Original Article Submission to the ICES Journal of Marine Science

Incorporating spatiotemporal variability in multispecies survey design optimization addresses tradeoffs in uncertainty

Zack S. Oyafuso*, Lewis A.K. Barnett, Stan Kotwicki

National Marine Fisheries Service

Alaska Fisheries Science Center

National Oceanic and Atmospheric Administration

7600 Sand Point Way NE

Seattle, WA 98115

*corresponding author: Zack S. Oyafuso

Email: zack.oyafuso@noaa.gov

Office Phone Number: (206) 526-4083

Keywords: Survey Optimization; Gulf of Alaska; Bottom-Trawl Survey; Stratified Random Sampling; Genetic Algorithm;

Abstract

In designing and performing surveys of animal abundance, monitoring programs often struggle to determine the sampling intensity and design required to achieve their objectives, and this problem greatly increases in complexity for multispecies surveys with inherent tradeoffs among species. To address these issues, we conducted a multispecies stratified random survey design optimization using a spatiotemporal operating model and a genetic algorithm that optimizes both the stratification (defined by depth and longitude) and the minimum optimal allocation of samples across strata subject to prespecified precision limits. Surveys were then simulated under those optimized designs and performance was evaluated by calculating the precision and accuracy of a resulting design-based abundance index. We applied this framework to a multispecies fishery-independent bottom trawl survey in the Gulf of Alaska, USA. Incorporating only spatial variation in the optimization failed to produce population estimates within the prespecified precision constraints, whereas including additional spatiotemporal variation ensured that estimates were both unbiased and within prespecified precision constraints. In general, results were not sensitive to the number of strata in the optimized solutions. This optimization approach provides an objective quantitative framework for designing new, or improving existing, survey designs for many different ecosystems.

1. Introduction

Productive and sustainable fisheries provide socioeconomic opportunities and ensure food and nutritional security. In the United States, commercial wild-capture fisheries totaled 4.3 million metric tons valued at $5.6 billion in 2018 (NMFS, 2020). Fisheries stock assessments provide the basis for managing these fisheries. Fishery-independent surveys are often the primary source of inputs for stock assessment models, providing information on the abundance and composition of fish populations. Thus, properly designed fisheries surveys are integral to ensuring that the most scientifically robust data products are used for fisheries management (Smith and Hubley, 2014; Zimmermann and Enberg 2016; Muradian et al. 2019). Survey data are also used to address a variety of research questions including species distributions over time (e.g., Thorson et al., 2015), ecosystem status indicators through environmental data collection (e.g., de Boois, 2019; Zador et al., 2019).

Accuracy and precision are the main quality metrics of a fisheries survey and are constrained by total sampling effort and budget. The precision of a survey, described as either a variance or a coefficient of variation (CV) is an important survey output commonly used for survey comparison studies (Overholtz et al., 2006), evaluations of survey outputs quality (Cao et al., 2014), and stock assessments (Francis, 2011). That said, fisheries surveys need to be flexible to many sources of logistical constraints and uncertainties while still maximizing the objectives of producing survey products with high accuracy and precision. Unavoidable survey effort reduction due to budgetary constraints, inclement weather, or vessel breakdowns pose serious implications to the reliability of fisheries surveys (ICES, 2020). Reductions in survey effort through a reduction in sampling intensity or frequency can compromise the precision and bias of abundance indices (ICES, 2020; Hutniczak et al., 2019; von Szalay, 2015). Additionally, fishery-specific constraints like gear type, coverage rate, and vessel type are other additional considerations when optimizing survey design (Miller et al., 2006). Given the high operating costs of fisheries-independent surveys and that these changes typically occur at time scales that leave little time for planning and quantitative evaluation, there is a need for rapid survey optimization tools to guide survey changes within a flexible framework.

The multispecies nature of many surveys means that invariably there are interspecific tradeoffs in designing a survey that optimizes over many species (and possibly life stages within species) with different spatiotemporal distributions and varying levels of directed targeting (Wang et al., 2018; Smith et al., 2011). The magnitude of variance in species abundance across space and/or time affects the optimal spatial extent and frequency of surveys (Lanthier et al., 2013; Rhodes and Jonzén, 2011). In some cases, there may be temporary needs for increased precision for certain species and/or regions (e.g., when a stock is close to a limit threshold or displays sudden declines in abundance; Laurel and Rogers, 2020; Barbeaux et al., 2017). Further, tradeoffs in survey design strategies can occur among data uses e.g., indices of abundance, compositional data, species distribution shifts, and population responses to marine reserve implementation (Smith et al. 2011; Miller et al., 2006). Thus, the evaluation of the effects of changes in total survey effort needs to also consider tradeoffs of quality metrics among species.

To illustrate the development of a fishery survey design optimization framework while addressing the above challenges related to survey effort reduction and tradeoffs among species, we focused on a case study involving the Gulf of Alaska (GoA) groundfish stratified random bottom-trawl survey (BTS). With a relatively long time series (nearly 40 yr in this case) of data on the distribution of these species, both spatiotemporal variability and/or species covariation can be incorporated into a more goal-driven and objective survey design optimization (e.g., Peel et al., 2012). The stratified survey optimization was conducted using a genetic algorithm that optimizes both the stratification of the spatial domain as to minimize total sample size subject to prespecified precision constraints for a given number of strata. We used a previously built multispecies spatiotemporal fish density distribution model as data inputs to the optimization. Surveys were then simulated under those optimized survey designs and the precision and bias of the population estimates were calculated as performance metrics. This framework for optimizing a stratified random survey design for estimating abundance with respect to a model-generated spatiotemporal distribution can be used to evaluate the multispecies tradeoffs of varying sampling intensities on the quality of fisheries survey estimates.

2. Methods

The framework of the optimization is presented in Figure 1. Section 2.1 is a brief overview of the multispecies spatiotemporal operating model, from which predicted densities are used as data inputs to the survey optimization algorithm. The optimization problem is defined in Section 2.2 and the algorithm used to solve the optimization problem is described in Section 2.3. Section 2.4 describes how the survey optimization is conducted in the GoA and 2.5 describes the simulation of those optimized survey designs against the operating model and the resulting performance metrics. The associated code can be found on the corresponding author's GitHub page (https://github.com/zoyafuso-NOAA/Optimal_Allocation_GoA_Manuscript).

Three types of CVs are defined in the following sections with slightly different interpretations and uses in this framework. In sections 2.2-2.4, CVs that incorporate variability in density across the domain and observed years for each species from the operating model described in section 2.1 and are used as prespecified constraints of precision to guide the optimization of a new multispecies stratified survey design. These CVs utilize population-level stratum variance statistics that integrate the many sources of process variability as specified in the OM in Section 2.1 with the exception of additional sources of measurement error. These CV constraints can be interpreted as the expectation of the sample CV for a given level of survey effort. The survey simulation in Section 2.5 is important in establishing precision levels more consistent with what would be observed in the sampling process. Within a simulation framework, the second CV defined in 2.5 describes the variability of an abundance index across many simulated surveys relative to the the true index, interpreted as the realized or "true" sampling CV (Kotwicki and Ono, 2019), a metric impossible to calculate when analyzing actual surveys. The sample CV is the third type of CV used in this analysis and refers to the CV associated with the abundance index calculated for one replicate survey. Unlike the CV constraints, these CV utilize sample-level statistics of stratum variance and are year-specific. The congruence of these sample CVs to the realized true CV is a performance metric defined in Section 2.5.

2.1 Operating Model

To serve as an operating model, we fitted a multispecies spatiotemporal distribution model to catch rate data using a vector-autoregressive spatiotemporal model (VAST; Thorson and Barnett, 2017). Readers are referred to the Supplementary S1 for more detail on the VAST operating model, but a brief description of the relevant outputs follows. We fitted the VAST model to catch-per-unit-effort data of GoA groundfishes collected from a fishery-independent BTS using a stratified random sampling design (von Szalay and Raring, 2018). We restricted the input data to the years 1996, 1999, and every other year from 2003 to 2019 to ensure consistency in sampling design and species identification (11 observed data years). Fourteen species and one species group were included to represent the groundfish complex in the GoA, based on commercial value and the dependence of stock assessment models on survey-derived abundance indices: *Atheresthes stomias*, *Gadus chalcogrammus*, *G. macrocephalus*, *Glyptocephalus zachirus*, *Hippoglossoides elassodon*, *Hippoglossus stenolepis*, *Lepidopsetta bilineata*, *L. polyxystra*, *Limanda aspera*, *Microstomus pacificus*, *Sebastes alutus*, *S. polyspinis*, *S. variabilis*, and *Sebastolobus alascanus*. Due to identification issues between two rockfishes, *Sebastes melanostictus* and *S. aleutianus*, the catches of these two species were combined into a species group (Sebastes spp.) we will refer to as "*Sebastes* B_R" (blackspotted rockfish and rougheye rockfish, respectively) hereafter.

