USE OF ADAPTIVE CLUSTER SAMPLING FOR HYDROACOUSTIC SURVEYS

M. Elizabeth Conners* and Steven J. Schwager

ABSTRACT

Resource managers are often required to estimate the size of a wildlife population based on sampling surveys. This problem is especially critical in fisheries, where stock size estimation forms the basis for key policy decisions. This study looks at design-based methods for a hydroacoustic fisheries survey, with the goal of improving estimation when the target stock has a patchy spatial distribution. In particular, we examine the efficiency and feasibility of a relatively new design-based method known as adaptive cluster sampling (ACS). A simulation experiment looks at the relative efficiency of ACS and traditional sampling designs in a hydroacoustic survey setting. Fish densities with known spatial covariance are generated and subjected to repeated sampling; distributions of the different estimators are compared.

Hydroacoustic data frequently display strong serial correlation along transects, so traditional designs based on one-stage cluster sampling are appropriate. Distribution of the estimator for these designs was markedly skewed. ACS designs performed better than traditional designs for all stocks with small-scale spatial correlation in fish density, yielding estimates with lower variance. ACS estimators were not skewed and had a lower frequency of large errors. For the most variable stock, use of ACS reduced the CV of the stock size estimate from over 0.9 to around 0.5. Differences between fixed-size and ACS designs were consistent over multiple realizations of each spatial covariance model.

A survey of rainbow smelt (*Osmerus mordax*) in the eastern basin of Lake Erie was used as a case study for development of a survey design. A field trial showed that use of ACS for the survey is feasible, but pointed out some areas for further research. The biggest drawback to use of ACS is uncertainty in the final sample size; this can be partially controlled by applying ACS within a stratified design. ACS retains the unbiased and non-parametric properties of design-based estimation, but allows increased sampling in high-density areas that are of greater biological interest. For stocks with an aggregated or patchy spatial distribution, ACS can provide a more precise estimate of stock size than traditional survey methods.

KEYWORDS: adaptive cluster sampling, survey design, spatial sampling, stock assessment, patchy distributions

ADDRESSES OF AUTHORS:

M. Elizabeth Conners* and Steven J. Schwager
Department of Biometrics, Cornell University,
Ithaca, NY 14853

*Current Address: Resource Ecology and Fishery Management Division
NMFS Alaska Fisheries Science Center
7600 Sand Point Way NE, Seattle WA 98115-0070

Correspondence should be addressed to M.E. Conners at liz.conners@noaa.gov

ACKNOWLEDGEMENTS

This work is part of a Master's thesis at Cornell University, written by MEC under the direction of SJS. Support for MEC was provided by National Institute of Health Sciences Training Grant in Statistics and Environmental Science, Grant number EHS-5-T32-ES070261. Data and ship time were provided by the New York State Department of Environmental Conservation, Lake Erie Fisheries Unit. Thanks to Naomi Altman, Lars Rudstam, Patrick Sullivan, and Neal Williamson for helpful reviews of the preliminary and final manuscripts.

INTRODUCTION

A persistent problem in management of natural resources is estimating the size of a population or stock from a limited sample. Determining the population size is particularly important in fisheries management, where key policy decisions are based on the estimated size of a stock. Historically, fish stocks were often assessed with methods yielding only a relative index of abundance (Gunderson 1995). Recently, there has been an increased effort not only to estimate stock sizes absolutely, but also to quantify the uncertainty in the estimate. Hydroacoustic methods provide both absolute abundance estimates and a much larger sample size than traditional gears, allowing wider application of statistical sampling theory to stock assessment surveys.

Statistical theory provides a number of design-based methods for estimating mean density or total population size. These methods are primarily designed for situations where sample units are independent and the underlying distribution is fairly normal, conditions rarely met by fisheries data. More often, fish density data show strong skewness, high kurtosis, and local correlation, resulting in a very large variance of estimation (Gilbert 1987, Patil and Rao 1994, Foote and Stefansson 1993). These problems are particularly acute when the target stock has an aggregated or "patchy" spatial distribution (Appenzeller and Leggett 1996, Barrange and Hampton 1997). This study looks at design-based methods for a hydroacoustic survey, with the goal of improving estimation when the target stock has a patchy spatial distribution. In particular, we examine the efficiency and feasibility of a relatively new (Thompson 1990, Thompson and Seber 1996) design-based method known as adaptive cluster sampling (ACS). While this study is based on fisheries applications, the results are also applicable to spatial surveys in many other fields, including forestry (Roesch 1993), wildlife ecology (Smith et al 1995), and epidemiology (Thompson 1996).

Design-Based Methods for Hydroacoustic Surveys

Hydroacoustic data processing provides a direct estimate of area- or volume-normalized fish density over a sampling unit. If densities in adjacent sampling units are independent, the variance of the stock size estimate can be estimated simply with the sample variance.

When adjacent sampling units are strongly correlated, however, the observed sample variance will grossly underestimate the true variance of estimation (Williamson 1982). The correct design-based approach in this situation is to use cluster sampling formulas, with the transect being the primary unit and the integrated segments of cruise track (referred to as EDSU's in some literature) as secondary units. Key early works on design of hydroacoustic surveys include Francis (1984), Gavaris and Smith (1987), Jurvelius and Auvinen (1989), and Jolly and Hampton (1990). A special ICES workshop held in 1992 (ICES 1993) reviewed design- and model-based approaches for hydroacoustic stock assessment. Current hydroacoustic survey designs are primarily based on cluster sampling with parallel transects across the study area, placed by either systematic or stratified random designs (Brandt et al 1991, Hampton 1996, Simmonds and Fryer 1996).

