THE PRECISION OF SIMULATED TRANSECT SURVEYS OF NORTHERN ANCHOVY, *ENGRAULIS MORDAX*, SCHOOL GROUPS

PAUL C. FIEDLER[1]

ABSTRACT

Simulated transect surveys of model anchovy populations were compared in terms of precision and efficiency. The precision of systematic surveys varies inversely with the distance between transects. Systematic surveys give more precise population estimates than random surveys, due to the large positive correlation between closely spaced transects. The precision of stratified systematic surveys is not significantly different from that of the unstratified surveys when the school groups are randomly distributed in the survey area. However, stratified systematic surveys are more precise when the school groups are clumped in one end of the survey area. The results of the simulations show that the patchy distribution of anchovy schools can be a major source of error in population estimates.

Any sampling program intended to estimate the size of a population is subject to a variety of errors which may reduce the accuracy or precision of the estimate. Precision is the reciprocal of the variation of replicate estimates. Successful management of a northern anchovy fishery in California will require the monitoring of changes in the population size. Acoustic survey techniques are currently being developed to obtain population and biomass estimates independent of the fishery (Hewitt et al. 1976). As in the study of any biological population, it will be important to avoid confusing the variation of a series of estimates due to sampling error with true fluctuations in the population size.

Precision may be affected by 1) the manner in which the sampled population is distributed in space, and 2) variations within the sampling method itself. Several studies have shown that the patchy distribution of individuals in a population may cause considerable variation in replicate population estimates and that the variation is related to sample design. Winsor and Clarke (1940) studied the variation of catches in series of plankton net tows. Although they did not separate the components of between-tow variation due to factors (1) and (2), it was observed that oblique tows were more precise than vertical or horizontal tows. Barnes and Marshall (1951) took an extensive series of replicate pump samples and attributed the considerable variation observed to the nonrandom distribution of the zooplankton since the volumes filtered were known accurately. Taft (1960) analyzed the variance of sardine egg counts in a grid of closely spaced stations. The distribution of eggs was extremely patchy (the densities between samples ranged over more than four orders of magnitude) and the relative 95% confidence limits for an estimate of the egg population in the area of the grid from a single sample were represented by a factor of 62. A simulation study by Wiebe (1971) showed that the precision of zooplankton population estimates depends both on the sampling design (net size and tow length) and the distribution of the population (size and location of patches).

Similar studies have investigated the precision of sampling fish populations. Taylor (1953) discussed the implications of the patchy distribution of fish for the optimum design of trawl surveys to estimate population size. Cram and Hampton (1976) demonstrated that the patchy distribution of pilchard schools can cause imprecision sufficient to render a population estimate useless for management.

The anchovy population is patchy on two levels: individual fish are aggregated in schools and schools themselves tend to be aggregated in school groups. This patchiness, or nonrandomness, is expected to be a major source of variation in population estimates. The present study simulated surveys of model anchovy populations to determine the effect of patchiness on the precision of population estimates. Three transect survey designs were compared: systematic, random, and stratified systematic. These are merely different

[1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093.

Manuscript accepted January 1978.
FISHERY BULLETIN: VOL. 76, NO. 3, 1978.

methods of selecting transects, or allocating sampling effort. The three types of simulated survey, with a range of sample sizes (numbers of transects), were run on 15 model anchovy populations.

METHODS

Anchovy populations were modeled as arrays with each element representing 1 n.mi.[2]. The array dimensions were 180 × 75, approximately the dimensions, in miles, of the Los Angeles Bight. Since a school is the population unit detected in an acoustic or aerial survey, the units of the model populations were schools. One hundred fifty thousand schools were distributed in the array resulting in a mean density of 11.1 schools/mi^2. Four acoustic surveys by the California Department of Fish and Game[2] in 1975 and 1976 yielded estimates ranging from 88,887 to 319,878 anchovy schools off southern California in the area of the bight. Mais (1974) gave a range of 21,920-343,070 ($\bar{x}$ = 150,996) schools off southern California and northern Baja California, most of which were within the bight.