The density (y_{git}) of each species or species group was predicted onto the GoA survey spatial domain at a resolution of 3.7 by 3.7 km ($i: 1, 2, \dots, N$ = 23 339 cells; some prediction grid cells had smaller area due to intersections with survey domain boundaries) for each species ($g: 1, 2, \dots, G = 15$ species) and observed year ($t: 1, 2, \dots, T = 11$ observed years). Figure 2 shows the average spatial distribution over time for each species. These predictions were taken to represent "true" densities values, which were used to generate optimal survey designs and evaluate the performance of simulated surveys given those designs. As the primary measure of survey performance is the accuracy and precision of the total abundance estimate, we define this by the proxy of mean density.

2.2 Survey Optimization Problem

The goal of the multispecies stratified survey design optimization is to jointly optimize the stratification and the sample allocation across strata (h: $1, 2, \dots H$) by finding that which minimizes total sample size, subject to prespecified precision constraints for each species. Specifically, the objective function is to minimize total sample size subject to G prespecified coefficient of variation (CV) constraints ($U_1, U_2, \dots, U_G$):

$$min \sum_{h=1}^{H} n_h \quad \text{[Equation 1]}$$

$$\begin{gathered} s.t. \\ CV(Y_1) < U_1 \\ \cdots \\ CV(Y_G) < U_G \, , \end{gathered} \quad \text{[Equation Set 2]}$$

$$CV(Y_g) = \frac{\sqrt{Var(Y_g)}}{Y_g} \quad \text{[Equation 3]}$$

$$Var(Y_g) = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \frac{S_{h,g}^2}{n_h} (1 - \frac{n_h}{N_h}) \text{ [Equation 4]}$$

where n_h and N_h are the sample sizes and number of sampling units in stratum h, respectively. By leveraging density predictions provided by the OM, this optimization is specified using population-level statistics. Y_g is the population mean of species g averaged over the cells in the spatial domain and over observed years. $Var(Y_g)$ in Equation 4 is the stratified random sampling variance associated with the population mean. Careful consideration is needed for this variance, specifically the stratum variance $S_{h,g}^2$, defined in Equation 4. The OM provides predicted densities across all cells and observed years for each species and integrates many sources of variation including temporal (year-to-year), habitat covariates (depth), species covariation, and additional spatial and spatiotemporal variation. A common issue in survey design optimization is how to integrate data from previous surveys (Francis, 2006), thus we investigated two types of stratum variances that incorporated the OM-derived densities predicted across the observed survey years in the GoA BTS:

1) Spatial-only stratum variance: The first method was to reduce the temporal dimension by averaging the predicted densities from the OM over the observed years for each cell in the spatial domain. In this "spatial-only" optimization, $S^2_{h,g}$ is the population stratum variance of density for species g in stratum h:

$$S^2_{hg} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(\overline{y_{gi\cdot}} - \overline{y_{hg}}\right)^2, [Equation\ 5]$$

where $\overline{y_{hg}}$ is the population mean density estimate of species g averaged across all observed years and cells contained within stratum h, and $y_{gi\cdot}$ is the predicted density of species g in cell i (where cell i is in stratum h) averaged across observed years. Note the use of the N_h term in Equation 5 denotes a population-level stratum variance.

2) Spatiotemporal stratum variance: A potential issue with the spatial-only version of the population stratum variance is underestimating the total "known" variability within a stratum by averaging over the year-to-year as well as spatiotemporal variation explicitly modeled in the OM. Thus, for this "spatiotemporal" optimization, the population-level stratum variance in Equation 5 was modified to incorporate both within-stratum (note the summation range between $i = 1$ to N_h) density variation and within-grid cell densities variation across years (note the summation range between $t = 1$ and T):

$$S^2_{hg} = \frac{1}{TN_h - 1} \sum_{t=1}^{T} \sum_{i=1}^{N_h} \left(y_{git} - \overline{y_{hg}}\right)^2 \text{[Equation 6]}$$

2.3 Optimization of Strata Boundaries and Sample Allocation

Comprehensive brute-force searches for the optimum stratification of the spatial domain and optimum allocation of samples are usually intractable for moderately sized problems. Thus, we searched for optimal stratifications and survey effort allocations via a genetic algorithm using the R package SamplingStrata (Barcaroli, 2014; Ballin and Barcaroli, 2013). The genetic algorithm uses evolutionary principles such as fitness-based selection, recombination, and mutation to iteratively search for an optimal stratification and sample allocation. Below, we provide a brief description of the algorithm and settings used but readers are referred to Ballin and Barcaroli (2013) for more technical details.

The optimization initializes with 30 random stratifications (a prespecified number of candidate solutions) based on two auxiliary variables, bottom depth (m) and longitude (eastings, km) for a user-defined number of strata. Here, we explore results from 5 to 60 strata to determine how the number of strata influences the precision of the abundance estimate. In the GoA, gradients across both depth and location have been observed to describe major patterns in demersal species composition (Mueter and Norcross, 2002). Longitude was used as a one-dimensional east-west location proxy. For each candidate solution, the Bethel algorithm (Bethel, 1989) is used to optimize the allocation of the minimum sample size across strata, subject to equations 1-2. Fitness is defined as the resultant sample size from the Bethel algorithm, with solutions with lower sample sizes having higher fitness. Elitism occurs by taking the solutions with highest fitness (defined *a priori* to be solutions in the top 10^{th} percentile for smallest sample size) and automatically advancing them to the next iteration of the algorithm. In the next iteration the remaining solutions are selected with probability proportional to their fitness values to "procreate" a new solution by applying a crossover of the solutions. Random changes in the stratifications, or mutations, are then applied at a given rate to the resultant solution. The mutation rate defines how often random changes to the solutions occur and was tuned to $1/(1+H)$ based on previous tuning guidelines (G. Barcaroli, personal communication) to reach reasonable convergence times. The process of procreation occurs until 30 candidate solutions are included in the next iteration of the algorithm. The algorithm is conducted for a total of 200 iterations, a value (along with the choice of 30 candidate solutions) chosen to ensure that, at least qualitatively, the algorithm reached an asymptotically optimal solution within a reasonable amount of computation time (see Supplementary S3 for an example of the algorithm output).

2.4 Optimization Schemes

In the GoA, total sampling effort is primarily determined by how many boats are available to conduct the survey, with all vessels operating for the same duration of time. These levels of sampling intensity correspond to approximately: 280 samples (1 boat), 550 (2 boats) and 820 (3 boats) (von Szalay and Raring, 2018; von Szalay et al., 2010). Thus, we focused on optimized survey designs under these three sample size scenarios for a given number of strata. The optimization does not maximize precision constrained by a total sample size, thus we needed to set the CV constraints (Equation Set 2) for each species to meet the three sample size scenarios regardless of which version of the stratum variance (spatial-only or spatiotemporal, Equations 5 or 6, respectively) is used. We implemented this systematically using two sets of rules depending on whether the CV constraint was constant or varying among species:

1) One-CV constraint scenario: CV constraints were set to the same value across species. Initially the CV constraint was set to some arbitrarily high value (e.g., 0.30) and the optimization was conducted to produce the optimal stratification and total sample size. Then, the population CV constraint is incrementally decreased (e.g.. 0.30 to 0.29) and the optimization was conducted again. By gradually decreasing the CV constraint, the optimized sample size slowly increases. This increment was chosen to be small enough to balance having adequate coverage over the three boat-effort scenarios (n = 280, 550, 820 stations) within a reasonable total computation time. This process was iterated until the range of considered sample sizes was captured (i.e., until the optimized sample size was ≥ 820).
2) Species-specific CV constraint scenario: CV constraints were allowed to differ across species. Similar to the one-CV constraint scenario, the CV constraint was initialized to be the same across species at some arbitrarily high value (e.g., 0.30). The optimization was conducted, and the optimized CVs across species (i.e., $CV(Y_1), CV(Y_2), \ldots\ CV(Y_G)$) were saved from the optimization. The CV constraints for the next instantiation were calculated by reducing the optimized CVs in the previous run by some proportional increment (e.g., 5%) for each species. Similar to the one-CV method, this process was iterated until the range of the three boat-effort scenarios was captured.

2.5 Simulation of data collection

For each combination of strata number and sample size scenario, the optimized survey was simulated *D* = 1000 times. r_{dgt} is the stratified random sample estimate of mean density of species *g* at time *t* for simulated survey *d*. $CV(r_{dgt})$ is the CV of the survey estimate and is similar to Equations 3-4 except using the sample stratum variance instead of the population stratum variance. To evaluate the precision and accuracy of the abundance estimates resulting from simulated surveys, we calculated the following performance metrics for each species.