There has also been considerable interest in model-based methods of estimation (Sullivan 1991, Swartzman 1992, Stolyarenko 1992, Steffanson 1996). Geostatistical methods have become popular for modeling data with spatial correlation (Guillard et. al 1992, Petitgas 1993a, 1993b, Pelletier and Parma 1994, Williamson and Traynor 1996). Model-based approaches forecast the total stock size by predicting fish density in unsampled regions of the study area, and allow calculation of the uncertainty of the total based on estimated variance of the error terms in the model (Ripley 1981, Foote and Steffanson 1993). In some situations, an appropriate model-based estimate can greatly improve precision over random sampling designs (Sullivan 1991). For stocks with very patchy distributions, however, a smooth surface trend model may be a very poor fit to the data (Foote and Steffanson 1993, Murray 1996). Comparison of ACS estimation with model-based methods is not included in this paper, but will be the subject of future work.

Adaptive Cluster Sampling (ACS)

Adaptive cluster sampling (ACS) is a design-based method that can be used when data are strongly correlated. The basic theory was put forth by Thompson (1990, 1991a, 1991b, 1993, 1996), Seber and Thompson (1994), and Thompson and Seber (1996). ACS was designed especially for situations where standard cluster sampling is ineffective; when the target stock tends to concentrate in a few dense clusters rather than being evenly distributed over the study area. Theoretical analysis shows that ACS reaches its greatest efficiency (relative to simple random sampling) when the target organisms are highly clustered, rare, or both (Thompson 1990, Thompson and Seber 1996, Christman 1997). Monte-Carlo simulations (Conners 1999) show that high relative efficiency of ACS is associated with frequency distributions of fish density that are strongly skewed, have high kurtosis, and have a large proportion of units with zero or very low densities. These types of distributions are frequently observed in fisheries data, especially with species that exhibit schooling behavior or strong microhabitat associations (Hampton 1996, Simmonds and Fryer 1996).

The Lake Erie Smelt Survey

As a framework and motivation for this study we use data from a hydroacoustic survey of rainbow smelt (*Osmerus mordax*) in the eastern basin of Lake Erie. Estimates of the total number of yearling and older (YAO+) smelt are used in formulating catch limits and stocking policies for smelt and their predators. The New York Department of Environmental Conservation (NYSDEC) Lake Erie Fisheries Unit in Dunkirk, New York provided assistance, data, and ship time for testing adaptive sampling techniques. The Lake Erie Fisheries Unit would like to optimize a survey design for estimating the total stock size, with an accurate estimate of the associated variance. Data from preliminary surveys suggest a "patchy" distribution of smelt density, with both small-scale spatial pattern and large-scale trends. Frequency distributions of the preliminary survey data are strongly skewed, and correlation coefficients between adjacent sampling units are 0.6-0.8, indicating strong local correlation. These features suggest that ACS may be particularly efficient for estimating the total size of this stock.

METHODS

ACS with Primary and Secondary Units

In a hydroacoustic survey, only transects may be randomly selected; each transect then includes a set of secondary sampling units in which the fish density is measured (by integrating the acoustic signal over a fixed distance). The type of ACS sampling that is applicable to hydroacoustic surveys, then, is ACS with primary and secondary units (Thompson 1991b, Thompson and Seber 1996, Pontius 1997). In this form of "Strip ACS", transects are the primary units and integrated sampling units are secondary units. The initial sampling design consists of one or more randomly or systematically placed transects. All of the sampling units in the initial design are measured; *secondary* units are added in the neighborhood of any *secondary* unit meeting the ACS criterion. The final sample includes all of the initial transects plus a "cloud" of adaptively added secondary units where any transect intersects a high-density cluster (Figure 1). The number of units added to the sample depends on the ACS criterion and neighborhood definition used, as well as on the scale of the secondary units and the spatial distribution of target fish.

ACS provides two methods for estimating the overall mean density (or total) and variance of the estimator (Table 1). Both estimators balance the total number of fish in a cluster or network (t_k) against the probability of detecting that network, based on the "width" (x_k) of the network relative to the initial sampling design. Estimation of the mean over the study area is based on the means within the sampled networks, including a large number of networks of size one and the few larger networks. The first estimator, referred to as the Hansen-Hurwitz-type (HH) estimator, is based on sampling with replacement and draw-by-draw selection probability for each transect. Variance estimation for the HH estimator is based on variance between the sampled networks. The second estimator, the Horvitz-Thompson-type (HT) estimator, may be used when sampling with or without replacement. The HT estimator uses a combinatoric argument to estimate individual and pairwise inclusion probabilities for each network (Table 1). This calculation, while conceptually straightforward, can be complex to implement (Conners 1999).

When the total number of transects in the study area (N) is large, the two estimation procedures yield nearly identical results [$\alpha_k \rightarrow (x_k/N)$, see Table 1]. For both estimators, the estimated probability of detecting a network is a function of the number of transects that intersect it and the probability of selecting those transects with the initial survey design. Thus, large clusters have a higher probability of being detected, and are downweighted in the estimation of the overall mean. Small clusters and units that do not meet the criterion have smaller inclusion probabilities, and contribute more to the overall mean. This weighting counteracts the positive bias in the estimated mean that would normally result from including a large number of high-density units in the sample.

