Estimating abundance of spatially aggregated populations: comparing adaptive sampling with other survey designs

Kathryn L. Mier and Susan J. Picquelle

Abstract: The main goal in estimating population abundance is to maximize its accuracy and precision. This is difficult when the survey area is large and resources are limited. We implemented a feasible adaptive sampling survey applied to an aggregated population in a marine environment and compared its performance with five classical survey designs. Specifically, larval walleye pollock (*Theragra chalcogramma*) in the Gulf of Alaska was used as an example of a widespread aggregated population. The six sampling designs included (*i*) adaptive cluster, (*ii*) simple random, (*iii*) systematic, (*iv*) systematic cluster, (*v*) stratified systematic, and (*vi*) unequal probability. Of the five different adaptive estimators used for the adaptive cluster design, the modified Hansen–Hurwitz performed best overall. Of the six survey designs, the stratified systematic survey provided the best overall estimator, given there was accurate prior information on which to base the strata. If no prior information was available, a systematic survey was best. A systematic survey using a single random starting point with a simple random estimator performed as well as and sometimes better than a systematic cluster survey with two starting points (clusters). The adaptive cluster survey showed no advantages when compared with these two designs and furthermore presented substantial logistical challenges.

Résumé : L'objectif principal poursuivi dans l'estimation de l'abondance d'une population est l'amélioration de l'exactitude et de la précision. Cela est difficile quand la surface inventoriée est grande et les ressources limitées. Nous avons mis au point un inventaire adaptatif d'échantillonnage pratique pour une population à distribution contagieuse dans un environnement marin et nous avons comparé sa performance en fonction de cinq plans d'inventaire classiques. Nous utilisons, en fait, des larves de goberges (*Theragra chalcogramma*) de l'Alaska du golfe de l'Alaska comme exemple spécifique d'une population à large répartition et à distribution contagieuse. Les six plans d'inventaire consistent en (*i*) un plan adaptatif avec regroupements, (*ii*) un plan aléatoire simple, (*iii*) un plan systématique, (*iv*) un plan systématique avec regroupements, (*v*) un plan systématique stratifié et (*vi*) un plan à probabilités inégales. Des cinq estimateurs adaptatifs utilisés dans le plan adaptatif avec regroupements, l'estimateur modifié de Hansen–Hurwitz donne le meilleur résultat global. Des six plans d'inventaire, l'inventaire systémique stratifié fournit le meilleur estimateur global, étant donné qu'il existe de l'information préalable précise pour déterminer les strates. Lorsqu'il n'y a pas de renseignements préalables, l'inventaire systématique fonctionne le mieux. Un inventaire systématique avec un seul point de départ aléatoire et avec un estimateur aléatoire simple fonctionne aussi bien, et souvent mieux, qu'un inventaire systématique avec regroupements avec deux points de départ (regroupements). L'inventaire adaptatif avec regroupements ne présente aucun avantage par rapport aux deux plans précédents et, de plus, il crée de sérieux problèmes de logistique.

[Traduit par la Rédaction]

Introduction

The most common requirement in resource management, studies of population dynamics, and many other subjects is to estimate mean abundance of spatially aggregated populations. The usefulness of these estimates are dependent on their bias and precision. Yet, in most field studies, traditional survey procedures are used without much background information on how to improve the bias and precision of the mean and standard error estimators. This can lead to inaccurate fish stock assessments or hypothesis tests and general misrepresentation of the spatial distribution.

The goal of this project was to compare the effectiveness of six sampling methods on patchy distributions (i.e., adaptive cluster, random, systematic, systematic cluster, stratified systematic, and unequal probability). The criteria for comparing designs were the bias, precision, and mean squared error (MSE) of the estimates. Because MSE is a measure of

Received 2 December 2006. Accepted 26 July 2007. Published on the NRC Research Press Web site at cjfas.nrc.ca on 17 January 2008.
J19684

K.L. Mier[1] and S.J. Picquelle. Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, WA 98115, USA.

[1] Corresponding author (e-mail: kathy.mier@noaa.gov).

doi:10.1139/F07-138

both bias and precision, it was used as a final criterion for determining the best estimator. We computed these estimates of the mean density and their standard error (SE) by simulating many patchily distributed populations with different degrees of patchiness and then repeatedly sampling these populations using each sampling method.

A primary motivation for this study was to decide if using an adaptive cluster design (Thompson 1990) would yield a more precise estimate. This design has received much attention in recent years with studies of patchily distributed populations, primarily because it allows concentrated sampling around initial samples that meet a predetermined criterion (typically areas of high density) and because it is reported to yield a smaller variance than random sampling. It can be particularly useful when more individuals of interest need to be collected for other studies.

In our study, we examine the performance of adaptive cluster sampling by sampling from realistic, simulated populations. Other examples of adaptive sampling applied to simulated patchily distributed populations include Christman (2000), Brown (2003), and Su and Quinn (2003). Christman (2000) looked at adaptive cluster sampling of rare, spatially clustered populations. In contrast, our population of interest is patchy or clustered, but not rare, and covers a large geographic extent. Our simulated populations are typical of many ubiquitous but patchy marine species, such as adult Pacific hake (*Merluccius productus*) and shortspine thornyhead (*Sebastolobus alascanus*) along the west coast of North America, adult arrowtooth flounder (*Atheresthes stomias*), rex sole (*Glyptocephalus zachirus*), and Pacific sandlance (*Ammodytes hexapterus*) in the Gulf of Alaska, and many species of ichthyoplankton; hence, our survey design comparisons have broad applicability. To further broaden the scope of our study, we simulated populations with a wide range of patchiness.

What also sets our paper apart from other simulations is the focus on a realistic implementation of an adaptive sampling design where the sampling unit is extremely small when compared with the entire area to be surveyed, a scenario that is common in large marine surveys. Our simulated population contained almost 9000 sampling units, whereas, for example, Thompson (1991) used a population with 400 sampling units, and Christman (2000) used a population with 75 sampling units. Our implementation includes setting the criterion for initiating adaptive sampling, controlling sample size, and avoiding negative variance estimates (which occurs in Horvitz–Thompson estimates). These issues have not been adequately addressed in real applications or in simulated samples. Much of the literature a priori uses species presence to initiate adaptive sampling, makes no attempt to control sample size, and seldom yields negative variances due to artificially simple populations. Those studies that have introduced methods to control sample size include Lo et al. (1997), Christman (2003), and Su and Quinn (2003). Each of these methods introduces bias into the estimator, as does our method. We measured the bias, or systematic error, by simulating realistic populations and then repeatedly sampling from these populations and comparing the true values to the repeated estimates.

Common estimators used for adaptive cluster sampling include the modified Hansen–Hurwitz (HH) and Horvitz–Thompson (HT) estimators of the mean and variance (Thompson and Seber 1996). In addition to these two, we also looked at two alternate estimators of the variance of the HT mean, the Yates–Grundy–Sen (Sen 1953; Yates and Grundy 1953) and the Brewer–Hanif (Brewer and Hanif 1983), as well as one other estimator of both the mean and variance, the Hajek–Sarndal (Hajek 1971; Sarndal et al. 1992). Patchy distributions usually yield large variances, regardless of survey design (Andrew and Mapstone 1987), but several studies suggest that using the modified HT estimator in an adaptive cluster survey yields the most precise estimate (Hanselman et al. 2003; Salehi 2003; Su and Quinn 2003). These conclusions, however, were dependent on the degree of aggregation, "neighborhood" definition, size of "networks", within-network variance, size of "initial sample", size of "sampling unit", and the "criterion" used to begin adaptive sampling (see Appendix A for definitions of terms that are in quotes).

We apply these methods to estimating mean density (number per 10 m^2) of larval fish in the ocean, and specifically to the patchily distributed walleye pollock (*Theragra chalcogramma*) (Stabeno et al. 1996). We patterned our simulated populations on the distribution of walleye pollock larvae as recorded from ichthyoplankton surveys in Shelikof Strait, Alaska, conducted by the National Oceanic and Atmospheric Administration (NOAA) Alaska Fisheries Science Center (AFSC). We built the framework used in this study by modeling the spatial distribution in terms of patches and using parameters that describe these patches based on actual historical data.

Materials and methods

Population simulations

Artificial populations were generated that have the same spatial characteristics as walleye pollock larvae in early May in Shelikof Strait, Alaska. The primary characteristic is the spatial pattern of its patches, that is, areas of high density (Stabeno et al. 1996). We examined the observed larval walleye pollock spatial distributions from nine spring ichthyoplankton surveys conducted by AFSC between 1986 and 1998 (Table 1). We chose surveys from the time interval shortly after the eggs hatched into larvae and before the larvae became widely dispersed, so that the patches were still coherent.

To examine the effect of patchiness on the performance of the various survey designs, we simulated two sets of populations: 50 populations that we considered to be very patchy and 50 populations that were much less patchy. We defined high-patchy populations to have patches that were small in geographic extent, but had large maximum densities within the patch. Conversely, we defined low-patchy populations to have patches that were large in geographic extent, but with maximum densities that were less extreme than the high-patchy population (Fig. 1). We adopted these definitions both because they are intuitive and because Brown (1996) found that these factors (along with the number of patches) were the most important in determining the benefits of adap-

Table 1. List of cruises used to describe spatial distribution of larval walleye pollock (*Theragra chalcogramma*).

Cruise	Dates	No. of stations
2MF86	4–17 May 1986	89
2MF89	26 April – 2 May 1989	45
3MF89	9–16 May 1989	94
2MF90	8–13 May 1990	51
3MF92	3–9 May 1992	64
4MF93	6–11 May 1993	66
5MF94	5–8 May 1994	50
1DI96	30 April – 5 May 1996	73
4MF98	3–5 May 1998	45

tive sampling over random sampling. By simulating populations with a wide range of patchiness, the conclusions drawn from these populations may be extrapolated to many other natural populations that have various degrees of patchiness.

The area of the simulated populations was defined by plotting the stations from cruises in Table 1 and delineating the area that was usually surveyed and contained larvae (Fig. 2). A grid of 8857 "cells", where each cell is 1 nautical mile × 1 nautical mile (1 n.mi. = 1.852 km) sea-surface area, was created inside this survey area. A density was simulated for each cell based on patch parameters estimated from historical data. This density was defined as the number of fish in the water column below a 10 m^2 surface area within the cell. We used the index i to refer to the ith cell; x_i and y_i to refer to the east–west and north–south location of the centroid within cell i, respectively, in units of meters; and $Z(x_i, y_i)$ to refer to the simulated density, in units of number per 10 m^2, within cell i. See Appendix B for details of the simulation of populations.

Survey simulations

The simulated surveys were based on quadrat sampling, which is common practice in ecological surveys. The geographic range of the simulated population was divided into a grid of 1 n.mi. × 1 n.mi. contiguous cells; this was our spatial sampling frame and each cell was a quadrat, the sampling unit. A sample of cells was selected and the population density within a cell was observed and multiplied by a sampling error term, ε_i. The density of fish "caught" in the ith sampled cell (number per 10 m^2) was designated as

$$z(x_i, y_i) = z_i = Z(x_i, y_i) \times \varepsilon_i$$

where $Z(x_i, y_i)$ is the simulated density of the population at that cell (see eq. B1 in Appendix B), and ε_i is a random lognormal variate, representing small-scale variability about the densities within a cell. The variance for the lognormal error term was obtained from the variance within stations with multiple hauls in the historic database, that is, where many samples were taken at the same location on the same day.

For each of the 100 simulated populations, 100 simulated samples were generated using each of the six survey designs. Each survey was designed to have approximately the same sample size, n, to facilitate the comparisons of the designs. Mean larval density was estimated from each simulated sample, and the ability of the survey method to recover the parameters of the simulated population was examined by measuring bias and precision.

For every survey design except random, we grouped the grid of 8857 cells into 10 n.mi. × 10 n.mi. "blocks", so that each block contained 100 cells (Fig. 3). For the unequal probability sampling and the stratified systematic sampling, we calculated the mean densities within each block from all 100 simulated populations and used these to create probabilities for sampling within each block and for creating strata boundaries. Averaging densities over all 100 populations is comparable to using data from a time series of previous surveys to set sampling parameters for a future survey, as opposed to conducting a pilot survey to determine sampling parameters.

