

Review of Statistical Aspects of Survey Sampling for Marine Fisheries

DANIEL K. KIMURA AND DAVID A. SOMERTON

National Oceanic and Atmospheric Administration, National Marine Fisheries
Service, Alaska Fisheries Science Center, Seattle, Washington, USA

Fishery surveys are an essential ingredient of modern fisheries stock assessment. To understand this, one must understand that the survey time series are the essential anchor that makes modern fishery stock assessment modeling possible. Without fishery surveys, fisheries stock assessment scientists would have great difficulty modeling absolute abundance, and therefore agencies would have difficulty setting levels of allowable catch. In this review, we explore the basic assumptions that need to be fulfilled for valid surveys to be accomplished. Although these assumptions are generally well understood, the complexity of survey sampling gear and the complexity of target animal behavior makes them difficult to fulfill in practice and can easily lead to problems when interpreting survey results from even the most carefully designed survey. In reviewing the literature surrounding fishery surveys, their sampling design, modeling, and methods of analysis, it became clear that fisheries scientists have long been preoccupied with coping with the intense variability found in fish catches. This variability is found within fishing hauls, between hauls, between area, time and depth strata. Coping with this variability, which is due to animal behavior, habitat variation, and the nature of fishing gear, will be a constant theme that ties together our review.

Keywords trawl surveys, hydroacoustic surveys, abundance indices, sampling design, data analysis, spatial sampling

Introduction

Marine fish and invertebrate populations are surveyed by fishery biologists using a variety of sampling techniques including capture with active fishing gears such as trawls or dredges, capture with passive gears such as baited hooks and traps, capture of pelagic eggs and larvae with plankton nets, or indirect sampling with acoustics or video photography (Gunderson, 1993). Although these techniques employ a diverse collection of technologies, they all are intended to produce a similar type of data; that is, indices of abundance that are typically used in conjunction with a mathematical model of the population to manage a fishery.

Indices of abundance are usually assumed to vary in proportion to the actual abundance of the population, where the proportionality constant is referred to as the catchability coefficient (q). For some sampling techniques, q is considered as the product of two quantities: 1) the availability (A_v) or proportion of the population occurring within the survey area,

Address correspondence to Daniel K. Kimura, National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way N. E., Seattle, WA 98115. E-mail: Dan.Kimura@noaa.gov

and 2) the efficiency of fishing gear (Eff), or proportion of the animals sampled within the water column (i.e., surface to bottom) bounded by the swept area. Availability is related to the population dynamics of the target species being surveyed, whose abundance can vary in response to seasonal and diurnal migrations, so that time, area, and depth of sampling are especially important. Efficiency can vary in response to factors such as environmental effects on the performance of the sampling gear and behavior of the target species, changes in the sampling vessel or its operation, or changes in the sampling gear or methodology. In cases where availability and efficiency, and therefore catchability, are experimentally measured, indices of abundance can be converted to estimates of biomass or numbers of individuals in the population (Somerton et al., 1999).

Catchability, however, is rarely experimentally estimated and indices of abundance are typically used as measures of relative abundance from one time period to the next. To be truly informative about trends in population size, such use requires that catchability is a constant or, at least, stationary (i.e., varying without trend) with time. Stationarity of catchability requires, in turn, that fishery biologists ensure that all aspects of survey methodology remain constant with time and, when changes in methodology are unavoidable (e.g., survey vessel replacement), that the new sampling methodology is sufficiently calibrated to the old. Provided that catchability is stationary, relative abundance estimates can be used to calibrate, or tune, age-structured models of fish and invertebrates subjected to a fishery.

Our review naturally divides itself into survey designs and model-based methods of analysis. Following Brewer et al. (1988), Smith (1990) recognized the dichotomy of sampling assumptions between design-based theory and model-based theory. The example used by Smith (1990) is a stratified random sampling design (design-based theory) and the delta-lognormal model that might be used to analyze these data (model-based theory). While the sample mean is an unbiased and consistent estimate of the finite population mean, the delta-lognormal model does not in general provide a consistent statistical estimate (using a criterion called asymptotic design consistency) based on a given stratified statistical design. Therefore, Smith (1990) exposes the possibility that sampling designs may not always fulfill the assumptions of model-based theory.

Model-based theory can lead to abstract statistical analyses which can be difficult to interpret. An example relevant to fishery surveys is the example of analyzing spatially correlated fisheries data. Here two model-based analyses, geostatistical kriging and random field linear models, are available to analyze similar data. Results from these analyses may seem to conflict when not viewed from the proper perspective. As will be shown, however, these methods are addressing quite different statistical questions, and the different results should be expected.

Relation between Fishery Stock Assessment and Fishery Surveys

Currently, the preferred method of stock assessment is through age-structured stock assessment models (see Quinn and Deriso, 1999). The defining characteristic of this type of model is that the population is modeled as numbers at age and that biomass estimates are calculated by converting these numbers to biomass. The model is then fit by comparing the trend in modeled biomass with the trend in survey abundance.

Let $t = 1, \dots, T$ be the year index, and $a = 1, \dots, A$ be the age index for some species which has A nonzero ages of catch. Let

C_{ta} = the observed catch in numbers at age,
 s_a = the selectivity for age a fish,

F_t = the instantaneous fishing mortality rate for fully available ages (i.e., ages for which $s_a = 1$).

The assumption of “separability” means that

$F_{ta} = F_t s_a$ - the instantaneous fishing mortality rate on age a fish during year t ,

N_{ta} = the total population size at age a in year t , and

U_{ta} = the exploitation rate on age a fish in year t ,

= $F_{ta}(1 - \exp(-M - F_{ta})) / (M + F_{ta})$ assuming an instantaneous natural mortality rate M .

It follows that

$$N_{t+1, a+1} = N_{t, a} \exp(-M - F_{ta}),$$

and predicted catch would be

$$\hat{C}_{ta} = U_{ta} N_{ta}.$$

This model contains $p = 2T + 2A - 1$ parameters:

$$b = \{F_1, \dots, F_T, s_1, \dots, s_A, N_{11}, \dots, N_{1A}, N_{21}, \dots, N_{T1}\}$$

with $m = TA$ observations (i.e., catch in numbers at age observations). Parameters can be estimated by minimizing the sum of squares (SS):

$$SS = \sum_{t, a} (\log(\hat{C}_{ta}) - \log(C_{ta}))^2.$$

This model was described rather completely by Doubleday (1976) and is the foundation of modern age-structured models.

Early use of the Doubleday model made it clear that unless residual error in this model was extremely small, the parameters in this model could not be estimated from catch data alone. Fournier and Archibald (1982), Deriso et al. (1985), and Methot (1986) showed how auxiliary information could be used in this model to improve model performance. *The most useful auxiliary information turned out to be a time series of fishable biomass* (i.e., biomass available and vulnerable to the fishing gear) estimates obtained from surveys $\{I_t^f\}$.

In the model, total and fishable biomass are estimated by

$$\hat{B}_t^{tot} = \sum_a \hat{N}_{ta} \bar{w}_a$$

and

$$\hat{B}_t^f = \sum_a \hat{N}_{ta} \hat{s}_a \bar{w}_a,$$

respectively, where \bar{w}_a is the average weight of a year-olds. Following Deriso et al. (1985), the augmented sum of squares (SST) to be minimized is:

$$SST = \sum_{t, a} (\log(\hat{C}_{ta}) - \log(C_{ta}))^2 + \sum_t \lambda (\log(\hat{q} \hat{B}_t^f) - \log(I_t^f))^2,$$

where λ is the relative weighting between the sum of squares.

If the catchability coefficient q is fixed to one, or determined experimentally, the surveyed biomasses, $\{I_t^f\}$, are said to estimate “absolute” abundance. This strong assumption significantly reduces the variance in model parameter estimates, but at the cost of potential bias. When q is estimated by minimizing SST , the biomass surveys are said to measure relative abundance. Parameters fit using this assumption will have larger variances, but results are less likely to be biased. Therefore, it can be helpful if survey biomass estimates can be accepted to be absolute measures of abundance. Somerton et al. (1999) consider how uncertainty in q should influence the stock assessment modeling.

It should also be noted that in modern stock assessment, modeling several indices of abundance can be fit simultaneously. Therefore, when fishing gear or survey vessels are changed, after the new gear has its own time series, the old and new time series may be entered as separate indices of abundance.

Bottom Trawl and Hydroacoustic Surveys

Although indices of abundance from essentially any type of assessment survey could be used to tune an age-structured model, we will focus on indices produced by bottom trawl and hydroacoustic surveys because of their widespread use and the extensive literature concerning analysis of the type of data they produce. Bottom trawl surveys (see Gunderson (1993) for a detailed description) sample fish and invertebrates by towing a large, conical net across the bottom for a standard unit of time or distance (referred to as a haul or tow). The resulting abundance indices are then expressed as either catch in numbers or weight per unit of time, distance, or area swept by the trawl. Trawling is typically conducted systematically over a grid of sampling sites or at randomly selected locations within the survey area. Each trawling event produces a single point estimate of relative abundance. By comparison, hydroacoustic surveys do not directly sample fish, but instead count them or estimate their aggregate biomass based on the strength of the acoustic energy reflected from a known volume of water that is insonified with an echosounder. Such echo integration data are collected continuously while the survey vessel completes a pattern of transects, but is usually discretized by averaging over some unit of distance (e.g., one nautical mile [nmi]). Both types of survey can therefore be considered as producing geo-referenced estimates of abundance that are proportional to local density.

Survey Designs

Most of the literature on fishery surveys concern trawls. However, much of what follows can be applied to the catch per effort of any standardized piece of fishing gear. Fishing effort for other gears might quantify measures such as number of hooks, fishing pots, and length of time fished in hours or days. For example, effort might equal 240 pot-hours. Often, catch data is standardized to catch per unit of effort (CPUE) so that observations having different levels of effort are comparable. However, greater statistical efficiency is obtained in surveys where each gear set represents a standard measure of fishing effort. In this review, *efficiency* generally refers to the variance of the estimated indices of abundance. This should not be confused with gear efficiency mentioned in the introduction.

Although CPUE sometimes seems to be the only possible index of abundance other possibilities exist. For a hook-and-line sport fishery for yellowtail snapper, *Ocyurus chrysurus*, Bannerot and Austin (1983) show that the square root of the frequency of zero CPUE was the best index of abundance.

The survey design considerations are sometimes general and common to many types of field experiments (e.g., stratified random sampling). At other times, they are somewhat unique to fisheries (e.g., optimum length of tow or fixing stations in different years).

Stratified Random and Systematic Trawl Survey Designs

For fishery surveys, trawl survey design is essentially the science of placing sampling points onto a two-dimensional layout. Fishery surveys take place in three dimensions, but location in two dimensions is generally a proxy for bottom depth at a particular location. Choosing stations (location of sampling) is of great practical importance. Given a broad space to be sampled, Cochran (1977) discusses three possible ways stations could be placed: totally randomly (ran), randomly within design strata (str), and systematically throughout the area (sys), each becoming progressively more restrictive on the placement of sample points.

In general, random sampling is a poor way to sample aggregated populations because samples may themselves be aggregated and the population aggregations over- or under-represented (Lai and Kimura, 2002; Lenarz and Adams, 1980) in smaller numbers of samples, even though averages from large numbers of samples would be unbiased. Furthermore, since the total population biomass is typically an estimate weighted by habitat area, it is strongly desired that different area types be sampled to some extent. Another way of characterizing spatial sampling is that ran, str, and sys sampling are likely to provide progressively spatially more uniform sampling. Str sampling may be best when we know *in detail* how abundance and variance vary among strata. This strategy can assure that a specified survey accuracy is achieved when the accuracy of only one target species is important. Sys sampling may be most appropriate when we are willing to admit spatial variability is likely to occur at the scale of the survey, but we have poor information concerning the location and variability of abundance, or interannual location varies greatly. In these situations the survey is probably best served by a grid of observations since detailed information regarding strata will be totally lacking. Lenarz and Adams (1980) found that at identical sampling effort for rockfish surveys, that $\text{var}(\bar{y}_{\text{sys}}) \leq \text{var}(\bar{y}_{\text{ran}})$ with $\text{var}(\bar{y}_{\text{str}})$ generally falling in between. However, the differences in precision there were generally judged to be slight.

Variance Estimates Under Systematic Sampling

Cochran (1977) and Thompson (2002) note the similarity between cluster sampling and systematic sampling. Systematic sampling can be viewed as the selection of a single cluster whose secondary samples are spread systematically through the sampling area. For this single sample (i.e., cluster) sampling is nonrandom, so strictly speaking, the random sampling formulas for sample variance do not apply. Thompson (2002) states: "From a sample size of one it is possible to obtain an unbiased estimator of the population mean, but it is not possible to obtain an unbiased estimator of its variance."

If there is more than one systematic sample prescribed in a broader sampling design, then the systematic samples (i.e., clusters) can be thought of as chosen at random and the variance between their means used to estimate a sampling variance (see Lenarz and Adams, 1980; Thompson, 2002). However, when only one systematic sample is available for analysis, it is usual to apply simple random sampling formulas to the observations. Thompson (2002) suggests that "With many natural populations, in which nearby units

tend to be similar to each other, this procedure tends to overestimate the variance of the estimator of the population mean or total.” This statement seems questionable since in this instance we would in fact be ignoring spatial correlation which would cause the variance to be underestimated.

Finite Population Corrections for Stratified Survey Designs

When the population of possible sampling units is finite, then variance of mean estimates use a finite population correction. For a finite population of size N , sampled with a random sample of size n , the variance of

$$\bar{y} = \sum_{i=1}^n y_i/n \text{ is } \text{Var}(\bar{y}) = (1 - n/N)\sigma^2/n \text{ rather than } \sigma^2/n.$$

So the question is “Should the collection of possible tows be considered a finite set?” One way this can be viewed is by thinking of the possible hauls as tiling the survey area. From this perspective, barring overlap of trawl hauls, the number of possible hauls is finite and the finite population correction can be applied. For large-scale surveys, such corrections are likely to be negligible.