Since our procedure does not optimize sample CVs directly, we evaluated the expected effect of a survey optimized with respect to population CVs on performance metrics of the sample CVs derived from simulated surveys. The "true" CV, $CV_{TRUE}(Y_{gt})$, describes the precision of the mean density estimate of species *g* at time *t* across replicate surveys and is the standard deviation of the simulated survey estimates (where $\overline{r_{\cdot gt}}$ is the mean density estimate of species *g* at time *t* averaged across the *D* surveys) relative to r_{gt}, the true mean density of species *g* at time *t*:

$$CV_{TRUE}\left(Y_{gt}\right) = \frac{\sqrt{(D-1)^{-1}\sum_{d=1}^{D}\left(r_{dgt} - \overline{r_{\cdot gt}}\right)^2}}{Y_{gt}}. [Equation\ 7]$$

Relative root mean square error of CV, $RRMSE(CV(r_{dgt}))$, is a measure of uncertainty of the replicate sample CVs of species *g* at time *t* and is a composite measure of the dispersion and bias of the replicate sample CVs about the true CV:

$$RRMSE\left(CV\left(r_{dgt}\right)\right) = \frac{\sqrt{D^{-1}\sum_{d=1}^{D}\left(CV\left(r_{dgt}\right) - CV_{TRUE}\left(Y_{gt}\right)\right)^2}}{D^{-1}\sum_{d=1}^{D} CV\left(r_{dgt}\right)}. [Equation\ 8]$$

Lastly, relative biases (RB) of the mean density and CV estimates relative to their respective true values were calculated as

$$RB(r_{dgt}) = 100\% \ \frac{\sum_{d=1}^{D}(r_{dgt} - Y_{gt})}{DY_{gt}} [Equation\ 9]$$

$$RB\left(CV(r_{dgt})\right) = 100\% \ \frac{\sum_{d=1}^{D}(CV(r_{dgt}) - CV_{TRUE}(Y_{gt}))}{D\ CV_{TRUE}(Y_{gt})} [Equation\ 10]$$

3. Results

3.1 Optimal stratification: The optimization solutions with the closest sample sizes to each of the three intended sample sizes were chosen as the representative solutions. Figure 3 shows those three representative solutions along with examples of simulated survey stations for five, ten, and fifteen strata. The longitudinal variable was generally cut into the west, central, and eastern parts of the spatial domain. Strata in the eastern part of the domain were often connected with the deeper continental slope strata. Sampling density was concentrated in the western and central parts of the spatial domain, with sparse sampling in the eastern portion. Solutions across boat-effort scenarios within a strata number scenario were generally consistent in the strata boundaries.

3.2 Tradeoff between sample size and CV constraint: The spatial-only optimization led to one, two, and three boat solutions with expected CV constraints of 0.19, 0.13, and 0.10, respectively (Figure 4). These CV constraints are from the one-CV constraint approach, meaning these values represent the maximum expected sampling CV that any one species can exhibit. The addition of spatiotemporal variability of the optimization increased the CV constraints across boat-effort scenarios to 0.28, 0.21, and 0.17, respectively. For a given CV constraint, the addition of spatiotemporal variability required roughly 2-3× more samples in the optimal solution. Figure 4 shows the relationship between sample size and CV for a five-strata scenario only, but this pattern was consistent across scenarios with different numbers of strata (Supplementary S4).

3.3. Expected vs realized precision: True CV encompasses the variability of the mean density estimates across realized survey replicates relative to the true mean density and is different from the prespecified (expected) CV constraints used to constrain the survey optimization algorithm. Simulation testing allows for the evaluation of the congruency of the true CV across years to the CV constraint. Simulated surveys under the spatial-only optimization failed to produce true CVs lower than the CV constraint consistent across observed years for some species (Figure 5). The median of the distribution of true CV across years for *Sebastes alutus, S. polyspinis*, and *S. variabilis* were 25-50% higher than the CV constraints specified in the optimization. When spatiotemporal variability was included in the optimization, all species were surveyed with true CVs lower than the CV constraints for the majority, if not all, years observed. Further, under the species-specific CV constraint scenario, all species were surveyed with true CVs at or slightly below their respective CV constraints. Additionally, the medians of the distributions of the true CVs were much closer to the expected CV than the one-CV constraint scenarios. These patterns were consistent across scenarios with different numbers of strata (Supplementary S5).

3.4 True CV across strata and sample sizes: Increasing sampling intensity reduced the true CV and the spread of the bias of the mean density estimate across species and strata scenarios (Figures 6-7). Estimates of mean density across species showed low bias (Figure 7), with slightly negative median biases up to 5%. Increased samples across species led to further reductions in bias and there were no noticeable differences in this effect across number of strata. There were also no noticeable trends in true CV across number of strata for either the one-CV constraint (Supplementary S6) or species-specific CV constraint optimizations (Figure 7).

3.5 Relative Root Mean Square Error of CV across strata and sample sizes: The RRMSE of CV encompasses both the bias and variability of the simulated sample CVs about the true CV.

Similar to true CV, increasing sampling reduced the uncertainty and spread of the bias of the sample CV estimates across species and strata scenarios with high consistency between both optimization types (Figures 8-9). An exception was the RRMSE of CV being higher for larger numbers of strata for a handful of species (e.g.*,* slope-dwellers such as *Sebastes* B_R and *Sebastolobus alascanus*) for the one-CV constraint optimization (Figure 8). There was less of a noticeable trend across strata in RRMSE of CV for the species-specific CV constraint optimization than for the one-CV constraint optimization (Supplementary S7). The species-specific CV constraint optimization was more consistent in demonstrating the pattern of lower true CV and RRMSE of CV with increasing sample sizes, particularly with *M. pacificus, Sebastolobus alascanus*, *Sebastes* B_R, *L. bilineata*, and *L. polyxystra*. Simulated sample CVs were slightly negatively biased relative to their respective true CV value with smaller magnitude and variability with increasing sampling intensity (Figure 9), regardless of the CV-constraint approach used.

4. Discussion

The inclusion of spatiotemporal variability in the population stratum variance calculation (Equation 6) led to CV constraints that were within the distribution of the true or realized CVs of abundance when surveys were simulated. These CV constraints are equivalent to those the user defines initially in Equation Set 2, thus the main goal of the survey simulation was the evaluate the congruency between the expected CV constraints and realized CVs in the form of the true CVs. In contrast, CV constraints using the spatial-only version of the population stratum variance (Equation 5) were not consistent with true CVs across species, with true CVs for some of the more variable *Sebastes* species vastly underestimated. The issue of including historical variation in the survey data has been discussed in detail previously (Francis, 2006), one complication being that incorporating year-to-year variation in our operating model may overestimate the within-stratum variability. In fact, the tradeoff of adding spatiotemporal variation to the stratum variance calculation (Equation 6) was a 2-3× increase in sample size for a given CV constraint (Figure 4), with many species' distributions of true CV lower than their respective CV constraints (Figure 5). However, the consistency between the true CVs and their respective CV constraints across species and years supports the use of this optimization to provide robust and consistent indices of abundance. Furthermore, future applications of this approach should also integrate within the optimization framework other important sources of observation error not included in this analysis, e.g., measurement error, untrawlable areas, detectability (Field et al., 2005), and sampling efficiency (Kotwicki and Ono, 2019), especially when realistically simulating surveys and assessing performance. The exclusion of additional sampling error in our framework limits the absolute interpretability of the CV constraints and true CVs, thus these CVs could be treated as the "best case" or lower limits of expected sampling CVs.

Specifying precisions constraints for each species is a clear advantage of this survey optimization framework and allows increased flexibility for survey planners to meet desired goals in their survey designs. When we initially used the one-CV constraint method to solve the optimization problem, there were some inconsistencies in simulated true CV (Supplementary S6) and RRMSE of CV (Figure 8) and sampling intensity for some species. With the one-CV constraint approach, a single CV constraint is defined for all species, thus the CV constraint imposed in the optimization is strict for some species and less so for others, which can produce these inconsistent findings. The species-specific CV constraint approach seemed to produce more consistent positive trends in the performance metrics with increasing sampling intensity by defining CV constraints for each species individually. By setting constraints for each species specifically and allowing the CV constraints to reduce proportionally for each species, solutions performed more consistently with increasing sampling intensity. Setting CV constraints for each species also gives survey planners more flexibility to emphasize or de-emphasize certain species within the optimization more explicitly while evaluating the resulting tradeoffs in precision for the other species. The CV constraint utilized in this optimization was a maximum constraint but additionally, minimum CV constraints can be also provided from stock assessment programs to provide additional constraints on the optimization. We naively assumed in the species-specific CV approach that the CV constraints need not be lower than 10%, but these values can be based on different priorities for different species. Work is currently being done for that purpose in the Gulf of Alaska stock assessments (ICES, 2020), based on how sampling precision affects uncertainty of assessment outputs such as estimated biomass. Ultimately, a cost-benefit analysis evaluating the relationship between total sampling effort, precision, and downstream management quantities like total allowable catch can more directly link the multispecies tradeoffs of surveys on the economic value of fisheries (Francis, 2006).