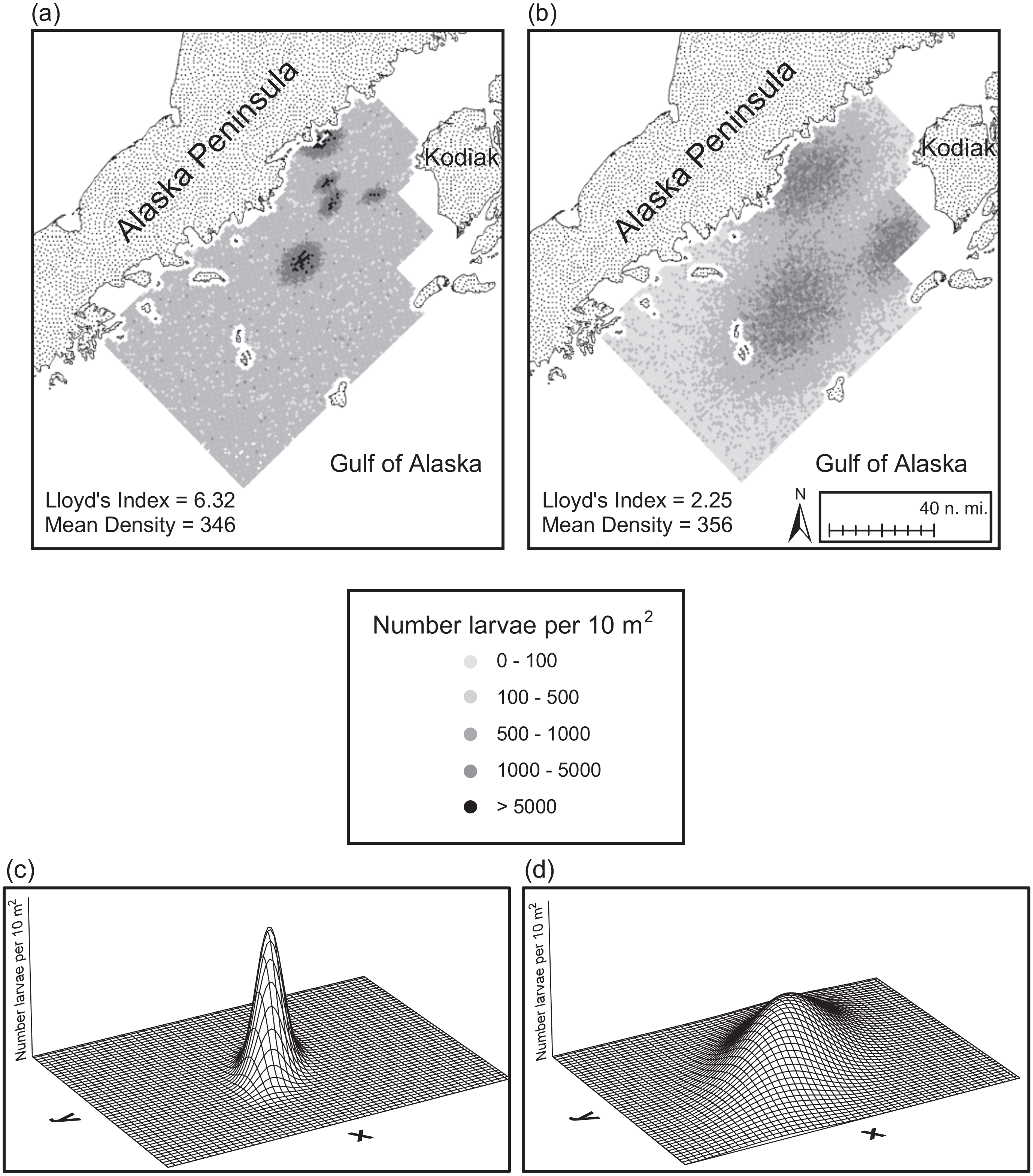
Graphic examples of each survey design are provided (Fig. 4). The formulas for the unbiased estimators of the mean (μ) and variance (σ) are provided for each classic survey design in Appendix C, Table C1 and for the adaptive cluster survey design in Appendix D, Table D1. For the sake of broad applicability, all estimates include the finite population correction factor, $(N - n)/N$, even though this number is very close to unity because our value of N is so large relative to n.

For systematic cluster, stratified systematic, and the adaptive cluster designs, we used conventional "cluster" sampling where each of the 100 cells within blocks is a single "primary unit" or cluster of cells. The cells within a primary unit are called "secondary units" and have the same position in each block. This follows Cochran's aligned or "square grid" systematic sample in two dimensions (Cochran 1977). A primary unit, or randomly chosen starting position, was sampled, then the corresponding secondary units were sampled per primary unit, in the same relative positions in each block.

Adaptive cluster sampling design

In adaptive cluster sampling, if the density at a sampled unit satisfies a preset criterion, then further sampling occurs around that unit (Thompson 1991). We followed Thompson's terminology, except for his definition of cluster. (Note: Thompson's definition of a cluster is a group of adaptively sampled stations, as opposed to Cochran's definition of cluster, as defined above.) The initial sample was a systematic cluster sample, similar to what we used for the systematic cluster survey; these stations were occupied independently of what previous observations were made. The sampling pattern for the additional adaptive stations was that of a square grid about the initial station, with adaptive stations 5 n.mi. immediately northeast, southwest, northwest, and southeast of the initial unit that met the criterion. These additional four stations comprise the neighborhood of the initial station. If any of these stations met the criterion, additional stations were added using the same pattern and spacing (Fig. 5). We chose this spacing because it (*i*) was smaller than the observed patch size (mean patch size was 10 $n.mi.^2$), (*ii*) helped limit the sample size, and (*iii*) overlapped the initial grid, insuring adjacent neighborhoods had coincident stations. At this point, our adaptive procedure ends and the next station on the initial grid is sampled, unlike Thompson's procedure, where the adaptive sampling continues until no more stations meet the criterion. The set of overlapping neighborhoods all spawned by a single initial

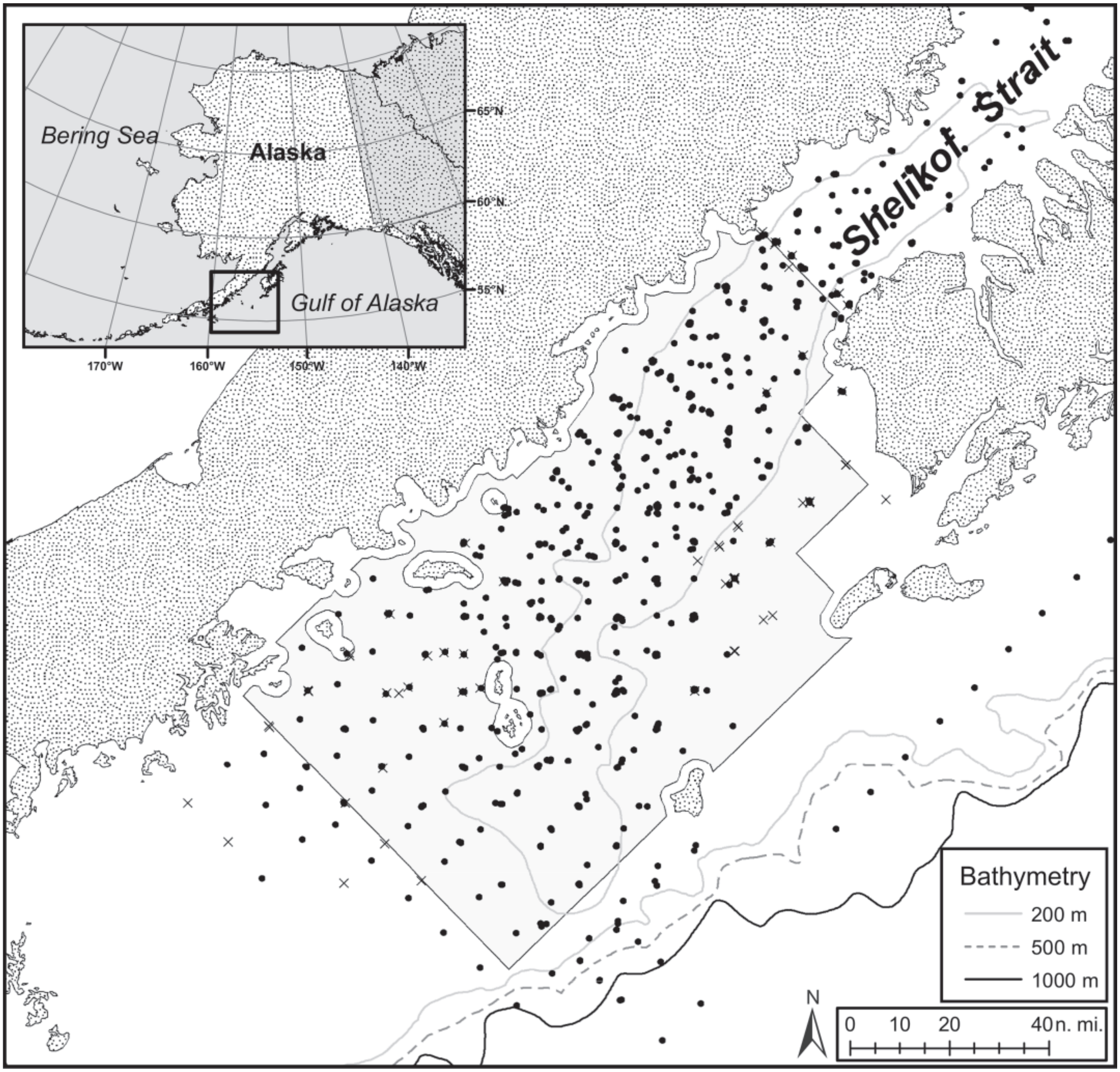
Fig. 1. Simulated populations of (*a*) high patchiness and (*b*) low patchiness. Single simulated patch of (*c*) high patchiness and (*d*) low patchiness based on bivariate normal probability distribution (see Appendix B for details of simulating a single patch).



station, excluding stations within the neighborhoods that do not meet the criterion, is called a network.

To control the sample size, we employed three methods: (*i*) restrict the number of iterations of testing for the criterion, (*ii*) use noncontiguous sampling units to form neighborhoods, and (*iii*) stop adaptive sampling and finish sampling the survey area with a systematic design in a second strata. In the first method, for each station in the initial sample, we allowed only two iterations of adding adaptive stations, that is, if an “A” station (initial sample) met the

Fig. 2. Map of Shelikof Strait in Alaska, USA, showing stations with 0 catch (shown with × symbols) and stations with positive catches (shown with solid dots), with boundary of simulated population.



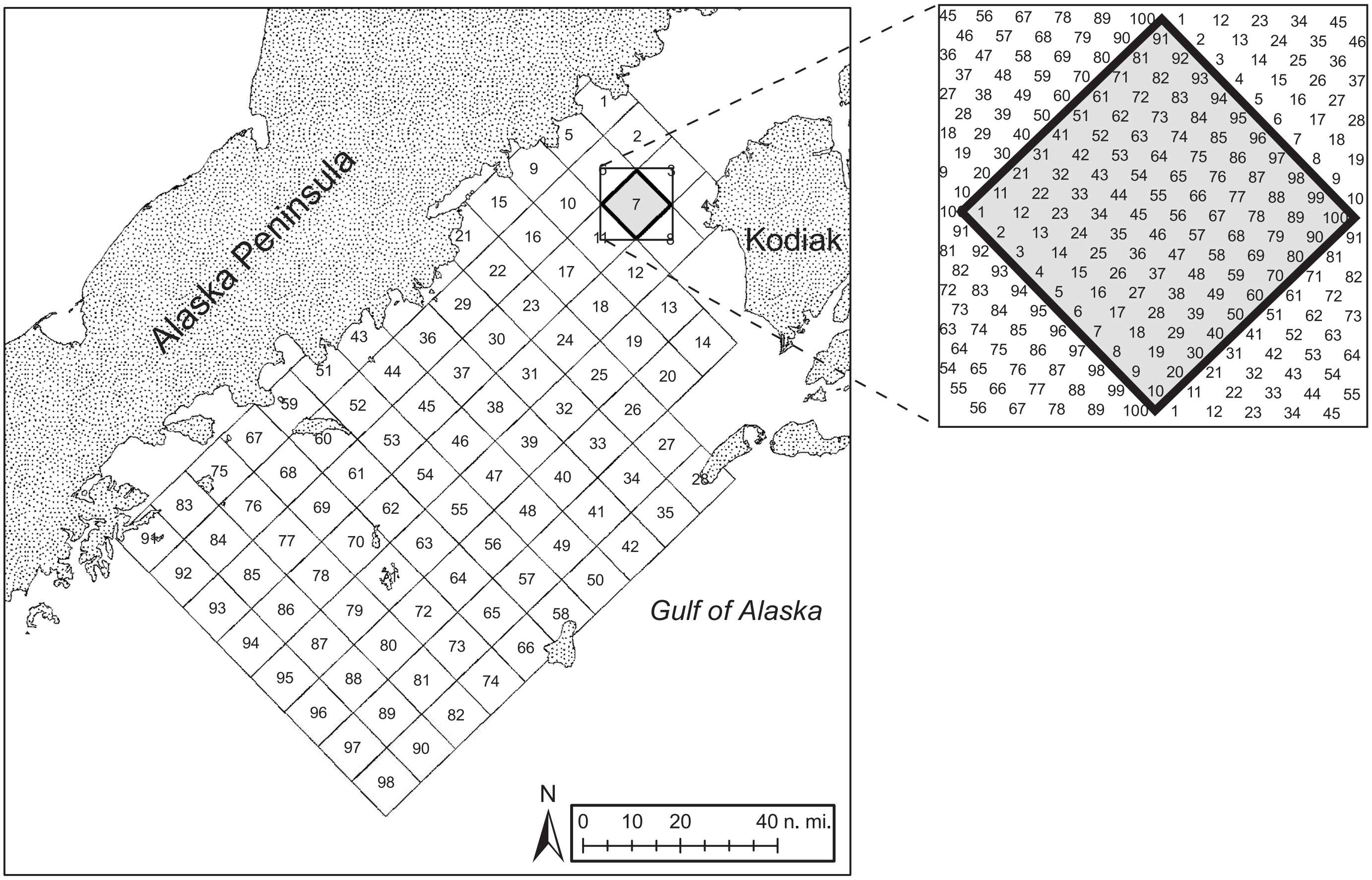
criterion, then the four stations in the neighborhood of A were sampled (called "B" stations). If any B station met the criterion, then the stations in the neighborhood of B stations were sampled (called "C" stations). Then we stopped the adaptive sampling and continued with the initial sampling plan. Quinn et al. (1999) called this a stopping rule with two iterations. This yielded possible sampling of 12 additional stations per A station (4 possible B stations plus 8 possible C stations). We called this a "district" (Fig. 5).