Allocating Sampling Effort for a Stratified Sampling Survey Design for a Single Species

The most common fisheries survey design may be a design stratified by geographic area. Let

$h = 1, \dots, L$ refer to area (A_h) strata

$i = 1, \dots, n_h$ refer to the hauls in stratum h

y_{hi} = catch per swept area of the i th haul in the h th stratum. Catches and density differ by only a constant factor when all hauls cover the same area.

$$\begin{aligned} E(y_{hi}) &= \mu_h \\ \text{Var}(y_{hi}) &= \sigma_h^2 \end{aligned}$$

Then the estimate of total biomass (B)

$$\begin{aligned} \hat{B} &= \sum_{h=1}^L A_h \bar{Y}_h, \text{ where} \\ \bar{Y}_h &= \sum_{i=1}^{n_h} y_{hi}/n_h. \end{aligned}$$

The expectation and variance of \hat{B} are:

$$\begin{aligned} E(\hat{B}) &= \sum_{h=1}^L A_h \mu_h, \text{ and} \\ \text{Var}(\hat{B}) &= \sum_{h=1}^L A_h^2 \sigma_h^2 / n_h. \end{aligned}$$

We can allocate the $n = \sum n_h$ observations among the L strata by minimizing $\text{Var}(\hat{B})$. Minimizing $\text{Var}(\hat{B})$ by Lagrange multipliers:

$$\begin{aligned} u &= \sum A_h^2 \sigma_h^2 / n_h + \lambda \sum n_h, \\ \partial u / \partial n_h &= -A_h^2 \sigma_h^2 / n_h^2 + \lambda = 0, \end{aligned}$$

with $\sum n_h = n$. Solving these last two equations gives the familiar Neyman allocation formula: $n_h/n = A_h \sigma_h / \sum A_h \sigma_h$ (Cochran, 1977).

Since the A_h are geographic area values, they are relatively easy to obtain. The σ_h^2 must be obtained from previous surveys, pilot surveys, commercial fishery data, or through modeling. Cochran (1977) noted that having too many strata can also prove counter-productive. His simulations indicated that for most problems, no more than six strata should be adequate.

Smith and Gavaris (1993) optimized an Atlantic cod (*Gadus morhua*) survey using average abundance measured from previous surveys. They noted that allocation of stations based on stratum abundance is equivalent to Neyman allocation if the stratum standard deviation is proportional to the average stratum mean density. They also noted that greatest gains in survey precision were from improved allocation of tows to strata, rather than improved strata definitions.

Allocating Sampling Effort for a Stratified Survey Design for Multiple Species

For surveys of multiple species, let the strata and notation be the same as for the single species analysis given above. However, assume that biomass estimates are being made for each of s different species. Recall that $CV(\hat{B}) = \sqrt{\text{Var}(\hat{B})}/\hat{B}$. So for a given CV to be attained, $\text{Var}(\hat{B}) \leq CV^2 \hat{B}^2$.

Suppose we have $s = 1, \dots, S$ species, surveyed in $h = 1, \dots, L$ strata.

$$\begin{aligned} \hat{B}_s &= \sum_{h=1}^L A_h \bar{Y}_{hs} \\ \text{Var}(\hat{B}_s) &= \sum_{h=1}^L A_h^2 \sigma_{hs}^2 / n_h \end{aligned}$$

where,

$$\bar{Y}_{hs} = \sum_{i=1}^{n_h} y_{hsi} / n_h, \text{ and } \sigma_{hs}^2 = \text{Var}(y_{hsi}).$$

Here, $i = 1, \dots, n_h$ refer to the hauls made in stratum h .

For the desired CV_s to be attained for each of $s = 1, \dots, S$ species, the following S inequalities must hold for each species:

$$\begin{aligned} \left[\sum_{h=1}^L A_h^2 \sigma_{h1}^2 / n_h \right] &\leq CV_1^2 \hat{B}_1^2 \\ &\cdot \quad \cdot \quad \cdot \\ \left[\sum_{h=1}^L A_h^2 \sigma_{hS}^2 / n_h \right] &\leq CV_S^2 \hat{B}_S^2. \end{aligned}$$

The minimum sample size and allocation that solves this problem can be obtained from the simplex algorithm where we define $x_1 = 1/n_1, \dots, x_L = 1/n_L$, all $x_h > 0$ and the goal is to maximize $\sum x_h$ subject to the S constraints given above.

Because the effect of q cancels in the CV, these results hold even if q varies by species, provided that the q for each species is constant for all area strata.

Fixing Stations on an Interannual Basis

The question we address here is whether stations should be fixed in location on an interannual basis. Since the goal of fishery surveys are to provide interannual trends in abundance, an argument can be made that interannual trends in abundance can best be estimated if the same stations are chosen every year. Warren (1994) notes that this can lead to biased estimates of both annual abundance and interannual trends. He argues that by randomizing a portion of the surveys each year, a balance can be made between estimating trends in abundance and reducing bias. The tool he uses to strike this balance is an “index of persistence,” a measure of how stable spatial distribution is between years. If a stock has high persistence arguably more stations should be fixed, while stocks with low persistence should be randomized each year. The middle ground would suggest “partial replacement.”

Adaptive Sampling Designs

We view adaptive sampling designs as those which allow the design to change in accordance with what is found as the sampling progresses. These are new methods in the area of fishery surveys, and much work needs to be carried out to determine if these methods can be truly effective.

Adaptive Cluster Sampling

For fishery surveys, adaptive cluster sampling (ACS) can be thought of as a sampling design, where stations are initially placed randomly. Those samples that find targets of interest with sufficient abundance (i.e., exceed the criterion value) then become the center points of additional search stations. Those additional stations can, if they exceed the criterion value, become the center points of additional search stations, and so forth, until the sampling algorithm is exhausted.

When animals are rare and clustered, adaptive sampling seems an attractive sampling strategy. Once these relatively rare animals are found, we are able to capture more of them. However, for groundfish surveys adaptive sampling seems problematic for making total biomass estimates for several reasons:

1. Vessel movement is extremely time-consuming and costly. If the initial random sample of stations is sampled first so that (a) we know what initial sampling stations qualify for further sampling, and (b) we don't use all of our survey effort on only a small portion of the area which needs to be surveyed, then the vessel must then go back to sample the adaptive sampling stations.
2. The adaptive sampling scheme must proceed to stations around the “edge” of concentrations which are costly to cover, but do not directly contribute to the mean estimate.
3. Is the spatial correlation of abundance something that can be detected over the high variability that exists in fishing observations? That is, we may see an occasional large catch, but how informative is this about what catches will be in neighboring sampling stations? If catches are essentially independent negative binomial random variables in these neighboring stations, spatial aggregations may not be detectable on the larger

scale. Under these circumstances, it would seem that no efficiency gain is possible with adaptive sampling.

4. How do we choose a criterion value for additional adaptive samples (Hanselman et al., 2003), and how does the criterion value affect the final survey estimates especially in light of point (3) above?

Lo et al. (1997) appears to be one of the first studies to use ACS to measure abundance from a marine fisheries survey. They measured the abundance of Pacific hake (*Merluccius productus*) larvae, using towed fishing gear (i.e., an ichthyoplankton survey). Lo et al. (1997) felt that ACS improved survey efficiency.

Hanselman et al. (2003) applied ACS to trawl surveys for Alaska rockfish. These surveys estimated the abundance of Pacific ocean perch (*Sebastes alutus*) and shortraker/rougheye (*S. borealis/S. aleutianus*). Hanselman et al. (2003) found there were gains using ACS for Pacific ocean perch (i.e., one-third less effort for the same standard error when tows were added using ACS rather than randomly), but not for shortraker/rougheye.

Adaptive cluster sampling presents some interesting possibilities for marine fishery surveys. However, the method seems to present as many questions as it does answers. Hanselman (2003) states, and we agree, the method needs further research.

ACS Estimators of Abundance. Because adaptive sampling adds stations in addition to those that can be considered a random sample, the simple mean of all observed catches (i.e., including ACS stations) is not an appropriate estimator of the population mean. Following Thompson (2002), consider a layout of N possible survey sampling units (i.e., stations) described as simple squares (Figure 1). Suppose a random sample of n stations is selected without replacement. When the catches of these stations exceed a criterion value, sampling at stations on the four sides of the square is then triggered. This is repeated until no other stations exceed the criterion, or until an artificial ceiling is reached (say 3). After all adaptive sampling (i.e., additional stations) has taken place, we will notice a grouping of stations into what Thompson calls *networks* Ψ_k . Thompson describes networks as containing all units that are above the criterion, but utilizing the edge units adds another layer of complexity that Hanselman et al. (2003) suggest will not materially change the estimates. In theory, under adaptive sampling, selection of any unit in the initial random sample will imply the network has been selected. However, it seems a large component of variance associated with sampling variability of observations in each unit makes this theory problematic when applied to fishery surveys.

Hansen-Hurwitz estimator. Let $w_i = \sum_{j \in \Psi_i} y_j / m_i$ be the sample mean of the network intersected by the i th initial sampling unit where the number of sampling units in this network is m_i . Then the Hansen-Hurwitz estimate (Thompson, 2002) of the population mean is $\mu_{HH} = \sum_{i=1}^n w_i / n$. Here the mean of the i th sampling unit has been replaced by the network mean.

Horvitz-Thompson estimator. Suppose network k has x_k units, then the probability of this network being included in the sample is

$$\alpha_k = 1 - \frac{\binom{N - x_k}{n}}{\binom{N}{n}}.$$

The estimated population mean based on the Horvitz-Thompson estimator (Thompson, 2002) is

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^K \frac{y_k^* z_k}{\alpha_k} \quad \text{where } y_k^* = \sum_{i \in \Psi_k} y_i$$

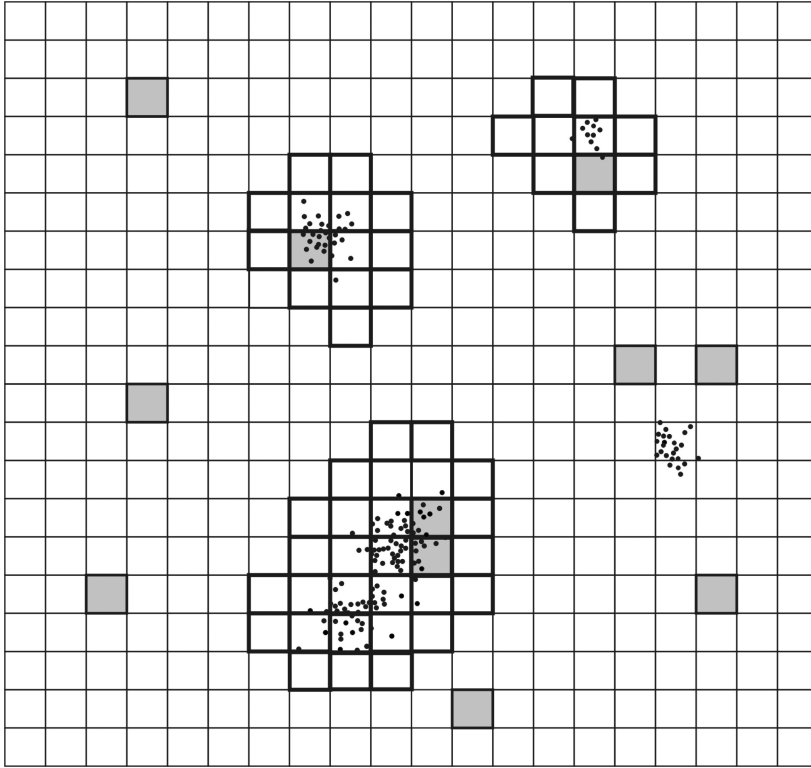


Figure 1. An illustration of adaptive cluster sampling using squares as the sampling units to sample target organisms (dots). Shaded squares represent squares selected in the initial random sample. Additional bolded squares represent squares that were selected as additional adaptive samples. Note that the process continues until all edge squares were found to be empty or below criterion (following Thompson, 2002).

is the sum of the y_i for the x_k units in the k th network. Also, z_k will be zero for networks not selected in the sample, single unit networks, and edge units not selected initially.

Francis' Two-Stage Sampling Method

Francis (1984) proposed an “adaptive strategy” for stratified random trawl surveys. Despite the similar name, the Francis method was essentially two-stage sampling for the stratified random trawl design, where the initial sample provided information that could be used in the second stage, providing additional tows with an optimal allocation of these tows.

Trawl and Acoustic Presence/Absence Survey

Everson et al. (1996) proposed trawl and acoustic presence/absence survey (TAPAS) surveys when “patches are small, mobile, ephemeral or unpredictable in location.” This ingenious proposal combines hydroacoustic and bottom trawl sampling to provide a biomass estimate based on fish patches (i.e., school structure). These authors recognized that transect sampling using hydroacoustics could provide information on patches: their size and number. We start with the usual trawl survey design by placing tows either randomly or systematically in a survey area A . We then connect these trawl stations as efficiently as possible with hydroacoustic transects, resulting in a total transect length L . The original trawl stations are

intended to estimate background density outside the patch areas. Therefore, if an original trawl station happens to fall in a patch it should be ignored. The vessel should proceed through the patch until the length of intercept of the transect with the patch is determined, say l_i . The vessel then goes back and places a single tow at random along the patch transect resulting in a density estimate d_i .

If the average density from the background tows is D_b , and there are p patches of total length $l = \sum_{i=1}^p l_i$, then the background biomass can be estimated as $B_b = AD_b(L - l)/L$. For the patches, the biomass can be estimated to be $B_p = A \frac{\sum d_i}{p} \times \frac{l}{L}$. Total biomass for area A can then be estimated as $B_t = B_b + B_p$. The attractive feature of TAPAS is that it seems like it can be a practical, time-efficient method. See Everson et al. (1996) for suggested methods for calculating variances for this method.

Statistical Distribution of Fish Abundance (Small-Scale Spatial Aggregation)

Generally fish can be thought of as being aggregated due to schooling behavior. Although catch data are rarely normally distributed, the availability of statistical methodology that assumes normality makes this assumption attractive, but often not wise. Fisheries catch and survey data often violate the normality assumption in three generally recognized ways: abundance of zero values, positive skewness, and occasional extremely large outliers. It is not unusual to have all three occurring in a single survey or dataset. As might be imagined, it is not easy to find methods of analysis that address these issues, yet allow an analysis that accomplishes a stated goal.

On the small scale, observations from a random nonaggregated process are usually thought of as being distributed as Poisson (P), where the mean and variance are equal to a single parameter (say μ). However, aggregated distributions in fisheries can often be approximated by the negative binomial (NB), where the mean (μ) and variance are related by $\text{var} = \mu + \mu^2/k$. The parameter k describes the degree of aggregation in the NB, small values imply extreme aggregation with large variance, while the Poisson variance is the limit as $k \rightarrow \infty$. Using the NB, observations can still be thought of as being independent, and identically distributed (*iid*).

The Negative Binomial Distribution Says that Shorter Tows are Better

Taylor (1953) recognized that for nonaggregated populations that are distributed as Poisson, the size of sampling unit should be irrelevant from a statistical point-of-view, and that for populations that are distributed NB, the shorter the tow, the more efficient the sampling design.