Simulation Study

A simulation study was conducted to test the efficiency of ACS for a fish stock similar to Lake Erie smelt. Simulated test stocks were created with known true total size and different levels of spatial aggregation. Selected stocks were sampled repeatedly, using both traditional and ACS designs. The study included four one-stage cluster sampling (traditional) designs and two ACS designs with different initial transect layouts. For each sampling replicate, the estimated total stock size and variance of the estimator were calculated. The experiment tabulated relative estimation errors (difference between the estimated and true total, $\left|\left(\hat{T}-T\right)/T\right|$) over 5,000 random samples of each design. Designs are compared based on the distribution of the estimated total $\hat{T}$ and the variance of the estimator over the sampling replicates. Programming was conducted in MATLAB 4.1.

An isotropic spherical variogram model was used to calculate the variance-covariance matrix for points on a 100x50 grid. This covariance matrix was then combined with randomly generated standard normal variables to generate a bivariate normal surface with the modeled covariance structure. In order to give the simulated data the strong skewness observed in the Lake Erie data, the generated densities were exponentiated, so that points on the final simulated grid had a lognormal distribution. After experimentation with several sets of variogram parameters, four model specifications were selected for sampling. The selected models (Table 2) represent stocks with no local correlation ("Random"), with strong local correlation over a large range ("Big Patches"), and with strong local correlation over a smaller range ("Small Patches"). A fourth stock ("Rare Patches") represents strong local correlation with relatively high background noise or nugget, which is most similar to the Lake Erie smelt data. All of the test stocks were generated with constant mean, which implies no large-scale spatial pattern. A number of realizations were generated for each of the spatial models; grids "typical" of each spatial model are shown in Figure 2a. The four grids in Figure 2 were standardized to have an equal true population total. For ease of interpretation, conclusions were based largely on comparison of these four standardized test stocks. These conclusions were, however, verified over 20 realizations of the stochastic spatial surface for each model.

Sampling and estimation for the traditional designs were performed using one-stage cluster sampling formulas (Cochran 1977, Thompson 1992). Equal allocation of transects to strata was used in stratified designs, since the simulated stocks had no large-scale spatial pattern. Traditional designs included random selection of 10 transects, systematic selection of 10 transects with a random start, and three stratified random sampling designs. Stratified designs divided the study grid along the long axis into two strata with five transects/stratum, into five strata with two transects per stratum, and along both axes of the study grid to form 10 strata, with two (half-length) transects in each stratum. Results were similar for the three types of stratification, so only the two-stratum design is presented here.

ACS sampling was performed according to Thompson and Seber (1996, Section 4.7) with transects as primary units and individual grid cells as secondary units (Figure 1). As in other work on ACS, a neighborhood definition of four-adjacent-cells (the four cells sharing a common boundary with the target cell) was used. For both ACS designs, the critical value defining networks was set at the 80th percentile of the true distribution of the grid points. Other studies (Conners 1999) have shown that this is near the optimal critical value for a skewed population, and that final estimates are not sensitive to small differences in critical value. In order to minimize computations, network statistics (t_k, w_k, α_k) were tabulated for the entire grid, so that each sampling simulation simply looked up values for the networks intersected by the sample. Calculation of two-way inclusion probabilities for the HT estimator proved to be complex; for a detailed algorithm and computer code, see Conners (1999). Output from sampling simulations for ACS designs included the final sample size after addition of adaptive units. Initial sample sizes for

ACS designs were selected to give expected final sample sizes as close as possible (in total number of secondary units sampled) to the fixed size of the traditional designs.

Field Trial of ACS for Hydroacoustic Surveys

In addition to the simulation study, we used one night of the 1998 Lake Erie survey to test the practicality of field implementation of an ACS design. Ship time was provided by the NYSDEC. The ACS trial consisted of one initial transect, followed by addition of adaptive sampling units. As a definition of the ACS "neighborhood", we added units along transect segments parallel to the initial transect at a spacing of 1.5 km, collecting hydroacoustic data over the range of latitude where units above the critical value were observed.

RESULTS

Simulation Study

The simulation study showed clear differences in the behavior of the estimators between traditional and ACS methods (Table 3). The estimators from traditional cluster sampling, based on the sample mean, were unbiased but did *not* have a symmetric distribution. Figure 3 shows the relative errors of estimation $\left[(\hat{T} - T)/T\right]$ from cluster sampling with stratified random transect selection. While the majority of the sampling replicates produced estimates close to the true total, relative errors close to –1 ($\hat{T}$ near 0) and above 1 ($\hat{T}$ more than twice T) were not uncommon. Estimators for the other cluster sampling designs had similar distributions.

The strong positive skewness in the traditional estimators was a result of a small effective sample size from skewed underlying distributions. Because the traditional designs use cluster-sampling algorithms, they are in effect a random sample of transect totals. The simulated transect totals vary widely, depending on whether the transect intersects a patch (Figure 2b). For the models with spatial correlation, this results in a strongly skewed distribution of transect totals. Smaller patch size and increased rarity of patches increase the variance and skewness of the transect totals (Figure 2). The effective sample size is equal only to the number of transects in the survey. For the simulations, a sample of 10 transects (10% of the total study area) was selected, so estimation of the stock size is based on ten transect totals. In many field surveys, this pattern would be compounded by large-scale trends in density and variation in the length of transects. While skewness in the estimator can be reduced by increasing the number of transects, for many surveys a larger number of transects would not be feasible.