While there are many approaches to optimizing survey design, the framework introduced provides a new approach to optimize a survey design that is particularly advantageous for estimating animal abundance time series. Previous simulation studies have shown that reductions in precision from lowered sampling can be alleviated by choosing a more optimal stratification scheme (Xu et al., 2015). Peel et al. (2012) developed a survey optimization based on a multispecies model-based (Generalized Additive Model) survey design. With the increasing usage of model-based spatiotemporal methods to develop indices of abundance (Thorson and Barnett, 2017; Thorson et al., 2015, 2017), it is becoming more relevant to develop formalized survey design optimizations in tandem with these model-based estimation methods. Other weighted multiple-criterion optimizations of stratified surveys focused on optimizing over additional data types like compositional and bycatch data (Miller et al., 2006). With emerging OMs like those presented in the SimSurvey R package (Regular et al., 2020), age- and spatially explicit OMs are becoming more accessible to incorporate other data types in a survey optimization.

The framework that we present can be used as a tool for long-term decision support for improving current surveys and resulting survey data products such as abundance indices and age or size composition estimates. For example, modifying the current stratified survey design in the GoA is a long-term process that will involve rigorous review and operational modifications over multiple years. Fortunately, the switch to a more efficient survey design would not require calibration, as the change would be between two stratified random designs which are inherently unbiased. Work is currently ongoing to compare the performance of this survey design framework versus the current GoA survey design via simulation testing. Currently the GoA BTS survey uses a stratified random design with 59 strata defined by bathymetry, bottom terrain, and statistical reporting designations (von Szalay and Raring, 2018). While upwards of 60 strata are not inherently too many strata, the delineations of the strata boundaries were subjectively chosen during a time where less information was known about the demersal species set. Furthermore, the existence of such numerous strata can cause problems computing sample-level stratum variances, as some strata can become undersampled to the extent that it is impossible to estimate a variance or variances are estimated with uncertainty too high to provide meaningful abundance estimates. From these preliminary results on the GoA survey design, an unbiased survey design can be optimized with less strata than used currently (e.g., 10-20 strata instead of 55-59). Integral to potentially changing the survey design in the GoA is understanding the current performance and tradeoffs of the present survey design. Metrics such as true CV, relative bias, and RRMSE of CV can be used to show any deficiencies in the current design and how to improve future survey designs and sampling allocations. The uncertainty associated with the sample CVs is related to its reliability as a data weight in some stock assessments (Francis, 2011) but is often overlooked in fisheries science despite such estimates themselves often being highly uncertain (Kotwicki and Ono, 2019). The slight negative bias in the sample CVs relative to the true CV, especially for highly variable species (*Sebastes* spp., Figures 8 and 9), contributed to the magnitude of the RRMSE of CV, and was expected given the patchy nature of these species' distributions. It is key to emphasize temporal variability in both the estimates and their associated uncertainties when evaluating and planning reliable and quality surveys.

These solutions are intended to objectively guide future survey designs we expect that the actual boundaries of the strata would be further modified based on expert opinion, logistical aspects of the survey operation, or other information sources prior to implementation. For example, the way the optimization partitioned depth and longitude resulted in unnatural longitudinal cuts that split islands, bays, and inlets. If this produces features that do not seem consistent with other data or knowledge of the system, other variables could be used to determine stratification and additional fine-scale habitat features could be incorporated as covariates in the operating model. Post-hoc, the shapes of the strata may also be changed to increase the feasibility and representation of the design. For example, some GoA groundfishes are managed within either three management areas or five management districts that roughly divide the domain into western/central/eastern areas. Work is currently ongoing to evaluate the effects of including these management strata either into the optimization as a separate stratum variable, conducting the optimization separately in each management strata, or through some post-stratification process. Survey teams may also be interested in the average distance among stations produced by optimal allocation, as logistical challenges may prevent certain parts of the spatial domain to be surveyed in a cost-efficient manner. For example in the current GoA BTS survey, one- and two-boat allocations currently do not sample the deepest strata due to time constraints. Survey design optimization packages like the SamplingStrata package (Barcaroli, 2014) can also incorporate survey costs with respect to survey duration per station or distance from port or limit the spatial domain to feasible depth ranges and trawlable (i.e., accessible) areas. The advantage of this systematic approach is that these modifications can be evaluated in a reproducible and transparent way to document the survey design process.

In addition to redesigning the stratification and sample allocation of existing surveys, the framework presented here could also be used to design surveys in new regions, or to optimize survey effort allocation within an existing stratification. However, applying this complete framework to optimize surveys may not always be feasible given the requirement of thorough species distribution modelling efforts to predict population density across the spatial domain at the resolution of the sampling unit. Fortunately, the optimization is a two-step process that first creates stratifications and then applies a multivariable optimal allocation algorithm (Bethel, 1989). Thus, in cases where a complete surface of density predictions is not available, the Bethel algorithm can be used on its own to provide optimal effort allocations given pre-specified strata boundaries and historical strata-level sampling means and variances. The framework of specifying CV constraints would be similar to our approach but without the implementation of a genetic algorithm to find optimal strata boundaries. For instance, we could have used the Bethel algorithm on the GoA survey example with the 59 previously defined strata, where data inputs would be the historical sample strata means and variances. This reduced version of the optimization framework could be applied as an intermediary approach, providing the time and additional data needed to complete the species distribution modeling necessary to perform the full optimization. Alternatively, survey planners could opt for one optimized stratified survey and adjust allocations using the Bethel algorithm based on potential future effort levels while making these new strata boundaries constant. We do not explicitly recommend that the stratification be changed between times with different sampling effort. However, if such changes were implemented, the survey time series would still be easily interpretable as we expect all stratified random sampling designs to produce unbiased estimates.

By leveraging the nearly 25-year time series of survey data, we can both incorporate the observed spatiotemporal variation to inform the design of the survey to meet a desired level of precision and continue to do so as data accrue over time. The updating of information over time reflects a major advantage of a survey design that can improve over time, and this framework is one way to provide an explicit but flexible framework for that process. That said, survey teams often have to contend with environmental changes that may cause species distributions to shift from their previously predicted distributions (e.g., Muhling et al., 2020). Such distribution shifts can influence both the optimality of the previous survey design and more fundamentally bias estimates due to changes in catchability and spatial availability. Survey designs can be flexibly optimized to account for environmental information and then updated based on short-term environmental forecasts. This could be done through an extension of our framework, by including the relevant dynamic environmental covariates in the operating model (e.g., Thorson, 2019). If such distribution shifts are recent or ongoing it may be prudent to conduct the optimization based on the predicted population densities in only the most recent years (e.g., Ault et al., 1999) .

Fisheries-independent surveys provide the foundation for scientifically sound fisheries management, thus the design of those surveys should be optimized for multiple scientific objectives. Using a heuristic approach, we designed a stratified survey design optimization that meets the objectives of producing precise abundance indices with minimal sampling intensity for multiple species. Major advantages of this approach are its explicit objectives of optimality and maximal precision, flexibility of inputs and constraints, and ability to communicate the expected impacts on the data products for downstream analyses. Systematically optimized survey designs can quickly accommodate rapid modifications in sample size or species prioritization that often arise as conditions change before or during a survey. The framework outlined here can be modified to incorporate different operational constraints (e.g., total sample sizes, inaccessible sampling units, and more detailed costs of sampling), species sets and species-specific precision constraints, and data inputs. Given the prevalence of multispecies surveys in fisheries and wildlife management among other applications, we hope that future survey design research will use and extend this approach for multispecies survey optimization to better balance objectives and further explore the tradeoffs inherent with surveying species with differing distributions of abundance.

Supplementary Material

The following supplementary material is available at *ICESJMS* online. Supplementary Material 1 contains technical details for the operating model. Supplementary Material 2 is the predicted spatial distributions for each species. Supplementary Material 3-10 contains additional result plots referred to in the main text.

Acknowledgements

We thank all staff of the Alaska Fisheries Science Center's Groundfish Assessment Program (GAP), program affiliates, volunteers, and charter captains and crew for collecting the data analyzed here during the Gulf of Alaska Groundfish Bottom Trawl Survey. This research was performed while ZSO held an NRC Research Associateship award at the NOAA Alaska Fisheries Science Center. We also thank the Gulf of Alaska and Aleutian Islands Team within GAP for their constructive feedback on earlier versions of this analysis. Finally, we thank James Thorson, Dana Hanselman and two anonymous reviewers for comments that improved the manuscript.

Data Availability Statement

The data and code underlying this article are available in the corresponding author's GitHub account (https://github.com/zoyafuso-NOAA/Optimal_Allocation_GoA_Manuscript).

References

Ault, J. S., Diaz, G. A., Smith, S. G., Luo, J., and Serafy, J. E. 1999. An efficient sampling survey design to estimate pink shrimp population abundance in Biscayne Bay, Florida. North American Journal of Fisheries Management, 19: 696-712.

Ballin, M., and Barcaroli, G. 2013. Joint determination of optimal stratification and sample allocation using genetic algorithm. Survey Methodology, 39: 369-393.