Suppose a survey consists of a single haul which covers an area A , and that the catch in numbers from this haul is distributed as $P(\mu)$. Then an observation from this tow, say X , will have $E(X) = \mu$ and $V(X) = \mu$. Here, the observations are in numbers of animals rather than biomass. Now divide A into r smaller hauls each covering an area of size a so that the total effort $E = ar = A$ is unchanged, but now we have X_1, \dots, X_r *iid* random variables with $X_i \sim P(\mu/r)$. As before $X = \sum_{i=1}^r X_i$ will have $E(X) = r \times \mu/r = \mu$, and $V(X) = r \times \mu/r = \mu$. It is evident that for the Poisson distribution the length of tow in the experimental design is irrelevant.

Consider the same question with the Poisson distribution replaced by the negative binomial $NB(\mu, k)$. Now for $E = A$, $E(X) = \mu$ with $V(X) = \mu + \mu^2/k$. Now as before, if we divide E into r tows of size a , then we have r *iid* random variables with $E(X_i) = \mu/r$ with $V(X_i) = \mu/r + \mu^2/(r^2k)$. Then for $X = \sum_{i=1}^r X_i$, $E(X) = r \times \mu/r = \mu$ and $V(X) = \mu + \mu^2/(rk)$. The variance of X is reduced substantially by dividing a single tow into r smaller independent tows. Note that as r becomes larger, the distribution variance of

the individual pieces tends towards the Poisson. Pennington and Grosslein (1978) noticed this and called this “locally random” behavior.

The analysis here in Sections 3.1, 3.2 and 3.3 assumes that k remains the same when the length of tow and sampling mean changes. Pennington and Grosslein (1978) provide instances where this is the case, but it is difficult to know how well this assumption holds in general.

Lenarz and Adams Index of Precision

Lenarz and Adams (1980) developed an index of precision (i.e., an inverse of CV) for estimated catch density \bar{d} which is helpful in evaluating the trade-off between the number and length of tows.

Let X_i be r iid $NB(\mu, k)$, where each haul is of size a . Let d be the mean catch per unit area, so that $\mu = da$. For the sampled values $\bar{d} = \bar{X}/a$, so that $V(\bar{X}) = (\mu + \mu^2/k)/r$ and $V(\bar{d}) = (\mu + \mu^2/k)/(ra^2)$. Substituting $\mu = da$ in this last expression gives

$$V(\bar{d}) = \frac{da}{ra^2} + \frac{d^2a^2}{kra^2} = \frac{d}{ra} + \frac{d^2}{kr},$$

so that

$$CV(\bar{d}) = \sqrt{\frac{1}{rad} + \frac{1}{rk}}.$$

This is the inverse of Lenarz and Adams (1980) precision index. Lenarz and Adams (1980) found most commercially important *Sebastes* sp. to be highly aggregated which implied shorter hauls should be favored. However, they noted that even for aggregated species, reducing sample element size can improve precision relatively little when a species has very low densities. This is due to fixed costs when increasing the number of tows. Note this index behaves as the Poisson if $k \gg \bar{d}$.

Optimum Size and Number of Sampling Units

Since the above argument that tows should be shorter can be repeated forever, the question arises as to just how short a haul should be made. This decision will largely be based on logistics because setting and hauling gear take considerable time, and during this time the fishing gear will behave in uncharacteristic ways. If fishing time is too short, then the catch will be too small and uncharacteristic of normal fishing operations. Pennington and Vølstad (1991) were able to solve this problem by minimizing the inverse of the Lenarz and Adams precision estimate subject to a cost constraint.

For a random survey, the total cost of a survey can be thought of as total time and can be expressed as $C = (c_1 + a)r + c_2\sqrt{r}$ where c_1 is the fixed cost of setting and hauling the gear and now a is the length of tow. The second term $c_2\sqrt{r}$ is the cost of moving between stations which is an approximation whether the stations are random or gridded. Then minimizing $CV(\bar{d})$ subject to the cost parameters gives:

$$\hat{r} = \left\{ \left[(c_2^2 + 4(c_1 + a)C)^{1/2} - c_2 \right] / [2(c_1 + a)] \right\}^2, \text{ and}$$

$$\hat{a} = \{ [c_1 + c_2/(2\sqrt{\hat{r}})] / (\bar{X}/k) \}^{1/2}.$$

An analysis of ocean pout (*Macrozoarces americanus*) (Pennington and Vølstad, 1991) supported the notion that tows should be of an optimum value, of rather short

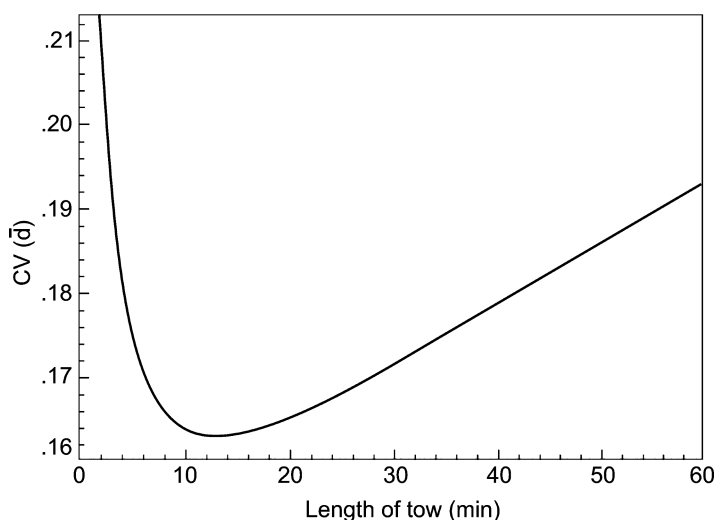


Figure 2. Relative precision of the density estimate $CV(\bar{d})$ as a function of trawl tow length in minutes. Results are for ocean pout assuming fixed costs for setting and hauling gear. Note that precision was optimized at an intermediate tow length (adapted from Pennington and Vølstad (1991).

duration, even taking into account the fixed cost for setting and hauling the gear (Figure 2).

The Log Transform is Useful for Taming Variability When Catches are Negative Binomial

The variation in magnitude of catches caused by the aggregation of fishes is the quintessential problem for fishery data analysts. This is often expressed as an excess of zero catches and/or a few very large catches that can dominate the total biomass estimate, skew the distribution to the right, and cause a great deal of heteroscedasticity. The usual normal theory tools are not applicable to such data.

The effectiveness of the log transformation for taming large catches, skewness, and heteroscedasticity makes it an indispensable tool for analysts of catch data whether this data is from scientific surveys or fisheries landings.

Snedecor and Cochran (1967) note that if the “standard deviation in the original scale varies directly as the mean” (i.e., when the CV is constant), then the logarithmic transformation will stabilize the variance. Snedecor and Cochran (1967) also goes on to note that it [lognormality] “is likely to be found when the effects are proportional rather than additive.” The proportionality argument is apparently from a derivation of the lognormal distribution using the accumulation of random proportional effects (Aitchison and Brown, 1957).

If aggregation is due to the negative binomial distribution $X \sim NB(\mu, k)$, then

$$\sigma = \left(\frac{1}{\mu} + \frac{1}{k} \right)^{1/2} \mu,$$

and the

$$CV(X) = \left(\frac{1}{\mu} + \frac{1}{k} \right)^{1/2}.$$

Therefore, if $\mu \gg k$, the standard deviation is directly proportional to the mean and

$$SD(\ln(X)) \approx \left(\frac{1}{\mu} + \frac{1}{k} \right)^{1/2}$$

by the delta method (see e.g., Seber, 1982, for the delta method). This is true for highly aggregated species. Although crude estimates of k can be estimated from the above expressions, iterative maximum likelihood estimates (Anscombe, 1950) are better.

Bissell (1972) extended the negative binomial model to allow for varying element size. This is useful when analyzing commercial trawl data where the lengths of tow may vary (Lenarz and Adams, 1980).

Bias When Applying the Logarithmic Transformation

Because of its importance we shall discuss in detail the use of the logarithmic transformation for analyzing fishery catch per unit effort data. The most natural introduction to this problem is to go through the basic problem of estimating the mean CPUE. This will be sufficient for providing a general basis for applying the log transformation. If survey data is lognormal, the log transformation is the natural method of analyzing data. However, estimating the mean abundance on the original scale requires attention because the simple antilog of the mean on the logarithmic scale is a biased estimate of the mean on the original scale.

Let the fishery catch per unit effort $Y_i = C_i/E_i$ be such that $X_i = \log(Y_i) \sim N(\xi, \sigma^2)$. It turns out the mean on the transformed scale, ξ , is the median, $\exp(\xi)$, on the original scale (Aitchison and Brown, 1957). Since the distribution on the original scale is typically skewed to the right, the median might be the preferred measure of central tendency for some problems. The mean on the original scale is $E(Y) = \mu = \exp(\xi + \sigma^2/2)$ (Finney, 1941).

Finney (1941) goes on to show that the mean and variance of Y_i , say (μ, v) can be estimated using the Bessel function:

$$g_n(t) = 1 + \frac{n-1}{n}t + \frac{(n-1)^3}{n^2 2!} \frac{t^2}{n+1} + \frac{(n-1)^5}{n^3 3!} \frac{t^3}{(n+1)(n+3)} + \dots$$

where

$$\hat{\mu} = \exp(\bar{X})g_n(s^2/2)$$

and

$$\hat{v} = \exp(2\bar{X}) \left(g_n(2s^2) - g_n\left(\frac{n-2}{n-1}s^2\right) \right)$$

are efficient estimators. In these equations \bar{X} and s^2 are simply the sample mean and variance of the $\{X_i\}$. Gavaris (1980) used these estimators to standardize fishing vessels.

Because $g_n(t)$ involves an infinite series, Finney (1941) offers an approximation. A first approximation that is commonly used is simply $\hat{\mu} = \exp(\bar{X} + s^2/2)$. Zhou and Gao (1997) report that Cox's method for providing confidence intervals provided relatively good performance in simulations:

$$\left\{ \exp\left(\bar{X} + s^2/2 - Z_{1-\alpha/2} \sqrt{\frac{s^2}{n} + \frac{s^4}{2(n-1)}}\right), \exp\left(\bar{X} + s^2/2 + Z_{1-\alpha/2} \sqrt{\frac{s^2}{n} + \frac{s^4}{2(n-1)}}\right) \right\}.$$

These equations are conceptually generally applicable when applying the log transformation. For example, suppose a linear regression analysis is performed on the log-scale yielding an abundance estimate \hat{X} with residual variance estimated as $\hat{\sigma}^2$. The estimate of abundance on the original scale can be estimated as $\exp(\hat{X})g_n(\hat{\sigma}^2/2)$, or one of the approximations.

The methods described here should be used with caution. Myers and Pepin (1990), and Syrjala (2000) have found a serious flaw in the methodology in the context of the delta-lognormal model. Whenever the mean on the logarithmic scale is back-transformed to the original scale, the potential for severe bias exists. Nevertheless, the logarithmic transformation can be used for inference on the logarithmic scale without back-transforming, and biomass indices can be used to fit a stock assessment model on the logarithmic scale without back-transforming. Alternatively, a constant c can be added to the observed y -values before taking the logarithmic transformation ($y' = \log(y + c)$). Ideally, c should be small compared with y , while avoiding small values of $y + c$ which inflate the sample variance caused by an excess of small values. When y is of moderate size $c = 1.0$ appears to be generally an adequate choice.

Estimating Mean Abundance from Highly Contagious Distributions

When observations are from highly contagious distributions, the usual arithmetic mean is not an efficient estimate of the population mean. Even interpreting the arithmetic mean causes pause. Application of the delta distribution to contagious distributions appears an attractive alternative to simple arithmetic means. In delta distributions zeroes are treated explicitly and positive values are handled as if they are from a highly contagious distribution.

Interpretation of the Arithmetic Mean When Spikes Occur

It is not that uncommon to have over 50% of the catch of some species in a trawl survey to come from a single haul. In the eastern Bering Sea this can be a serious problem particularly for surveys of some crab and rockfish species. Syrjala (2000) notes that while the arithmetic mean is an unbiased estimator of abundance, there will be a large number of surveys over time that underestimate the true mean, with an occasional survey that greatly overestimates the true mean when a large magnitude low frequency value occurs. McConnaughey and Conquest (1992) rightly observe that these large observations are the result of spatial aggregation of the species and should not be discarded. In this case, fishery managers have the proper view of the problem when they take moving averages that mollify the current impact of the observation, but allow the large value to contribute over several years allowing an increase in the survey estimates.

Kappenman (1999) suggests reducing the magnitude of extreme values by replacing them with their expected values (i.e., the expected values of order statistics). For example, if the k largest values are too large, they would be replaced with the expectation of these k largest order statistics before calculating the arithmetic mean.

Delta-Lognormal Distribution Approach

The delta-lognormal distribution (Pennington, 1983, 1986) attempts to solve the skewness problem, large outliers problem, and large numbers of zero values problem in one model. The assumption of the delta-lognormal distribution is that there is a positive

but unknown probability that the observation is zero and, given that the observation is positive, it is drawn from an unknown lognormal distribution. This appears to attack all of the problems: not only are the proportion zeros taken care of, but the logarithmic transformations reduce both the right skewness of the distribution and the effect of outliers.

Let the total sample size be n , with m of these $Y_i > 0$. Aitchison and Brown (1957) provide estimators of the mean (c) and variance (d) of the delta-lognormal distribution similar to those provided earlier by Finney (1941). These are:

$$\hat{c} = \begin{cases} \frac{m}{n} \exp(\bar{x}) g_m(s^2/2), & m > 1 \\ \frac{Y_1}{n}, & m = 1 \\ 0, & m = 0 \end{cases}$$

$$\hat{d} = \begin{cases} \frac{m}{n} \exp(2\bar{x}) \left(g_m(2s^2) - \left(\frac{m-1}{n-1} \right) g_m \left(\frac{m-2}{n-1} s^2 \right) \right), & m > 1 \\ \frac{Y_1^2}{n}, & m = 1 \\ 0, & m = 0 \end{cases}$$

where \bar{x} and s^2 are the sample mean and variance of $\{X_i = \log(Y_i)\}$, where m of the $Y_i > 0$. In these equations $g_m(t)$ is the Bessel function described earlier.

The usefulness of the delta-lognormal distribution comes from the observation that $\text{var}(\hat{c})$ is always smaller than the $\text{var}(\bar{y})$, where \bar{y} is the average of all sampled values when the delta-lognormal distribution is in fact correct (Pennington, 1986; Smith, 1988).