Adaptive cluster sampling both increased the efficiency of estimation over traditional sampling designs and produced estimators with a more symmetric distribution. Table 3 shows the coefficient of variation for the estimate of total stock size over 5,000 replicates of each sampling design. For the stock without spatial correlation, all of the traditional sampling designs and the stratified ACS design had CV's in the range of 0.32-0.37. For the correlated stocks, however, the stratified ACS design reduced the CV compared to all three traditional designs. The stock with small patches had a CV of 0.37-0.38 for traditional sampling designs but 0.21-0.23 for the ACS designs. For the stock with big patches, ACS reduced the CV from 0.65-0.70 for traditional designs to 0.31-0.32. The stock with rare patches showed a reduction in CV from 0.90-0.99 down to 0.39-0.44.

In order to be sure that results were not an artifact of particular test stocks, simulations were repeated over stocks from 20 realizations of each of the four spatial models. A check of variance components showed that uncertainty in estimation due to sampling was greater than that due to stochastic variation in the underlying model. Increased efficiency of ACS over traditional designs was consistent over the repeated realizations of the correlated spatial models. For each model, the stratified ACS estimator was the optimal design, in that it had the smallest overall Bayes risk of estimation (average error of estimate over variation due to both the spatial model and random transect selection).

Table 3 also shows the "Relative Efficiency" for each sampling design. This measure compares the variance of the sampling estimator to the theoretical variance of a simple random sample of "equivalent" size:

$$RE = \frac{Var_{SRS}(n_{eq})}{Var(\hat{T})} \qquad \hat{V}_{SRS}(n_{eq}) = \frac{(N - n_{eq})}{N} * \frac{\sigma^2}{n_{eq}}$$

where σ^2 is the true variance of the simulated test stock and n_{eq} is the size of the "equivalent" sample. Following Seber and Thompson (1996, Table 4.4), the"equivalent" sample size is 500 units for the fixed-size designs and the average final sample size (see Table 3) for ACS designs. A Relative Efficiency greater than one indicates that the sampling estimator has a smaller variance than expected from simple random sampling, or that it is more efficient than simple random sampling. Fixed-size cluster designs were more efficient than SRS only when systematic and stratified cluster designs were applied to the stock with "big" patches. These two designs are known to be efficient for situations where the local covariance (in this case covariance between the transect totals) is monotonically decreasing (Ripley 1981, page 25). ACS designs were more efficient than "equivalent" SRS for both the small and the rare target stocks. This is consistent with the results of previous authors (Thompson 1991b and Christman 1997), who have demonstrated that ACS is more efficient than SRS when the target stock is rare, highly aggregated, or both. For "rare patches" stock, the relative efficiency of ACS designs is over three times that of SRS. The ACS designs were not efficient for the "big patches" stock because the large final sample size made the equivalent SRS variance small.

Thompson and Seber (1996, page 129) and Christman (1997) compare the relative efficiencies of the two ACS estimation procedures (the HH and HT estimators). These authors also found that both estimators had relative efficiency >1 for target stocks that were rare and/or highly aggregated, but lower efficiency for more dispersed stocks. Both authors also note that the HT estimator has better efficiency (lower variance) than the HH estimator, with the difference becoming more pronounced as the initial sample size is increased. In our simulations the initial sample size never exceeded 10%, and the two estimators gave very similar results.

Both estimation procedures also gave reasonable estimates of the variance of T. Simulated ACS designs included both systematic and stratified random designs for the initial transect selection. Both initial designs showed symmetric distributions and reduced error frequency relative to traditional designs. ACS with the initial systematic design performed less efficiently than ACS from a stratified random design, as a result of stronger limitation on final sample sizes (discussed below).

One of the greatest concerns about implementaion of ACS is the random nature of the final sample size, and the possibility that the final sample might grow too large to be feasible. Distributions of final sample size (in secondary units) for some of the simulated ACS samples are shown in Table 4. The difference between initial and final sample size was strongly related to the spatial distribution of the stock. For the stock with no spatial correlation, average final sample sizes were only 11%-13% larger than the initial sample, regardless of the initial sample size or design. For this stock, none of the final sample sizes exceeded 1.25 times the initial sample size. The "Small" and "Rare" stocks, which both have strong correlation over a short range, had a greater increase in average final sample size over initial size. Final sample size for these stocks was 1.5-1.7 times initial sample size using stratified random ACS and 1.7-2.1 times the initial sample size when starting from a systematic sample. The stock with a few "Big" clusters showed the greatest increase in final sample size; average final sample sizes for this stock were 2-4 times the size of the initial sample.

For all of the test stocks, the increase from initial to final sample size was larger when the initial design was systematic that when a stratified random initial design was used. In the case of the "Small" and "Rare"stocks, this effect produced substantial differences. These differences occur because our algorithm for stratified ACS treated each stratum as a separate entity; adaptive sampling for clusters located near the edge of the stratum was stopped at the stratum boundary and not allowed to expand into adjacent strata. Thompson and Seber (1996, pg 134) state that terminating a network at the stratum boundary is slightly less efficient than using complete networks, but ACS estimators will still be design-unbiased for the stratum totals and may be combined into an overall estimate assuming independence of the strata. Defining networks in this way makes the strata into "partition boundaries" that limit the potential size of any network; this is one of the strategies suggested by Thompson and Seber (1996, pg 161) as a means to control final sample size. In our simulations, this stratified design clearly acted to reduce the final sample sizes, which gave slightly higer efficiencies than for the sytematic design.

Results of the Field Trial

One night of the 1998 Lake Erie smelt survey was used to test the practicality of field implementation of ACS. The trial showed that use of an ACS design is feasible, but identified some potential problems. Real-time data analysis capability and differentially corrected GPS are needed to identify units meeting the ACS criterion during the survey and accurately position adaptive units. Capabilities of the survey vessel, especially the relative travel speeds with and without data collection, are important.

