (*Atheresthes stomias*) (ATF), Pacific cod (*Gadus macrocephalus*) (COD), Pacific ocean perch (*Sebastes alutus*) (POP), and sablefish (*Anoplopoma fimbria*). The values of metrics, coefficient of variation (CV), relative bias, and relative root mean square error (rRMSE), are compared across 4 levels of survey effort. Data used in analysis are from the bottom-trawl surveys conducted in the Gulf of Alaska during 1996–2015 by the NOAA Alaska Fisheries Science Center. The lowest rRMSE value for each species is indicated with an asterisk (*).

		Survey effort: no. of stations (no. of vessels)							
		Design-based estimator				Model-based estimator			
Species	Performance metric	280 (1)	550 (2)	820 (3)	1094 (4)	280 (1)	550 (2)	820 (3)	1094 (4)
Estimates of biomass									
ATF	CV	0.110	0.080	0.070	0.060	0.063	0.044	0.044	0.045
	Rel. bias	−0.007	0.003	0.000	−0.001	−0.032	−0.025	−0.030	−0.031
	rRMSE	0.120	0.080	0.070	0.060	0.070	0.049*	0.052	0.054
COD	CV	0.100	0.072	0.059	0.050	0.058	0.042	0.043	0.033
	Rel. bias	−0.002	−0.003	−0.003	−0.004	−0.016	−0.021	−0.032	−0.025
	rRMSE	0.100	0.072	0.060	0.050	0.062	0.046	0.052	0.041*
POP	CV	0.210	0.160	0.150	0.120	0.120	0.100	0.100	0.090
	Rel. bias	−0.005	−0.020	0.000	0.005	0.010	0.040	−0.050	0.030
	rRMSE	0.220	0.170	0.150	0.120	0.130	0.100	0.100	0.090*
Sablefish	CV	0.230	0.210	0.200	0.230	0.550	1.300	0.600	1.900
	Rel. bias	−0.500	−0.550	0.000	0.030	0.100	0.750	0.450	1.400
	rRMSE	0.600	0.580	0.200*	0.240	0.570	1.550	0.750	2.400
Estimates of variance in biomass									
ATF	CV	0.33	0.25	0.23	0.21	0.30	0.10	0.10	0.75
	Rel. bias	−0.07	−0.01	0.00	0.00	0.24	0.09	0.09	−0.18
	rRMSE	0.35	0.26	0.23	0.22	0.39	0.19*	0.19*	0.79
COD	CV	0.22	0.17	0.15	0.13	0.18	0.09	0.11	0.13
	Rel. bias	−0.03	−0.02	−0.02	0.00	−0.02	0.00	0.03	0.10
	rRMSE	0.23	0.18	0.15	0.13	0.21	0.09*	0.13	0.18
POP	CV	0.51	0.39	0.45	0.35	0.34	0.32	0.28	0.30
	Rel. bias	−0.12	−0.10	−0.10	−0.04	0.00	0.12	0.00	0.20
	rRMSE	0.52	0.39	0.45	0.35	0.34	0.36	0.28*	0.35
Sablefish	CV	0.89	0.89	0.71	0.81	1.50	1.52	1.15	1.80
	Rel. bias	−0.50	−0.48	−0.30	−0.42	−0.55	−0.56	−0.47	−0.57
	rRMSE	1.02	0.96	0.75*	0.88	1.60	1.65	1.25	1.85

was also considerably greater between the 2 hypothetical survey effort levels than between the 2 realized survey effort levels (Figs. 6–8). As expected, the increases in both the CV and the relative bias of variance estimates from the design-based estimator were consistent and substantially greater for all of these species under the hypothetical reduction scenario than under the reduction in survey effort from 3 vessels to 2 vessels. The CV of variance estimates increased from a value indicating desirable estimates (0.13) to a value indicating acceptable estimates (0.22) for Pacific cod, remained at levels indicating acceptable estimates (0.21 and 0.33) for arrowtooth flounder, and increased from a value indicating acceptable estimates (0.35) to a value indicating poor estimates (0.51) for Pacific ocean perch. The relative bias in variance estimates from the design-based estimator increased from levels that indicate that estimates were well within the desirable category (0.00–0.04) to considerably higher levels (0.03–0.12) (Table 1).

The effect of this effort reduction scenario on the CV and relative bias of variance estimates, however, was inconsistent and not always intuitive for the model-based estimator. The CV of variance estimates increased modestly for Pacific ocean perch (from 0.30 to 0.34) and Pacific cod (from 0.13 to 0.18); therefore, the category for the quality of estimates did not change for either of these species. In contrast, the CV unexpectedly decreased from 0.75 to 0.30 for arrowtooth flounder because of an outlier value at the 4-vessel effort level (with the outlier excluded, the CV is 0.10 at this effort level), resulting in an