Recall, however, that Myers and Pepin (1990) and Syrjala (2000) found a serious flaw in the methodology. These authors note that the delta-lognormal distribution is not robust to relatively small, undetectable departures from the model assumptions. Syrjala (2000) describes how an excess of small values drive poor behavior in the delta-lognormal distribution. Because the transformation $\log(y)$ approaches minus infinity as y approaches zero, it is easy to see the problem an excess of small observations can cause. Myers and Pepin (1990) suggest that the gamma distribution is superior to the lognormal for fisheries data. However, this may only be true when there is an excess of small values. Firth (1988) generally found little advantage in using the gamma distribution over the lognormal. One possible antidote would be to add 1.0 to all positive values prior to making the logarithmic transformation. The resulting mean can then be adjusted for the added one after estimates have been made using:

$$\frac{m}{n} (\exp(\bar{x}) g_m(s^2/2) - 1).$$

Delta-Gamma Distribution Approach

Stefánsson (1996) combined the delta-gamma distribution with generalized linear model (GLM)-type modeling to analyze groundfish survey data. Like the more familiar delta-lognormal distribution, this model combines the advantage of simultaneously analyzing

zero and nonzero haul data. The method used is to model 0/1 presence absence data as binomial, and the nonzero haul data as gamma. The resulting likelihood function can be written as:

$$L = \left(\prod_{s,t} (1 - p_{st})^{n_{st}-r_{st}} p_{st}^{r_{st}} \right) \left(\prod_{s,t: y_{st} > 0} \frac{y_{st}^{\alpha-1} e^{-\alpha y_{st}/\mu}}{\Gamma(\alpha) \left[\frac{\mu}{\alpha}\right]^\alpha} \right)$$

where s, t refer to station and year, where y_{st} is the catch in number per tow, n_{st} is the number of replications of station s in year t , and r_{st} is the number of replications with catches greater than zero (usually $n_{st} = 1$ and $r_{st} = 0/1$). The GLMs are used to model the probability that a station tow is greater than zero (p_{st}), and the gamma parameters for μ can be linear models of latitude, depth, etc. The gamma distribution parameter α is the gamma shape parameter. These parameters are estimated using maximum likelihood estimation.

Steffanson applied this model to estimate the annual abundance of haddock (*Gadus aeglefinus*) in Icelandic groundfish surveys.

Larger Scale Spatial Correlation

On the larger scale, fish aggregations can be thought of as densities varying in two-dimensional space. From this perspective, one can expect that observations will be spatially correlated. Typically, this correlation is assumed to be dependent only in the distance between observations, statistically termed a stationary process. That is, the closer the observations are to each other, the higher the expected correlation. If data are sampled at a scale that is greater than that which spatial correlation occurs, then the usual univariate statistics with independent observations should be applied.

We shall consider two classes of statistical analysis that are designed to analyze spatially correlated data and seem particularly applicable to fishery survey data. The first is geostatistics (Cressie 1989; Pelletier and Parma, 1994; Rivoirard et al., 2000), which as the name implies, was invented to create predictive maps of ore grade from mine samples. The second is random field linear models (Zimmerman and Harville, 1991), which has roots in the evaluation of crop yields from agricultural experiments. On the surface these methods may seem similar, but in fact they are extremely different in the kinds of statistical inference that they allow.

Implications of Serial and Spatial Correlation

Up until now, we have assumed that survey observations are *iid* random variables. Although surveys inevitably take place in the three-dimensional real world, for simplicity we can assume that they take place in a plane. In many, if not most surveys, it would seem that survey abundance would show at least some spatial correlation. That is, observations near each other can be expected to be positively correlated. However, the degree of spatial correlation depends greatly on the species being surveyed and the scale of the survey. Here, scale refers to the distance between sampling stations. If the distance between sampling stations is large, spatial correlation will be less noticeable.

Serial and spatial correlations have similar effects on mean estimates. Roughly speaking, when serial correlation is present, the marginal information content per

observation for estimating a mean is reduced. This can be expressed as an effective sample size, which would be the sample size that describes the shrinkage in variance when estimating the mean. For the simple serial correlation model where neighboring observations have a correlation ρ , Cressie (1993) gives the formula as

$$n' = n/[1 + 2\{\rho/(1 - \rho)\}\{1 - (1/n)\} - 2\{\rho/(1 - \rho)\}^2(1 - \rho^{n-1})/n].$$

The variance of the sample mean would be σ_0^2/n' instead of σ_0^2/n when observations are correlated. For large values of n , $n' \approx n(1 - \rho)/(1 + \rho)$. As an example when $\rho = 0.2$, then $n' \approx (2/3)n$ so that the reduction in effective sample size can be significant. Nevertheless, this correction is usually ignored when analyzing fisheries survey data.

Geostatistics

Geostatistical methods (see Cressie, 1993) have been found to be useful in fishery surveys for estimation and mapping (Pelletier and Parma, 1994; Rivoirard et al., 2000; Petitgas, 2001). Geostatistics is essentially a synonym for kriging, the statistical method of interpolation used to estimate survey density on a uniform grid, where data were collected from a more arbitrary spatial sampling. Since the total biomass estimate (B) is the sum of density weighted by area

$$\sum_{i=1}^n a_i \bar{X}_i,$$

the use of geostatistics (i.e., kriging) is intended to improve estimates from data having nonuniform sampling. Because hydroacoustic surveys by necessity sample along transects, sampling is quite nonuniform relative to two-dimensional space. For this reason, the use of kriging is a common way of estimating abundance from hydroacoustic data (Petitgas, 2001). It is expected that kriging will improve estimates from nonuniform continuous sampling common when sampling with hydroacoustics. Also, Rivoirard and Wieland (2001) used kriging with external drift to correct trawl survey data for daylight effects.

Consider fish abundance to be a two-dimensional spatial process $Z(x_i)$, where $Z(x_i)$ is fish abundance at a location x_i in two-dimensional space. It makes sense that $Z(x) = \mu(x) + \varepsilon(x)$ can be modeled as some trend $\mu(x)$ and a stochastic component $\varepsilon(x)$ which is spatially correlated error. As Cressie (1989) notes, it is somewhat confusing as to what is trend and what is spatially correlated error: "What is one person's mean structure could be another person's correlation structure." However, Sullivan (1991) found that detrending hydroacoustic data relative to depth prior to kriging substantially decreased kriging variance. We assume $\mu(x) \equiv \mu$ so that all structure can be modeled as spatial stochasticity. If we assume the spatial stochasticity is second-order stationary (i.e., $E(Z(x)) \equiv \mu$ and $\text{Cov}(Z(x+h), Z(x)) = C(h)$ where h is the Euclidean distance between observations), the semivariogram can be written as $\gamma(h) = \sigma^2 - C(h)$. Here σ^2 is the variance of the process at distance zero. In geostatistics, the semivariogram is usually estimated empirically, followed by smoothing after a choice is made regarding the

parametric covariance function $C(h)$. The choice is usually made from among (Cressie, 1993):

$$\text{exponential : } C(h) = \sigma^2 \exp(-h/r)$$

$$\begin{aligned} \text{spherical : } C(h) &= \sigma^2 \left(1 - \frac{3h}{2r} + \frac{h^3}{2r^3} \right), \text{ if } 0 \leq h \leq r \\ &= 0, \quad \text{if } h > r \end{aligned}$$

$$\text{Gaussian : } C(h) = \sigma^2 \exp(-h^2/r)$$

where the exponential and spherical functions appear to be more robust than the Gaussian.

Simple Kriging

Making the unrealistic assumption that $\mu(x_i)$ is known at every x_i , along with $\text{cov}(Z(x_i), Z(x_j))$, estimation can be accomplished using simple kriging (Cressie, 1993). Although this assumption is unrealistic, it provides some intuitive insight into just what kriging is. The goal is to estimate the abundance index at any point x_0 , $\hat{Z}(x_0) = \beta_0 + \sum_{i=1}^n \beta_i Z(x_i)$ by minimizing the expected prediction error $\text{Var}(\hat{Z}(x_0)) = E(\hat{Z}(x_0) - Z(x_0))^2$. Let $Z'_1 = \{Z(x_1), \dots, Z(x_n)\}$, $\mu'_1 = \{\mu(x_1), \dots, \mu(x_n)\}$, $\mu_0 = \mu(x_0)$, $\Sigma_{01} = \{\text{cov}(Z(x_0), Z(x_1)), \dots, \text{cov}(Z(x_0), Z(x_n))\}$ be a $1 \times n$ vector, and $\Sigma_{11} = \{\sigma_{ij}\}$ be an $n \times n$ matrix whose i, j th entry is $\sigma_{ij} = \text{cov}(Z(x_i), Z(x_j))$. The minimum prediction error estimate and estimation variance from simple kriging are then: $\hat{Z}(x_0) = \mu_0 + \Sigma_{01} \Sigma_{11}^{-1} (Z_1 - \mu_1)$ and $\text{var}(\hat{Z}(x_0)) = \text{var}(Z(x_0)) - \Sigma_{01} \Sigma_{11}^{-1} \Sigma'_{01}$. This is the same result as if we assumed that $Z(x_0)$ and Z_1 are multivariate normal and calculate the multiple regression estimate $E(Z(x_0)|Z_1)$ (Morrison, 1967).

Ordinary Kriging

Ordinary kriging is characterized by the more realistic assumption that μ is an unknown constant. As in the case of simple kriging, the goal of ordinary kriging is to provide a linear unbiased estimate $\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$ of $Z(x_0)$, where $E(\hat{Z}(x_0)) = \mu$, and $\{\lambda_i\}$ are parameters to be estimated.

The unbiasedness of $\hat{Z}(x_0)$ is assured by the constraint that $\sum_{i=1}^n \lambda_i = 1$. Using Lagrange multipliers to ensure this constraint, Cressie (1993) derives the $\{\lambda_i\}$ for the minimum mean square error predictor as:

$$\begin{aligned} \lambda_0 &= \Gamma_0^{-1} \gamma_0 \text{ where,} \\ \lambda_0' &= \{\lambda_1, \dots, \lambda_n, m\}, \\ \gamma_0' &= \{\gamma(x_0 - x_1), \dots, \gamma(x_0 - x_n), 1\}, \text{ and} \\ \Gamma_0 &= \begin{cases} \gamma(x_i - x_j) & i, j = 1, \dots, n \\ 1, & i = n+1, j = 1, \dots, n \\ 0, & i = n+1, j = n+1 \end{cases} \end{aligned}$$

Here m is the Lagrange multiplier, and γ is the semivariogram. The minimum mean square error of the predictor is then $E[\hat{Z}(x_0) - Z(x_0)]^2 = \lambda_0' \gamma_0$.

Interpreting Geostatistical Estimation Error

A serious misinterpretation can occur when the prediction error described above as $\text{Var}(\hat{Z}(x_0))$ is interpreted as the sampling error. The easiest way to understand the problem is to note that geostatistical analysis is conditioned on one realization of a stochastic process. The kriging and prediction variance described above have to do only with this one realization. For example, the kriging estimation variance can be used to construct a confidence interval about the predicted value that describes what is the probability that $Z(x_0)$ (for this one realization) is contained in $\hat{Z}(x_0) \pm z_{1-\alpha/2} \sqrt{\text{Var}(\hat{Z}(x_0))}$. Here, $z_{1-\alpha/2}$ is the $1 - \alpha/2$ th percentile of the standard normal distribution.

However, if one is interested in the broader issue of what is the confidence interval about $\hat{\mu}$ estimated from the spatial observations $Z(x_i)$, when the spatial sampling is repeated, then one must require a probability distribution that captures the variability in repeated sampling of the $Z(x_i)$. Usually this is the probability distribution associated with the presumed likelihood function of the data, and the analysis is performed using standard likelihood methods or approximations. Geostatistical methods cannot be used to make this broader comparison. If the kriging prediction variance over a grid of values $\sum_{i=1}^n \hat{Z}(x_{0i})/n$ is used to estimate the variance of the survey mean estimate, it is likely to seriously underestimate the variance of the mean from repeated surveys.

Random Field Linear Models

Zimmerman and Harville (1991) first proposed a random field linear model (RFLM) for the analysis of spatially correlated data. Lai and Kimura (2002) suggested applying RFLM to the analysis of fisheries survey data, and Cressie (1993) briefly discusses this method as “another example of why statistics for spatial data deserves a place in the statistician’s repertoire.”

Random field linear model is essentially the usual fixed effects normal linear model, but with the covariance matrix parameterized to allow for spatial correlation. Random field linear models are therefore appropriate for estimation and statistical inference. Recall that geostatistics (i.e., kriging) is essentially the linear predictor of a value on a surface, conditioned on observing other values on that surface. The usage of geostatistics is therefore more narrowly applicable to surface estimation.

Random field linear model analysis utilizes the standard likelihood approach. First, linear models can be used to model the data set. Once a spatial covariance function is selected, based on only the distance d between observations, model parameters can be estimated using maximum likelihood estimation.

For the RFLM, a likelihood analysis can be used assuming normality (i.e., $Y \sim N[X\beta, \Sigma(\theta|D)]$) and starting with the negative log-likelihood defined up to a constant by:

$$-\log(L) = \frac{1}{2} \log(|\Sigma(\theta|D)|) + \frac{1}{2} (Y - X\beta)' \Sigma(\theta|D)^{-1} (Y - X\beta).$$

Here $D = \{d_{ij}\}$ is the distance between the i th and j th observations, $\Sigma(\theta|D)$ is the covariance matrix based on one of the covariance functions described earlier, θ are the parameters associated with this covariance function, X is the design matrix, and β are the regression coefficients.

Maximum likelihood (ML) estimates $(\hat{\theta}, \hat{\beta})$ are simultaneously estimated by minimizing $-\log(L)$. Tests of k linear constraints on β (i.e., the null hypothesis) are carried out using the standard likelihood ratio test. Under the null hypothesis $(\tilde{\theta}, \tilde{\beta})$ are estimated, leading to the likelihood ratio test:

$$\hat{\chi}_k^2 = 2[\log(L(\hat{\theta}, \hat{\beta})) - \log(L(\tilde{\theta}, \tilde{\beta}))]$$

which under the null hypothesis is asymptotically distributed as a chi-square distribution with k -degrees of freedom.

The covariance matrix of parameter estimates can be estimated from the inverse Hessian matrix of minus the log-likelihood, and the variance of selected functions of estimated parameters can be calculated using the delta method (Seber, 1982). Therefore, either the delta method or the likelihood ratio method can be used to construct hypothesis tests.