The schools were placed in circular school groups located at random. Schools were distributed uniformly within a school group. School group radii and densities were chosen randomly and independently from log-normal approximations of observed frequency distributions based on 52 school groups from six California Department of Fish and Game Sea Survey acoustic surveys (MacCall et al.[3]) (Figure 1). There was no significant correlation between the density of schools within a school group and the size of the school group in these observations. Where school groups overlapped, the densities were simply added together, although this effectively increased both the mean radius and density. In one model population illustrated in Figure 2, 16 school groups containing 150,303 schools covered about 14% of the survey area. Fifteen model populations were used, each with the same total number of schools, but different locations, sizes, and densities of school groups.

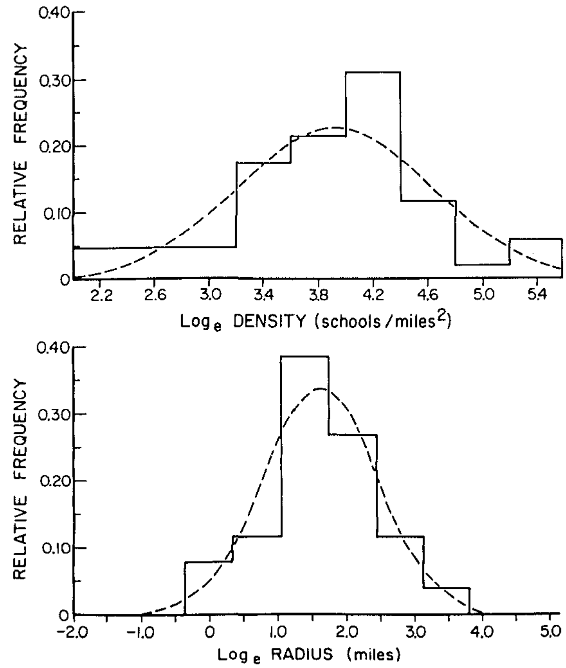

FIGURE 1.—Comparison of distributions of northern anchovy school group density and size observed in the California Current during California Department Fish and Game surveys (solid lines) to log-normal approximations used in simulations (dashed lines).

A simulated survey consisted of a series of transects across the survey area. There were 180 possible transects, each 1 mi wide. Acoustic surveys currently run by the Southwest Fisheries Center, National Marine Fisheries Service, NOAA, used a transect width of 0.14 mi (250 m). Aerial transect widths were typically 0.2 to 0.5 mi. A larger transect width was used in the simulations to hold the model population array down to a reasonable size. We assumed that the general results of the simulations would not change by using a smaller transect width. Since all schools were counted within a transect, the only source of error in the survey estimate was the large variance in the number of schools per transect. For instance, in the model population in Figure 2, the mean number of schools per transect was 835.0, while SD was 920.4 (variance = 8.47 $\times 10^5$).

Systematic surveys were simulated by counting the schools within a series of transects separated by a constant transect interval. A population estimate was calculated simply by dividing the survey count by the fraction of the survey area covered by the transects. Transect intervals of 2, 3,

[2]S. J. Crooke. 1975. Cruise reports 75-A-1 and 75-A-6. K. F. Mais. 1976. Cruise reports 76-A-3 and 76-A-9. State of California - The Resources Agency, Department of Fish and Game, Marine Resources Region, Long Beach, CA 90802.

[3]MacCall, A., P. E. Smith, G. Stauffer, J. Squire, J. Zweifel, and S. Crooke. Report of CalCOFI anchovy workshop working group on methods of estimating anchovy abundance. Unpubl. manuscr. Southwest Fisheries Center, National Marine Fisheries Service, NOAA, La Jolla, CA 92038.