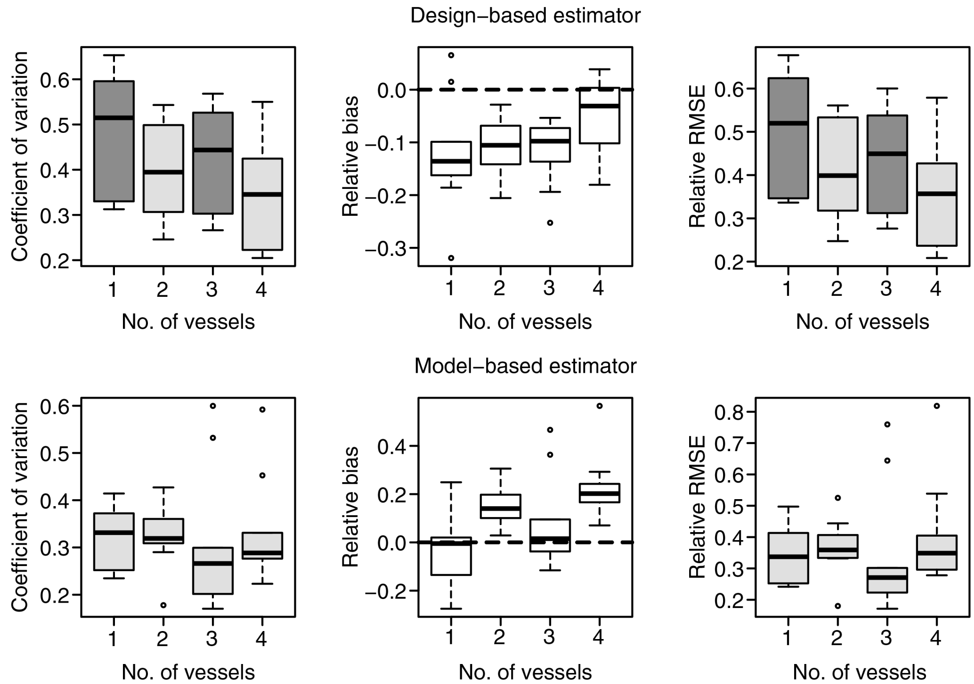

Figure 6

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of variance in biomass of Pacific ocean perch (*Sebastes alutus*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The white, light gray, and dark gray colors of individual boxes indicate that metric values are within the ranges at which estimates are classified as desirable (<0.20), acceptable (0.20–0.40), and unreliable (>0.40), respectively.

improvement in the quality of variance estimates from the poor to the acceptable category (Table 1). Furthermore, relative bias increased for arrowtooth flounder (from 0.18 to 0.24), resulting in a change in the categorization of estimates from desirable to acceptable, but relative bias decreased for both Pacific cod (from 0.10 to 0.02) and Pacific ocean perch (from 0.20 to 0.00), with estimates for both species at both effort levels considered desirable.

For sablefish, no effect on the CV of biomass estimates from the design-based estimator is discernable under this effort reduction scenario, with the same value (0.230), which indicates acceptable estimates, found for both effort levels. However, as with the scenario in which survey effort was reduced from 3 vessels to 2 vessels, relative bias was very low and indicates that the quality of biomass estimates is desirable at the high (4-vessel) effort level (0.030) but large and negative (–0.500, low enough to categorize the estimates as poor) at the low (1-vessel) level. For estimates of both biomass and variance from the model-based estimator, CV (0.55–1.90) and relative bias (0.10–1.40) were both generally high and put estimates in the poor category at both survey effort levels except for one outlier value for relative bias of biomass estimates (0.100) at the 1-vessel level (Table 1).

Trends in total error across all survey effort levels

As expected, the total error of the biomass estimates for the design-based estimator, as measured by using the rRMSE, consistently decreased with increasing sampling density for arrowtooth flounder, Pacific cod, and Pacific ocean perch, although the quality of estimates remained in the desirable category for all species and at all effort levels, except for Pacific ocean perch at the 1-vessel effort level (0.220). Because the CVs responded more to differences in effort level than the relative bias values, the CV was the primary source of variation in the rRMSE with effort level. The magnitude of the decline in the rRMSE