An RFLM Example

A comparative fishing experiment was conducted during the synoptic 1989 sea scallop research survey by the Northeast Fisheries Science Center to evaluate the comparability of scallop catches among three research vessels (RVs) (Lai and Kimura, 2002). Identical tows, using 8'-wide scallop dredges towed at 3 knots for 15 min, were made by each of the three RVs at 13 stations off Long Island, New York. Longitude and latitude were converted to distance in nautical miles using standard formulas.

Random field linear model analysis was used to test the null hypothesis that the catch rates for the RVs were the same (i.e., $H_0 : \mu_a = \mu_b = \mu_c$). The design matrix for this analysis is the matrix of indicator variables:

$$X = \begin{pmatrix} 1 & 0 & 0 \\ \cdot & \cdot & \cdot \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \cdot & \cdot & \cdot \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix}.$$

The analysis was carried out using the three covariance functions $(\theta = (\sigma^2, r))$, and the ML method, with the results given in Table 1. The spherical function gave the best fit to the data indicated by the smallest $-\log(L)$. The second best fitting covariance function was the Gaussian model, and the exponential model fit the worst. Only the Gaussian RFLM fits indicate a significant difference in the vessel treatments, rejecting $H_0 : \mu_a = \mu_b = \mu_c$ at the $\alpha = 0.05$ level (Table 1). Because the spherical function gave the best fit under RFLM and indicated no significant difference in the vessel treatments, this is the principal conclusion using RFLM. The spherical function also produced the smallest $\hat{\theta}_1 = \hat{\sigma}^2$, which is a measure of expected residual mean square error.

Table 1

Results when applying an RFLM model to sea scallop data, using ML with exponential, spherical, and Gaussian covariance functions. Tests for $H_0 : \mu_a = \mu_b = \mu_c$ were carried out using the likelihood ratio χ^2 statistic

	ML		
	Exponential	Spherical	Gaussian
$\hat{\theta}_1$	14356.918	11765.989	14403.327
$\hat{\theta}_2$	6.339	8.791	4.792
$\hat{\mu}_1$	107.225	109.899	118.281
$\hat{\mu}_2$	124.762	127.540	138.335
$\hat{\mu}_3$	100.277	102.996	101.973
$-\log(L)$	181.019	179.864	180.840
$\chi^2(2df)$	1.992 ^{n.s.}	2.191 ^{n.s.}	6.994*

n.s. and * indicate failed to reject H_0 and reject H_0 at the 0.05 significance level.

Modeling Relative Abundance Using Covariates

Catch rates from fishery surveys and fishing gears can be greatly influenced by particulars of fishing gear, and environmental parameters under which fishing is occurring. Survey standardization can therefore encompass both types of parameters.

The difference between a scientific fishery survey and commercial fishery CPUE is less conceptual, and more one of sampling design. Therefore, similar tools have been found useful for analyzing both survey and commercial fisheries data. The goal of analysis will generally be to (1) identify factors influencing CPUE (these may be vessel or gear characteristics), or factors related to fish habitat (e.g., depth, salinity, etc.), and/or (2) estimate relative abundance indices that have been adjusted for these factors that may have differed in different years. It is generally recognized that factors such as salinity, temperature, and depth (Smith, 1990; Swain et al., 2000) and daylight (or time of day) and vessel speed (Rivoirard and Wieland, 2001; Adlerstein and Ehrich, 2002) can have a great effect on catch rates even when fishing gear are unchanged. Therefore, the methods described below use not only the catch or CPUE data, but also vessel and/or environmental covariates which can be related to catch rates.

Historically, the general log-linear model appears to be one of the first to relate CPUE data to covariates. More recent approaches appear to follow a GLM-type approach, where the response can have a variety of possible sampling distributions. The generalized additive model (GAM) again allows a variety of possible sampling distributions, but can fit covariates as smoothed functions whose forms may not be specified.

Generalized linear models, GAMs, and generalized linear mixed models (GLMMs) are complex and difficult to fully understand. Fortunately, a recent special volume in Fisheries Research (2004, Vol. 70 (2–3)) provided applications of these methods to fisheries research in general, and to the analysis of CPUE data in particular (see Maunders and Punt, 2004). Among these articles, an authoritative review of the theory surrounding these methods aimed at fisheries applications (Venables and Dichmont, 2004) should be of interest to a broad audience. Venables and Dichmont (2004) generally favor GLMs over GAMs, because of the inability of GAMS to incorporate interactions.

General Log-Linear Models

Gavaris (1980) and Kimura (1981) showed how data from a fleet of vessels could be standardized to that of a selected vessel. The general approach is to take the log (CPUE) and model the transformed observations using linear models, with the implicit assumption that data are lognormal. These methods can be used on a time series of fishery abundance estimated from fishery surveys, or CPUE data from a fleet of commercial fishing vessels. Commercial fishery CPUE data may be considered a fleet performing a fishery survey but with extremely poor gear standardization, and extremely poor spatial design. The goal is to use linear models to remove effects that are caused by differences in fishing gear among vessels, and differences in the time and area covered by the fishing operations. The idea is that when all of these other factors are accounted for, the remaining year effect is due to differences in annual abundance.

Generalized Linear Model-Type Approaches

In many ways, the generalized linear model approach described by McCullagh and Nelder (1989) and Chambers and Hastie (1992) seems the most flexible way of analyzing survey data with covariates. The method allows the response variable Y to have a variety of possible distributions, e.g., Poisson, binomial, gamma, etc. This is done by providing an appropriate link function g and variance function V , such that $E(Y) = \mu$, $\eta = g(\mu)$, $Var(Y) = \phi V(\mu)$, and $\eta = X\beta$ where ϕ is a constant and X is a matrix of covariates and β are the regression coefficients. Parameter estimation follows some method of calculating maximum likelihood estimates.

Power and Moser (1999) proposed a direct application of the negative binomial distribution. Assume that a survey haul or set has a negative binomial probability of catching y animals:

$$\Pr(Y = y \mid \alpha, k) = \frac{(y + k - 1)!}{(k - 1)!y!} \frac{\alpha^y}{(1 + \alpha)^{(y+k)}}$$

where

$$\mu = E(Y) = k\alpha \text{ and } var(Y) = \mu + \mu^2/k.$$

They used the GLM link for the negative binomial distribution,

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j = \ln \left(\frac{\mu_i}{\mu_i + k} \right),$$

where $\{x_{ij}\}$ are covariates. Maximum likelihood solutions were estimated using the Newton-Raphson algorithm. The model was applied using ichthyoplankton catch data, which were shown to have highly contagious negative binomial distributions.

Stefánsson (1996) used the delta-gamma distribution, a GLM link, and maximum likelihood methods to estimate parameters.

Generalized Additive Model Approach

Swartzman et al. (1992) used GAMs (Chambers and Hastie, 1992) to analyze survey data for flatfish in the eastern Bering Sea with relation to depth and bottom temperature. The

goal of this analysis was to (1) detect trends in groundfish abundance, and (2) improve abundance estimates by including trend.

Intuitively, the GAM approach can be viewed as a sort of regression of variable Y on the smoothed residual responses on explanatory variables x_j , say $S_j(x_j)$, providing estimates of $E(Y|x_1, \dots, x_p)$. Swartzman et al. (1992) assume a Poisson process, so that the catch in numbers (y) per fixed area say, are transformed as $Y = \log(y)$. In their study, Y was related to location, depth, and bottom temperature. The regression above can be mathematically described as

$$E(Y | x_1, \dots, x_p) = S_0 + \sum_{j=1}^p S_j(x_j),$$

where S_0 is a constant, and the $\{S_j\}$ represent the p explanatory variables. The explanatory variables may seem a little vague because they are based on smoothed functions (see Chambers and Hastie, 1992).

The model is fit using the backfitting algorithm (Chambers and Hastie, 1992), also called the Gauss-Seidel iterative method. Swartzman et al. (1992) describe the algorithm (for the Poisson process) as holding all the other covariate functions fixed, and then for the j th covariate, calculating the partial residual

$$r_j = z - S_0 - \sum_{k \neq j} S_k(x_k)$$

where the adjusted dependent variable z is estimated from

$$z_m = \eta_m + \frac{(y - \exp(\eta_m))}{\exp(\eta_m)}$$

on the m th iteration. The regression function S_j is updated by smoothing r_j on x_j , and then the process is repeated for each covariate. Following updating of $\{S_j\}$, η_m can be updated to

$$\eta_{m+1} = S_0 + \sum_{j=1}^p S_j(x_j)$$

followed by the next iteration. The iterations continue until the deviance

$$Dev(y, \mu) = 2 \sum_{i=1}^n \{y_i \log(y_i / \mu_i) - (y_i - \mu_i)\}$$

where $\mu = \exp(\eta)$, is reduced to a minimum.

Confidence intervals for predictors can be calculated using the bootstrap, and hypothesis tests for significance of covariates can be made using either the bootstrap or permutation tests. Results of the analysis provided in Swartzman et al. (1992) include plots of annual survey abundance, standard errors, plus contour plots that show the preferred habitat of different flatfish species.

Vessel Calibration (Estimating Relative Fishing Power)

Because the fishing gear or the vessel towing them can have a great impact on relative catch rates, it may be necessary or useful to investigate fishing power differences among

vessels. This may be critical when more than one vessel is required to complete a survey in a given year, or a new research vessel is brought on and the old one is retired. Since the main value of fishery surveys is to have a time series of abundance (or an index of abundance) over many years, a lack of standardization of gear over years can call into question the validity of the comparisons being made. Modern statistical treatment of relative fishing power naturally begins with Robson (1966) who used log catch rates and regression analysis to estimate relative fishing power. Pelletier (1998) reviewed the literature on the inter-calibration of research survey vessels and found both independent hauls and paired hauls were commonly used. The independent haul method was applied to confined areas where fish abundance could be assumed to be uniform. However, it was generally found that paired tows better controlled for heterogeneity in fish abundance over space and time. The efficiency of randomized complete block designs will generally be preferred by statisticians for this type of research. In its simplest form, two or more vessels fish side by side with the fishing gears that we wish to compare. Under this circumstance the catch ratios can be compared directly as simple means (or totals).

Pelletier (1998) developed a quasi-likelihood model (see Chambers and Hastie, 1992) for estimating relative fishing power (intercalibration) between two vessels with paired tows. Pelletier explicitly derived the intuitive relative fishing power estimate $\hat{x} = C_{2T}/C_{1T}$, where C_{iT} represents the total catch of vessel i over all paired tows. Pelletier also suggests bootstrapping this statistic (using paired catches for observations) which can provide estimates of variance, confidence intervals, or statistical tests.

The classical analysis of this same data would be to perform a randomized block analysis on the $y_{ij} = \log(C_{ij})$ where C_{ij} is the catch of individual hauls, i refers to vessel 1 or 2, and j refers to the block (pair). The anova model would be the usual: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, with relative fishing power estimated as

$$\hat{R}_{i,i'} = \exp(\hat{\tau}_i - \hat{\tau}_{i'}),$$

and confidence intervals estimated on the log-scale as

$$\hat{\tau}_i - \hat{\tau}_{i'} - t_{1-\alpha/2} \sqrt{2s^2/J} < \hat{\tau}_i - \hat{\tau}_{i'} < \hat{\tau}_i - \hat{\tau}_{i'} + t_{1-\alpha/2} \sqrt{2s^2/J},$$

where s^2 is the mean square error (i.e., interaction mse) from the two-way anova, and $t_{1-\alpha/2}$ is the $1 - \alpha/2$ th percentile of the t -distribution with the degrees of freedom associated with the error term. Overall, bootstrapping Pelletier's \hat{x} seems superior to the classical block anova described here since the parametric normal assumption and log-transformation would not be required.

Bergh et al. (1990) noted a 4-to-10-fold reduction in necessary sample size when comparing fishing gears when gears were alternated within a fishing trip, as opposed to assigning a single gear to each trip. This study showed the value of using a complete block design where each fishing trip represented a block.

Kimura and Zenger (1997) in a data-rich situation analyzed 5 years of long-line comparative fishing data. An analysis of catch ratios using linear models allowed corrections for site, depth, fish size, and site:depth and site:size interactions.

Wilderbuer et al. (1998) used Kappenman's method based on a robust estimate of scale parameters to estimate relative fishing power. Wilderbuer et al. (1998) showed this new method gave results similar to the more traditional methods described here. As a warning, Munro (1998) noted that applying fishing power corrections may actually increase the mean square error of estimates.

Although relative fishing power corrections can be useful, caution should be exercised when applying estimated coefficients to surveys made under conditions different than those from which the corrections were calculated. As noted earlier, for modeling purposes, it may be better to treat surveys using different vessels as separate abundance indices.

Hydroacoustic Survey Designs

Hydroacoustic echo integration is a tool used by fishery biologists to investigate the abundance of fish in midwater (the middle layers of depth). It is impractical to present this complex technology in detail here (see MacLennan and Simmonds, 1992; Gunderson, 1993). However, only a brief coverage of the rudiments from Traynor¹ is sufficient to understand the basic issues surrounding hydroacoustic survey sampling design. The method is based on insonifying the water column from a moving vessel with acoustic energy at a frequency of around 38 KHZ. **The timing and strength of return echoes from this insonification is dependent on the species, depth, size, position, and orientation of fish in the water column. The species and size of fish must be investigated directly by midwater trawls.**

The echo integration system provides as output, the average acoustic backscattering area SA which are in units m^2/nmi^2 . That is, for every nmi^2 insonified, there were so many m^2 of return echo. The amount of return echo provided per target is the individual target strength $\langle\sigma\rangle$ in m^2 . Therefore, the total number of targets per nmi^2 is given by $\rho = SA/\langle\sigma\rangle$. The biomass can be determined as the number of targets times an average weight. Ignoring the many difficulties contained in these calculations, the hydroacoustic echo integration system provides a mean average acoustic backscatter area (or mean biomass density) averaged over, say 0.5 nmi of transect.

The question of hydroacoustic survey design is mainly a question of where to place transects (i.e., what path should the vessel take). This returns us to the question of spatial correlation, and how we should deal with this. For the large-scale spatial variability models (kriging and RFLM) we discuss how to deal with spatial correlation *ex post facto*. The question of hydroacoustic sampling design is how transects should be placed to provide the most efficient estimates of average biomass.