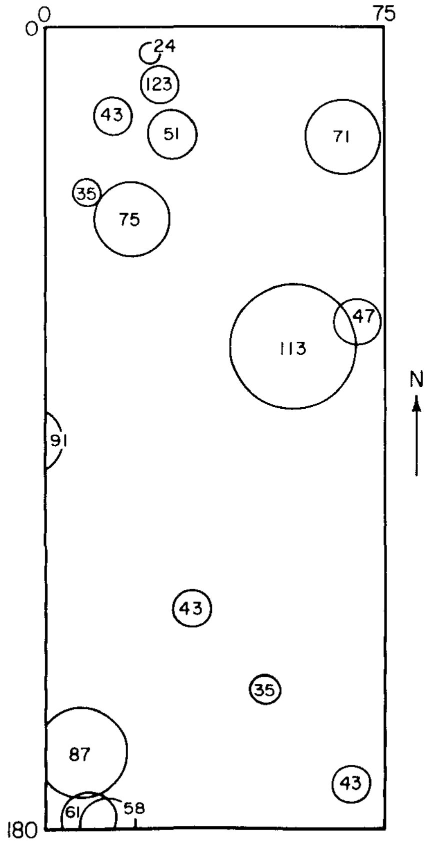

FIGURE 2.—A model northern anchovy population. Densities of school groups in schools per square mile. Simulated survey transects are oriented horizontally. The numbers on the axes are the coordinates of the array and the dimensions, in miles, of the survey area it represents.

4, 5, 7, 10, 12, 15, 20, 25, 30, and 40 mi were used. A survey with a transect interval of d miles consisted of $180/d$ transects. For each transect interval, 20 replicate surveys were run by randomly choosing the initial transect from the first d transects in the survey area. The replicate survey estimates were used to calculate an unbiased mean population estimate and a coefficient of variation (standard deviation of the replicate estimates divided by the mean), which is a measure of the precision of the estimate (Wiebe 1971). This procedure was repeated on the 15 different model populations.

Random surveys were simluated in an analogous manner to allow a direct comparison of sampling errors. For each of the transect intervals (d) of the systematic surveys, 20 replicate surveys were run consisting of $180/d$ transects chosen at random without replacement (a transect was not repeated within a survey). Coefficients of variation were calculated as a measure of precision.

Stratified systematic surveys were simulated after dividing each of the model populations into four 45-mi wide strata. The schools along three transects in each of the four strata were counted to obtain a preliminary estimate of relative population sizes. Then a total of 60, 36, 18, 12, 9, 7, or 6 transects were divided among the strata according to the estimated population fractions. For example, if a stratum contained one-half of the schools counted in the preliminary survey, one-half of the total number of transects was allocated to that stratum for the stratified survey. At least one transect was allocated to each stratum to avoid biasing the final population estimate. For each total number of transects, 20 replicate systematic surveys were run by randomly choosing the initial transect and simulating a systematic survey within each stratum with the allocated number of transects. Once again, coefficients of variation of the replicate population estimates were calculated for each of the 15 model populations.

RESULTS

The results of the systematic survey simulations indicate that the sampling error represented

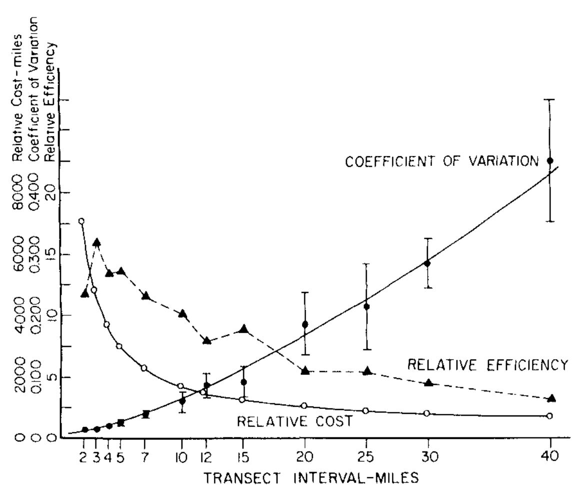

FIGURE 3.—Results of the simulations of systematic surveys of the 15 models for northern anchovy populations. Relative efficiency is proportional to precision (the reciprocal of the coefficient of variation) divided by relative cost (see text). Averages and 95% confidence limits are given for coefficients of variation.