Barbeaux, S., Aydin, K., Fissel, B., Holsman, K., Laurel, B., Palsson, W., Shotwell, K., Yang, Q., and Zador, S. 2018. Assessment of the Pacific cod stock in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska. North Pacific Fishery Management Council.

Barcaroli, G. 2014. SamplingStrata: An R package for the optimization of stratified sampling. Journal of Statistical Software, 61: 1-24.

Bethel, J. 1989. Sample allocation in multivariate surveys. Survey Methodology, 15: 47-57.

Cao, J., Chen, Y., Chang, J. H., and Chen, X. 2014. An evaluation of an inshore bottom trawl survey design for American lobster (*Homarus americanus*) using computer simulations. Journal of Northwest Atlantic Fishery Science, 46: 27-39.

de Boois, I. J. (Ed). 2019. Moving towards integrated ecosystem monitoring. ICES Cooperative Research Report No. 347. 44 pp.

Field, S. A., Tyre, A. J., and Possingham, H. P. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. The Journal of Wildlife Management, 69: 473-482.

Francis, R. I. C. C. 2006. Optimum allocation of stations to strata in trawl surveys. New Zealand Fisheries Assessment Report, 23.

Francis, R. I. C. C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences, 68: 1124-1138.

Hutniczak, B., Lipton, D., Wiedenmann, J., and Wilberg, M. 2019. Valuing changes in frequency of fish stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1640-1652.

ICES. 2020. ICES Workshop on unavoidable survey effort reduction (WKUSER). ICES Scientific Reports, 2: 72. 92 pp.

Kotwicki, S., and Ono, K. 2019. The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys. Fish and Fisheries, 20: 760-774.

Lanthier, G., Boisclair, D., Bourque, G., Legendre, P., Lapointe, M., and Angers, B. 2013. Optimization of temporal versus spatial replication in the development of habitat use models to explain among-reach variations of fish density estimates in rivers. Canadian Journal of Fisheries and Aquatic Sciences, 70: 600-609.

Laurel, B. J., and Rogers, L. A. 2020. Loss of spawning habitat and prerecruits of Pacific cod during a Gulf of Alaska heatwave. Canadian Journal of Fisheries and Aquatic Sciences, 77: 644-650.

Miller, T. J., Skalski, J. R., and Ianelli, J. N. 2006. Optimizing a stratified sampling design when faced with multiple objectives. ICES Journal of Marine Science, 64: 97-109.

Mueter, F. J., and Norcross, B. L. 2002. Spatial and temporal patterns in the demersal fish community on the shelf and upper slope regions of the Gulf of Alaska. Fishery Bulletin, 100: 559-581.

Muhling, B. A., Brodie, S., Smith, J. A., Tommasi, D., Gaitan, C. F., Hazen, E. L., Jacox, M. G., *et al.* 2020. Predictability of species distributions deteriorates under novel environmental conditions in the California Current System. Frontiers in Marine Science, 7.

Muradian, M. L., Branch, T. A., and Punt, A. E. 2019. A framework for assessing which sampling programmes provide the best trade-off between accuracy and cost of data in stock assessments. ICES Journal of Marine Science, 76: 2102-2113.

NMFS, 2020. Fisheries of the United States, 2018. U.S. Department of Commerce, NOAA Current Fishery Statistics No. 2018. 167 pp.

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

Peel, D., Bravington, M., Kelly, N., Wood, S. N., and Knuckey, I. 2012. A model-based approach to designing a fishery-independent survey. Journal of Agricultural, Biological, and Environmental Statistics, 18: 1-21.

Regular, P. M., Robertson, G. J., Lewis, K. P., Babyn, J., Healey, B., and Mowbray, F. 2020. SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations. PLoS One, 15: e0232822.

Rhodes, J. R., and Jonzén, N. 2011. Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? Ecography, 34: 1040-1048.

Smith, S. G., Ault, J. S., Bohnsack, J. A., Harper, D. E., Luo, J., and McClellan, D. B. 2011. Multispecies survey design for assessing reef-fish stocks, spatially explicit management performance, and ecosystem condition. Fisheries Research, 109: 25-41.

Smith, S. J., and Hubley, B. 2013. Impact of survey design changes on stock assessment advice: sea scallops. ICES Journal of Marine Science, 71: 320-327.

Thorson, J.T., 2019. Measuring the impact of oceanographic indices on species distribution shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnology and Oceanography, 64: 2632-2645.

Thorson, J. T., and Barnett, L. A. K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74: 1311-1321.

Thorson, J. T., Fonner, R., Haltuch, M. A., Ono, K., and Winker, H. 2017. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. Canadian Journal of Fisheries and Aquatic Sciences, 74: 1794-1807.

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

v on Szalay, P.G. 2015. The impact of sample size reduction on the precision of biomass estimates in the Gulf of Alaska. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-297. 17 pp.

v on Szalay, P. G., and Raring, N. W. 2018. Data report: 2017 Gulf of Alaska bottom trawl survey. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-374. 260 pp.

v on Szalay, P. G., Raring, N. W., Shaw, F. R., Wilkins, M. E., and Martin, M. H. 2010. Data report: 2009 Gulf of Alaska bottom trawl survey. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-208. 245 p.

Wang, J., Xu, B., Zhang, C., Xue, Y., Chen, Y., and Ren, Y. 2018. Evaluation of alternative stratifications for a stratified random fishery-independent survey. Fisheries Research, 207: 150-159.

Xu, B., Ren, Y., Chen, Y., Xue, Y., Zhang, C., and Wan, R., 2015. Optimization of stratification scheme for a fishery-independent survey with multiple objectives. Acta Oceanologica Sinica, 34: 154-169.

Zador, S., Yasumiishi, E., and Whitehouse, G. A. 2019. Ecosystem Status Report 2019 Gulf of Alaska. North Pacific Fishery Management Council. 232 pp.

Zimmermann, F., and Enberg, K., 2016. Can less be more? Effects of reduced frequency of surveys and stock assessments. ICES Journal of Marine Science, 74; 56-68.

Figures

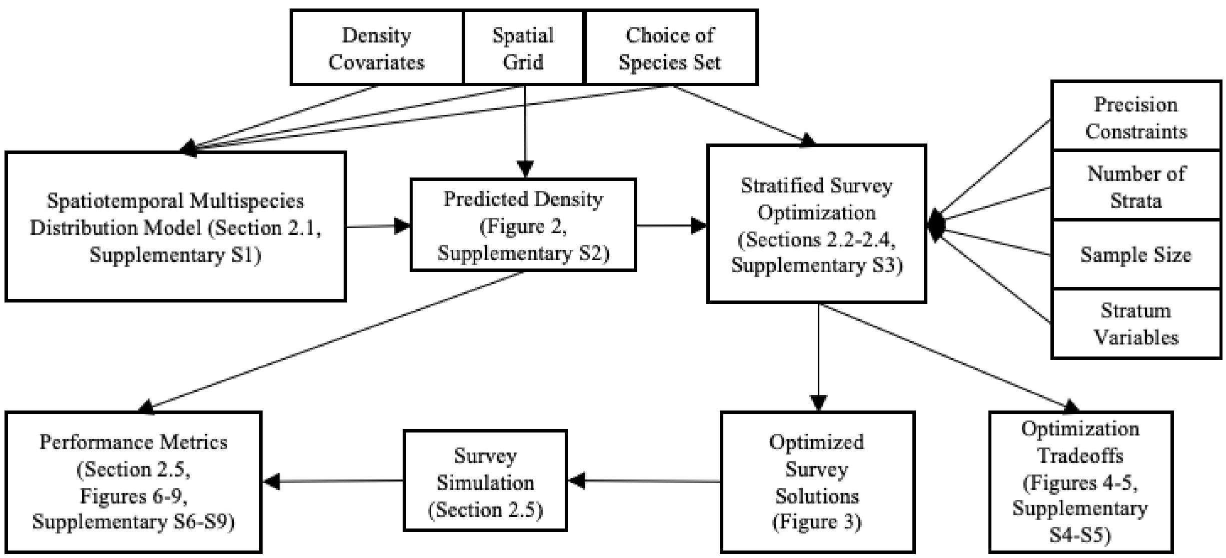

Figure 1: F lowchart of the multispecies stratified survey optimization.