There are three obvious transect patterns that will reasonably cover most sampling areas (Figure 3): zig-zag, stratified parallel, and random parallel. **Kimura and Lemberg (1981) using simulation, concluded that zig-zag transects were superior to stratified parallel transects at low transect densities, but stratified parallel transects were superior to zig-zag transects at higher transect densities.** Transect density refers to the number of crossings of the sampling area. Regions of turnaround are oversampled by zig-zags, and regions between the turnarounds are undersampled. **Francis (1985) preferred stratified parallel transects because it is apparent that they provide maximum separation of transects and therefore the least correlation between transects.** Randomly placed transects, although providing valid confidence intervals, probably provide the poorest precision because they can more easily miss spatial concentrations. Jolly and Hampton (1990) proposed using quasi-random transects that are separated into strata. Simmonds and Fryer (1996) concluded that a systematic survey with kriging provided the most precise survey estimate, but if reliable variance estimates were desired, then a stratified random survey with at least two transects per strata should be used.

¹J. Traynor's brief introduction, *Midwater Fish Surveys at the, Alaska Fisheries Science Center (AFSC) Quarterly Report*, Jan.–March 1997, provides an excellent introduction to hydroacoustics as a fishery survey method. For a reprint write the RACE Division, Alaska Fisheries Science Center, 7600 Sand Point Way N.E., Seattle, WA 98115.

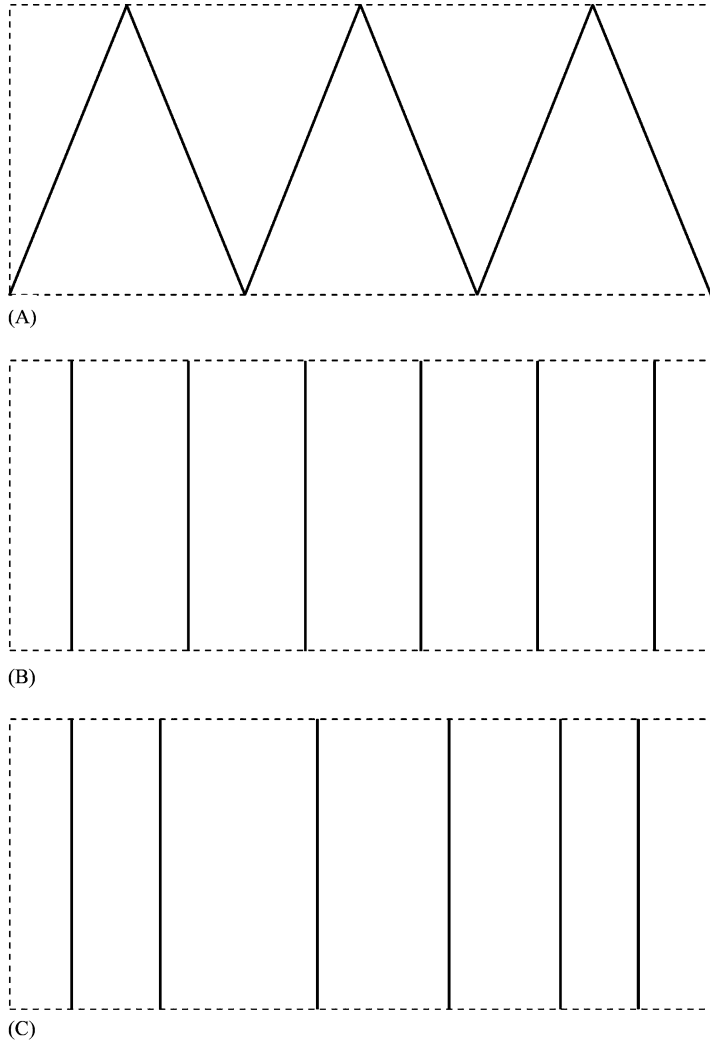


Figure 3. An illustration of the main types of hydroacoustic transects: (a) zig-zag, (b) stratified parallel, and (c) random parallel.

Aglen's Index of Sampling Coverage

The question arises how we should measure sampling coverage when an area is sampled with transects. Suppose an area A is sampled with transects of total length T . Three measures of sampling coverage (SC) seem most obvious:

1. $SC_1 = T/A$: the total length of transect divided by area.
2. $SC_2 = (T\varepsilon)/A$: the area of water sampled divided by total sampling area. Here ε is the width of the hydroacoustic swath that is sampled.
3. $SC_3 = T/\sqrt{A}$: the total length of transect divided by the square root of area (Aglen's Index).

The first index, SC_1 , is scale-dependent and therefore is dependent on the unit of measurement being used. Therefore, SC_1 measured in m is different than SC_1 measured in km. The second and third indices, SC_2 and SC_3 are scale-independent. The second index is the proportion of the area sampled by the survey. Francis (1985) argued that SC_2 is best because if two areas have the same SC , then their combined area would have the same SC . This is true with SC_1 and SC_2 , but not SC_3 (the value of SC_3 is larger than for the individual areas). Aglen (1989) argued that this is appropriate “because the sampled proportion required for a given precision will decrease as the area increases.”

Aglen (1989) noted that a simple, but insightful way to look at this problem is to look at each transect mean backscatter as an independent estimator of the average abundance in a particular area. Let the survey consist of n transects, where M_A is the average acoustic backscatter of a randomly selected transect. Then $var(\bar{M}_A) = var(M_A)/n$ where $var(M_A)$ is the between transect variability. Since fish abundance (t) can be scaled from the backscatter using target strength $\langle\sigma\rangle$, it follows that $CV(t) = CV(M_A)/\sqrt{n}$. If the sampling area is square, then $n = T/\sqrt{A} = SC_3$ so that $CV(t) = CV(M_A)/\sqrt{(T/\sqrt{A})} = CV(M_A)/\sqrt{SC_3}$. Therefore, SC_3 appears to be an appropriate index of sample coverage, playing a similar role to sample size n for random sampling.

CV of Survey Abundance Estimates

The calculations above assume that mean backscatter is uncorrelated from transect to transect. Under this assumption, the variance and coefficient of variation of mean abundance can simply be calculated from random sampling formulas where the sampling unit is the individual transect. Williamson (1982) suggests analyzing the resulting data using cluster sampling formulas, where each transect is viewed as an independent cluster sample. Although Francis (1985) criticizes this approach because the spatial process being sampled is not stationary, it should be pointed out that geostatistics and RFLM analyses are also generally applied assuming spatially stationary processes (i.e., a constant mean process).

Aglen (1989) was able to compare the CV estimates from repeated surveys, and compare these with CV estimates assuming that individual transect means were independent random variables. He found that on average, for the over 70 cases he was able to observe, that the CV s calculated both ways compared well (Figure 4). This is strong evidence that on average, transect-to-transect variability can provide reasonable confidence intervals about survey means.

For illustrative purposes we compare coefficients of variation of mean hydroacoustic adult walleye pollock (*Theragra chalcogramma*) biomass estimates using kriging (Wilson et al., 2003), estimates from RFLM and estimates using the Williamson (1982) cluster sampling model (see Table 2). These results illustrate the underestimation of CV when using kriging estimation variance, when in fact between survey sampling variance is desired. Geostatistical estimation error and sampling models such as RFLM are answering different statistical questions. Kriging deals with surface estimation and its variance and RFLM deals with variability between independent samplings.

Applying the Bootstrap to Fishery Survey Data

The basic idea behind the bootstrap (Efron, 1982; Efron and Tibshirani, 1986) is to substitute the true sampling cumulative distribution function (CDF) F , with the empirical CDF \hat{F} , and then proceed as if \hat{F} were the true F . Thus, the functional $g(F)$ (e.g., an integral), can

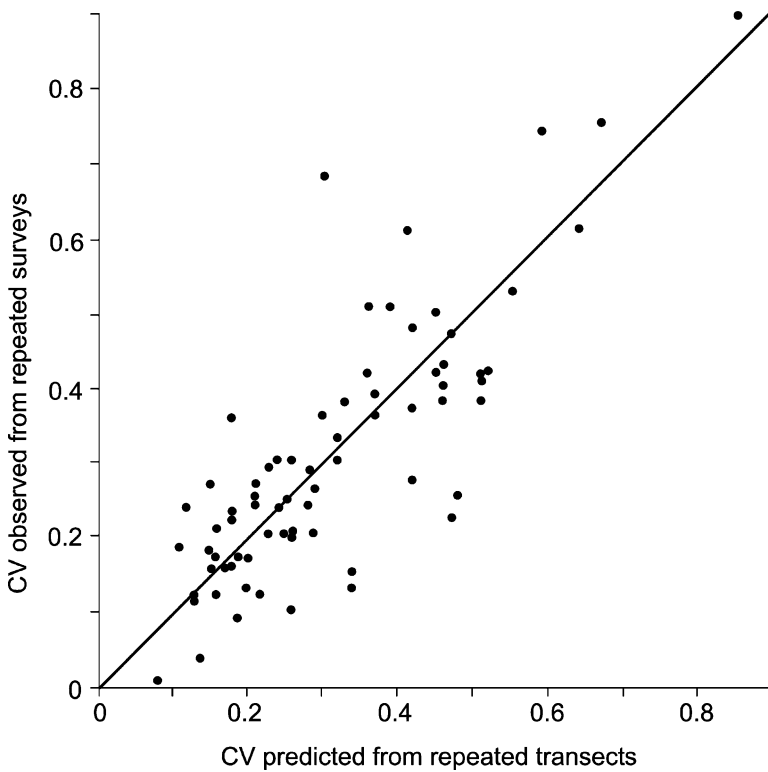


Figure 4. A comparison of hydroacoustic survey abundance CVs, where the x-axis represents CV estimated from transect-to-transect variability assuming transects are statistically independent, and the y-axis is CV estimated from replicated surveys. This illustrates that between-transect variability can sometimes be used as a basis for estimating survey variability (adapted from Aglen 1989).

be estimated by $g(\hat{F})$. This can be mathematically paraphrased as follows: given the results of an initial sampling of a population, say (X_1, X_2, \dots, X_N) , randomly resample this finite population with sample size n (say X_1^*, \dots, X_n^*), with replacement, say $B = 1,000$ times, and then calculate some statistic or parameter for each of the $B = 1,000$ resamplings—say, for example, von Bertalanffy growth parameters. Such a resampling is equivalent to sampling from the empirical CDF \hat{F} , and then estimating $g(\hat{F})$ from this sample. Calculate the mean and standard deviation of these parameters, and use these to construct a mean estimate

Table 2
Coefficient of variation of Barnabas Gully mean walleye pollock hydroacoustic abundance estimates in 2001

Pass	Random sampling (0.5 nmi)	Kriging	RFLM	Cluster sampling
1	0.29	0.12	0.34	0.31
2	0.13	0.07	0.35	0.28
3	0.17	0.11	0.32	0.31

and confidence intervals about this mean. The extent to which this resampling process has turned out to be a reasonable approximation for practical problems has made bootstrapping a standard tool in fisheries work as well as many other applied areas. Although bootstrapping can be used to estimate the bias of estimates (say $g(\hat{F}) - g(F)$), the method is most useful for estimating variance and confidence intervals, and the trade-off of sampling intensity at various levels of survey designs.

The flexibility of bootstrap sampling makes it a particularly valuable tool for analyzing fishery survey and catch sampling data. There are several reasons for this: (1) it is relatively easy to use compared with performing traditional statistical analysis; (2) the method does not require selection of a probability sampling model except that all observations were equally likely; (3) it is adaptable to sampling structures imposed by the fishery survey gear (e.g., fishing effort may come in "strings of pots" or "strings of hooks," so catches on neighboring pieces of gear may be highly correlated); and (4) it can be focused on any statistic calculated from the sampled data. These are powerful recommendations. Also encouraging is the range of fishing gears to which the bootstrap has been applied: groundfish trawls (Smith, 1997), pot gear for sablefish (Kimura and Balsiger, 1985), longlines for sablefish (Sigler and Fujioka, 1988), and fishery hydroacoustics (Robotham and Castillo, 1990).

Infinite Versus Finite Population Bootstraps

When sampling can be viewed as occurring from an infinite population, the application of the bootstrap seems most straightforward. For passive fishing gears such as traps and longlines, it is difficult to estimate the area being fished so an infinite population theory seems applicable. Bootstrapping applied to pot (trap) surveys (Kimura and Balsiger, 1985) and longline surveys (Sigler and Fujioka, 1988) show how the bootstrap can take into account both the random nature of selecting stations, and the correlation that results from capture gear being on a fishing line. The pot index survey (Kimura and Balsiger, 1985) typically had three or four stations sampled at five depth intervals with five sets within each interval. A straightforward bootstrapping method was used which selected a station at random (with replacement), then systematically sampled each of the depth strata (i.e., all depth strata were sampled once). Within each depth stratum, a set was randomly selected (with replacement) consisting of the catches from a string of 10 pots. Analysis of the bootstrap results clearly showed (Figure 5) that increasing the number of sampling locations was much more effective at increasing survey precision than increasing the number of sets made within each depth.

Smith (1997) thoroughly investigated the use of bootstrap sampling within the context of a stratified random trawl fishery survey. He showed that if survey strata were small, that finite population considerations and the usual infinite population (naïve) bootstrap would provide unbiased estimates of the mean, but biased estimates of the variance. Applying the infinite population bootstrap to finite populations overestimates the variance, so in that sense is a conservative strategy. This bias was described earlier by Rao and Wu (1988). Smith (1997) and Davison and Hinkley (1997) describe several alternatives to the infinite population bootstrap that would largely eliminate this bias. The simplest conceptually may be the finite population bootstrap described by Booth et al. (1994), which "constructs" a finite population of desired size by replicating the observed data a desired number of times and then sampling this constructed population with a fixed sample size, say n , without replacement, a large number of times. The resulting bootstrapped variances will have substantially reduced bias.

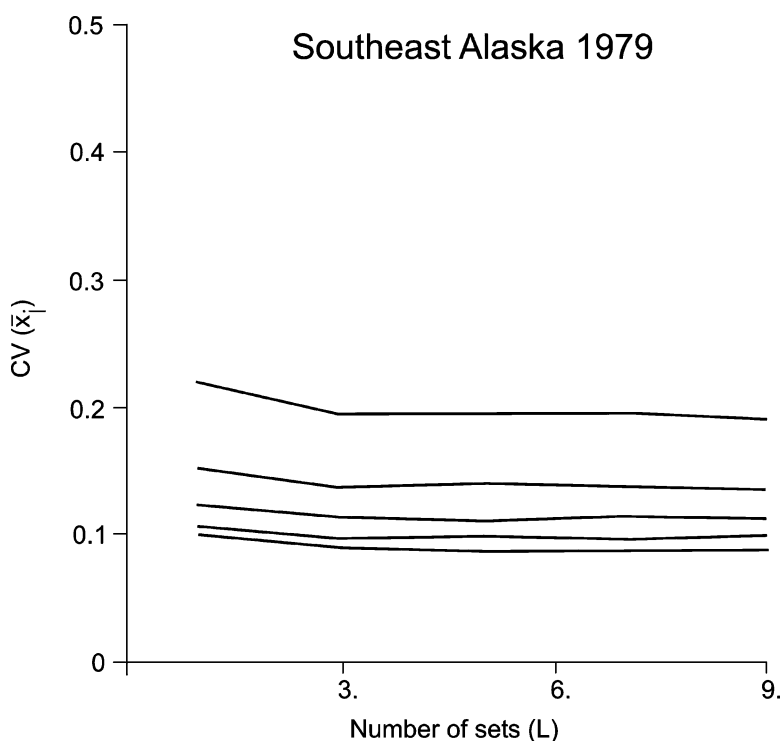


Figure 5. Bootstrap estimates of the CV of sablefish survey abundance estimated from pot index surveys. On each graph, curves from top to bottom are for 3, 6, 9, 12, and 15 locations sampled. Flat curves illustrate that increasing sampling intensity within selected stations were not effective at improving survey precision (adapted from Kimura and Balsiger 1985).