by the coefficient of variation increased as the transect interval increased and sample size decreased (Figure 3). The cost of a survey was assumed to be proportional to the distance covered along the transects plus 360 mi to and from port. Relative efficiency was 10^3 times the reciprocal of the product of the coefficient of variation (C.V.) and relative cost, i.e., precision divided by cost. Efficiency generally decreased as the transect interval increased, but peak efficiency was obtained at a transect interval of 3 mi. By interpolation, it can be seen that a population estimate may range 10 and 25% (2 × C.V.) from the true population size when surveys are run with transect intervals of 8.5 and 16 mi, respectively.

Systematic sampling gave a consistently lower coefficient of variation, or greater precision than random sampling (Figure 4). The variability between model populations, indicated by the confidence limits on the mean coefficient of variation, was greater for the random sampling error ($F_{(14,14)} \geq 5.44, P<0.05/12$) for 8 of the 12 sample sizes. Also represented in Figure 4 are the expected coefficients of variation for random sampling calculated from the model population parameters (σ^2 and μ) by the following equation with a finite population correction:

$$\text{C.V.} = \frac{1}{\mu}\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N}\right)}$$

where σ^2 = the average variance of the number of schools per transect in the 15 model populations = 1,154,636

μ = the mean number of schools per transect = 835.3

n = the number of transects in the survey

N = total number of transects in the survey area = 180.

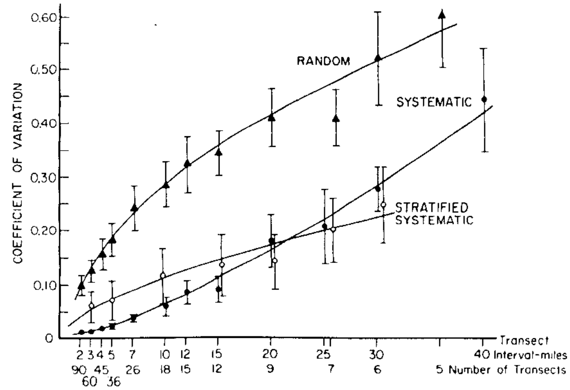

FIGURE 4.—Comparison of the results of the simulations of random, systematic, and stratified systematic surveys. For random surveys, the curve represents the expected coefficients of variation calculated from the parametric variance of the model populations (see text).

For 11 of the 12 sample sizes, the expected value was within 95% confidence limits of the mean observed coefficient of variation. This close agreement supports the validity of the method used to obtain the coefficients of variation in the simulations. There was apparently no significant difference between the coefficients of variation of the systematic and stratified systematic surveys (Figure 4). This was confirmed by analysis of variance (Table 1, $P>0.25$ that there was no added variance due to survey design). However, there were significant interaction effects between survey design and model population ($P<0.01$ that there was no added variance from this source) and between survey design and the number of transects ($P<0.05$).

An attempt was made to elucidate the interactions involving survey design by performing analyses of variance on subsets of the data. It was found that for large sample sizes (transect interval ≤15 mi, or number of transects ≥12) unstratified

TABLE 1.—Analysis of variance of the coefficients of variation from simulated systematic and stratified systematic surveys of model northern anchovy populations. This is a mixed model analysis of variance (Sokal and Rohlf 1969): survey design and number of transects are fixed treatment effects and model population is a random effect.