The second method we used to reduce sample size was to define the neighborhood to be noncontiguous grid cells. The spacing of the adaptive stations five cells apart allowed networks to grow beyond the extent of a single district, that is, C stations overlapped A stations, therefore allowing networks to overlap and become one large network (Fig. 5).

The third method for controlling sample size, that is dividing the survey area into two strata, was accomplished by not fixing the strata boundary before the survey; instead, we ceased adaptive sampling when the number of stations left to sample before reaching the desired sample size equaled the number of blocks yet to be sampled. This is a form of poststratification; therefore, the appropriate stratified estimator was used.

The criterion value used to initiate adaptive sampling was a density greater than the 85th percentile of the densities from the 100 simulated populations. This is equivalent to

Fig. 3. Map showing the survey area gridded into 10 nautical mile (1 n.mi. = 1.852 km) × 10 n.mi. blocks, labeled 1–98, each of which is divided into 1 n.mi. × 1 n.mi. cells, labeled 1–100.



using order statistics as recommended by Thompson and Seber (1996). Basing the criterion on all 100 simulated populations is equivalent, in practice, to using past survey data, as opposed to conducting a pilot survey, to compute the criterion, similar to how we computed the sampling parameters for unequal probability sampling and stratified systematic sampling. We determined this percentile by calculating the proportion of A stations that we could allow to be adaptively sampled to meet the desired sample size.

The estimators commonly used for adaptive sampling are the Hansen–Hurwitz (HH) and the Horvitz–Thompson (HT) estimators. We considered three additional estimators, and all five are described in Appendix D. Each of the five adaptive estimators was applied to stratum 1 only. The final estimate for each sample was computed by combining the two strata estimates using the usual stratified estimators (see Appendix C, Table C1). The variance of stratum 2 was estimated using the estimator from systematic sampling, treated as random, because it had only one cluster.

Classic sampling designs

Random sampling is sampling with equal probability, while unequal probability sampling assigns different selection probabilities to the individual sampling units. Sampling with probability proportional to a measure of size is a special type of unequal probability sampling (Jessen 1978). In our case, the probability was proportional to the mean density within each 10 n.mi. × 10 n.mi. block, averaged over all 100 simulated populations. This mean density was our measure of size.

The systematic survey design had only one primary unit (starting point) to demonstrate a very common survey design, because multiple starting points for systematic sampling are seldom used in practice. Traditionally, the AFSC ichthyoplankton surveys use a modified systematic sampling design. It is common in marine science to sample systematically along a grid rather than randomly to get even coverage of the survey area and to make the most efficient use of ship time. This is often done with a single random start, which theoretically would yield a “cluster” of one sample. Because no variance can be calculated from a sample of one cluster (Cochran 1977), a simple random sample estimator is used with this design, where each station is considered a sampling unit, with the assumption that there is no spatial periodicity in the data that coincides with the systematic spacing of the sample (Stehman and Overton 1994). Any deviation from this assumption will appear as bias in the estimates of the mean and variance. If one uses more than one random starting point, resulting in multiple clusters, then a variance can be calculated (Cochran 1977). We call this systematic cluster design, and we simulated this design using two random starting points.

Stratified systematic survey designs are obtained by dividing the elements of the sampling frame into groups, or strata, such that the variability within strata is believed to be less than the variability between strata. For simplicity, we

Fig. 4. Maps showing examples of the spatial arrangement of six different survey designs: (*a*) adaptive cluster (adaptive stratum is dark gray, nonadaptive stratum is light gray, and dots of same shade represent clusters), (*b*) random, (*c*) unequal probability (intensity of shading indicates sampling probability within the block), (*d*) systematic using single cluster, (*e*) systematic cluster (dots of same shade represent clusters), (*f*) stratified systematic showing two strata (dots of same shade represent clusters; blocks of same shade indicate strata).

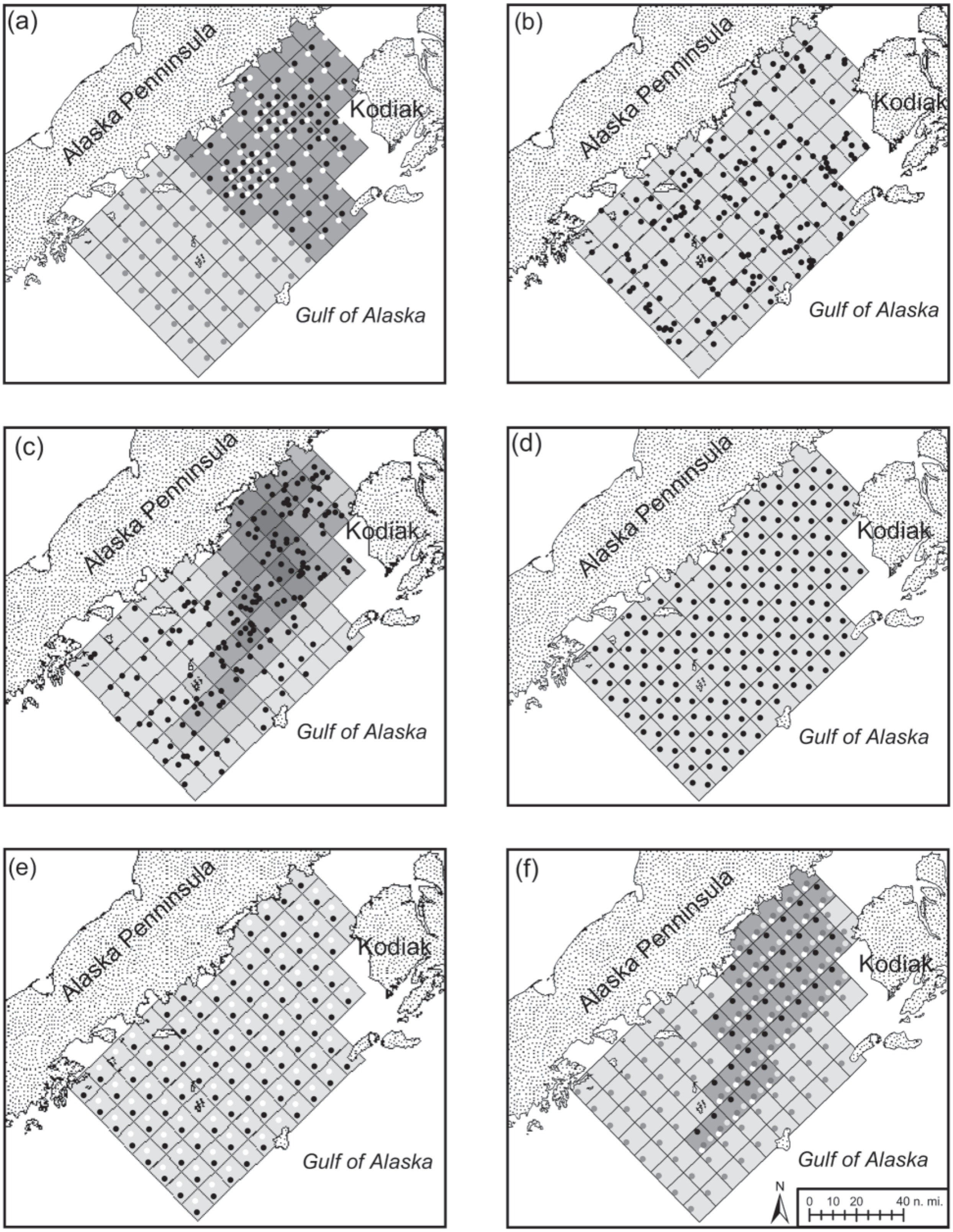
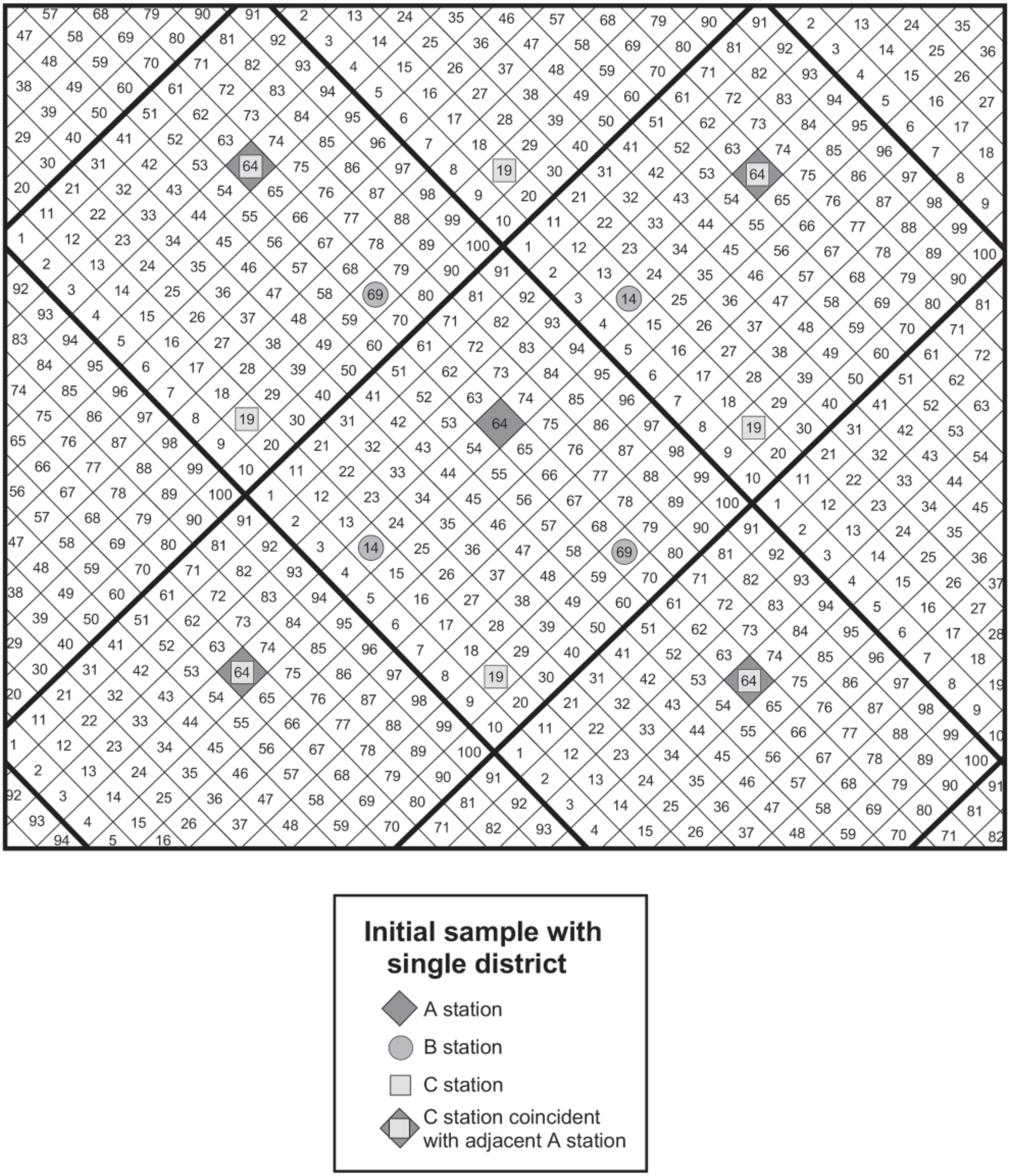


Fig. 5. Adaptive sampling design. Map shows a “neighborhood” (i.e., an *A* station with four surrounding *B* stations) and a “district” (i.e., 13 *A*, *B*, and *C* stations) for the adaptive survey. This is the result of using a stopping rule of $s_0 = 2$ (2 iterations of adaptive sampling).



chose to use just two strata. Appropriate strata boundaries were chosen using the mean densities within each 10 n.mi. × 10 n.mi. block, averaged over all 100 simulated populations. Blocks with means larger than the grand mean of all cells in all 100 populations were assigned to stratum 1 and the other blocks to stratum 2. Allocation of number of sampled cells to the strata was proportional to the standard deviation of the simulated densities within each stratum, averaged over the 100 simulated populations; thus, more stations were sampled in strata with higher average standard deviations.

Survey design comparisons

In our simulation, we based our comparison of survey designs on the bias, precision, and MSE of the estimated mean

Table 2. Formulas used to compare properties of estimators among different survey designs.