The simulation experiment and theoretical work use an ACS "neighborhood" definition of four-adjacent-cells, but this definition is not practical for a hydroacoustic survey. We used a neighborhood definition of parallel transect segments, using Loran navigation lines as approximate parallels. Adaptive units for ACS were identified as segments of parallel transects over the same latitude as units above the critical density. Four of the 24 sampling units in the initial segment met an ACS criterion of density greater than 5,000 smelt/ha; three of these were near the southern shoreline and one was further north (Figure 4). Adaptive transect segments were surveyed on either side of the initial transect over the latitude range of both "patches", so data were collected on some extra units not strictly needed for ACS. These extra units may be useful for other purposes (e.g. mapping of the high-density patches), but they increase the final sample size and decrease the efficiency of ACS. A survey vessel with a high traveling speed (when not sampling) would increase the efficiency of ACS by reducing the travel time between sampled units and facilitating collection of data only at needed points. Addition of adaptive segments was halted to the west at the boundary of the sampling stratum, and to the east by approaching daylight.

The field trial illustrated the greatest concern with application of ACS: detection of a large patch that results in a large final sample size. With only one night of sampling, we were unable to complete all of the adaptive sampling of the detected network. The amount of adaptive sampling could be decreased by using greater spacing between adaptive segments, but too great a distance may affect accurate estimation of the patch total. More research is needed on the best "neighborhood" definition for use with ACS, depending on the expected size and shape of patches.

CONCLUSIONS

The advent of hydroacoustic stock assessment has resulted in dramatic increases in the amount of data that can be collected, but lack of independence between adjacent sampling units can restrict the applicability of design-based theory in this setting. Cluster sampling designs are effective for a spatially dispersed stock, when transect totals are reasonably consistent over the study area. A target stock with a spatially patchy distribution, however, will have a strongly skewed distribution of transect totals, leading to high variance and poor performance of traditional estimators. Patchy spatial distributions can be caused by irregular distributions of microhabitat, by behavioral traits (schooling), or by combination of factors. Strong local correlation is common in fisheries and many other environmental applications.

Adaptive cluster sampling (ACS) was designed for spatially patchy and/or rare events. In simulations, ACS performed better than traditional cluster sampling designs whenever local correlation was present. ACS estimators exhibited an unbiased, symmetric distribution with a consistently lower variance than traditional designs. The coefficient of variation for the most variable of the test stocks was reduced from 0.9 for traditional cluster sampling to 0.4 using ACS. ACS also provided greater protection against gross mis-estimates of total stock size. Using traditional sampling, more than 50% of estimates for the most variable stock were "poor" (relative error more than 50%), and only 6-10% were "good" (within 10% of the true total). ACS sampling of this stock reduced the frequency of "poor" estimates to 20-30% and increased the frequency of "good" estimates (Figure 5).

The greatest limitation to practical use of ACS is the uncertainty in the final sample size; there is always a possibility that the final sample will outgrow available budget or schedule. Some recent research has addressed methods to limit the final size of an ACS sample (Salehi and Seber 1997a,b; Brown and Manly 1998). Thompson and Seber (1996) discuss general ways to limit the final sample size. One practical way is to implement ACS within a stratified initial design, with the adaptive sample ending at stratum boundaries. ACS design parameters and sampling intensity can be adjusted during the survey, as long as parameters are consistent over each stratum. A large sample in one stratum could be partially offset by using greater transect spacing or a higher critical value in subsequent strata. It may also be possible to use a form of post-stratification to analyze portions of the survey that are interrupted by weather or equipment problems. A field trial on Lake Erie demonstrated that ACS is feasible for hydroacoustic surveys. More research is needed on optimal definition of the ACS neighborhood in a line-transect setting and on tradeoff effects between the number and length of transects used.

ACS retains the unbiased and non-parametric properties of design-based estimation, but allows increased sampling in high-density areas that are of greater biological interest. For many fish stocks, the majority of the population is located in a few high-density areas, so increasing the sampling effort in these areas makes both statistical and biological sense. In these circumstances, ACS provides improved precision of stock estimation and is less sensitive to errors caused by the highly skewed distribution of density data. The greater the degree of spatial aggregation exhibited by the stock, the greater is the potential efficiency gain from using ACS. Strong skewness or kurtosis in density data is a good indication that ACS designs may be effective for a particular stock, worth the extra effort in survey design and field execution.

REFERENCES CITED

Appenzeller, A.R., and Leggett, W.C. 1996. Do diel changes in patchiness exhibited by schooling fishes influence the precision of estimates of fish abundance obtained from hydroacoustic trawl surveys? Archiv fur Hydrobiologie, 135(3): 377-391.

Barange, M. and Hampton, I. 1997. Spatial structure of co-occurring anchovy and sardine populations for acoustic data: Implications for survey design. Fisheries Oceanography, 6(2): 94-108.

Brandt, S.B., Mason, D.M., Patrick, E.V., Argyl, R.L., Wells, L., Unger, P.A., and Stewart, D.J. 1991. Acoustic measures of the abundance and size of pelagic planktivores in Lake Michigan. Canadian Journal of Fisheries and Aquatic Sciences, 48: 894-908.

Brown, J. A. and Manly, B.J.F. 1998. Restricted adaptive cluster sampling. Environmental and Ecological Statistics, 5(1): 49-63.