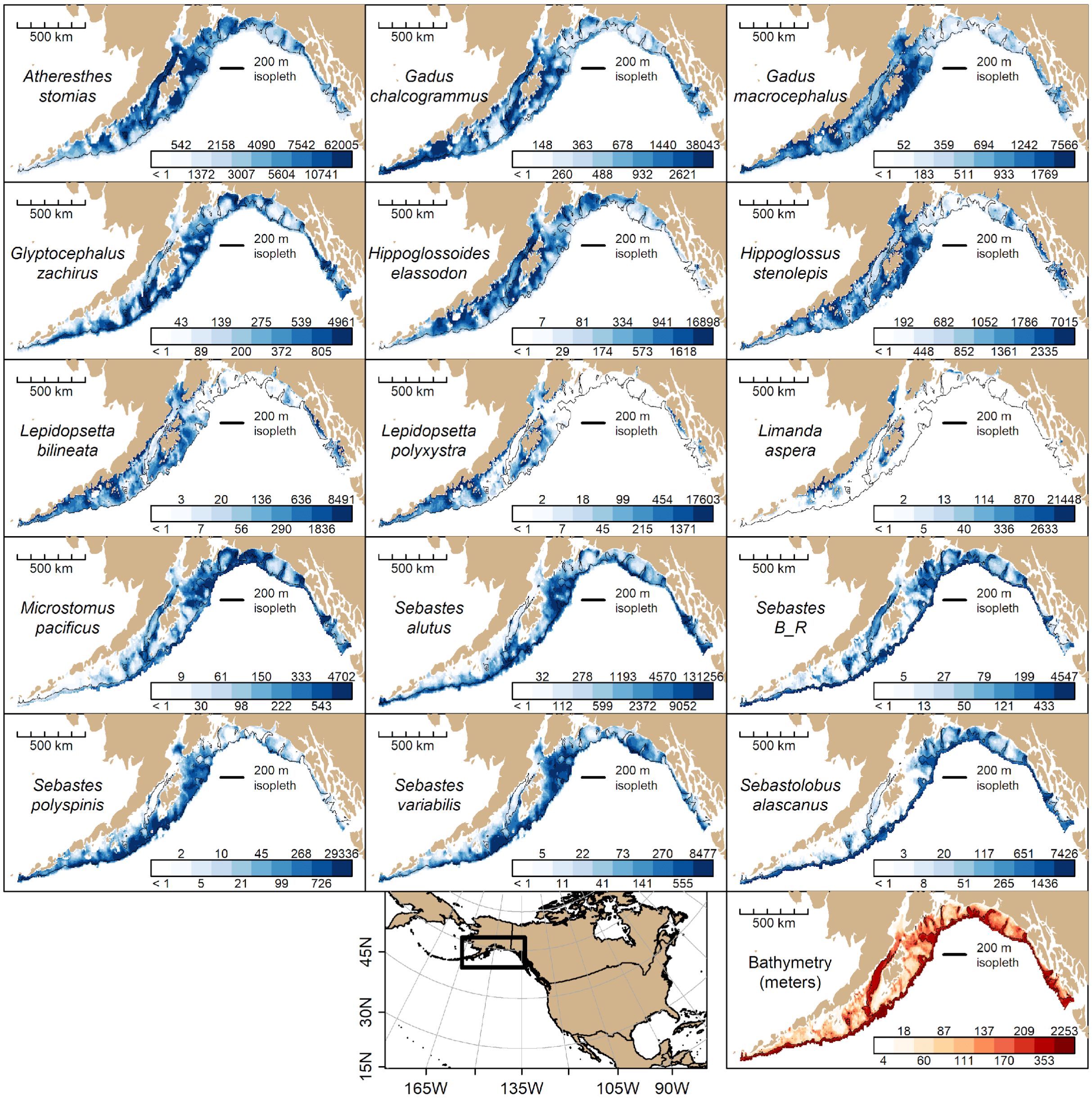

Figure 2: Predicted mean density across years (kg km^{-2}) for each species included in the survey optimization across the Gulf of Alaska. Bottom right panel shows the bathymetry within the survey footprint along with the 200 m isobath , which is a general delineation of species distributions. Refer to the Supplementary S1 for a brief explanation of the operating model used to produce these predicted densities and Supplementary S2 for predicted densities by year.

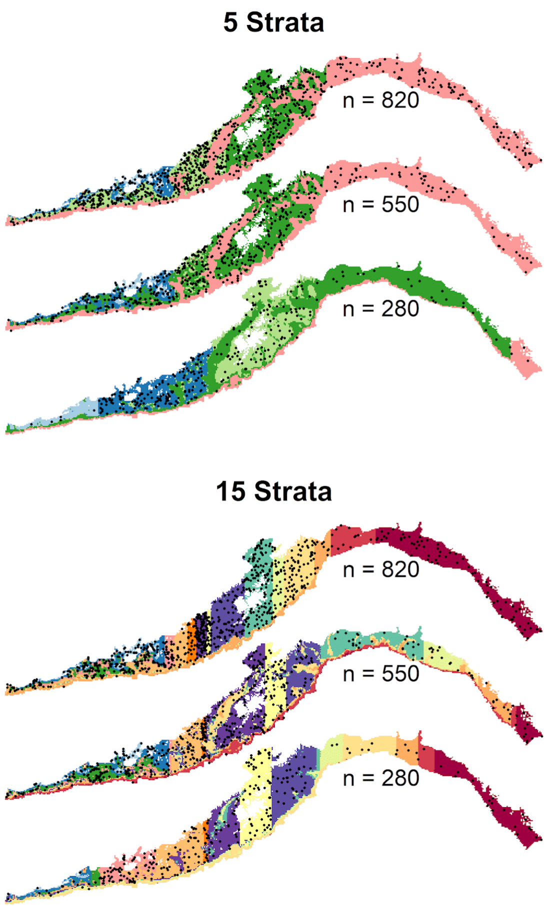

Figure 3: Representative examples of strata boundary maps arising from solutions for the species-specific CV constraint optimization for five, ten, and fifteen strata across the three effort (boats) scenarios with simulated stations randomly sampled according to each optimized stratified survey superimposed. The colors represent different strata .

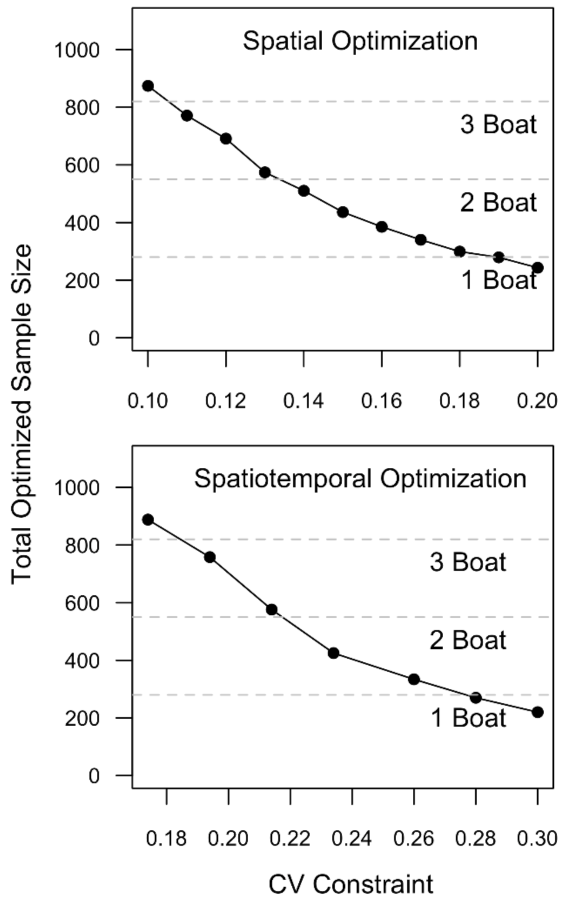

Figure 4: Total optimized sample size (number of stations) across coefficient of variation (CV) constraint, accounting only for spatial variability (top) or both spatial and temporal variability (bottom). The five-strata optimization solutions are shown, but qualitative results were consistent across strata (Supplementary S4). Both optimizations were conducted under the one-CV constraint approach where all species have the same CV constraint in the optimization. Horizontal dotted grey lines indicate the sampling levels for one, two, and three boat-effort scenarios .