	Mean	SE
True population	$\mu_l = \frac{\sum_{i=1}^{N} Z_l(x_i, y_i)}{N}$	$\sigma_l = \sqrt{\mathrm{Var}(\hat{\mu}_l)} = \sqrt{\frac{\sum_{s=1}^{S}(\hat{\mu}_{sl} - \bar{\mu}_l)^2}{S}}$ where $\bar{\mu}_l = \frac{1}{S}\sum_{s=1}^{S}\hat{\mu}_{sl}$
Relative bias	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{(\bar{\mu}_l - \mu_l)}{\mu_l}\right]\times 100$	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{(\bar{\sigma}_l - \sigma_l)}{\mu_l}\right]\times 100$ where $\bar{\sigma}_l = \frac{1}{S}\sum_{s=1}^{S}\hat{\sigma}_{sl}$
Relative imprecision	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{\sqrt{\mathrm{Var}(\hat{\mu}_l)}}{\mu_l}\right]\times 100$ where $\mathrm{Var}(\hat{\mu}_l) = \sigma_l$	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{\sqrt{\mathrm{Var}(\hat{\sigma}_l)}}{\mu_l}\right]\times 100$ where $\mathrm{Var}(\hat{\sigma}_l) = \frac{\sum_{s=1}^{S}(\hat{\sigma}_{sl} - \bar{\sigma}_l)^2}{S}$
Relative $\sqrt{\mathrm{MSE}}$	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{\sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\mu}_{sl} - \mu_l)^2}}{\mu_l}\right]\times 100$	$\frac{1}{L}\sum_{l=1}^{L}\left[\frac{\sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\sigma}_{sl} - \sigma_l)^2}}{\mu_l}\right]\times 100$

Note: Z_l (x_i, y_i) = density at the ith cell of the lth population; $\hat{\mu}_{sl}$ and $\hat{\sigma}_{sl}$ are the estimated mean and SE of the mean from the sth sample of the lth population, respectively; N = total number of cells; L = total number of populations; S = total number of samples per population; and MSE = mean squared error.

population density, $\hat{\mu}$, and its estimated SE, $\hat{\sigma}$, for each survey design. The MSE is often used to compare estimators having different amounts of bias and precision, that is

$$\begin{aligned}\mathrm{MSE} &= E(z_i - \mu)^2 = E[(z_i - \hat{\mu}) + (\hat{\mu} - \mu)]^2 \\ &= E(z_i - \hat{\mu})^2 + (\hat{\mu} - \mu)^2 \\ &= (\text{Variance of } z_i) + (\text{Bias})^2\end{aligned}$$

where μ is the true mean of the population, and $\hat{\mu}$ is the mean of the sampled values of z_i (Cochran 1977). For ease of comparison, we put all these parameters in the same units as the population mean, that is, for precision we used SE instead of variance and we took the square root of the MSE. All measures of survey performance were expressed as a percentage of the population mean density (Table 2), thus making them relative. Results were reported for two separate groups, i.e., 50 populations with a high level of patchiness and 50 populations with a low level of patchiness. Note that the true mean population density, μ_l, is the average value of all cell densities for the lth population; however, the "true" population SE for the lth population, σ_l, is actually an approximation, because it is the square root of the variance about the 100 sample means for the lth population rather than the variance about sample means from all possible samples. The relative bias, precision, and $\sqrt{\mathrm{MSE}}$ for a particular survey design were computed for each population and then averaged over the L = 50 populations within each population group (high patchiness and low patchiness).

Hypothesis testing was performed to detect if relative bias of the mean differed significantly between survey designs within each population group (high patchiness and low patchiness) by performing a one-way analysis of variance (ANOVA), where the dependent variable was the absolute value of the relative bias for each population, the single fixed factor was survey design, and populations were replicates. If the survey design factor was significant, then Fisher's LSD (least significant difference) multiple comparisons (Milliken and Johnson 1992) were performed to see which were different. This testing was done for all six performance measures, that is, the relative bias, SE, and $\sqrt{\mathrm{MSE}}$ for estimates of both the mean and SE for both the high- and low-patchy populations.

Results

Simulated populations

We simulated two sets of 50 populations: one set had a high level of patchiness and the other exhibited a low level of patchiness. Examples of simulated populations with high and low patchiness are provided (Fig. 1). The simulated populations had values of Lloyd's index of patchiness (Lloyd 1967; Pielou 1969) that ranged between 6.2 and 7.7 and between 2.0 and 2.2 for the high- and low-patchy populations, respectively. The range of Lloyd's index for our nine historical cruises was 1.6–3.3, with most of these falling within the range of the low-patchy simulated populations. We wanted to examine the performance of the survey designs on populations even more patchy than our historical cruises, so we used values for each of the individual population parameters, such as patch size and maximum patch density, that were at the extreme of the observed range of values from the historical cruises to simulate the more patchy populations. This resulted in populations with a higher Lloyd's index than the historical data because none of the observed cruises had extreme values for all of the population parameters, whereas the simulated high-patchy populations did. Thus, the results from sampling the low-patchy populations are relevant to populations with spatial distributions similar to larval walleye pollock, while the results from sampling the high-patchy populations apply to populations with more exaggerated patchiness.

After simulating the populations, we then sampled the populations to obtain estimates using both adaptive estimators and classical estimators. We first looked at the results of the adaptive cluster sampling, chose the best adaptive estimator, then compared this with the classical estimators to determine which was the best overall survey design.

Comparison of adaptive cluster estimators

Detailed results of the comparison of the five adaptive cluster estimators are provided in Appendix D. We concluded that HH and HT perform equally well and better than the other three estimators considered. However, we recommend the HH estimator over HT because the HH estimator never produces a negative variance estimate and it is much easier computationally.

Comparison of survey designs

We compared the six survey designs using the HH estimator for the adaptive cluster survey. In general, the estimators for high-patchy populations performed worse than the low-patchy populations, with the difference being most pronounced in the imprecision and $\sqrt{\mathrm{MSE}}$ of both the mean and SE (Fig. 6).

Although the adaptive cluster estimator (HH) had the greatest bias of the mean compared with the other survey designs, it still underestimated the mean by less than 0.8% of the mean for high-patchy and 1.4% for low-patchy populations (Fig. 6). The biases of the other estimators were considered negligible according to Cochran (1977), because they fall below 10% of the SE. Comparing precision about the mean, the random estimator performed significantly worse than all others (Table 3) for high-patchy populations (at about 18%).

Looking at the $\sqrt{\mathrm{MSE}}$ values of the mean, the differentiation amongst the estimators was greatest for the high-patchy populations, with random being significantly higher than all others and systematic and stratified systematic being significantly lower (Table 3). Within low-patchy populations, the $\sqrt{\mathrm{MSE}}$ of the mean for the random estimator was again the largest, but not significantly higher than that of unequal probability. The $\sqrt{\mathrm{MSE}}$ values of the systematic and systematic cluster means were significantly lower than those of all others (Table 3).

The bias in the SE was greatest for the systematic estimator at about 7.5% for high-patchy populations and 3.5% for low-patchy populations (Fig. 6). The bias about the SE for the adaptive cluster design was significantly lower than that of the systematic and systematic cluster but significantly higher than that of the rest of the estimators for the high- patchy populations (Table 3). For low-patchy populations, the adaptive cluster bias about the SE was not significantly different from the stratified systematic bias and was significantly lower than the systematic, but significantly higher than all others.

The systematic cluster and adaptive cluster estimates of the SE were significantly the least precise for both high- and the low-patchy populations (Fig. 6, Table 3). The $\sqrt{\mathrm{MSE}}$ values of the SE indicated that the stratified systematic and unequal probability survey designs delivered the best estimators of SE for high-patchy population groups, and the unequal probability and random survey designs produced the best estimators of SE for the low-patchy populations.

Clearly, there is no obvious best design for both high and low-patchy populations. Based on the $\sqrt{\mathrm{MSE}}$ of the mean, the random design was a very poor choice for all populations and the systematic, systematic cluster, and stratified systematic designs' performances were superior for all populations. Looking at the $\sqrt{\mathrm{MSE}}$ of the SE, the two systematic designs produced the worst estimates of the SE for both population groups, while stratified systematic, unequal probability, and random all performed admirably. We conclude that the stratified systematic survey design was the best overall estimator, regardless of the level of patchiness. This conclusion is well supported by examining Fig. 6. Even though the hypothesis tests in Table 3 do not clearly show superior performance of the stratified systematic design for the low-patchy populations, it performs consistently better in terms of the $\sqrt{\mathrm{MSE}}$ of the mean and SE.

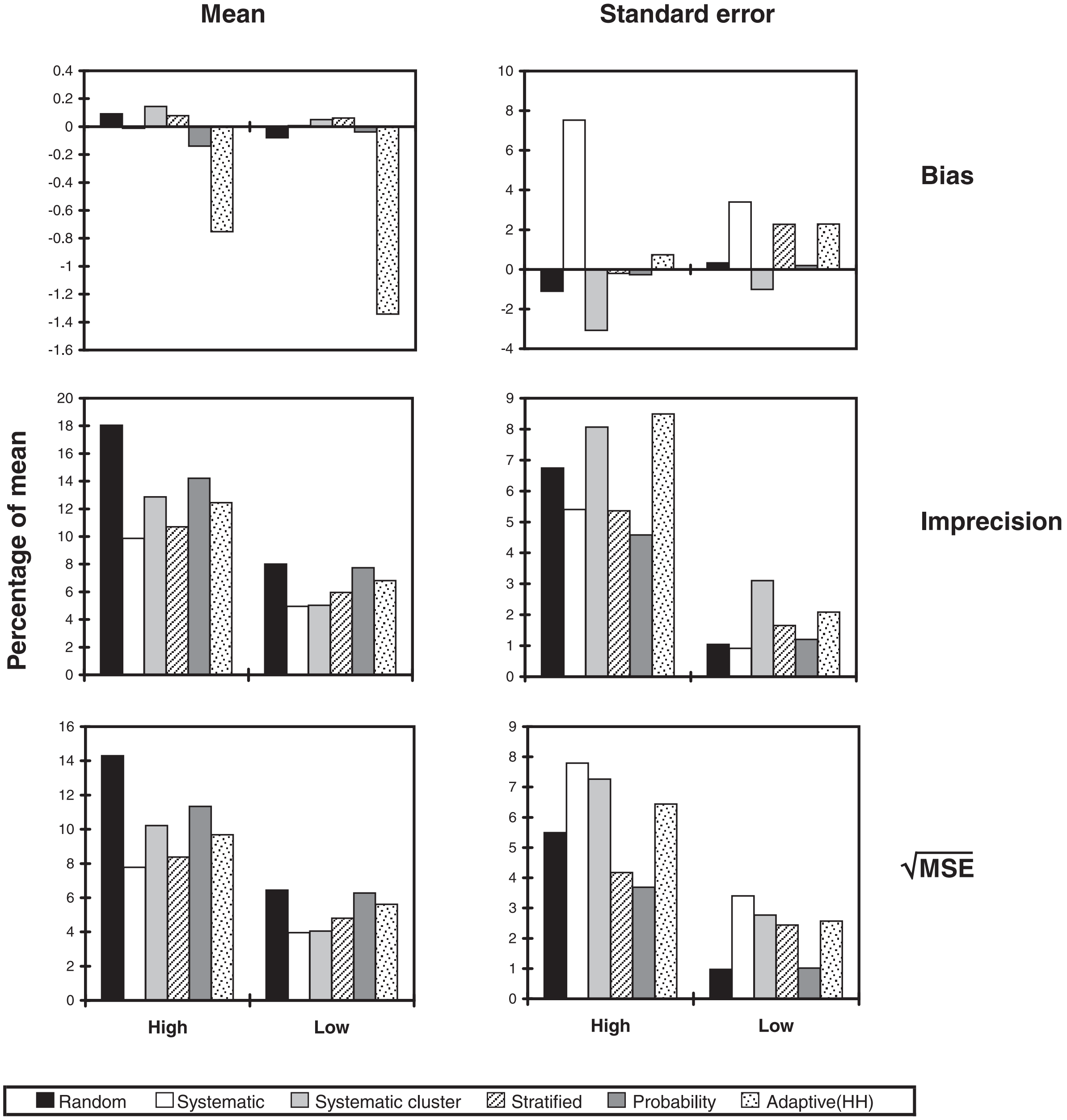
Another important observation is that using a single cluster with a systematic design did not significantly decrease the performance, based on $\sqrt{\mathrm{MSE}}$ of the mean, when compared with a systematic survey with two clusters (when they have the same sample size); in fact, the systematic was significantly better for the high-patchy populations. Only in the case of the $\sqrt{\mathrm{MSE}}$ values of the SE for low-patchy populations did the systematic cluster perform significantly better than the systematic (Table 3). This observation contradicts what Christman (2000) found; her simulated systematic surveys had a much higher bias in the variance, perhaps because her populations contained fewer primary units, and the probability of missing a cluster was large.

A valid concern with using a single cluster is that the SE estimator will be biased, and indeed it was, but the precision of the SE was sufficiently better for the single cluster systematic design than for the two-cluster design so that these two effects balanced each other out. As the number of clusters increase in the systematic cluster design, we expect the recision of the SE to improve, but this benefit would be tempered by the reduced number of samples within each cluster, given the total sample size did not change.