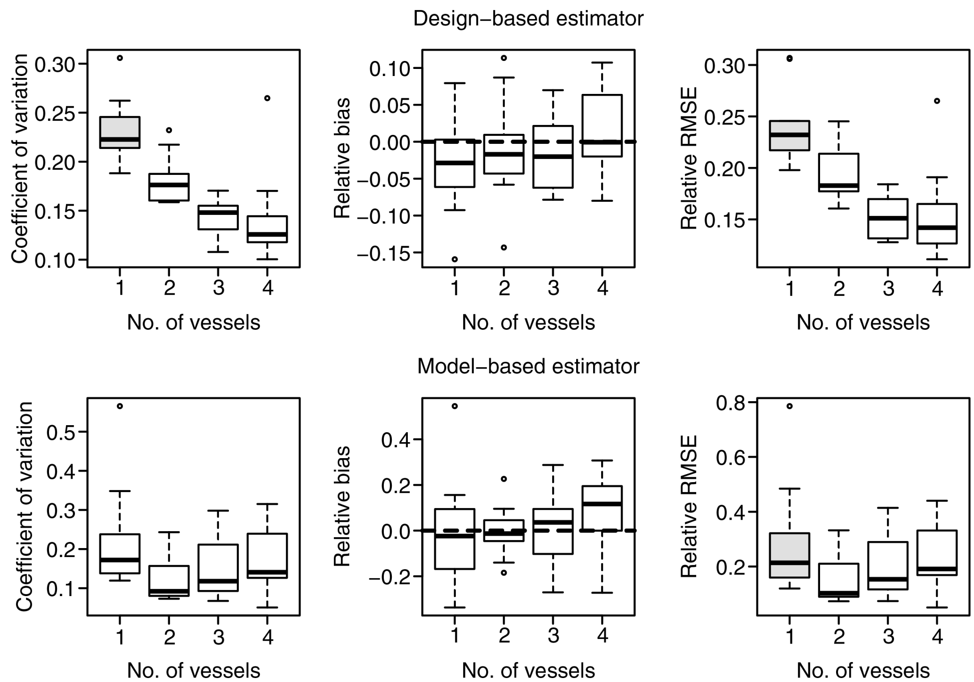

Figure 7

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of variance in biomass of Pacific cod (*Gadus macrocephalus*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The white and light gray colors of individual boxes indicate that metric values are within the ranges at which estimates are classified as desirable (<0.20) and acceptable (0.20–0.40), respectively.

between the 1-vessel and 4-vessel effort levels was relatively similar for all 3 species, ranging from 45% (from 0.220 to 0.120) for Pacific ocean perch to 50% (from 0.100 to 0.050) for Pacific cod (Figs. 3–5). For sablefish, the rRMSE was similar and indicates poor estimates for the 1-vessel and 2-vessel effort levels (0.580–0.600). It reached a minimum at the borderline of the desirable and acceptable categories for estimate quality (0.200) for the 3-vessel effort level and was at an intermediate value (0.240) that indicates acceptable estimates for the 4-vessel level (Table 1). The lack of a decreasing trend in the rRMSE with increasing sampling density for sablefish was due to a large and abrupt increase in relative bias for effort levels with fewer than 3 vessels and to the nonintuitive increase in CV at the 4-vessel effort level (Fig. 9).

The rRMSE of the variance estimates from the design-based estimator consistently decreased with increasing sampling density for arrowtooth flounder and Pacific cod but not for Pacific ocean perch and sablefish (Figs. 6–8 and 10). For Pacific ocean perch, the decreasing trend was interrupted slightly between the 2-vessel and 3-vessel levels but resumed between the effort levels with 3 vessels and 4 vessels, reaching a minimum at the 4-vessel level. For sablefish, the pattern of the rRMSE of the variance estimates from the design-based estimator was similar to that for the corresponding biomass estimates but indicates that variance estimates are firmly in the poor category at all effort levels (0.75–1.02).

Although the rRMSE of biomass estimates from the model-based estimator did not decline as consistently with increasing sampling density for arrowtooth flounder, Pacific cod, and Pacific ocean perch as the rRMSE of estimates from the design-based estimator, the magnitudes of the rRMSE were consistently lower for the model-based estimator than for the design-based estimator at all effort levels, indicating that the model-based estimator was generally the better performing estimator (Table 1, Figs. 3–5). However, this trend was not found for sablefish, for which the magnitudes of the rRMSE of biomass estimates were substantially higher

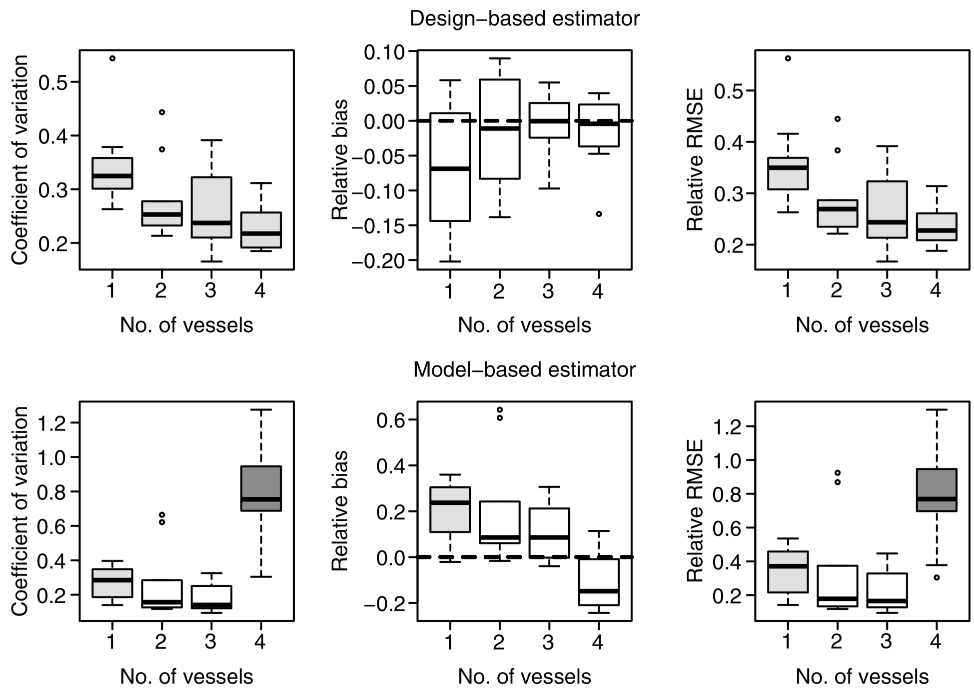

Figure 8

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of variance in biomass of arrowtooth flounder (*Atheresthes stomias*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The white, light gray, and dark gray colors of individual boxes indicate that metric values are within the ranges at which estimates are classified as desirable (<0.20), acceptable (0.20–0.40), and unreliable (>0.40), respectively.

for the model-based estimator at all but 1 of the 4 effort levels (the 1-vessel level) (Table 1). Also, as with the design-based estimator, there was no decline in the rRMSE with sampling density for the model-based estimator for this species. Both estimators performed poorly for sablefish compared to their performances for the other 3 species (Fig. 9).