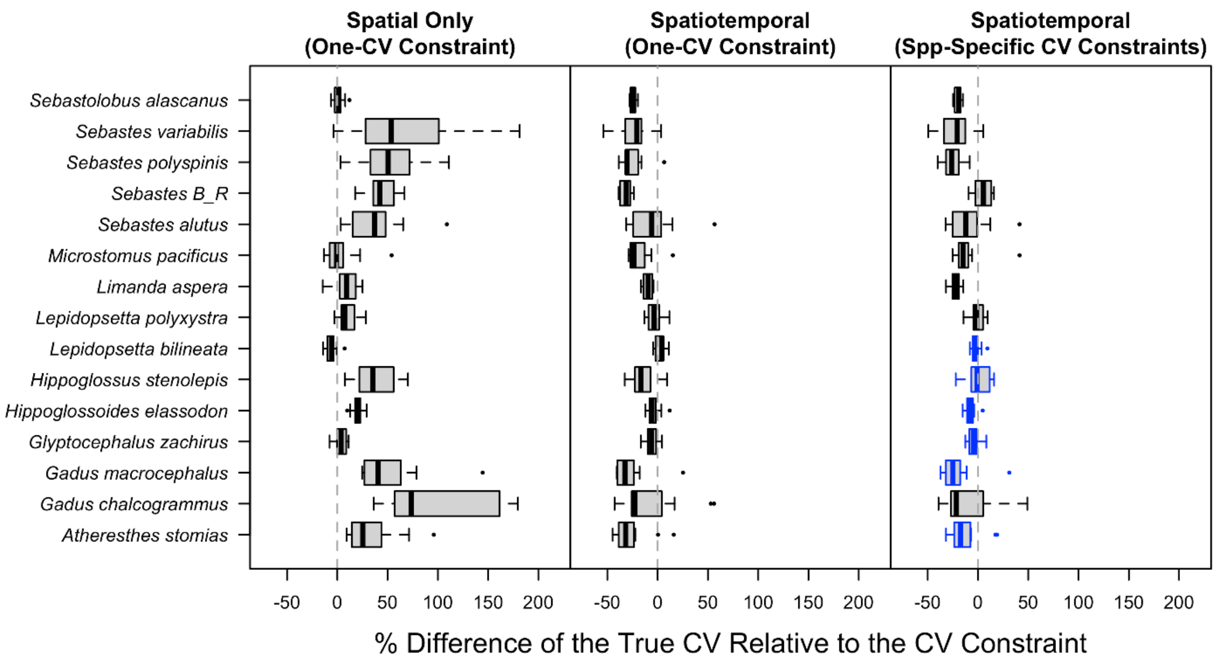

Figure 5: Comparison of the relative difference between expected and realized coefficient of variation (CV) of abundance. Specifically, this shows the distribution of percent differences of the true CVs, calculated for each year, relative to the CV constraint associated with a five-strata, two boat -effort scenario (n = 550) for all included species. The left and center plots show optimizations using the one-CV constraint approach. The right plot shows an optimization using the species-specific CV constraint approach (refer to the main text for how CV constraints were specified across species). For the species-specific CV constraint approach, a value of 0.10 was chosen as the lowest a population CV constraint could be specified (indicated by the blue borders). A positive value indicates that the observed true CV is greater than the CV constraint that was specified in the optimization. A negative or near-zero value indicates that the observed true CV is within the CV constraint specified in the optimization. Results were qualitatively consistent with other total effort and strata scenarios.

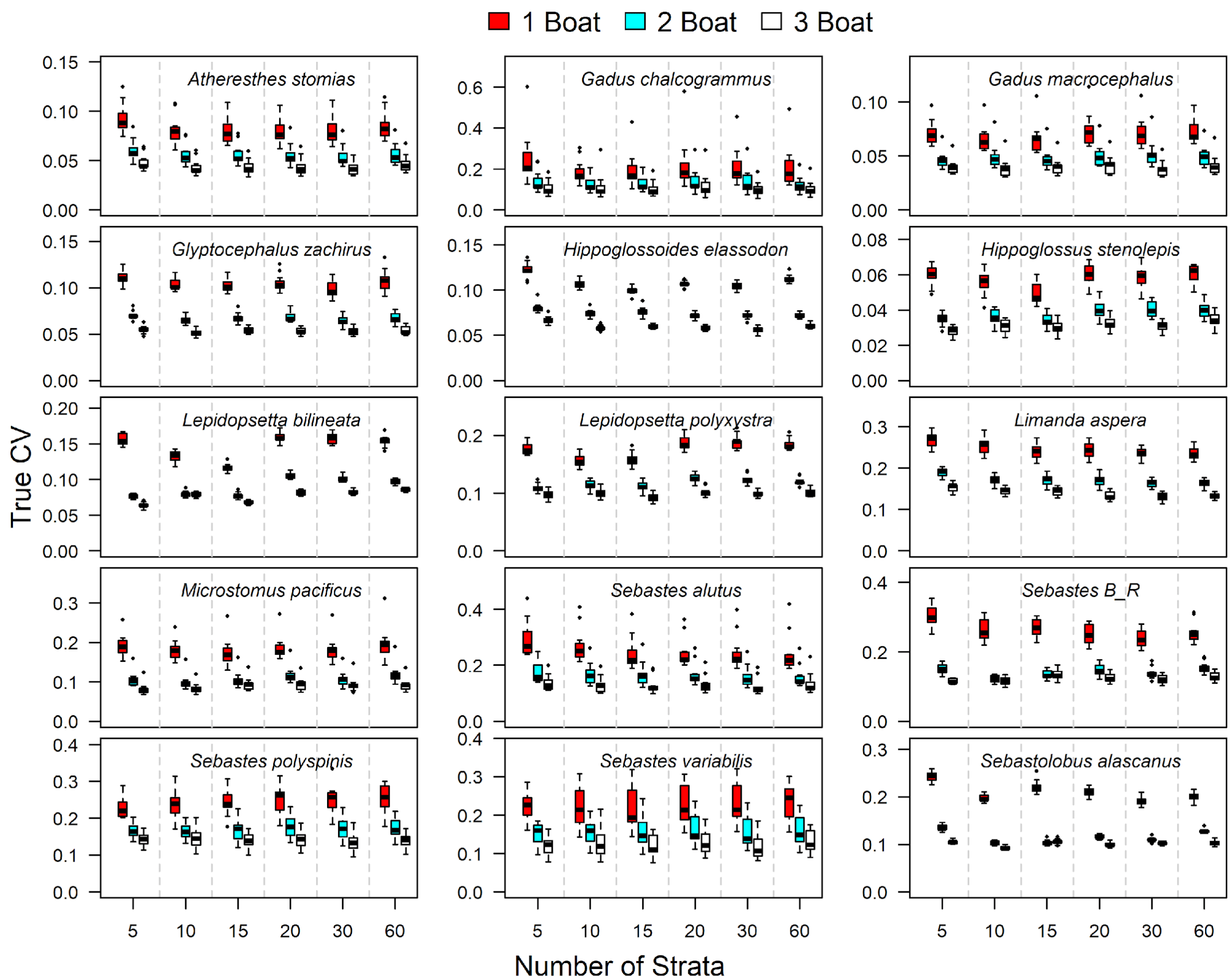

Figure 6: Distribution of true coefficient of variation (CV) across observed years for each species, level of sampling effort (color) and number of strata for the species-specific CV constraint approach.