Discussion

Choosing the best survey design to sample a patchy distribution should be based on minimizing bias and maximizing precision of the estimates of both the mean and the SE, for a given cost, which we equated to sample size. Bias, or the

Fig. 6. The bias, imprecision, and $\sqrt{\mathrm{MSE}}$ of the mean and SE averaged over the 50 high- and 50 low-patchy populations for the six survey designs.



difference between an estimate and the true value, is distinguished from precision, which is the differences among repeated measures (Sokal and Rohlf 1995). Precision is measured by the variance, that is, the smaller the variance, the more precise the estimator (we used SE instead of variance of the mean to keep all performance measures in the same units). The MSE is a measure of both precision and bias combined. Our results showed that the MSE of the mean of a patchily distributed population is determined more from precision (variance) than from bias.

The consequences of an inaccurate estimate of the mean are obvious, but the consequences of an inaccurate estimate of the SE are also important. If the estimated SE is inaccurate (i.e., overestimated or underestimated because of either bias or lack of precision or both), then the researcher will have less (if overestimated) or more (if underestimated) confidence in the estimate of the mean than warranted. This impacts the decisions based on the mean. If the mean is used in hypothesis testing, then the test will have incorrect power, resulting in incorrect test results. If the mean is used to set harvest rates (e.g., fishing quotas) and the SE is overestimated, then the resource manager may be cautious and set the quota lower than necessary. Conversely, if the SE is underestimated, then the resource manager may be less cautious and set the quota higher than the stock can support, potentially leading to overfishing.

Table 3. Results of multiple comparison tests showing significant differences (at 5% significance level) among survey design estimators for both (*i*) high-patchy and (*ii*) low-patchy populations.

	Adaptive (HH)	Unequal probability	Random	Stratified systematic	Systematic cluster	Systematic
High-patchy						
Bias of mean	B	B	B	A	A	A
SE of mean	B	C	D	A	B	A
MSE of mean*	B	C	D	A	B	A
Bias of SE	B	A	A	A	C	D
SD of SE	D	A	C	B	D	B
MSE of SE*	C	A	B	A	D	D
Low-patchy						
Bias of mean	D	C	B	A	A	A
SE of mean	C	D	D	B	A	A
MSE of mean*	C	D	D	B	A	A
Bias of SE	C	A	A	C	B	D
SD of SE	E	C	B	D	F	A
MSE of SE*	BC	A	A	B	C	D

Note: The adaptive estimator is the Hansen–Hurwitz (HH) estimator. Estimators with the same letters indicate no significant difference.

*MSEs capture both bias and precision.

There are factors other than bias and precision that influence which survey design is used. One is the availability of prior information on the spatial distribution of a population. Of the six designs that we considered, three require prior information (stratified systematic, unequal probability, and adaptive cluster), and the other three do not (random, systematic, and systematic cluster). One might expect that designs that require prior information will perform better, but this is dependent on how well the prior knowledge applies to the present survey. Another factor that directly impacts the survey design choice is whether there is a single target species or whether many species are involved. If there is no single target species, there most likely will not be a single best design for all species, in which case designs that do not require prior information are the best compromise.

Our study shows that of the three applicable designs that require no prior knowledge (random, systematic, and systematic cluster), the random survey design was the least effective, that is, the $\sqrt{\mathrm{MSE}}$ of the mean was the poorest of all the survey designs, although the $\sqrt{\mathrm{MSE}}$ of the SE was moderate. Random sampling has inherent disadvantages as well, including (*i*) no way to deal with high variability in the data and its effect on the estimators other than increasing the sample size, (*ii*) difficult implemention in the field because of irregular spacing of sampled units, (*iii*) missed patches because of uneven spatial coverage, and (*iv*) data are poorly suited for spatial analysis, such as interpolation. Systematic sampling is much easier to implement and provides even coverage; however, multiple starting points (clusters) are seldom used in practice. Rather, each station sampled is considered to be a separate sampling unit, instead of part of a single cluster, and the sample is treated as if it were random (Schopka 1994; Wertheimer and Celewycz 1996; Edgar et al. 1997). From our simulation, we conclude that there was very little difference in bias, precision, and $\sqrt{\mathrm{MSE}}$ between the use of a systematic design with a single cluster and using a systematic cluster with two clusters. Although the SE is slightly overestimated by the systematic design, this is mitigated by the improved precision of the SE. Therefore, our simulation shows that the convenience of sampling a single cluster does not seriously impact the bias and precision of the estimates.

Our interest in adaptive cluster sampling is primarily due to advantages reported in the literature. For example, adaptive cluster sampling (*i*) is reported to yield more precise estimates for patchy distributions (Brown 2003; Hanselman et al. 2003; Turk and Borkowski 2005), (*ii*) allows for increased observations of interest by concentrating additional sampling effort in those areas where there is higher density, and (*iii*) makes more efficient use of ship time by adding samples that are near the initial sample (Turk and Borkowski 2005).

The first two advantages mentioned above also apply to stratified systematic and unequal probability sampling. Unequal probability sampling is more precise if the selection probabilities are proportional to the values being observed and allows for concentration of sampling in areas of historically high fish densities. An example of unequal probability sampling in fisheries can be found in Dressel and Norcross (2005). Stratified systematic surveys yield more precise estimators if the strata create more homogeneous areas and may reduce the cost per observation by sampling more convenient groups. Stratified systematic sampling has the added advantage of providing separate estimates for each stratum without having to take another sample. However, if the strata are not appropriately chosen, the precision of the estimator may decrease. An example where a stratified systematic survey is used in fisheries can be found in Pennington et al. (2002).

In comparing adaptive cluster sampling with stratified systematic and unequal probability survey designs, the adaptive cluster survey is more stringent in that it requires absolute abundance information to determine the criterion that initiates adaptive sampling, and it requires information on the spatial distribution to specify the spacing of the added adaptive samples. Stratified systematic and unequal probability surveys only require relative abundance, which may be more constant from year-to-year than absolute abundance. For all three of these survey designs, we used the average absolute abundances and average spatial distributions, averaged over the 100 simulated populations. One possible reason that the adaptive cluster survey did not show marked improvement over the other designs is that the criterion that initiated adaptive sampling was the same for every population, that is, the 85th percentile of population densities from all 8857 cells, averaged over the 100 populations. This is equivalent in practice to using historical data to specify the criterion, as opposed to conducting a pilot survey. The consequence of this is that populations with a high mean abundance would exhaust the quota of adaptive samples early in the survey and leave much of the area to be systematically sampled, thus diluting the benefit of adaptive sampling. In practice, if we were to choose a percentile of a presurvey as the criterion for adaptively sampling, the adaptive estimator might perform better. However, the number of sampling units in the adaptive survey would have to be reduced by the number of units sampled in the presurvey, which would obviously worsen the adaptive estimator's performance.

One of the greatest weaknesses in adaptive sampling is the inability to control the sample size, and this is especially critical in marine surveys over large areas employing ships on a fixed project schedule and budget, which accrue large costs per sample. We modified Thompson's adaptive cluster design in three ways to limit sample size and allow for the greatest coverage: (1) we stopped adaptive sampling after two iterations, (2) we skipped four sampling units when adaptive stations were sampled, allowing networks to overlap, and (3) we divided the survey area into two strata and only adaptively sampled one stratum. Methods 1 and 3 introduce bias into the estimates, the more serious of which is method 1, the restriction of adaptive iterations. Using a stopping rule ($s_0 = 2$) is a significant deviation from Thompson's design, and as a result our networks were incomplete. However, the second method for controlling sample size, that is, spacing the adaptive stations five cells apart, helped to counteract this bias by allowing for overlapping districts. We think that our very small negative bias (less than 1%) for the HT estimator compared with Su and Quinn's (2003) positive bias of 6%–8% (with or without a stopping rule) and Cabral and Murta's (2004) extreme bias is due to our overlapping neighborhood structure that allowed the formation of large networks. This obviously produces smaller sample sizes than contiguous neighborhoods, but is only viable if the patch size is sufficiently large relative to the area spanned by the neighborhood. Our procedure of skipping over units is similar to adaptively sampling only cells midway between secondary units (Woodby 1998). Although Woodby applies this "neighborhood restriction" to a one-dimensional systematic adaptive cluster sample design, he mentions that it could be extended to two dimensions, as we did. A two-dimensional example where one unit is skipped in all four directions when sampling adaptively is given by Christman (1997).

By stopping the adaptive sampling when we reached a predetermined sample size limit minus the number of initial secondary units yet to be sampled, we successfully covered the entire survey area and kept the sample size approximately equal to that of the other survey designs. As a result, we poststratified our sample into an adaptively sampled strata and a nonadaptively sampled strata, which resulted in a biased estimate of the variance, but this bias decreases as the average sample size for each strata increases (Jessen 1978), and it had a negligible effect in our case.

Other approaches in the literature deal with these challenges in adaptive sampling, that is, choosing the criterion for initiation of adaptive sampling, and limiting sample size. In Hanselman et al. (2003), three different strategies for specifying the criterion were investigated. One strategy used the 80th percentile of historical data, which is similar to using the 85th percentile of all 100 simulated populations in this paper. This avoids the need for a presurvey and the introduction of bias by the criterion value being dependent on the sampling. However, using a percentile of historical data would only be useful for populations with low interannual variability.

In a simulation study, Su and Quinn (2003) used both order statistics and a stopping rule ($s_0 = 3$) to help limit the size of the sample. They conducted a presurvey and adaptively sampled the stations with highest density, where the number of these (r) was dictated by the time available for sampling. They found that the HH estimator had relatively high positive bias if the level of aggregation was high, the stopping rule (s_0) was small, and r was large. The bias of the HT estimator was not affected much by the stopping rule unless r was small, in which case the bias was reduced.

Lo et al. (1997) used a stratified, two-stage adaptive cluster design, which also restricted the number of adaptively sampled stations. Instead of a presurvey, an iterative method based on past cruises was used to determine the criterion value that would provide the right number of stations that time and resources would allow. However, this value proved to be too high and had to be changed during the survey, requiring the sample to be poststratified, as we did. Their results indicated that a simple stratified estimator, using all stations sampled in a restricted adaptive cluster survey design, was more precise than the stratified HT estimator.

Christman (2003) controls sample size with a Markov chain one-per-stratum design, i.e., a two-stage sampling design where the area is first divided into strata, a random-start systematic sample of one is taken, then wherever the sampled unit satisfies the specified criterion, a secondary systematic sample is taken within that stratum. For the designs that she describes, the HT estimator can only be evaluated when the criterion is $z_i > 0$ (rather than $z_i > c$, where c is a value other than 0).

Brown (1996, 2003) states that to have an efficient adaptive survey, one must design the survey in such a way as to balance the network size (not too big) with the within-network variance (not too small), because the within-network variance generally decreases with network size. Brown's simulations found adaptive sampling to be most efficient with networks containing two to four units. She recommends using neighborhood definitions and the criterion value to limit the network size, as was done in this paper.

Another important consideration in sampling is cost. Several papers address this issue. Brown (1996) found that the average distance traveled for adaptive sampling was less than that for random sampling for all simulated populations, regardless of the degree of patchiness. Hanselman et al. (2003) measured travel time efficiency and found that adaptive sampling was more efficient than random sampling for sampling patchily distributed Pacific ocean perch (*Sebastes alutus*). We did not specifically compare distance traveled between our survey designs, but any survey design with a grid pattern (i.e., systematic, systematic cluster, stratified systematic, and adaptive cluster) has logistical advantages over unevenly spaced stations (i.e., random and unequal probability). Typically, time is required to process a sample after collection, and a grid pattern provides a consistent amount of time before collecting the next sample.

In conclusion, our simulation suggests that of the adaptive cluster estimators, HH and HT were the best overall, and because HH was easier to compute than HT and never yields a negative variance, the HH estimator is preferred. Christman (1997) also states that although her simulation study showed that the HT estimator was often more efficient, HH was less sensitive to changes in population. Nevertheless, our adaptive cluster survey design did not perform as well as the stratified systematic design and introduced many logistical problems compared with other designs. If prior information is available for a population, a stratified systematic survey design produces the best estimator. Without any information on the spatial distribution, a systematic survey design is best, and sampling a single cluster (using a simple random estimator) rather than multiple clusters does not appreciably impact the bias or precision of the estimated SE. These conclusions apply only to populations with spatial structure similar to the populations that we simulated, that is, populations that are widespread throughout the survey area with a few patches with a broad range of size and density and with considerable small-scale patchiness. However, both Christman (2000) and Cabral and Murta (2004) found that stratified designs out-performed adaptive sampling, and because there were fundamental differences in our simulated populations compared with theirs, these conclusions apparently apply to a broader range of clustered populations.