The rRMSE values of variance estimates were somewhat similar and consistently indicate that the quality of estimates from the design-based and model-based estimators is desirable or acceptable for the arrowtooth flounder and Pacific cod, except at the 4-vessel effort level for arrowtooth flounder, where the value for the model-based estimator unexpectedly spiked to its highest level (0.79) (Figs. 7–8). For Pacific ocean perch, the rRMSE of variance estimates from the model-based estimator was smaller than that of estimates from the design-based estimator at all sampling densities, but there was no clear trend in the rRMSE with increasing sampling density for either estimator (Fig. 6). The rRMSE values for estimates for Pacific ocean perch at all survey effort levels consistently indicate not only that estimates from the design-based estimator range from barely acceptable (0.35–0.39) to poor (0.45–0.52) but also that estimates from the model-based estimator are acceptable (0.28–0.36). Although the design-based estimator consistently outperformed the model-based estimator at all effort levels for sablefish, the rRMSE values were unacceptably high for both estimators, indicating that neither estimator performed well for this species (Table 1, Fig. 10).

Discussion

The recent reduction of sampling density from 820 to 550 stations for the bottom-trawl survey conducted in the GOA (von Szalay, 2015) was expected to only modestly affect the 3 performance metrics for design-based and

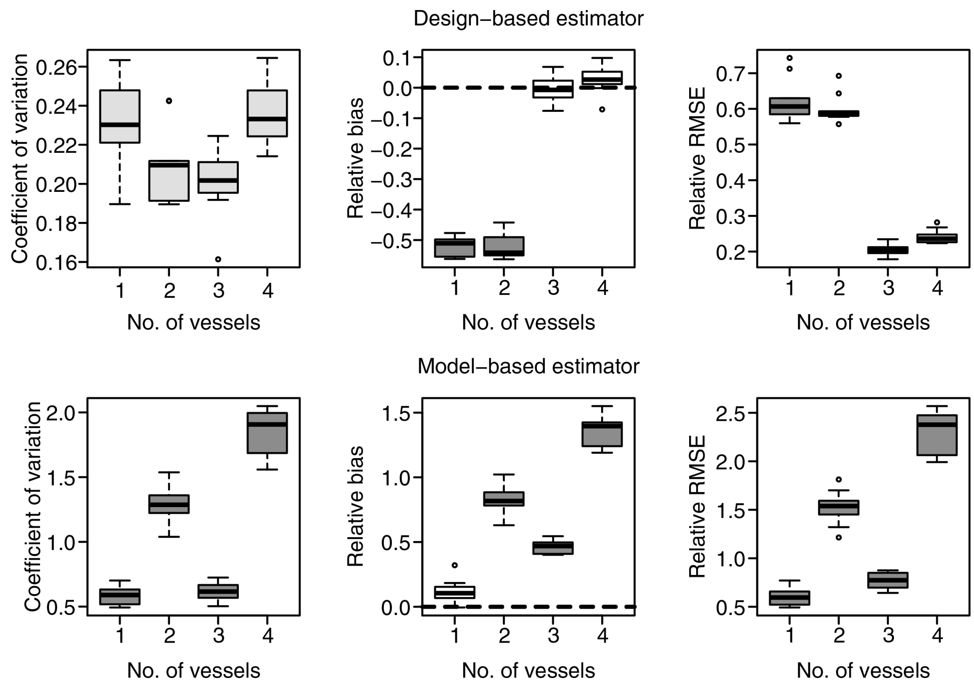

Figure 9

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of biomass of sablefish (*Anoplopoma fimbria*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The white, light gray, and dark gray colors of individual boxes indicate that metric values are within the ranges at which estimates are classified as desirable (<0.20), acceptable (0.20–0.40), and unreliable (>0.40), respectively.

model-based estimators of biomass and variance, and the results of this study support this hypothesis for 3 of the 4 species examined: arrowtooth flounder, Pacific cod, and Pacific ocean perch. This outcome is consistent with the findings of similar simulation studies on sampling effort (Xu et al., 2015; Meng et al., 2019). For example, Xu et al. (2015) analyzed the relationship between performance metrics and effort levels in a simulation study to optimize sampling effort for a survey with a stratified random design, and they achieved relatively high precision and accuracy of abundance indices for most species at a broad range of effort levels, indicating that the quality of abundance indices can be robust despite changes in effort levels. However, the effect on the performance metrics for sablefish was substantial (rRMSE changed from a value indicating that the quality of estimates is desirable [0.200] to a value indicating that the quality is poor [0.580]); in particular, the relative bias for biomass estimates from the design-based estimator changed from unbiased (0.000) to highly biased (−0.550).