Source of variation	SS	df	MS	Significance level
Main effects:				
Survey design	0.0157	1	0.0157	$F_{(1,14)} = 0.908, P>0.25$
Number of transects	1.2872	6	0.2145	$F_{(6,84)} = 56.681, P<<0.001$
Model population	0.5270	14	0.0376	$F_{(14,84)} = 8.000, P<0.001$
Interactions:				
Design-number of transects	0.0748	6	0.0125	$F_{(6,84)} = 2.627, P<0.05$
Design-population	0.2418	14	0.0173	$F_{(14,84)} = 3.681, P<0.001$
Number of transects-population	0.3179	84	0.0038	$F_{(84,84)} = 0.819, P>0.50$
Error	0.3988	84	0.0047	
Total	2.8633	209		

systematic surveys were significantly more precise than stratified systematic surveys ($P<0.025$), although there is still a significant interaction between survey design and model population ($P<0.001$). For smaller sample sizes, there was no significant difference between the precision of the two designs.

In the model populations, school groups were located randomly within the survey area. However, the distribution of schools between strata was never random because of the wide range of school group sizes and the small number of school groups in a population. The 15 model populations were divided into three groups (low, intermediate, and high nonrandomness) based on the index of dispersion of the number of schools per stratum. Analysis of variance revealed that for highly nonrandom populations, stratified systematic surveys were significantly more precise than unstratified surveys ($P<0.025$). On the other hand, there was no significant difference between the survey designs for populations of intermediate or low nonrandomness. The effect of the nonrandomness of the populations, in the limited sense used here, is illustrated more dramatically below.

In summary, these results indicate that both the number of transects and the spatial distribution of the population can affect the precision of a survey estimate. The effect of survey design involves complex interactions with the other two factors. These factors should be considered, if possible, when choosing the optimum design for a survey.

DISCUSSION

In general, systematic sampling may result in considerable gains or losses in precision compared with simple random sampling. The greatest increase in precision occurs when there is a high degree of correlation between adjacent sampling units and the correlation decreases as the interval between units increases. In this situation, systematic sampling resembles stratified sampling. On the other hand, precision may be greatly reduced when there is a periodic variation in the population and the sampling interval is equal to this period or a multiple of it (Hansen et al. 1953).

Correlograms between sampling units (transects) in five of the model anchovy populations indicated that transects <10 mi apart had a high positive correlation, while the correlation tended to be slightly negative at distances >20 mi (Figure 5). This autocorrelation structure was due to the frequency distribution of school group sizes. The mean distance at which the autocorrelation function passed through zero was 15.0 mi, while the mean diameter of the individual school groups in the five model populations was 11.8 mi. Distribution of school groups within the model populations was random. However, real populations are likely to be nonrandom in this respect and additional correlations would be expected from this factor. The strong positive correlation between transects separated by short distances explains why systematic surveys with small transect intervals were more precise than random surveys with an equivalent number of transects. As the transect interval increased, the correlation between transects decreased to near zero and the imprecision of systematic sampling approached that of random sampling (Figure 4).

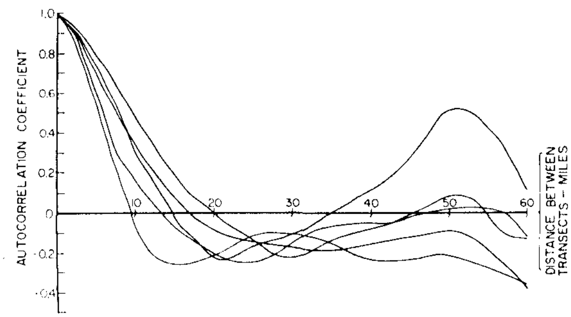

FIGURE 5.—Autocorrelation of transect counts in five model northern anchovy populations.