Acknowledgements

We thank the staff at the Fisheries–Oceanography Coordinated Investigations (FOCI) program and the Recruitment Processes Task at the Alaska Fisheries Science Center (AFSC) for providing carefully checked historical data for this project. We also thank the crews of the NOAA ships *Miller Freeman* and *Discoverer* for their help in collecting the data. Samples were processed at the Polish Plankton Sorting Center and at the Ichthyoplankton Laboratory at AFSC. Other scientists who have provided useful comments include Nancy C.H. Lo of the Southwest Fisheries Science Center, as well as David A. Somerton, Jeffrey M. Napp, and Steven E. Syrjala, all from AFSC. We also appreciate the insightful and thorough reviews provided by the two anonymous reviewers. This research is contribution FOCI-0578 to NOAA's Fisheries–Oceanography Coordinated Investigations.

References

Andrew, N.L., and Mapstone, B.D. 1987. Sampling and the description of spatial pattern in marine ecology. Oceanogr. Mar. Biol. Annu. Rev. **25**: 39–90.

Brewer, K.R.W., and Hanif, M. 1983. Sampling with unequal probabilities. Springer Verlag, New York.

Brown, J.A. 1996. The relative efficiency of adaptive cluster sampling for ecological surveys. *In* Mathematical and information sciences reports. Series B. *Edited by* R.H. Morton and A. Swift. Massey University, New Zealand.

Brown, J.A. 2003. Designing an efficient adaptive cluster sample. Environ. Ecol. Stat. **10**: 95–105.

Cabral, N., and Murta, A.G. 2004. Effect of sampling design on abundance estimates of benthic invertebrates in environmental monitoring studies. Mar. Ecol. Prog. Ser. **276**: 19–24.

Christman, M.C. 1997. Efficiency of some sampling designs for spatially clustered populations. Environmetrics, **8**: 145–166.

Christman, M.C. 2000. A review of quadrat-based sampling of rare geographically clustered populations. J. Agr. Biol. Environ. Stat. **5**(2): 168–201.

Christman, M.C. 2003. Adaptive two-stage one-per-stratum sampling. Environ. Ecol. Stat. **10**: 43–60.

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

Dressel, S.C., and Norcross, B.L. 2005. Using poststratification to improve abundance estimates from multispecies surveys: a study of juvenile flatfishes. Fish. Bull. **103**: 469–488.

Edgar, G.J., Moverley, J., Barrett, N.S., Peters, D., and Reed, C. 1997. The conservation-related benefits of a systematic marine biological sampling programme: the Tasmanian reef bioregionalisation as a case study. Biol. Conserv. **79**(2–3): 227–240.

Hajek, J. 1971. Discussion of an essay on the logical foundations of survey sampling, Part one, by D. Basu. *In* Foundations of statistical inference. *Edited by* V.P. Godambe and D.A. Sprott. Holt, Rinehart, Winston, Toronto, Ont. p. 236.

Hanselman, D.H., Quinn, T.J., II, Lunsford, C., Heifetz, J., and Clausen, D. 2003. Applications in adaptive cluster sampling of Gulf of Alaska rockfish. Fish. Bull. **101**: 501–513.

Jessen, R.J. 1978. Statistical survey techniques. John Wiley and Sons, New York.

Lloyd, M. 1967. Mean crowding. J. Anim. Ecol. **36**: 1–30.

Lo, N.C.H., Griffith, D., and Hunter, J.R. 1997. Using a restricted cluster sampling to estimate Pacific Hake larval abundance. CalCOFI Rep. **38**: 103–113.

Milliken, G.A., and Johnson, D.E. 1992. Analysis of messy data. Chapman and Hall, London, UK.

Pennington, M., Burmeister, L., and Hjellvik, V. 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. **100**: 74–80.

Pielou, E.C. 1969. An introduction to mathematical ecology. Wiley, New York.

Quinn, T.J., II, Hanselman, D.H., Clausen, D.M., Heifetz, J., and Lunsford, C. 1999. Adaptive cluster sampling of rockfish populations. *In* Proceedings of the American Statistical Association, Joint Statistical Meeting, Biometrics Section, Baltimore, Md., 8–12 August 1999. *Edited by* C. Brooks. American Statistical Association, Alexandria, Va. pp. 11–20.

Raj, D. 1956. Some estimators in sampling with varying probabilities without replacement. J. Am. Stat. Assoc. **51**: 269–284.

Rao, J.N.K., and Singh, M.P. 1973. On the choice of estimator in survey sampling. Austr. J. Stat. **15**: 95–104.

Salehi, M.M. 2003. Comparison between Hansen–Hurwitz and Horvitz–Thompson estimators for adaptive cluster sampling. Environ. Ecol. Stat. **10**: 115–127.

Sarndal, C.E., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer Verlag, New York.

Schopka, S.A. 1994. Fluctuations in the cod stock off Iceland during the twentieth century in relation to changes in the fisheries and environment. ICES Mar. Sci. Symp. **198**: 175–193.

Sen, A.R. 1953. On the estimate of the variance in sampling with varying probabilities. J. Ind. Soc. Agr. Stat. **5**: 119–127.

Sokal, R.R., and Rohlf, F.J. 1995. Biometry. 3rd ed. W.H. Freeman and Company, New York.

Stabeno, P.J., Schumacher, J.D., Bailey, K.M., Brodeur, R.D., and Cokelet, E.D. 1996. Observed patches of walleye pollock eggs and larvae in Shelikof Strait, Alaska: their characteristics, formation and persistence. Fish. Oceanogr. **5**(Suppl. 1): 81–91.

Stehman, S.V., and Overton, W.S. 1994. Environmental sampling and monitoring. *In* Handbook of statistics. *Edited by* G.P. Patil and C.R. Rao. Elsevier Science B.V., Amsterdam, The Netherlands. pp. 263–306.

Su, Z., and Quinn, T.J., II. 2003. Estimator bias and efficiency for adaptive cluster sampling with order statistics and a stopping rule. Environ. Ecol. Stat. **10**: 17–41.

Thompson, S.K. 1990. Adaptive cluster sampling. J. Am. Stat. Assoc. **79**: 781–790

Thompson, S.K. 1991. Adaptive cluster sampling designs with primary and secondary units. Biometrics, **47**: 1103–1115

Thompson, S.K. 1992. Sampling. John Wiley & Sons, Inc., New York.

Thompson, S.K., and Seber, G.A.F. 1996. Adaptive sampling. John Wiley & Sons, Inc., New York.

Turk, P., and Borkowski, J. 2005. A review of adaptive cluster sampling: 1990–2003. Environ. Ecol. Stat. **12**: 55–94.

Wertheimer, A.C., and Celewycz, A.G. 1996. Abundance and growth of juvenile pink salmon in oiled and non-oiled locations of western Prince William Sound after the Exxon Valdez oil spill. *In* Symposium 18: Proceedings of the Exxon Valdez Oil Spill Symposium, 2–5 February 1993, Anchorage, Alaska. *Edited by* S.D. Rice, R.B. Spies, D.H. Wolfe, and B.A. Wright. American Fisheries Society, Bethesda, Md. pp. 518–532.

Woodby, D. 1998. Adaptive cluster sampling: efficiency, fixed sample sizes, and an application to red sea urchins (*Strongylocentrotus franciscanus*) in southeast Alaska. *In* Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. *Edited by* G.S. Jamieson and A. Campbell. Can. Spec. Publ. Fish. Aquat. Sci. No. 125. pp. 15–20.

Yates, F., and Grundy, P.M. 1953. Selection without replacement from within strata with probability proportional to size. J. R. Stat. Soc. B, **15**: 253–261.

Appendix A. Definitions and notations

Subscripts

i = cell, 1 through N
j = primary unit, 1 through N_s
p = patch, 1 through P
k = network, 1 through κ
h = stratum, 1 through H
l = simulated population, 1 through L
s = simulated sample, 1 through S
b = block, 1 through 98

All survey designs

cell i = 1 nautical mile (n.mi.) × 1 n.mi. of surface area (1 n.mi. = 1.852 km); the simulated population covers 8857 cells, $i = 1,2,\ldots,8857$
sampling unit = station = a sampling event at the center of each cell = column of water under the surface area of the bongo sampler frame, that is, a circle of 60 cm diameter or 0.283 m^2
N = total number of cells = 8857
L = number of populations = 100
S = number of samples per population = 100
(x_i, y_i) = location of the centroid of cell i; x_i is the east–west position and y_i is the north–south position
block = 10 n.mi. × 10 n.mi. square containing 100 of the 1 n.mi. × 1 n.mi. cells; there are 98 blocks in the population
$Z_i = Z(x_i, y_i)$ = simulated average number of larvae in a column of water under 10 m^2 of surface area, averaged over area of cell i; the simulated population consists of N = 8857 population units
$\mu = \bar{Z}$ = population average = average number of larvae per 10 m^2 of surface area; this is the parameter to be estimated by each survey
$z_i = z(x_i, y_i)$ = simulated population unit, Z_i, multiplied by a sampling error term; this simulates the density of larvae, that is, number of larvae in the water column below a 10 m^2 of surface area, sampled somewhere within cell i
ε_i = sampling error around mean density in cell i, which is a random lognormal variate
P = number of patches in population
$\boldsymbol{\mu}_p = (\mu_{x,p}, \mu_{y,p})$ = location of the centroid of patch p
$\boldsymbol{\sigma}_p^2 = \begin{pmatrix} \sigma_{x,p}^2 & \rho_p \\ \rho_p & \sigma_{y,p}^2 \end{pmatrix}$ = parameters controlling the size and shape of patch p
$f_p(x_i, y_i \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)$ = proportion of patch p that occurs in cell i
M_p = maximum density of larvae (number per 10 m^2) in patch p
$B(x_i, y_i)$ = number of larvae at location (x_i, y_i) that are not part of any patch
η = random lognormal variate multiplied to the simulated density for every cell to incorporate variability about the modeled densities
LI = Lloyd's index of patchiness
n = number of cells sampled = number of stations in the sample

Unequal probability sampling

π_i = probability that cell i is sampled
$\bar{z}'_b$ = the mean density of larvae in block b, averaged over all 100 simulated populations
$\bar{z}'_i = \bar{z}'_b$ for every cell i that is in block b, this assigning the same value to each cell within each block

Systematic cluster sampling

primary unit = observations that are systematically selected based on the same starting point within each block, also called clusters in the literature. For systematic sampling, the simulated population is partitioned into 100 primary units.

N_s = number of primary units in population

n_s = number of primary units in sample

secondary unit = each cell within a primary unit. Each primary unit has on average 89 secondary units.

M_j = number of secondary units in primary unit j

$\sum_{j=1}^{N_s} M_j$ = total number of secondary units in population (Thompson's MN) = 8857

$\bar{z}_j$ = average observed density of larvae in primary unit j

z_{ij} = density at ith sampling unit in primary unit j

Stratified sampling

H = number of strata

N_h = number of cells in stratum h

N_s = number of primary units, same value for all strata

n_h = number of primary units sampled in stratum h

M_{hj} = number of secondary units in primary unit j in stratum h

z_{hij} = density at ith sampling unit in primary unit j in stratum h

$\bar{z}_{hj}$ = average observed density of larvae in primary unit j in stratum h

$\bar{z}_h$ = average observed density of larvae in stratum h

Adaptive cluster sampling

(These definitions are generally consistent with those given in Thompson (1991), where he uses a systematic sampling design for the initial sample.)

criterion = rule for initiating adaptive sampling. We used the criterion that the observed density is greater than 85% of all densities in all 100 simulated populations.

initial sample = systematic sample with n_s starting points.

neighborhood = the neighborhood of a sampling unit consists of itself and a group of stations around the initial sampling unit following a prespecified pattern. For our simulated adaptive survey, the neighborhood is a systematic grid of four units around the initial unit. The distance between the central unit and each of its four neighbors is 5 n.mi. (Fig. 5).

district = In this application of adaptive sampling, we restrict the number of iterations of testing for the criterion of interest and adding more neighborhoods to two. The district for a unit consists of the initial station, labeled *A*, and two possible levels of neighborhoods, labeled *B* and *C*. Specifically, the district for a secondary unit in the initial sample (an *A* unit) consists of itself, its neighborhood of four *B* units, and each *B* unit's neighborhood of three *C* units. (Note that there are only three *C* units in each *B* unit's neighborhood because the *A* unit is at the same location as one of *B*'s four neighbors.) We stop with the *C* units because each *C* unit's neighborhood includes an adjacent *A* unit that is already sampled (Fig. 5).

network = the subset of stations within a cluster (as defined by Thompson 1991) that satisfies the criterion.

κ = number of networks

z_k = the total density in the kth network

n_k = the number of primary units that intersect the kth network (Thompson (1991) uses x_k)

I_{jk} = 1 if the jth primary unit in the initial sample intersects the kth network and 0 otherwise

I_k = 1 if the initial sample intersects the kth network and 0 otherwise

π_k = probability that the initial sample intersects the kth network, that is, $P(I_k = 1)$

π_{k_1,k_2} = probability that the initial sample intersects both networks k_1 and k_2

s_0 = stopping rule = number of iterations of adaptive sampling before stopping. This paper uses a stopping rule of $s_0 = 2$.

N_s, n_s, and M_j have the same definitions as for systematic cluster sampling.