Effect on estimates for sablefish

The disproportionate effect of reduced sampling density on abundance indices for sablefish in comparison to those for the other species analyzed in our study can be attributed to differences in depth distribution. Of the 4 species examined in our study, the sablefish occupies the broadest and deepest range of depths, from less than 100 m to 1500 m (Kimura et al., 1998). In contrast, the Pacific ocean perch, which occupies the next widest depth range, rarely occurs at depths greater than 500 m (von Szalay and Raring, 2016). Furthermore, the depth distribution of sablefish is size-specific because fish migrate ontogenetically to deep water, where the largest fish live and, hence, a substantial portion of the biomass of sablefish occurs (Jacobson et al., 2001). For example, approximately 37% of the total biomass of GOA sablefish was estimated to occur in the deepest stratum (700–1000 m) sampled by the GOA bottom-trawl survey in 2015 (von Szalay and Raring, 2018). Because this deepest stratum was eliminated from

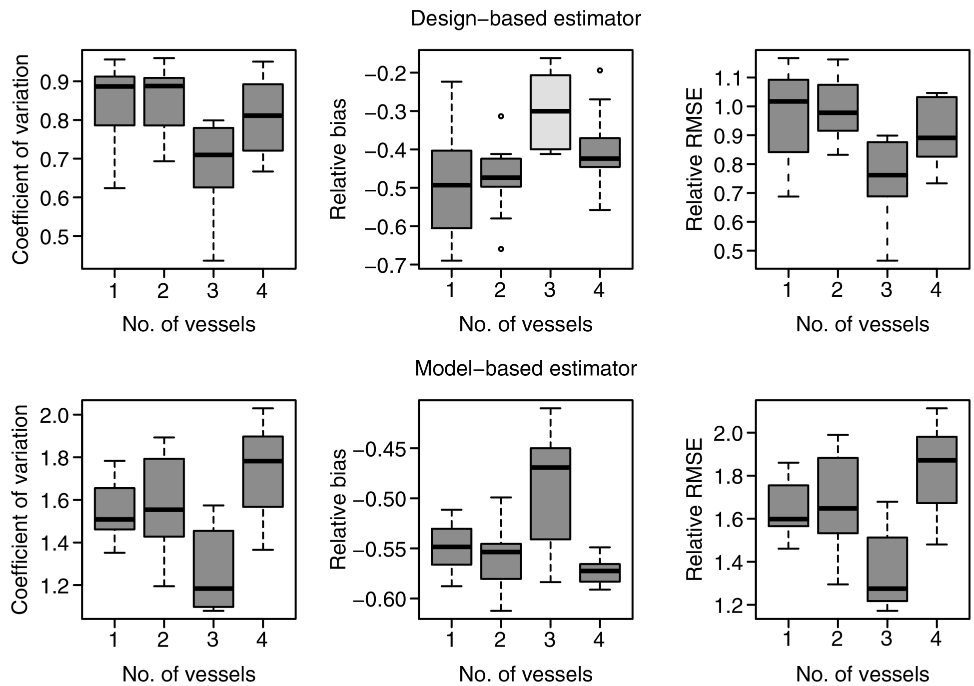

Figure 10

Relative performance of the design-based estimator and the estimator based on a vector autoregressive spatiotemporal model in estimation of variance in biomass of sablefish (*Anoplopoma fimbria*) in simulated bottom-trawl surveys of the Gulf of Alaska at 4 survey effort levels. Each box plot represents the distribution of a performance metric, coefficient of variation, relative bias, or relative root mean square error (RMSE), for a survey effort level (number of vessels) over the 10 survey years between 1996 and 2015. In each box plot, the thick black line is the median. The upper and lower parts of each box represent the first and third quartiles (the 25th and 75th percentiles). The whiskers extend above and below the box no more than 1.5 times the interquartile range. Data points that appear beyond the end of the whiskers are outliers. The light and dark shades of gray of individual boxes indicate that metric values are within the ranges at which estimates are classified as acceptable (0.20–0.40) and unreliable (>0.40), respectively.

the sampling protocol when the effort for this survey was changed from 3 vessels to 2 vessels, a large portion of the sablefish population in the GOA is not sampled at the reduced effort level, causing the increase in the relative bias of estimates found for this species. The estimates for the other 3 species are not affected by the removal of this stratum from the sampling protocol because these species do not occur at the depths in that stratum. Therefore, survey teams should consider evaluating how changes in sampling density affect not only survey data products but also survey footprint and, therefore, spatial availability of the survey in different years.