Christman, M.C. 1997. Efficiency of some sampling designs for spatially clustered populations. Environmetrics, 8(2): 145-166.

Conners, M.E. 1999. Use of adaptive cluster sampling for fisheries hydroacoustics. M.S. Thesis, Cornell University.

Cochran, W.G. 1977. Sampling Techniques. John Wiley and Sons, NY.

Einhouse, D., Tyson, J., et al. 1997. Report of the Lake Erie forage task group, Lake Erie Committee, Great Lakes Fishery Commission, March 1997. 24pp.

Foote, K.G. and Stefansson, G. 1993. Definition of the problem of estimating fish abundance over an area from acoustic line-transect measurements of density. ICES Journal of Marine Science, 50: 369-381.

Francis, R.I.C.C. 1984. An adaptive strategy for stratified random trawl surveys. New Zeland Journal of Marine and Freshwater Research, 18: 59-71.

Gavaris, S. and Smith, S.J. 1987. Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic Cod *Gadus-morhua* on the eastern Scotian shelf, Canada. Journal of Northwest Atlantic Fishery Science, 7(2): 137-144.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, Inc., NY.

Guillard, J., Gerdeaux, D. et al. 1992. The Use of Geostatistics to Analyze Data From an Echo-Integration Survey of Fish Stock in Lake Sainte-Croix. Fisheries Research, 13(4): 395-406.

Gunderson, D.R. 1995. Surveys of Fisheries Resources. John Wiley & Sons, NY.

Hampton, I. 1996. Acoustic and egg-population estimates of South African anchovy biomass over a decade: Comparisons, accuracy, and utility. ICES Journal of Marine Science, 53(2): 493-500.

ICES 1993. Report of the workshop on the applicability of spatial statistical techniques to acoustic survey data. ICES Cooperative Research Report 195.

Jolly, G.M. and Hampton, I. 1990. A Stratified Random Transect Design For Acoustic Surveys of Fish Stocks. Canadian Journal of Fisheries and Aquatic Sciences, 47(7): 1282-1291.

Jurvelius, J. and Auvinen, H. 1989. Fish echo-sounding in a large lake a systematic and unsystematic sampling of pelagic fish stock. Aqua Fenn, 19(2): 123-128.

Murray, A.W.A. 1996. Comparison of geostatistical and random sample survey analyses of Antarctic krill acoustic data. ICES Journal of Marine Science, 53: 415-421.

Patil, G. P. and Rao, C.R. Eds. 1994. Environmental Statistics. Elsevier Science Publishers, North-Holland, NY.

Pelletier, D. and Parma, A.M. 1994. Spatial distribution of Pacific halibut *(Hippoglossus stenolepis*): an application of geostatistics to longline survey data. Canadian Journal of Fisheries and Aquatic Sciences, 51(7): 1506-1518.

Petitgas, P. 1993a. Geostatistics for fish stock assessments a review and an acoustic application. ICES Journal of Marine Science, 3:285-298.

Petitgas, P. 1993b. Use of a disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. Aquatic Living Resources, 6(3): 201-209.

Pontius, J.A. 1997. Strip adaptive cluster sampling: Probability proportional to size selection of primary units. Biometrics, 53(3): 1092-1096.

Ripley, B.D. 1981. Spatial statistics. John Wiley & Sons, NY.

Roesch, F.A., Jr. 1993. Adaptive cluster sampling for forest inventories. Forest Science, 39: 655-669.

Salehi, M. and Seber, G.A.F. 1997. Two-stage adaptive cluster sampling. Biometrics, 53(3): 959-970.

Salehi, M. and Seber, G.A.F. 1997. Adaptive cluster sampling with networks selected without replacement. Biometrika, 84(1).

Seber, G.A.F. and Thompson, S.K. 1994. Environmental adaptive sampling. *In* Handbook of Statistics 12: 201-220. Ed. by G. P. Patil and C. R. Rao. Elsevier Science Publishers, North Holland, NY.

Simard, Y.Legendre,P., et al. 1992. Mapping, estimating biomass, and optimizing sampling programs for spatially autocorrelated data case study of the northern shrimp *Pandalus borealis*. Canadian Journal of Fisheries and Aquatic Sciences, 49(1): 32-45.

Simmonds, E.J. and Fryer, R.J. 1996. Which are better, random or systematic acoustic surveys? A simulation using North Sea herring as an example. ICES Journal of Marine Science, 53(1): 39-50.

Smith, D.R., Conroy, M.J. et al. 1995. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. Biometrics, 51(2): 777-788.

Stefansson, G. 1996. Analysis of groundfish survey abundance data: Combining the GLM and delta approaches. ICES Journal of Marine Science, 53(3).

Stolyarenko, D.A. 1992. Two-dimensional spline approximation on local integral. Computational Statistics & Data Analysis, 14: 509-514.

Sullivan, P.J. 1991. Stock abundance estimation using depth-dependent trends and spatially correlated variation. Canadian Journal of Fisheries and Aquatic Sciences, 48(9): 1691-1703.

Swartzman, G.H. 1992. Spatial-analysis of Bering Sea groundfish survey data using generalized additive models. Canadian Journal of Fisheries and Aquatic Sciences, 49(7).

Thompson, S.K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association, 85(412): 1050-1059.

Thompson, S.K. 1991a. Stratified adaptive cluster sampling. Biometrika, 78(2): 389-398.

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

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

Thompson, S.K. 1996. Adaptive cluster sampling based on order statistics. Environmetrics 7(2).