Appendix B. Simulation of populations

Estimating patch parameters

A patch was defined for our purposes as a region that has densities greater than one standard deviation (SD) above the grand mean over all densities (Fig. B1) (Stabeno et al. 1996). The large-scale spatial distribution (on the order of tens of nautical miles) of larvae was characterized by a few large patches close to the spawning area and lower densities of larvae distributed more homogeneously over the rest of the survey area. The small-scale distribution (on the order of metres) of larvae was characterized by a large, multiplicative variance due to small-scale patchiness.

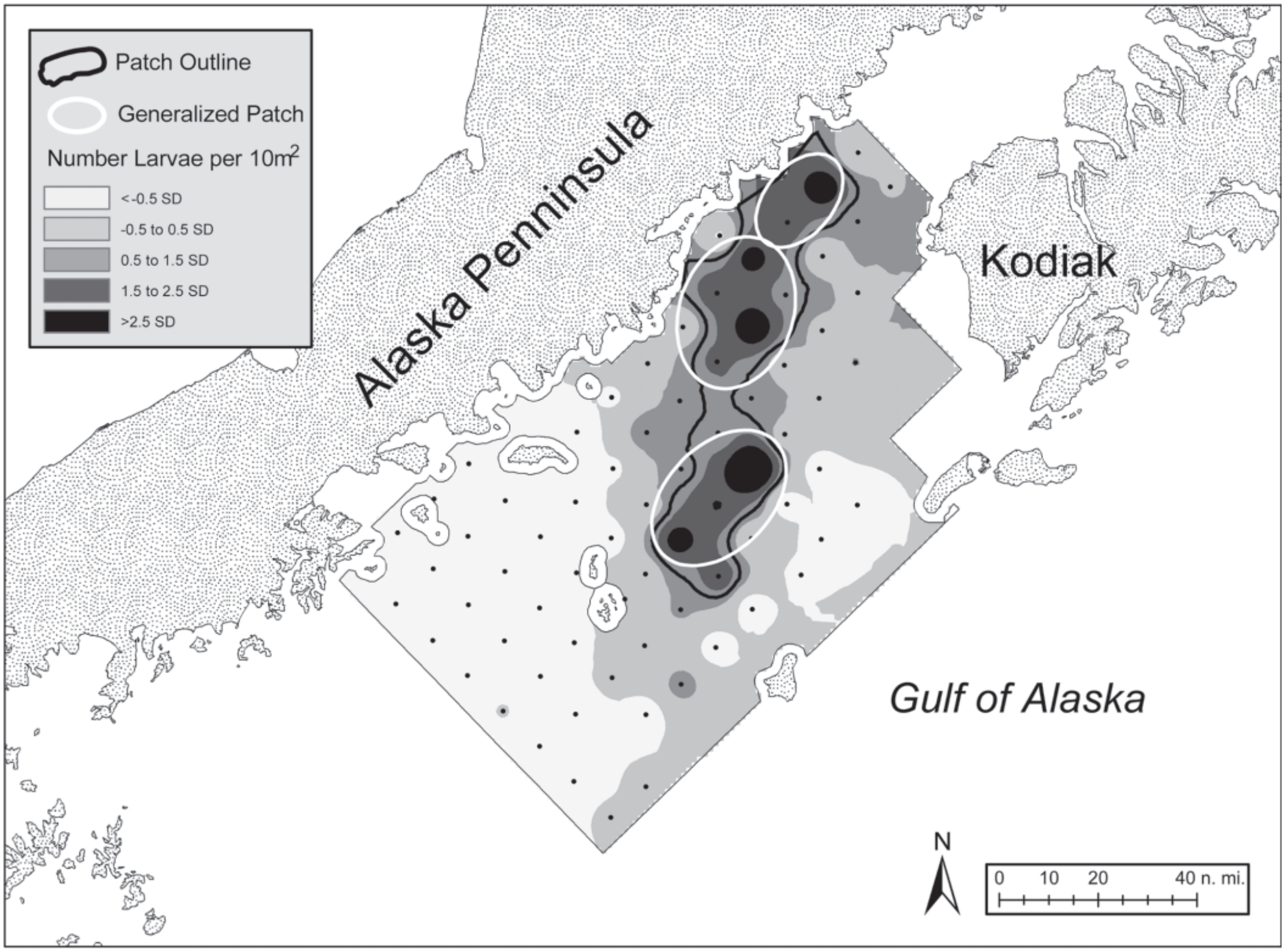
Large-scale patchiness was defined by the number, location, size, and shape of the patches; maximum densities inside the patches; and densities outside of the patches. These characteristics were observed from the historical data and incorporated into the simulation model as parameters that each have a stochastic component.

The spatial patterns for all nine historical cruises were summarized individually by first interpolating densities over the survey area and then standardizing within years (subtracting the mean and dividing by the SD) so that each cruise had a mean of 0 and SD of 1. The interpolation method used was the inverse distance weighted method in the Spatial Analyst extension of ArcGIS (version 8.3, Environmental Systems Research Institute, 2002). The interpolated densities from each cruise were mapped, and the patches were identified and delineated. Irregularly shaped patches were conceptualized as two or more slightly overlapping elliptical patches (Fig. B1). The nine spatial distributions of standardized densities were then analyzed to determine values or probability distributions that best described each of the following parameters: (*i*) number of patches, (*ii*) location of the patches, and (*iii*) size and shape of the patches; the untransformed densities were used to determine (*iv*) maximum densities inside the patches and (*v*) densities outside of the patches. These parameters are defined as follows (all symbols are defined in Appendix A):

(*i*) The number of patches, *P*, was simulated by a random variable from a uniform probability distribution, ranging between three and six patches to be consistent with the observed number of patches in the historical data.

(*ii*) The location of each patch was simulated by selecting the location of the center of the patch. A center location

Fig. B1. Historical distribution of patch size and shape. Irregularly shaped patches were conceptualized as two or more slightly overlapping elliptical patches, outlined in white.



for each patch was randomly chosen with predetermined selection probabilities based on the total number of times that each cell was inside a patch in any of the cruises. Once a patch location was selected, subsequent patches within the population were located so that they did not substantially overlap the previous patches.

(*iii*) The size and shape of the *P* patches were modeled by letting the portion of the *p*th patch (i.e., $f_p(x_i, y_i)$) that occurs at a location be proportional to the probability density of a bivariate normal probability distribution at x_i and y_i, where the density has been rescaled so that the maximum density has the value of 1.0, that is

$$f_p(x_i, y_i \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2) = \exp\left\{\frac{-1}{2(1-\rho_p^2)} \times \left[\left(\frac{x_i-\mu_{x,p}}{\sigma_{x,p}}\right)^2 - 2\rho_p\left(\frac{x_i-\mu_{x,p}}{\sigma_{x,p}}\right) \times \left(\frac{y_i-\mu_{y,p}}{\sigma_{y,p}}\right) + \left(\frac{y_i-\mu_{y,p}}{\sigma_{y,p}}\right)^2\right]\right\}$$

where

$$\boldsymbol{\mu}_p = (\mu_{x,p}, \mu_{y,p})$$

and

$$\boldsymbol{\sigma}_p^2 = \begin{pmatrix} \sigma_{x,p}^2 & \rho_p \\ \rho_p & \sigma_{y,p}^2 \end{pmatrix}$$

This function was chosen because of its three-dimensional bell shape (Figs. 1*c*, 1*d*). Note that we used this function to describe the number of fish at each location, x_i and y_i; we did not use it as a frequency distribution of the number of fish, which definitely does not have a Gaussian probability distribution. Instead, the x_i, y_i locations were modeled with a bivariate normal distribution where the probability density at each location was proportional to the number of fish at the location. The bivariate mean of the bivariate normal density function, $\mu_{x,p}$ and $\mu_{y,p}$, is the location of the cell at the center of the *p*th patch, which was determined in the previous step. The other parameters of the function are SDs in the *x* and *y* direction, $\sigma_{x,p}$ and $\sigma_{y,p}$, respectively, and correlation between *x* and *y*, ρ_p. The values for $\sigma_{x,p}$, $\sigma_{y,p}$, and ρ_p determined the size, shape, and orientation of the patch and were simulated to mimic the generalized patches (Fig. B1). The simulated values for $\sigma_{x,p}$ and $\sigma_{y,p}$ were small for the high-patchy populations and large for the low-patchy populations, with the constraint that all simulated values resulted in patch sizes that were within the range of the observed patch sizes in the historical data.

(*iv*) M_p is the maximum density within the *p*th patch. The observed values of M_p were correlated between patches within the historical cruises but varied widely between cruises. To replicate this, we simulated M_p for every patch within a population using a lognormal random variable with the same mean, but allowed this mean to vary between populations. The high-patchy populations had simulated values of M_p that were over an order of magnitude higher than the values for the low-patchy populations, again with the constraint that all simulated values were within the range of observed values in the historical data.

(*v*) Background numbers of fish, $B(x_i, y_i)$, at locations x_i and y_i that do not belong to any patch were simulated by a constant value for each population. The observed mean density outside of a patch in the historical data varied widely between cruises but was highly correlated with the mean of the maximum patch densities within the cruise ($r = 0.90$). The constant, *B*, was simulated for each population as a linear function of the mean of the maximum patch densities for that population.

Simulating density at each location

The simulated density of fish, $Z(x_i, y_i)$, or number per 10 m^2 at each location (x_i, y_i), was the sum of the densities at that location from all overlapping simulated patches, with the added background density level, that is

$$\text{(B1)} \quad Z(x_i, y_i) = \left[\sum_{p=1}^{P} f_p(x_i, y_i \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2) M_p + B(x_i, y_i)\right]\eta$$

where *P* is the total number of patches, $f_p(x_i, y_i \mid \boldsymbol{\mu}_p, \boldsymbol{\sigma}_p^2)$ is the simulated relative density within the *p*th patch, M_p is the maximum density within a patch, and $B(x_i, y_i)$ is the background density, all as defined above. A random lognormal variate, η, was then multiplied to the simulated density for every cell to incorporate variability about the modeled densities.

Measuring high and low patchiness

Our populations were classified as high-patchy or low-patchy based on the values of the parameters used to simulate them. In an attempt to quantify the patchiness with a single metric, we computed Lloyd's index of patchiness (LI) (Lloyd 1967; Pielou 1969):

$$\text{LI} = 1 + \frac{\text{Var}(Z) - \bar{Z}}{\bar{Z}^2}$$

where

$$\bar{Z} = \frac{1}{8857}\sum_{i=1}^{8857} Z(x_i, y_i)$$

and

$$\text{Var}(Z) = \frac{1}{8857}\sum_{i=1}^{8857} [Z(x_i, y_i) - \bar{Z}]^2$$

Examples of simulated high- and low-patchy populations are shown in Figs. 1*a* and 1*b*.

Appendix C

Appendix C appears on following page.

Table C1. Equations used to calculate nonadaptive estimators of population mean and variance.

Survey design	$\hat{\mu}$	$\hat{\text{V}}\text{ar}(\hat{\mu})$
Random	$= \sum_{i=1}^{n} z_i / n$	$= \left(\frac{N-n}{N}\right)\frac{\sum_{i=1}^{n}(z_i-\hat{\mu})^2}{n(n-1)}$
Unequal probability	$= \frac{1}{n}\sum_{i=1}^{n}\frac{z_i}{N\pi_i}$ where $\pi_i = \frac{\bar{z}_i'}{\sum_{i=1}^{N}\bar{z}_i'}$	$= \left(\frac{N-n}{N}\right)\frac{\sum_{i=1}^{n}\left(\frac{z_i}{N\pi_i}-\hat{\mu}\right)^2}{n(n-1)}$
Systematic cluster	$= \frac{1}{n_s}\sum_{j=1}^{n_s}\bar{z}_j$ where $\bar{z}_j = \frac{1}{M_j}\sum_{i=1}^{M_j} z_{ij}$	$= \frac{N_s-n_s}{N_s}\sum_{j=1}^{n_s}\frac{(\bar{z}_j-\hat{\mu})^2}{n_s(n_s-1)}$
Systematic	$= \sum_{i=1}^{n} z_i / n$	$= \left(\frac{N-n}{N}\right)\frac{\sum_{i=1}^{n}(z_i-\hat{\mu})^2}{n(n-1)}$
Stratified systematic	$= \frac{1}{N}\sum_{h=1}^{H}N_h\,\bar{z}_h$ where $\bar{z}_h = \frac{1}{n_h}\sum_{j=1}^{n_h}\bar{z}_{hj}$ $\bar{z}_{hj} = \frac{1}{M_{hj}}\sum_{i=1}^{M_{hj}} z_{hij}$	$= \frac{1}{N^2}\sum_{h=1}^{H}N_h^2\,\hat{\text{V}}\text{ar}(\bar{z}_h)$ where $\hat{\text{V}}\text{ar}(\bar{z}_h) = \frac{N_s-n_h}{N_s}\sum_{j=1}^{n_h}\frac{(\bar{z}_{hj}-\bar{z}_h)^2}{n_h(n_h-1)}$

Note: All symbols are defined in Appendix A.