Estimates of biomass and variance for sablefish were considerably more affected by the reduction in survey effort from 3 vessels to 2 vessels than estimates for any of the other species. In addition, the quality of estimates from both the design-based and model-based estimators for this species was consistently the worst by far on the basis of values for all 3 performance metrics. This difference in estimate quality for sablefish can be explained in part by the much wider depth distribution of this species, with a large portion of the biomass of the GOA population occurring in the 3 deepest depth strata, all on the continental slope (Fig. 2) (von Szalay and Raring, 2016). Two main factors cause the modified Neyman allocation scheme to effectively result in under-sampling of these strata. First, the sampling allocation is proportional to stratum area, an approach that is not ideal because it is preferable to allocate resources with the aim to reduce variance in survey estimates (Oyafuso et al., 2022). Second, because the strata on the continental slope are narrow and encompass small areas, relatively few stations are allocated to the slope strata (Equation 3). For species like sablefish, which have large concentrations of biomass in these strata, the taking of fewer samples there reduces the quality of estimates of abundance and variance.

Other factors also cause under-sampling of strata. The number of stations allocated is inversely proportional to the cost of sampling a stratum. The cost, measured in terms of the time required to both find a viable towing location and the time spent conducting a trawl tow in a given stratum, is greater for deep strata on the continental slope than for strata on the comparatively shallow continental shelf. This difference in cost also leads to less sampling allocation to strata where biomass of sablefish is high. Therefore, including cost of sampling in the allocation algorithm should be considered in conjunction with the value of the data because consideration of cost alone may result in inadequate sampling of valuable species. The relatively low sampling effort on the continental slope is also due to necessary trade-offs among species, an inherent aspect of the design of all multispecies surveys. Because a dedicated longline survey for sablefish on the continental slope already exists (Sigler and Zenger, 1989), and because of the extremely high commercial value of this species, sablefish receive no weight when optimal sampling allocations are calculated for each species and a weighted average, based on commercial value, is generated. This tack is taken to prevent the sablefish from unnecessarily dominating the sampling allocation scheme at the expense of other species that lack a dedicated survey.

The consistently poor performance of the model-based estimator for sablefish at all effort levels and for estimates of both biomass and variance can likely be attributed to the effectively unbalanced sampling design for this species, a design that results in under-sampling in some areas for the reasons discussed earlier in this section (Fig. 2). The unbalanced nature of the stratified random design with respect to sablefish is analogous to the poor performance of the same estimator when coupled with a survey design in which the survey stations, rather than fish distributions, were heavily concentrated in just a few strata (von Szalay et al., 2023). Although the stratified random sampling design was far superior to all the other designs to which it was compared when coupled with a design-based estimator, it did not perform well when coupled with the model-based estimator, which is sensitive to highly unbalanced sampling.

Effect on estimates for Pacific ocean perch

With the exception of the variance estimate from the model-based estimator for arrowtooth flounder at the 4-vessel effort level, both estimators performed relatively well for arrowtooth flounder and Pacific cod. In contrast, variance estimates were only marginally acceptable to unacceptable for Pacific ocean perch, at all sampling densities, although values of the 3 performance metrics for biomass estimates were almost always better than those for the corresponding variance estimates. Of these 3 species, Pacific cod consistently had the best biomass and variance estimates, and Pacific ocean perch consistently had the worst estimates, on the basis of rRMSE values.

The comparatively poor estimates for Pacific ocean perch can be attributed to the relatively patchy spatial distribution of this species compared with those of the other 2 species (Lunsford et al., 2001; Clausen and Fujioka, 2007). The bulk of the biomass of Pacific ocean perch is confined to a relatively small depth interval between 150 and 300 m, in a narrow band along the upper continental slope, and the distribution of this species within this band is patchy (von Szalay and Raring, 2016). Furthermore, the traditional stratified sampling design coupled with a design-based estimator used for the GOA bottom-trawl survey has been shown to be suboptimal for Pacific ocean perch, for which larger sample sizes in patchy areas of high density are required to reduce variance estimates of biomass (von Szalay et al., 2023). This problem could potentially be addressed through implementation of a hybrid survey design that incorporates an element of adaptive sampling in areas where catches of Pacific ocean perch are very large. However, a more consistent approach for multispecies surveys, such as the GOA bottom-trawl survey, would be to use a different sampling design altogether.

One approach is to consider other stratification methods as alternatives to the traditional Neyman allocation scheme, in which stratum boundaries are defined by depth as well as by management area. In one alternate stratification procedure, which is not constrained by predefined strata boundaries, abundance and variability from historical survey data are used to derive an empirical metric (an *information score*) for fine-tuning the traditional stratification process. This method of stratification can be used to ensure that more sampling is done in areas of high fish density and has proven to be particularly effective at reducing both the CV and the rRMSE of estimates for Pacific ocean perch vis-á-vis the traditional Neyman approach (von Szalay et al., 2023). The improvements in both the CV and rRMSE were particularly noteworthy for the variance estimates (at the 3-vessel level of survey effort), which changed from borderline acceptable to highly reliable (CV: from 0.40 to 0.03; rRMSE: from 0.40 to 0.10). Variance homogeneity within strata was considerably greater with stratification based on information scores than with the traditional stratification coupled with the Neyman allocation algorithm. This improvement in homogeneity can be attributed to the variance component of the information score, which allows variance to be targeted more directly, and consequently makes the stratification method based on information scores more efficient than the traditional stratification approach.