In order to reduce total sampling error, a common strategy is to allocate effort proportional to the sampling error within parts of a sampling program. The variation observed in the population estimates of the simulated surveys was caused by the large variance in the number of schools per transect. It can be shown in the model populations, as in many biological populations, that the standard deviation was positively correlated with the mean number of schools per transect in a stratum. Therefore, it was thought that the stratified systematic surveys would reduce the total sampling error by allocating more transects where the variance was large. The simulations failed to show any gains in precision from this strategy. This result was not expected, but is possibly due to the random distribution of school groups. The model populations may have been ideal in this sense, but we had relatively little information on the distribution of school groups within the range of the northern anchovy. As stated above, stratified systematic surveys were significantly more precise than unstratified surveys for the five model populations with the most nonrandom distribution of schools between strata. If the school groups themselves are aggregated, it is reasonable to expect an increase in precision by stratifying the survey.

To test this possibility, the simulations were repeated on model populations in which the school groups were limited to only one-half of the survey area. An analysis of variance (Table 2) indicated in this case that the stratified systematic surveys were more precise than the unstratified systematic surveys ($P<0.005$ that there was no added variance due to survey design). The overall mean coefficients of variation were 0.095 and 0.133, respectively. However, there were significant interaction effects involving survey design, indicating that the advantage of stratifying the survey will depend on the number of transects and the spatial distribution of the population. The additional cost of the preliminary survey in the stratified design must also be considered when comparing it with the unstratified design.

The results of the simulated systematic surveys showed that the patchy distribution of schools was an important source of error in estimates of the anchovy population size. Acoustic surveys run by the Southwest Fisheries Center have used transect intervals of 6.6 and 40 mi. The simulations gave evidence that the population estimates from these surveys could be expected to range at least 8 and 90% (2 × C.V.), respectively, from the true population size. The most efficient simulated sampling, in terms of precision per unit cost, occurred at a transect interval of 3 mi. This would require a cruise grid of 4,860 mi, equivalent to a 34-day acoustic survey at 12 kn and 12 h per day, to reduce the coefficient of variation (due to the patchy distribution of schools) to 1.4%. Maximizing efficiency is not a valid goal, however, when the precision gained is greater than that required for the problem of managing the fishery, when other sources of error become more important, and when there are absolute limits on cost. Anchovy population estimates within 25% of the true value might be considered sufficient for management, at least to allow confidence that a consistent change observed over several years is real (pers. commun., P. E. Smith, Southwest Fisheries Center, National Marine Fisheries Service, NOAA, La Jolla, Calif., Oct. 1977).

As stated before, the anchovy population is patchy on two levels: individuals are aggregated into schools and schools are aggregated into school groups. The simulations have quantified the sampling error due to the second level of patchiness only. Although little is known about the distribution of anchovy school groups, it was also demonstrated that their aggregation is potentially an important consideration in designing a survey. The acoustic survey methods currently used by the National Marine Fisheries Service and the California Department of Fish and Game do little more than count the number of anchovy schools (Hewitt et al. 1976; Mais 1974). The Department of Fish and Game calculates a biomass estimate by multiplying the observed school area by a constant factor thought to represent an average biomass per unit area. More sophisticated methods of estimating biomass from the acoustic signal received from a school are now being explored at the Southwest Fisheries Center. For these reasons, the problem of sampling error due to a varying number of fish per school (the first level of patchiness) was not addressed here.

Many sources of error may be involved in an anchovy biomass estimate. Patchiness is important in any type of sampling program. Other sources of error that may be important in an

TABLE 2.—Analysis of variance of the coefficients of variation from simulated systematic and stratified systematic surveys of model northern anchovy populations when the model population school groups are clumped in one-half of the survey area.