Appendix D. Estimators for adaptive cluster sampling

Methods

The estimators commonly used for adaptive sampling are the Hansen–Hurwitz (HH) and the Horvitz–Thompson (HT) estimators. These estimators divide each observation by the draw-by-draw selection probability or the inclusion probability, respectively. Because these probabilities are not known for every unit in the sample, a modified HH estimator is often used, based on the number of times a unit in the initial sample intersects the *k*th network (all terms are defined in Appendix A). In contrast, a modified HT estimator is based on the partial inclusion probability, that is the probability that the initial sample intersects the *k*th network (Table D1) (Thompson and Seber 1996). One problem with the HT estimator is that the formula can yield a negative variance. Therefore, we also computed two other modifications of the HT variance estimator. The Yates–Grundy–Sen (YGS) estimator of the variance of the HT mean is usually nonnegative (Raj 1956; Thompson and Seber 1996), but is unbiased only for fixed sample sizes (Sarndal et al. 1992). Although our sample sizes were not fixed, they vary little because of our stopping rule and our stratification. Another estimator of the variance of the HT mean is the Brewer–Hanif (BH) estimator. It is conservative (slightly greater than the actual variance) and invariably nonnegative (Thompson 1992). The last estimators of the mean and variance that we considered were the Hajek–Sarndal (HS) estimators. Hajek's estimation of the mean is recommended when π_k is not proportional to z_k. Sarndal et al. (1992) derived the equation for the variance of this estimator (Table D1).

Results

In looking at the bias of the mean of the three adaptive cluster estimators of the mean, the HS estimator has a positive bias of about 12% and 8% for high- and low-patchy populations, respectively, whereas HH and HT have a negative bias of less than 2% (Fig. D1). The bias in the HS estimator was probably due to our particular survey design, because the sum of the reciprocal of probabilities of selection consistently underestimated *N*. The HS estimator had considerably more bias than HH and HT, was slightly more imprecise, and was no better in terms of estimating the SE; hence, the HS estimator was dropped from all further consideration.

The bias of the mean in the HH and HT estimators was only slightly greater than Cochran's (1977) rule of thumb for negligible bias, which is one-tenth of the SE. In terms of precision of the mean, both estimators performed worse for

Table D1. Equations used to calculate adaptive cluster estimators of population mean and variance.

Estimator*	$\hat{\mu}$	$\hat{\mathrm{V}}\mathrm{ar}(\hat{\mu})$
HH	$= \frac{1}{n_s} \sum_{j=1}^{n_s} \frac{1}{M_j} \sum_{k=1}^{\kappa} \left(\frac{z_k I_{jk}}{n_k} \right)$	$= \left(\frac{N_s - n_s}{N_s} \right) \frac{1}{n_s(n_s - 1)} \sum_{j=1}^{n_s} \left[\left(\frac{1}{M_j} \sum_{k=1}^{\kappa} \frac{z_k I_{jk}}{n_k} \right) - \hat{\mu} \right]^2$
HT	$= \frac{1}{\left(\sum_{j=1}^{N_s} M_j \right)} \sum_{k=1}^{\kappa} \frac{z_k I_k}{\pi_k}$	$= \left(\frac{N_s - n_s}{N_s} \right) \frac{1}{\left(\sum_{j=1}^{N_s} M_j \right)^2} \sum_{k_1=1}^{\kappa} \sum_{k_2=1}^{\kappa} \left(\frac{z_{k_1} z_{k_2} I_{k_1} I_{k_2}}{\pi_{k_1 k_2}} \right) \left(\frac{\pi_{k_1 k_2}}{\pi_{k_1} \pi_{k_2}} - 1 \right)$
	where $\pi_k = 1 - \binom{N_s - n_k}{n_s} \Big/ \binom{N_s}{n_s}$ and I_k is as defined above except it is independent of the primary unit	where $\pi_{k_1 k_2} \neq 0$ is given by $\pi_{k_1 k_2} = 1 - \left[\binom{N_s - n_{k_1}}{n_s} + \binom{N_s - n_{k_2}}{n_s} - \binom{N_s - n_{k_1} - n_{k_2} + n_{k_1,k_2}}{n_s} \right] \Big/ \binom{N_s}{n_s}$
YGS	Same as for HT	$= \left(\frac{N_s - n_s}{N_s} \right) \frac{-1}{2 \left(\sum_{j=1}^{N_s} M_j \right)^2} \sum_{k_1=1}^{\kappa} \sum_{k_2=1}^{\kappa} \left(1 - \frac{\pi_{k_1} \pi_{k_2}}{\pi_{k_1 k_2}} \right) \left(\frac{z_{k_1}}{\pi_{k_1}} - \frac{z_{k_2}}{\pi_{k_2}} \right)^2$
BH	Same as for HT	$= \left(\frac{N_s - n_s}{N_s} \right) \frac{1}{\left(\sum_{j=1}^{N_s} M_j \right)^2} \frac{\sum_{k=1}^{\kappa} \left[\frac{\kappa z_k}{\pi_k} - \left(\sum_{j=1}^{N_s} M_j \right) \hat{\mu} \right]^2}{\kappa(\kappa - 1)}$
HS	$\mu_{\mathrm{HT}} = \sum_{k=1}^{\kappa} \frac{z_k I_k}{\pi_k} \Big/ \sum_{k=1}^{\kappa} \frac{1}{\pi_k}$	$= \left(\frac{N_s - n_s}{N_s} \right) \frac{1}{\left(\sum_{j=1}^{N_s} M_j \right)^2} \sum_{k_1=1}^{\kappa} \sum_{k_2=1}^{\kappa} \left(\frac{\pi_{k_1 k_2} - \pi_{k_1} \pi_{k_2}}{\pi_{k_1} \pi_{k_2} \pi_{k_1 k_2}} \right) (z_{k_1} - \mu_{\mathrm{HT}})(z_{k_2} - \mu_{\mathrm{HT}})$

Note: All symbols are defined in Appendix A.
*HH, Hansen–Hurwitz; HT, Horvitz–Thompson; YGS, Yates–Grundy–Sen; BH, Brewer–Hanif; HS, Hajek–Sarndal.

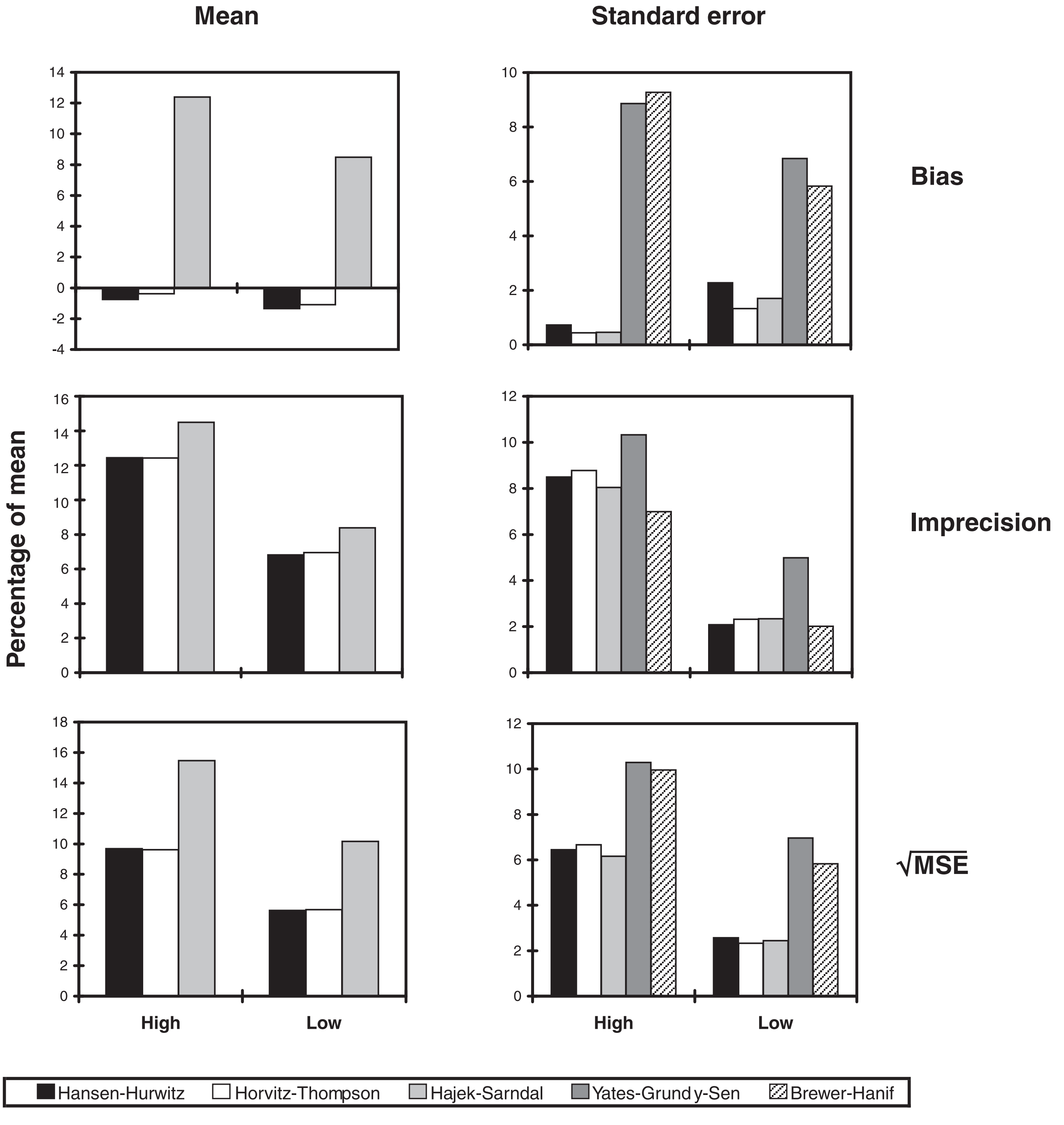
high-patchy populations compared with low-patchy populations (12% and 7%, respectively) (Fig. D1). The ANOVAs and multiple comparison tests failed to detect significant differences between these two estimators in terms of bias, precision, or $\sqrt{\mathrm{MSE}}$ about the mean for either the high- or low-patchy populations (Table D2).

The HH and the HT estimators of the SE were similar to each other, overestimating the SE by a negligible amount (Fig. D1). YGS and BH were by far the poorest estimators of the SE in terms of bias, overestimating by around 9% for high-patchy and 6% for low-patchy populations (Fig. D1). Although YGS and BH were the poorest estimators in terms of bias and $\sqrt{\mathrm{MSE}}$ of the SE, BH was the most precise. However, because $\sqrt{\mathrm{MSE}}$ captures both bias and precision, this was used as the final criterion for dismissing YGS and BH as desirable estimators for these data sets.

Two variance estimators (YGS and HT) produced negative estimates of variance, which are obviously invalid, although computationally possible. The frequency of negative variances for the HT estimator was higher overall for the high-patchy populations (6% for the high and 1% for the low). The incidence of negative variances for the YGS estimator was similar (7% for the high and 1.5% for the low). The occurrence of negative variances for HT has been observed by others (Rao and Singh 1973), but YGS has been reported to be usually positive (Thompson and Seber 1996). All negative variances were excluded for purposes of computing measures of bias and precision of the SE for these two estimators.

Of the two remaining variance estimators (having eliminated the HS, YGS, and BH estimators), there were no significant differences between the HT and the HH variance estimators for relative bias, SD, or $\sqrt{\mathrm{MSE}}$ of the SE for high-patchy populations. For low-patchy populations, the HT is significantly smaller in terms of relative bias of the SE (Table D2). However, in terms of the relative $\sqrt{\mathrm{MSE}}$ for both population groups for both mean and SE, there were no sig-

Fig. D1. The bias, imprecision, and $\sqrt{\mathrm{MSE}}$ of the mean and SE averaged over the 50 high- and 50 low-patchy populations for the five different adaptive cluster estimators.



nificant differences (Table D2). Because this is a measure of both bias and precision combined and because the HH estimator has some advantages over the HT estimator (i.e., it never produces a negative variance estimate and it is much easier computationally), we recommend the HH estimator in practice.

Table D2. Results of multiple comparison tests showing significant differences (at 5% significance level) among adaptive cluster survey design estimators for both (*i*) high-patchy and (*ii*) low-patchy populations.

	Adaptive HH	Adaptive HT	Adaptive YGS**	Adaptive BH**	Adaptive HS
High-patchy					
Bias of mean	A	A	A	A	B
SE of mean	A	A	A	A	B
MSE of mean*	A	A	A	A	B
Bias of SE	A	A	B	B	A
SD of SE	B	B	C	A	AB
MSE of SE*	A	A	B	B	A
Low-patchy					
Bias of mean	A	A	A	A	B
SE of mean	A	A	A	A	B
MSE of mean*	A	A	A	A	B
Bias of SE	B	A	D	C	A
SD of SE	AB	B	C	A	B
MSE of SE*	A	A	C	B	A

Note: HH, Hansen–Hurwitz; HT, Horvitz–Thompson; YGS, Yates–Grundy–Sen; Brewer–Haniff; and HS, Hajek–Sarndal. Estimators with same letters indicate no significant difference.

*MSEs capture both the bias and precision.

**The adaptive estimators BH and YGS are different estimators of the variance for the HT estimator of the mean. Therefore, the bias, SE, and MSE of the mean are the same as the HT estimator.