Thompson, S.K. 1993. Multivariate aspects of adaptive cluster sampling. *In* Multivariate Environmental Statistics, pp. 561-572. Ed. By G. P. Patil and C. R. Rao. Elsevier Science Publishers, North Holland, New York.

Thompson, S.K., Ramsey, F.L., et al. 1992. An adaptive procedure for sampling animal populations. Biometrics, 48(4): 1195-1199.

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

Williamson, N.J. 1982. Cluster sampling estimation of the variance of abundance estimates derived from quantitative echo sounder surveys. Canadian Journal of Fisheries and Aquatic Sciences, 39: 229-231.

Williamson, N.J. and Traynor, J.J., 1996. Application of a one-dimensional geostatistical procedure to fisheries acoustic surveys of Alaskan pollock. ICES Journal of Marine Science, 53(2): 423-428.

Table 4. Distribution of final sample size (in secondary units, including adaptive units), for ACS sampling.

Stock	Initial Design	Initial Size	Avg. Final Size (F)	90th %tile	Std. Dev.	Final/ Initial	90% / Initial
Random	Strat	400	449	462	10.0	1.12	1.16
	System	400	455	461	5.9	1.14	1.15
Small	Strat	400	674	736	50.4	1.68	1.84
	System	250	521	572	42.0	2.09	2.29
Rare	Strat	400	602	651	38.4	1.50	1.63
	System	250	473	550	48.3	1.89	2.20
Big	Strat	400	1087	1158	89.2	2.72	2.90
	System	100	684	863	276.7	6.84	8.63

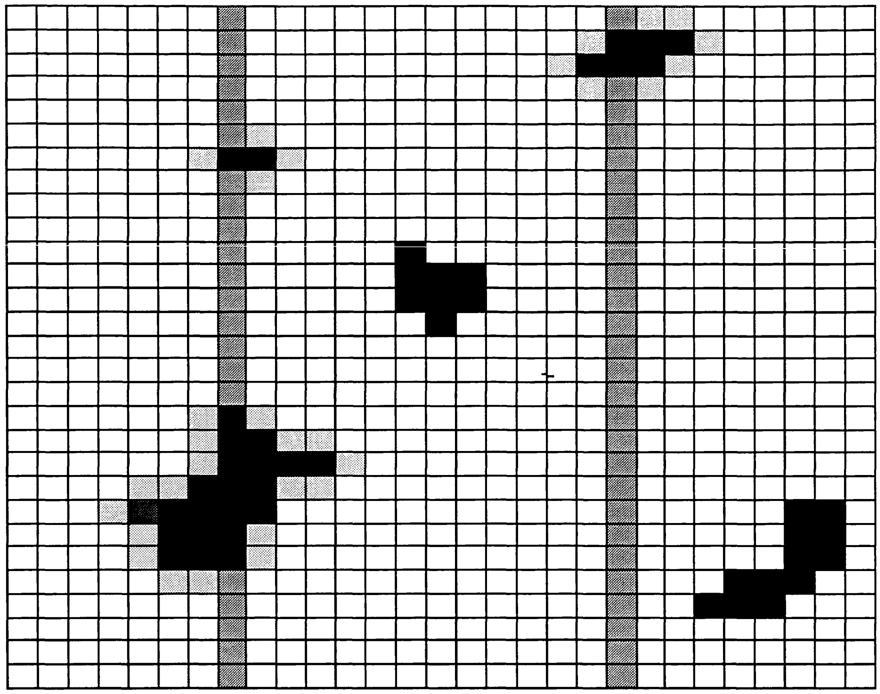
Figure 1. Example of strip adaptive cluster sampling for a patchy population. Black squares represent fish densities above the critical value. The initial sample consists of two transects, which detect three patches. Units adjacent to high densities are added adaptively until densities drop below the critical value.

Figure 2. Simulated target stocks for sampling comparison: a) greyscale map of "true" density grid; b) transect totals for each stock (sum of densities from each column); c) variance of estimation (MSE) for total stock size for different sampling designs.

Figure 3. Distribution of Relative Errors $(\hat{T} - T)/T$ over 5000 sampling replicates. Results are shown for the four standardized test stocks. Traditional is stratified random sampling and ACS is ACS sampling from a stratified random design. Note the high skewness of the traditional estimator

Figure 4. Data from field trial of ACS on Lake Erie. Each circle represents a 5-minute integrated sampling unit; shaded units meet ACS criterion of y_{ij}>5000 fish/hectare. The long center transect is the initial transect.

Figure 5. Comparison of the frequency of "Good" estimates (within 10% of the true total) and "Poor" estimates (more than 50% off of the true total) from traditional and adaptive cluster sampling designs.