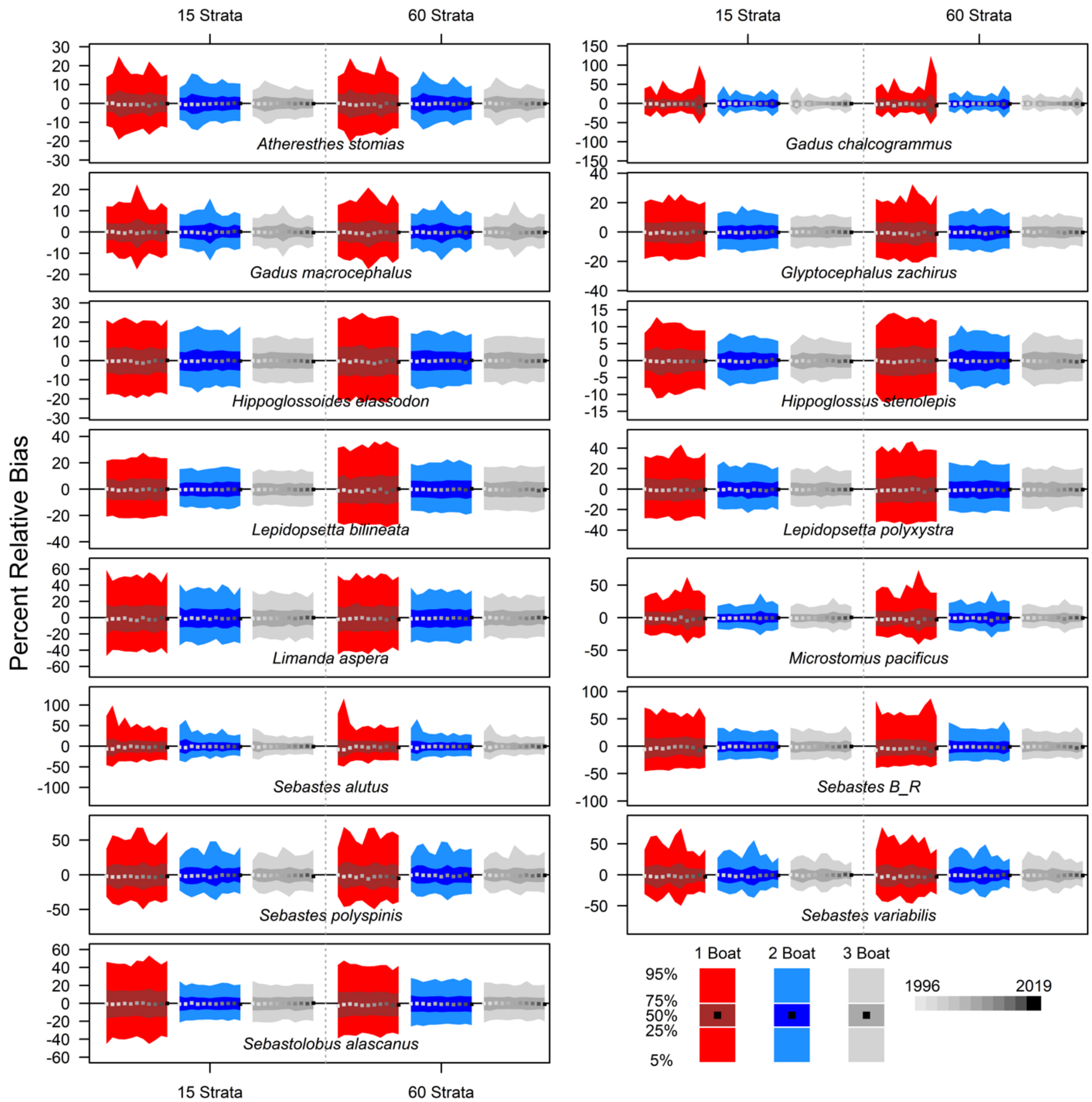

Figure 7: Distribution of percent relative bias in the simulated mean density estimates across years relative the true mean density for each species, level of sampling effort (color) and two strata levels (15 and 60 to represent the range investigated) for the species-specific CV constraint approach. Results were similar for the one-CV constraint approach (Supplementary S8).

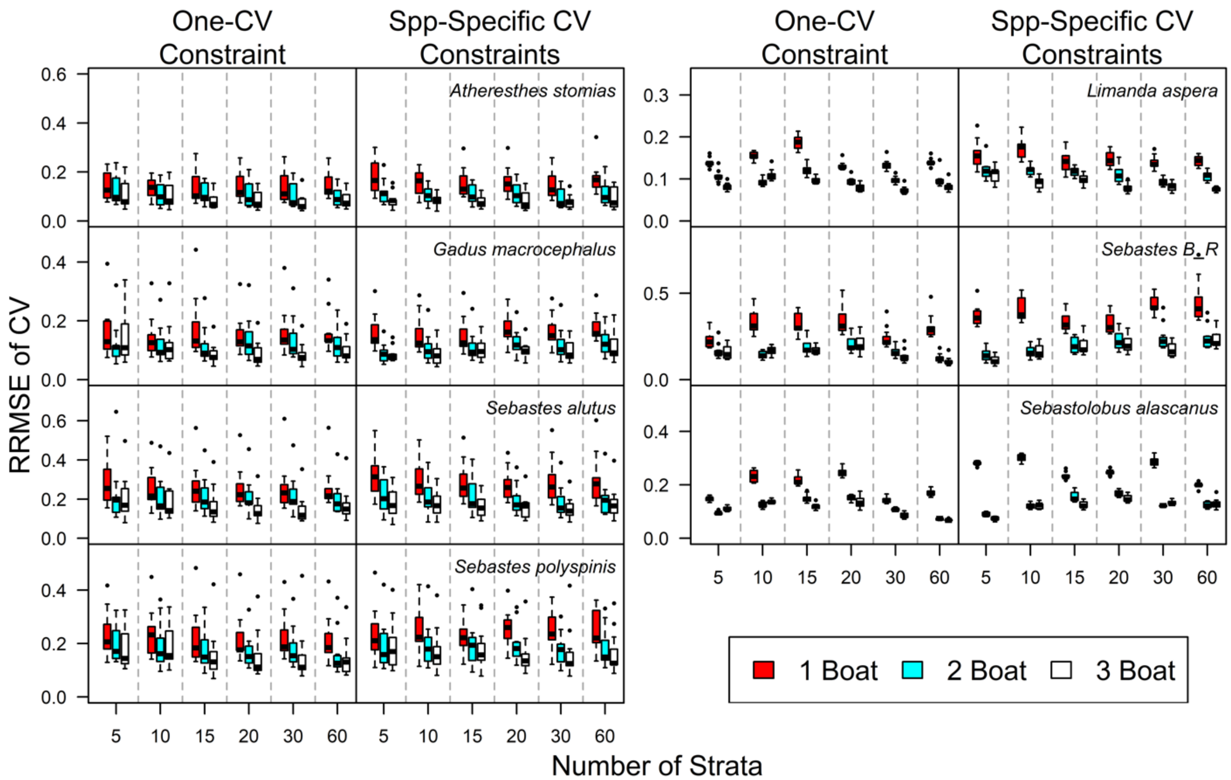

Figure 8: Distribution of relative root mean square error (RRMSE) of the coefficient of variation (CV) across observed years for a subset of species (see Supplementary S7 for a full version), level of sampling effort (color) and number of strata for the one-CV constraint approach (left set of plots) and species-specific CV constraint approach (right set of plots) .

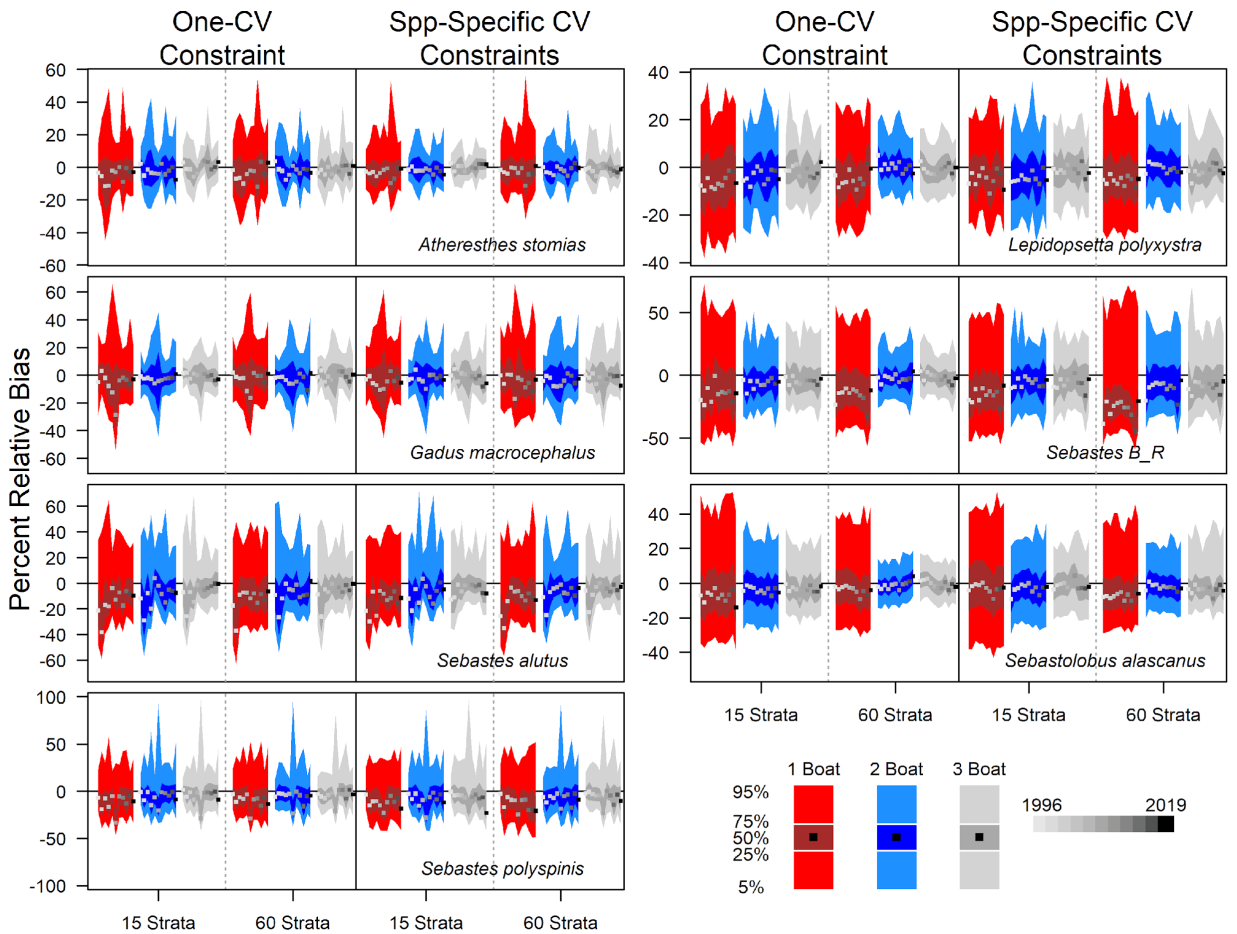

Figure 9: Distribution of percent relative bias in the simulated coefficient of variation (CV) estimates across observed years relative the true CV for a subset of species (see Supplementary S9-10 for a full version), level of sampling effort (color) and two strata levels (15 and 60 to represent the range investigated) for the species-specific CV constraint approach.