The method for stratification based on information scores also can be used to ensure adequate sampling in areas of high density but low temporal variance. This additional focus on areas of high density contrasts with the modified Neyman allocation approach, which can be used to account for only variance, stratum area, and a cost variable but not directly for density. Consequently, in the extreme case of a stratum with consistently high but uniform fish density, no sampling stations would be allocated with the Neyman allocation scheme.

Comparison of results from estimators

The model-based estimator consistently outperformed the design-based estimator in terms of both CV and rRMSE of biomass estimates at all sampling densities for arrowtooth flounder, Pacific cod, and Pacific ocean perch. To a lesser extent, the opposite was true for relative bias, on the basis of which the design-based estimator consistently outperformed. However, because relative bias was consistently low for all of these species, its contribution to the rRMSE of biomass estimates was much smaller than the contribution of the CV; therefore, the model-based estimator performed best overall. An important short-term implication of this finding is that the relatively modest negative effect on the rRMSE of the recent reduction in survey effort can be completely mitigated by switching to spatiotemporal model-based methods, such as those in the VAST package, to estimate biomass.

Although the model-based estimator did not outperform the design-based estimator as consistently at all sampling densities for variance estimates as it did for biomass estimates, it did so at the only 2 sampling densities implemented in practice and likely to be implemented in the future, the 2-vessel and 3-vessel effort levels, for arrowtooth flounder, Pacific cod, and Pacific ocean perch. Therefore, the model-based estimator can also be used to mitigate the negative effect on variance estimates resulting from the recent reduction of sampling density in the GOA. The advantage of the model-based approach is so large that, on the basis of CV and rRMSE, it provides better biomass and variance estimates for Pacific ocean perch, even at the low 1-vessel effort level, than the design-based estimator and provides approximately the same quality of biomass estimates for arrowtooth flounder and Pacific cod at the 3-vessel effort level.

In contrast to its performance for arrowtooth flounder, Pacific cod, and Pacific ocean perch, the model-based estimator did not perform well for sablefish, even at the 3-vessel and 4-vessel effort levels at which sampling included the deepest strata. This difference in performance is likely due to the relatively high concentration of large fish (>70 cm in fork length) along the narrow strata on the continental slope. The model in the VAST package has been shown to perform very poorly when coupled with an unbalanced sampling design for arrowtooth flounder, Pacific cod, and Pacific ocean perch (von Szalay et al., 2023). It is likely that the high concentration of biomass of sablefish along the narrow continental slope has a similar effect on VAST estimates as an unbalanced survey design because the knots created in the VAST model to extrapolate fish density are often spaced too far from the sampling locations to generate valid density estimates. In the case of sablefish, the knots tend to be concentrated on or near the continental slope where fish density is greatest. This concentration is problematic because, as discussed earlier in the "Effect on estimates for sablefish" section, the stratified random sampling design results in under-sampling of this part of the survey area. Similarly, a VAST model coupled with an unbalanced survey design for a more evenly distributed species, such as arrowtooth flounder, also performs poorly when the bulk of sampling locations are concentrated in just a few strata. This poor performance occurs because the sampling locations tend to be located close to just a few knots where fish density may not be representative of the entire survey area.

Effect of climate change on validity of sampling design

Although the findings of this study indicate that the effects on the quality of biomass and variance estimates for most species are only modest when survey effort is reduced by as much as a third from historical maxima, this benign outcome is predicated on the validity of using historical survey data as the basis for allocating sampling locations in a stratified random sampling design. This assumption may be problematic because of shifts in distribution of fish species that may occur under climate change, with future distributions deviating from historical norms as species migrate in response to increased water temperatures and other variables (Hobday and Evans, 2013; Maureaud et al., 2021; Hollowed et al., 2022). For example, if fish move abruptly to strata that previously had low abundance, the data used as input for the Neyman allocation scheme will not reflect the new distribution, and these strata consequently will be under-sampled. In order to maintain the quality of abundance estimates in this scenario, it may be necessary to maintain or increase the overall sampling effort to characterize distribution shifts and use new survey data as an input for the sampling design process.

Conclusions

The reduction of sampling density from 820 to 550 stations for the GOA bottom-trawl survey only modestly affected the 3 performance metrics for biomass and variance estimates from both the design-based and model-based estimators for arrowtooth flounder, Pacific cod, and Pacific ocean perch. However, the effect on the relative bias of abundance indices for sablefish was profound. The performance metric of rRMSE consequently put the quality of abundance estimates for sablefish firmly in the poor category, meaning that estimates are unreliable at the reduced sampling densities. The disproportionate effect of reduced sampling density on estimates of abundance for sablefish can be attributed to the wider and deeper depth distribution of this species compared with those of the other 3 species, and estimates for this species are more sensitive to changes in survey effort as well as to concomitant changes in sampling area.