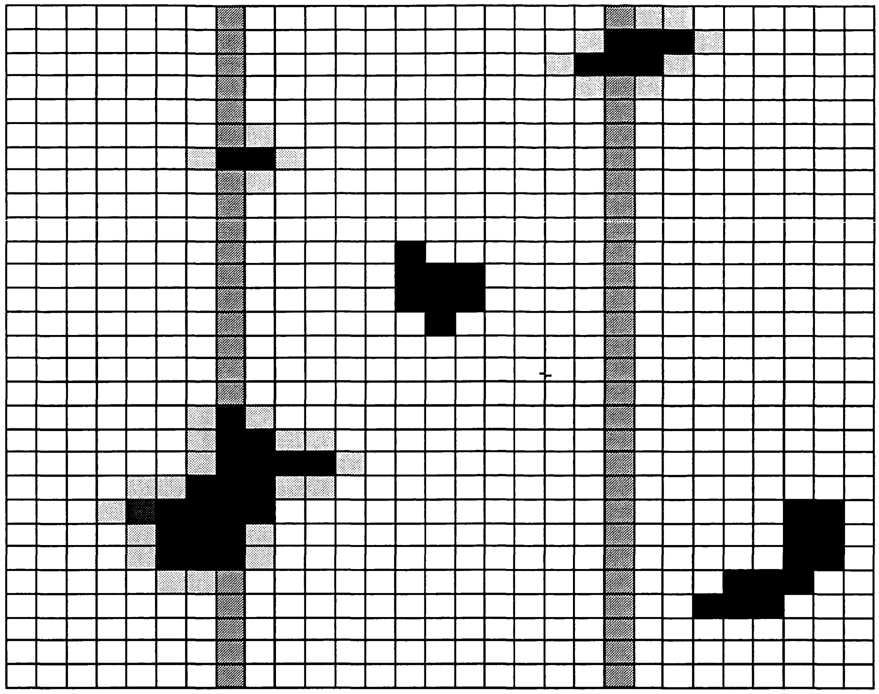


Figure 1. Example of strip adaptive cluster sampling for a patchy population. Black squares represent fish densities above the critical value. The initial sample consists of two transects, which detect three patches. Units adjacent to high densities are added adaptively until densities drop below the critical value.

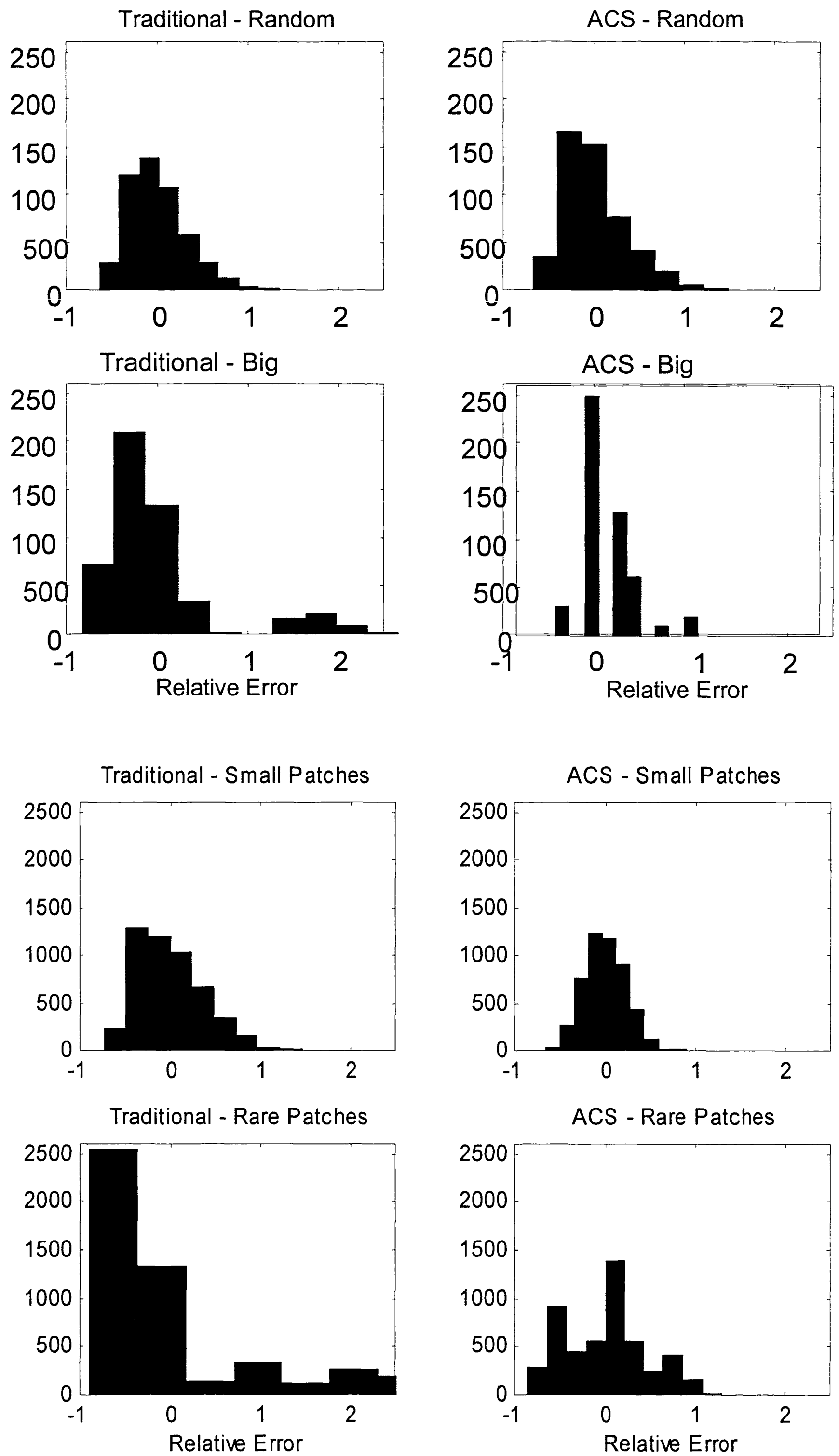


Figure 3. Distribution of Relative Errors $(\hat{T} - T)/T$ over 5000 sampling replicates. Results are shown for the four standardized test stocks. Traditional is stratified random sampling and ACS is ACS sampling from a stratified random design. Note the high skewness of the traditional estimator.