Example: Bottom Trawl and Hydroacoustic Surveys for Alaskan Walleye Pollock

Walleye pollock (*Theragra chalcogramma*) occur at all water depths from near the surface to the bottom and often rest directly on the bottom. This distribution presents a challenge to the process of conducting an assessment survey because no single sampling tool is capable of sampling over the entire depth range. As a consequence, the Alaska Fisheries Science Center uses two types of survey techniques for walleye pollock in the eastern Bering Sea: 1) a bottom trawl survey to sample individuals directly in contact with the bottom and in the water within 2 m of the bottom, and 2) a hydroacoustic survey to sample individuals from 2 m above the bottom to 14 m below the surface.

The cost of this survey is very high. The charter and fuel costs for the two 40-m commercial fishing vessels required for the 65-day bottom trawl survey alone in 2003 was \$730,267. This does not include the cost of the scientific party, and the hydroacoustic portion of the survey which is made from National Oceanic and Atmospheric Administration (NOAA) vessels. However, this must be weighed against the value of the eastern Bering Sea groundfish fishery which is one of the most valuable in the world. In 2002, the ex-vessel value of this fishery (all groundfish species) was estimated to be approximately \$484,600,000.²

²Ex-vessel value was provided by Dr. Joseph Terry, an economist with the Alaska Fisheries Science Center.

The chartered commercial fishing vessels conduct one-half hour bottom trawl tows at approximately 350 sampling sites. Most of the sites are arranged at the centers of a grid of sampling squares measuring 20 nmi on a side. In two locations, however, sampling sites are also placed at the corners of the grid to increase the sampling density for blue king crab (*Paralithodes platypus*) which has a limited spatial distribution. The survey area is stratified into three depth zones: 20–50 m, 50–100 m, and 100–200 m, with divisions that correspond to the positions of oceanographic fronts, and sharp gradients in bottom temperature and species catch composition. The survey is also stratified into a northern and southern zone with a boundary line running northeast-southwest near the center of the survey area that corresponds to a transition in growth rate displayed by several species. Along with the high density sampling areas, the survey area is effectively partitioned into 10 strata (Figure 6). Except for the high-density strata, allocation of sampling sites to strata is proportional to stratum area. Although optimal allocations have been considered, the bottom trawl survey is used to assess the relative abundance of more than 30 species of fish and invertebrates and no allocation scheme has proven better (i.e., has lower variance) for all species combined than simple allocation by area. A systematic spatial arrangement of stations is used, rather than a random arrangement, primarily because of consideration about the efficiency of vessel usage (e.g., the travel time between stations is approximately equal to the time needed to process that catch and record the data). The relative abundance of walleye pollock is shown in Figure 7 for the 2000 bottom trawl survey.³

The hydroacoustic survey uses a single research vessel to conduct acoustic transects totaling approximately 6000 nmi in length (Figure 8). These transects run in a north-south direction and are spaced 20 nmi apart (i.e., stratified parallel) so that they are aligned with the sampling grid used by the bottom trawl survey. Fish density is assessed continuously along the track lines, and then averaged over each 0.5-nmi section. Since walleye pollock is the only commercially important species assessed by the hydroacoustic survey, these transects do not cover the entire area surveyed by the bottom trawl survey but instead cover only the areas where walleye pollock predictably occur. The relative abundance of walleye pollock is shown in Figure 8 for the 2000 hydroacoustic survey.⁴

Although both the bottom trawl and hydroacoustic surveys occur during the summer, sampling stations are generally not sampled synchronously because the speed of conducting the surveys differs. However, both surveys recognize that pollock can change their depth distribution from night to day; therefore both surveys are conducted only during daylight hours.

Because the closest between-sample distance on the hydroacoustics survey (0.9 km) is much less than on the bottom trawl survey (26.0 km), the potential for spatial correlation is much greater on the hydroacoustics survey. Semi-variograms based on the 2000 surveys, show that the hydroacoustic survey has spatial correlation at the spatial scales sampled (i.e., 0.5 nmi along transects, 20 nmi between transects) (Figure 9) while the bottom trawl survey

³Details of the 2000 bottom trawl survey are in: Nebenzahl, D., and P. Goddard. 2000. 2000 bottom trawl survey of the Eastern Bering Sea Continental Shelf. AFSC Processed Report 2000-10. Available from: Alaska Fisheries Science Center, 7600 Sand Point Way NE., Seattle, WA. 98115. USA.

⁴Details of the 2000 hydroacoustic survey are in: Honkalehto, T., N. Williamson, S. de Blois, and W. Patton. 2002. Echo integration-trawl survey results for walleye pollock (*Theragra chalcogramma*) on the Bering Sea shelf and slope during summer 1999 and 2000. Department of Commerce NOAA Technical Memorandum NMFS-AFSC-125,126. Available from: Alaska Fisheries Science Center, 7600 Sand Point Way NE., Seattle, WA. 98115. USA.

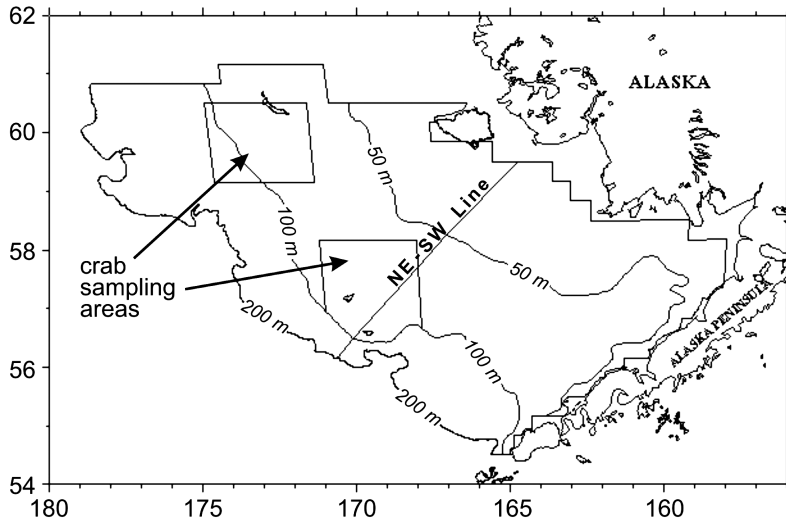


Figure 6. Strata boundaries used for groundfish surveys in the eastern Bering Sea. Strata boundaries are defined by depth contours, the northeast-southwest dividing line, and high density crab sampling areas.

has strong spatial correlation only at the spatial scale (26.0 km) used in the high density sampling areas (Figure 10). Because of this difference in the sample spacing, the analysis of the relative abundance data to produce an overall index of relative abundance proceeds differently for the two surveys.

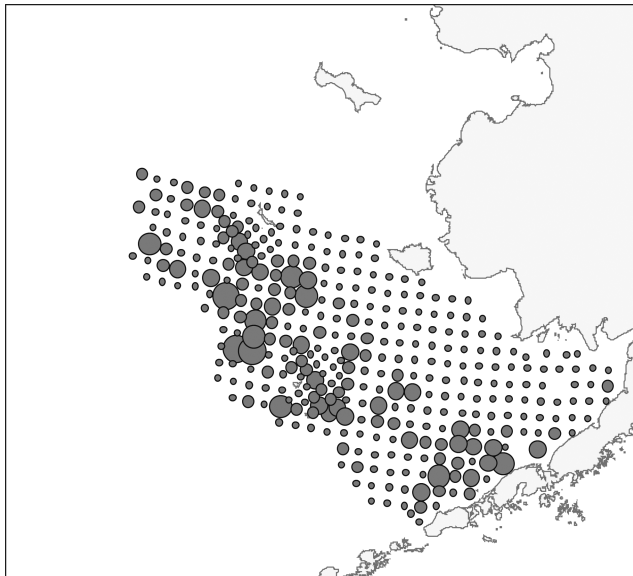


Figure 7. Spatial distribution of sampling sites for the 2000 Bottom Trawl Survey in the eastern Bering Sea, conducted by the National Marine Fisheries Service. The relative abundance of walleye pollock is represented by the circle diameter at the sampling point.

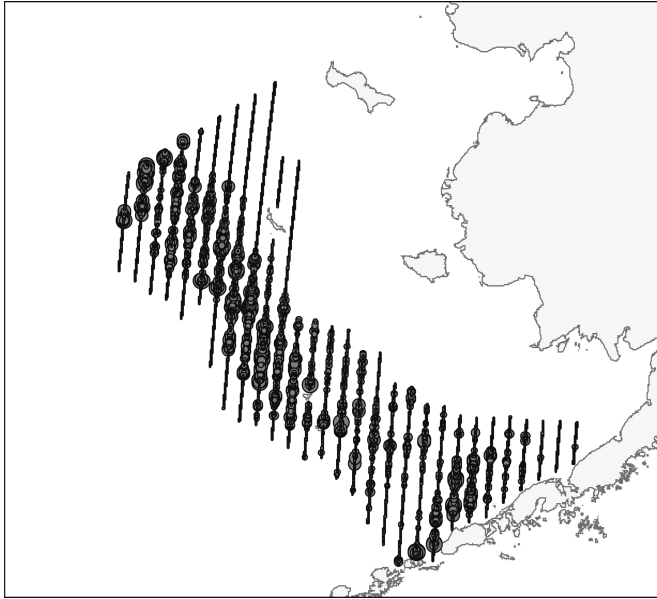


Figure 8. Spatial distribution of sampling sites for the 2000 Hydroacoustic Survey in the eastern Bering Sea, conducted by the National Marine Fisheries Service. The relative abundance of walleye pollock is represented by the circle diameter at the sampling point.

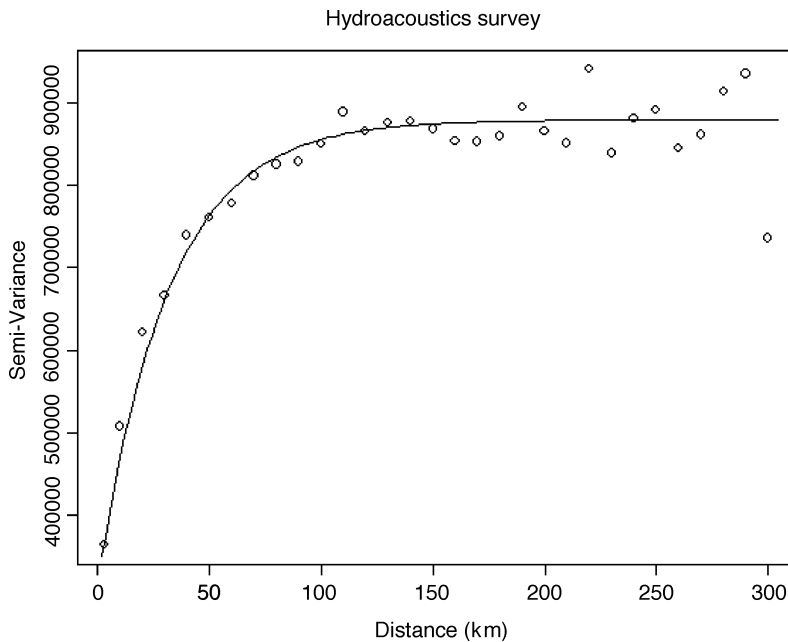


Figure 9. Semi-variogram showing the spatial covariance in relative abundance between sampling sites for the hydroacoustic survey, as a function of between-site distance. The hydroacoustic survey has relatively small minimum between-site distances and shows substantial covariance even at distances much larger than the minimum distance.

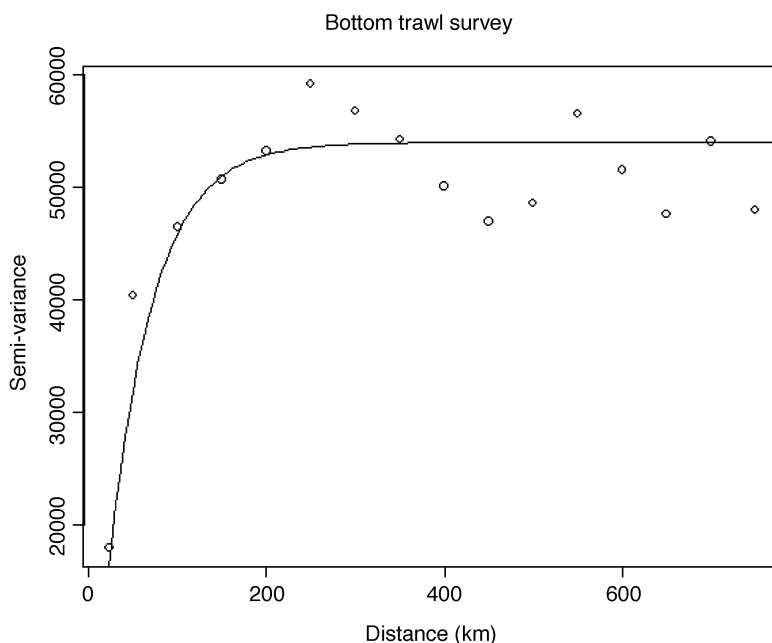


Figure 10. Semi-variogram showing the spatial covariance in relative abundance between sampling sites for the bottom trawl survey, as a function of between-site distance. The bottom trawl survey has relatively large minimum between-site distances and shows substantial covariance only at the smallest between-site distance (25.9 km).