Resumen

Se evaluó el efecto de reducir el esfuerzo de prospección sobre la exactitud y precisión de las estimaciones de abundancia de 4 especies comerciales o ecológicamente importantes con distribuciones diferentes observadas en una prospección de arrastre de fondo realizado en el Golfo de Alaska. Se utilizaron simulaciones de un modelo mixto lineal generalizado espaciotemporal con base en observaciones históricas de las densidades de capturas para evaluar la robustes estadística, medida en términos de coeficiente de variación, sesgo relativo y error cuadrático medio relativo, de las estimaciones de abundancia y sus varianzas. Estas métricas se utilizaron para comparar las estimaciones entre el estimador tradicional basado en el diseño y el estimador alternativo, basado en un modelo vectorial autorregresivo espaciotemporal, en 4 densidades de muestreo diferentes, que representan 2 niveles de esfuerzo de muestreo históricos y 2 teóricos a cada lado del intervalo histórico. La reciente reducción en la densidad de la prospecion de 820 a 550 estaciones sólo tuvo un efecto modesto en las medidas de rendimiento de ambos estimadores para el lenguado *Aesthes stomias*, el bacalao del Pacífico (*Gadus macrocephalus*) y la perca *Sebastes alutus*. Sin embargo, el efecto sobre las estimaciones de abundancia del bacalao negro (*Anoplopoma fimbria*) fue considerable. Atribuimos esta diferencia de resultados al mayor rango de profundidad utilizado por el bacalao negro, que ocupa preferentemente los estratos profundos relativamente poco muestreados (>500 m), y al área de estudio truncada en los niveles de muestreo reducidos, en los que se han eliminado los estratos más profundos (>700 m).

Acknowledgments

We are deeply grateful for J. Thorson's development of the R package VAST and for the valuable feedback and inspiration provided by the working group formed during the International Council for the Exploration of the Sea Workshop on Unavoidable Survey Effort Reduction. We also greatly appreciate the labor-intensive work of the many scientists and volunteers who have contributed to the GOA bottom-trawl surveys over the years. Finally, we thank M. Siple, P. Hulson, and 2 anonymous reviewers for their informative comments and suggestions, which helped to substantially improve the readability and clarity of this paper.

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. 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. Fairbanks, Fairbanks, AK.

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

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. Computer 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.]
2023. Workshop on Unavoidable Survey Effort Reduction 2 (WKUSER2). ICES Sci. Rep. vol. 5, issue 13, 115 p. [Available from website.]

Jacobson, L. D., J. Brodziak, and J. Rogers.
2001. Depth distributions and time-varying bottom trawl selectivities for Dover sole (*Microstomus pacificus*), sablefish (*Anoplopoma fibria*), and thornyheads (*Sebastoloubs alascanus* and *S. altivelis*) in a commercial fishery. Fish. Bull. 99:309–327.

Kimura, D. K., A. M. Shimada, and F. R. Shaw.
1998. Stock structure and movement of tagged sablefish, *Anoplopoma fimbria*, in offshore northeast Pacific waters and the effects of El Niño-Southern Oscillation on migration and growth. Fish. Bull. 96:462–481.

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

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. 73:423–498. Crossref

Lunsford, C. R., L. Haldorson, J. T. Fujioka, and T. J. Quinn II.
2001. Distribution patterns and survey design considerations of Pacific ocean perch (*Sebastes alutus*) in the Gulf of Alaska. *In* Spatial processes and management of marine populations. 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. Fairbanks, Fairbanks, AK.

Maureaud, A. A., R. Frelat, L. Pécuchet, N. Shackell, B. Mérigot, M. L. Pinsky, K. Amador, S. C. Anderson, A. Arkipkin, 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. Global Change Biol. 27:220–236. Crossref

Meng, X., J. Wang, C. Zhang, Y. Xue, B. Xu, and Y. Ren.
2019. Effects of sampling transect number on estimation of abundance index for bottom trawl surveys of fisheries resources in the Yellow River estuary. J. Fish. China 43:1507–1517. [In Chinese.] 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 estimation. NOAA Tech. Memo. NFMS-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. Crossref

R Core Team.
2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available from website, accessed March 2020.]

Rue, H., S. Martino, and N. Chopin.
2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc., B 71:319–392. Crossref

Sigler, M. F., and H. H. Zenger Jr.
1989. Assessment of Gulf of Alaska sablefish and other groundfish based on the domestic longline survey, 1987. NOAA Tech. Memo. NMFS F/NWC-169, 54 p.

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 Autoregressive 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., 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.
2016. Data report: 2015 Gulf of Alaska bottom trawl survey. NOAA Tech. Memo. NMFS-AFSC-325, 249 p.
2018. Data report: 2017 Gulf of Alaska bottom trawl survey. NOAA Tech. Memo. NMFS-AFSC-374, 260 p.

von Szalay, P. G., S. Kotwicki, L. A. K Barnett, L. J. Rugolo, and K. Ono.
2023. Reducing uncertainty in survey abundance estimates by considering alternative designs and estimators: a case study with 3 species in the Gulf of Alaska. Fish. Bull. 121:50–66. Crossref

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 investigations 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.]

Xu, B., C. Zhang, Y. Xue, Y. Ren, and Y. Chen.
2015. Optimization of sampling effort for a fishery-independent survey with multiple goals. Environ. Monit. Assess. 187:252. Crossref

Zimmermann, M., M. M. Prescott, and P. J. Haeussler.
2019. Bathymetry and geomorphology of Shelikof Strait and the western Gulf of Alaska. Geosciences 9:409. Crossref