Source of variation	SS	df	MS	Significance level
Main effects:				
Survey design	0.0732	1	0.0732	$F_{(1,14)} = 12.908, P<0.005$
Number of transects	1.4995	6	0.2499	$F_{(6,84)} = 119.403, P<<0.001$
Model population	0.1327	14	0.0095	$F_{(14,84)} = 4.318, P<0.001$
Interactions:				
Design-number of transects	0.0476	6	0.0079	$F_{(6,84)} = 3.589, P<0.005$
Design-population	0.0794	14	0.0057	$F_{(14,84)} = 2.591, P<0.01$
Number of transects-population	0.1758	84	0.0021	$F_{(84,84)} = 0.955, P>0.50$
Error	0.1858	84	0.0022	
Total	2.1941	209		

acoustic survey are as follows (Instituto del Mar del Peru 1974; Cram and Hampton 1976; P. E. Smith pers. commun.):

1) Failure to discriminate between anchovy schools and other acoustic targets.
2) Unschooled fish and small schools not detected.
3) Vessel avoidance.
4) Inability to survey in shallow inshore waters.
5) Movement of school groups relative to the survey grid.
6) Fish in the top surface layer missed by the acoustic beam.
7) Errors in the factor for conversion of the acoustic signal information to a biomass estimate.
8) Effect of varying hydrographic conditions on the acoustic signal.
9) Blocking of signal to and from fish far from the ship by fish nearer to the ship.

The magnitude of the error caused by these factors can now only be roughly estimated. They may affect either or both the precision and accuracy of a population estimate. Corrections to reduce the biases are conceivable. The present study has demonstrated the magnitude of the error associated with the patchiness of the anchovy population. Although the model population distributions may be crude approximations to the real distribution, the general conclusions reached here are not likely to be changed by adding further levels of complexity to the model. The sampling error due to patchiness can be reduced by properly designing a survey, but never eliminated. Temporal and spatial differences in population estimates must be interpreted with an awareness that the error exists.

ACKNOWLEDGMENTS

I thank Paul E. Smith for his guidance in completing this project, as well as Alec McCall, Keith Parker, Jim Zweifel, and Roger Hewitt for valuable discussions. Ian Hampton and two anonymous reviewers provided helpful comments on earlier drafts of the manuscript. This work was carried out under a State of California Marine Research Committee contract at the Southwest Fisheries Center, La Jolla, Calif.

LITERATURE CITED

BARNES, H., AND S. M. MARSHALL.
1951. On the variability of replicate plankton samples and some applications of "contagious" series to the statistical distribution of catches over restricted periods. J. Mar. Biol. Assoc. U.K. 30:233-263.

CRAM, D. L., AND I. HAMPTON.
1976. A proposed aerial/acoustic stragety for pelagic fish stock assessment. J. Cons. 37:91-97.

HANSEN, M. H., W. N. HURWITZ, AND W. G. MADOW.
1953. Sample survey methods and theory, Vol. 1. John Wiley and Sons, N.Y., 638 p.

HEWITT, R. P., P. E. SMITH, AND J. C. BROWN.
1976. Development and use of sonar mapping for pelagic stock assessment in the California Current area. Fish. Bull., U.S. 74:281-300.

INSTITUTO DEL MAR DEL PERU.
1974. Report of the fourth session of the panel of experts on stock assessment on Peruvian anchoveta. Bol. Inst. Mar Peru (Callao) 2:605-723.

MAIS, K. F.
1974. Pelagic fish surveys in the California Current. Calif. Dep. Fish Game, Fish Bull. 162, 79 p.

SOKAL, R. R., AND F. J. ROHLF.
1969. Biometry. W.H. Freeman and Co., San Franc., 776 p.

TAFT B. A.
1960. A statistical study of the estimation of abundance of sardine (Sardinops caerulea) eggs. Limnol. Oceanogr. 5:245-264.

TAYLOR, C. C.
1953. Nature of variability in trawl catches. U.S. Fish Wildl. Serv., Fish. Bull. 54:145-166.

WIEBE, P. H.
1971. A computer model study of zooplankton patchiness and its effects on sampling error. Limnol. Oceanogr. 16:29-38.

WINSOR, C. P., AND G. C. CLARKE.
1940. A statistical study of variation in the catch of plankton nets. J. Mar. Res. 3:1-34.