The bottom trawl survey calculates a survey-wide index of relative abundance for walleye pollock using standard expressions for stratified random designed surveys. This is true despite the fixed, systematic placement of the sampling sites within the strata because it is believed that year-to-year variation in the spatial distribution of the primary species of interest provides sufficient randomization (Warren, 1994). The stratified design did not decrease the overall variance estimated for walleye pollock, although it did account for the geographic size and age composition of this population. The stratification that was used is also appropriate for some of the other species in the survey that are selective regarding area and depth. Analyses based on generalized additive modeling of relative abundance data considering depth as a covariate has been applied to the 2000 bottom trawl data for walleye pollock, but no significant decrease in variance, relative to that obtained with stratified sampling design, was achieved. This is likely true because the stratification scheme was also based, in part, on depth, thus the effect of depth variation on the variance of mean relative abundance is partially taken into consideration by the stratification scheme. For 2000, the bottom trawl estimate of pollock abundance within 2 m of the bottom was 5.13 million t with a 95% confidence interval of ± 1.17 million t.

The hydroacoustic survey, by comparison, calculates a survey-wide index of relative abundance of pollock using a one-dimensional geostatistical procedure (Williamson and Traynor, 1996) which accounts for the spatial correlation in the samples. For 2000, the hydroacoustic estimate of pollock abundance between 14 m from the surface and 2 m of the bottom was 3.05 million t with a 95% confidence interval of ± 0.18 million t. The much greater precision of the hydroacoustic estimates is due to the greater number of sample

observations, and the use of kriging estimation variance. In 2000, the bottom trawl survey had 352 sample observations, while the hydroacoustic survey had, without correction for spatial correlation, 8937 sample observations.

Acknowledgments

The authors are pleased to acknowledge the reviews of Ms. Susan J. Picquelle and Dr. Stephen E. Syrjala of the Alaska Fisheries Science Center, and Dr. Donald R. Gunderson of the University of Washington. The authors are also in debt to two anonymous reviewers for sharing their insights and pointing out errors on many of the topics presented here. We thank all of these reviewers for providing many constructive comments that lead to significant improvements. We thank Dr. Asgeir Aglen of the Institute of Marine Research, Bergen, Norway, for permission to reproduce a figure from his doctoral dissertation. The errors that remain are of course the responsibility of the authors.

References

- Adlerstein, S., and S. Ehrich. Effect of deviations from target speed and of time of day on catch rates of some abundant species under North Sea international bottom trawl survey protocol conditions. *ICES Journal of Marine Science*, **59**: 594–603 (2002).
- Aglen, A. Reliability of acoustic fish abundance estimates. Doctorate of Science thesis, Department of Fisheries Biology, University of Bergen, Norway (1989).
- Anscombe, F. J. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**: 358–382 (1950).
- Aitchison, J., and J. A. C. Brown. *The Lognormal Distribution*. Cambridge University Press (1957).
- Bannerot, S. P., and C. B. Austin. Using frequency distributions of catch per unit effort to measure fish–stock abundance. *Transactions of the American Fisheries Society*, **112**: 608–617 (1983).
- Bergh, M. O., E. K. Pikitch, J. R. Skalski, and J. R. Wallace. The statistical design of comparative fishing experiments. *Fisheries Research*, **9**: 143–163 (1990).
- Bissell, A. F. A negative binomial model with varying element sizes. *Biometrika*, **59**: 435–441 (1972).
- Booth, J. G., R. W. Butler, and P. Hall. Bootstrap methods for finite populations. *Journal of the American Statistical Association*, **89**: 1282–1289 (1994).
- Brewer, K. W. R., M. Hanif, and S. M. Tam. How nearly can model-based and design-based estimation be reconciled? *Journal of the American Statistical Association*, **83**: 128–132 (1988).
- Chambers, J. M., and T. J. Hastie. *Statistical Models*. S. Wadsworth & Brooks, CA (1992).
- Cochran, W. G. *Sampling Techniques*, Third Ed. New York: John Wiley & Sons (1977).
- Cressie, N. Geostatistics. *The American Statistician*, **43**: 197–202 (1989).
- Cressie, N. A. C. *Statistics for Spatial Data*. John Wiley and Sons (1993).
- Davison, A. C., and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press (1997).
- Deriso, R. B., T. J. Quinn, and P. R. Neal. Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**: 815–824. (1985).
- Doubleday, W. G. A least squares approach to analysing catch at age data. *Research Bulletin of the International Commission for Northwest Atlantic Fisheries*, **12**: 69–81 (1976).
- Efron, B. The jackknife, the bootstrap, and other resampling plans. *Soc. Ind. Appl. Math. CBMS National Science Foundation Monograph*, 38 (1982).
- Efron, B., and R. J. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with Discussion). *Statistical Science*, **1**: 54–96 (1986).
- Everson, I., M. Bravington, and C. Goss. A combined acoustic and trawl survey for efficiently estimating fish abundance. *Fisheries Research*, **26**: 75–91 (1996).

- Finney, D. J. On the distribution of a variate whose logarithm is normally distributed. *Journal of the Royal Statistical Society Suppl.*, **7**: 155–161 (1941).
- Firth, D. Multiplicative errors: log-normal or gamma? *Journal of the Royal Society B*, **50**: 266–268 (1988).
- Fournier, D. A., and C. P. Archibald. A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**: 1195–1207 (1982).
- Francis, R. I. C. C. An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research*, **18**: 59–71 (1984).
- Francis, R. I. C. C. Two acoustic surveys of pelagic fish in Hawke Bay, New Zealand, 1980. *New Zealand Journal Marine and Freshwater Research*, **18**: 59–71 (1985).
- Gavaris, S. Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**: 2272–2275 (1980).
- Gunderson, D. R. *Surveys of Fishery Resources*. New York: John Wiley and Sons (1993).
- Hanselman, D. H., T. J. Quinn, C. Lunsford, J. Heifetz, and D. Clausen. Applications in adaptive cluster sampling of Gulf of Alaska rockfish. *Fishery Bulletin*, **101**: 501–513 (2003).
- Jolly, G. M., and I. Hampton. A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**: 1282–1291 (1990).
- Kappenman, R. F. Trawl survey based abundance estimation using data sets with unusually large catches. *ICES Journal of Marine Science*, **56**: 28–35 (1999).
- Kimura, D. K. Standardized measures of relative abundance based on modelling log (c.p.u.e.) and their application to Pacific ocean perch (*Sebastes alutus*). *Journal du Conseil*, **39**: 211–218 (1981).
- Kimura, D. K., and J. W. Balsiger. Bootstrap methods for evaluating sablefish pot index surveys. *North American Journal Fisheries Management*, **5**: 47–56 (1985).
- Kimura, D. K., and N. A. Lemberg. Variability of line intercept density estimation (a simulation study of the variance of hydroacoustic biomass estimates). *Canadian Journal of Fisheries and Aquatic Sciences*, **38**: 1141–1152 (1981).
- Kimura, D. K., and H. H. Zenger. Standardizing sablefish (*Anoplopoma fimbria*) long-line survey abundance indices by modeling the log-ratio of paired comparative fishing cpues. *ICES Journal of Marine Science*, **54**: 48–59 (1997).
- Lai, H. L., and D. K. Kimura. Analyzing survey experiments having spatial variability with an application to a sea scallop fishing experiment. *Fisheries Research*, **56**: 239–259 (2002).
- Lenarz, W. H., and P. B. Adams. Some statistical considerations of the design of trawl surveys for rockfish (Scorpaenidae). *Fishery Bulletin*, **78**: 659–674 (1980).
- Lo, N. C. H., D. Griffith, and J. R. Hunter. Using a restricted adaptive cluster sampling to estimate Pacific hake larval abundance. *California Cooperative Oceanic Fisheries Investigations Reports*, **38**: 103–113 (1997).
- MacLennan, D. N., and E. J. Simmonds. *Fisheries Acoustics*. London: Chapman and Hall, (1992).
- Maunder, M. N., and A. E. Punt. Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, **70**: 141–159 (2004).
- McConnaughey, R. A., and L. L. Conquest. Trawl survey estimation using a comparative approach based on lognormal theory. *Fishery Bulletin*, **91**: 107–118 (1992).
- McCullagh, P., and J. A. Nelder. *Generalized Linear Models*. London: Chapman & Hall (1989).
- Methot, R. D. Synthetic estimates of historic abundance and mortality for northern anchovy, (*Engraulis mordax*). Southwest Fisheries Center Administrative Report: LJ 86–29 85 p. (1986).
- Morrison, D. F. *Multivariate Statistical Methods*. McGraw Hill 338 p (1967).
- Munro, P. T. A decision rule based on the mean square error for correcting relative fishing power differences in trawl survey data. *Fishery Bulletin* **96**: 538–546 (1998).
- Myers, R. A., and P. Pepin. The robustness of lognormal-based estimators of abundance. *Biometrics*, **46**: 1185–1192 (1990).
- Pelletier, D. Intercalibration of research survey vessels in fisheries: A review and an application. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**: 2672–2690 (1998).

- Pelletier, D. and A. M. Parma. Spatial distribution of Pacific halibut (*Hippoglossus stenolepis*): An application of geostatistics to longline survey data. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**: 1506–1518 (1994).
- Pennington, M. Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, **39**: 281–286 (1983).
- Pennington, M. Some statistical techniques for estimating abundance indices from trawl surveys. *Fishery Bulletin*, **84**: 519–525 (1986).
- Pennington, M., and J. H. Vølstad. Optimum size of sampling unit for estimating the density of marine populations. *Biometrics*, **47**: 717–723 (1991).
- Pennington, M. R., and M. D. Grosslein. Accuracy of abundance indices based on stratified random trawl surveys. ICES CM 1978/D:13, 35p. (1978).
- Petitgas, P. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. *Fish and Fisheries*, **2**: 231–249 (2001).
- Power, J. H., and E. B. Moser. Linear model analysis of net catch data using the negative binomial distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**: 191–200 (1999).
- Quinn, T. J., and R. B. Deriso. Quantitative Fish Dynamics. Oxford University Press, 542 p. (1999).
- Rao, J. N. K., and C. F. J. Wu. Resampling inference with complex survey data. *Journal of the American Statistical Association*, **83**: 231–241 (1988).
- Rivoirard, J., K. Simmonds, P. Foote, and N. Bez. Geostatistics for Estimating Fish Abundance. Iowa State Press (Blackwell) 206p. (2000).
- Rivoirard, J., and K. Wieland. Correcting daylight effect in the estimation of fish abundance using kriging with external drift, with an application to juvenile haddock in North Sea. *ICES Journal of Marine Science*, **58**: 1272–1285 (2001).
- Robothom, H., and J. Castillo. The bootstrap method: An alternative for estimating confidence intervals of resources surveyed by hydroacoustic techniques. *Rapp. P.-v. Reun. Cons. Int. Explor. Mer*, **189**: 421–424 (1990).
- Robson, D. S. Estimating the relative fishing power of individual ships. *Research Bulletin for the International Commission for Northwest Atlantic Fisheries*, **3**: 5–14 (1966).
- Seber, G. A. F. *The Estimation of Animal Abundance and Related Parameters*, Second Ed. New York: Macmillan (1982).
- Sigler, M. F., and J. T. Fujioka. Evaluation of variability in sablefish, *Anoplopoma fimbria*, abundance indices in the Gulf of Alaska using the bootstrap method. *Fishery Bulletin*, **86**: 445–452 (1988).
- Simmonds, E. J., and R. J. Fryer. Which are better, random or systematic acoustic surveys? A simulation using North Sea herring as an example. *ICES Journal of Marine Science*, **53**: 39–50 (1996).
- Smith, S. J. Evaluating the efficiency of the delta-distribution mean estimator. *Biometrics*, **44**: 485–493 (1988).
- Smith, S. J. Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**: 894–903 (1990).
- Smith S. J., and S. Gavaris. Improving the precision of abundance estimates of eastern Scotian shelf Atlantic cod from bottom trawl surveys. *North American Journal of Fisheries Management*, **13**: 35–47 (1993).
- Smith, S. J. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**: 616–630 (1997).
- Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. Ames, IA: The Iowa State University Press (1967).
- Somerton, D., J. Ianelli, S. Walsh, S. Smith, O. R. Godo, and D. Ramm. Incorporating experimentally derived estimates of survey trawl efficiency into the stock assessment process: A discussion. *ICES Journal of Marine Science*, **56**: 299–302 (1999).
- Stefansson, G. Analysis of groundfish survey abundance data: Combining the GLM and delta approaches. *ICES Journal of Marine Science*, **53**: 577–588 (1996).
- Sullivan, P. J. Stock abundance estimation using depth-dependent trends and spatially correlated variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **48**: 1691–1703 (1991).

- Swain, D. P., G. A. Poirier, and A. F. Sinclair. Effect of water temperature on catchability of Atlantic cod (*Gadus morhua*) to the bottom-trawl survey in the southern Gulf of St. Lawrence. *ICES Journal of Marine Science*, **57**: 56–68 (2000).
- Swartzman, G., C. Huang, and S. Kaluzny. Spatial analysis of Bering Sea groundfish survey data using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**: 1366–1378 (1992).
- Syrjala, S. E. Critique on the use of the delta distribution for the analysis of trawl survey data. *ICES Journal of Marine Science*, **57**: 831–842 (2000).
- Taylor, C. C. Nature of variability in trawl catches. *Fishery Bulletin of the U.S. Fish and Wildlife Service*, **54**: 145–166 (1953).
- Thompson, S. K. *Sampling*, Second Ed. New York: John Wiley & Sons (2002).
- Venables, W. N., and C. M. Dichmont. GLMs, GAMs and GLMMs: An overview of theory for applications in fisheries research. *Fisheries Research*, **70**: 319–337 (2004).
- Warren, W. G. The potential of sampling with partial replacement for fisheries surveys. *ICES Journal of Marine Science*, **51**: 315–324 (1994).
- Wilderbuer, T. K., R. F. Kappenman, and D. R. Gunderson. Analysis of fishing power correction factor estimates from a trawl comparison experiment. *North American Journal of Fisheries Management*, **18**: 11–18 (1998).
- Williamson, N. J. Cluster sampling of the variance of abundance estimates derived from quantitative echo sounder surveys. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**: 229–231 (1982).
- Williamson, N. J., and J. J. Traynor. Application of a one-dimensional geostatistical procedure to fisheries acoustic surveys of Alaskan pollock. *ICES Journal of Marine Science*, **53**: 423–428 (1996).
- Wilson, C. D., A. B. Hollowed, M. Shima, P. Walline, and S. Stienessen. Interactions between commercial fishing and walleye pollock. *Alaska Fishery Research Bulletin*, **10**: 61–77 (2003).
- Zhou, X. H., and S. Gao. Confidence intervals for the log-normal mean. *Statistics in Medicine*, **16**: 783–790 (1997).
- Zimmerman, D. L., and D. A. Harville. A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics*, **47**: 223–239 (1991).