Not to be cited without permission of the authors[1]

Ne pas citer sans autorisation des auteurs[1]

Canadian Atlantic Fisheries Scientific Advisory Committee

Comité scientifique consultatif des pêches canadiennes dans l'Atlantique

CAFSAC Research Document 85/46

CSCPCA Document de recherche 85/46

TWO PHASE SAMPLING OF THE HERRING FISHERY IN NAFO DIVISION 4T

by

Martin Ahrens
Marine Fisheries Division
Department of Fisheries and Oceans
Fisheries Research Branch, Gulf Region
P.O. Box 5030
Moncton, New Brunswick E1C 9B6

[1] This series documents the scientific basis for fisheries management advice in Atlantic Canada. As such, it addresses the issues of the day in the time frames required and the Research Documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research Documents are produced in the official language in which they are provided to the Secretariat by the author.

[1] Cette série documente les bases scientifiques des conseils de gestion des pêches sur la côte atlantique du Canada. Comme telle, elle couvre les problèmes actuels selon les échéanciers voulus et les Documents de recherche qu'elle contient ne doivent pas être considérés comme des énoncés finals sur les sujets traités mais plutôt comme des rapports d'étape sur les études en cours.

Les Documents de recherche sont publiés dans la langue officielle utilisée par les auteurs dans le manuscrit envoyé au secrétariat.

ABSTRACT

Two phase sampling and one phase random sampling of a herring fishery are compared. Equations for mean weights, proportions at age, total catch, and catch at age and their variances are presented. The results of a demonstration sampling program show that neither scheme is intrinsically biased. This result is compatible with statistical theory.

RESUME

L'échantillonnage en deux phases et l'échantillonnage au hasard à une phase d'une pêcherie de hareng sont comparés. Les équations pour les poids moyens, les proportions à l'âge, la capture totale, et les captures à l'âge, ainsi que leurs variances, sont présentées. Les résultats d'un projet démonstration d'échantillonnage montrent qu'aucun des deux approches comporte un biais intrinsèque. Ce résultat est compatible avec la théorie des statistiques.

INTRODUCTION

Since 1973, the sampling of commercial herring landings in NAFO Division 4T for stock assessment purposes has been based upon a standard random sample of 50 fish. Each such sample has been assumed to be representative of the composition of the catch of a single vessel. The sample is frozen in the port, and all measurements and examinations are conducted in the laboratory. On the same day as these "detailed" samples are collected, approximately 250 additional herring are measured for length only. These "length frequency" samples are usually taken from the same vessel. However, to date they have not been used for stock assessment purposes.

The collection of "detailed" samples from catches of other species has followed a different procedure. In this case, detailed samples have been composed of a specified number (usually one or two) per length interval (usually one to three centimetres). Data obtained from these fish have been extrapolated to the whole catch using the larger length frequency sample. Fisheries biologists sometimes refer to this procedure as "length stratified sampling". Statistically, it is a case of "double sampling", or (more correctly) "two phase sampling" (Kendall et. al., 1983).

The examination of herring detailed samples is costly in manpower. It is thus desirable to avoid redundancy in the expenditure of this effort. The use of simple random samples to determine both the length composition and detailed biological characteristics of herring may be hypothesized to provide an unnecessarily large amount of data about fish in the most common length intervals, while generating very little information about the less abundant lengths.

For most fish stocks, the only essential assessment parameter measured in the second phase sample is the age. However, herring in Division 4T are assumed to belong to two biologically-distinct "spawning groups", and several characteristics of the fish examined in the laboratory are used to partition catches-at-length into these two groups. The need to sort out two distinct catch matrices provides a presumptive justification for more intensive sampling of herring than would be required for the establishment of a single age-length key.

Notwithstanding the above, it would be desirable to reduce the total number of fish examined in detail in any one sample. This would permit the processing of a larger number of samples with the same laboratory effort.

It would also be desirable to obtain more data from fish at the less common lengths, in order to reduce the variance about growth and maturation curves. The most obvious way to do this would be to sample relatively fewer fish at the more common lengths.

Finally, it would be desirable to take advantage of the existence of the length data in the larger length frequency samples when extrapolating ratios observed in the detailed samples. There appears to be no statistical requirement that the latter be a subset of the former (Cochran, 1977), as long as they are drawn from the same population. Thus, length frequency data from previous years might be used to improve the accuracy of catch matrices for those years. Furthermore, it is proposed that two phase sampling in the future may result in increased accuracy and precision in the determination of catch compositions.

This paper has a dual purpose:

1. Outline the appropriate statistical equations for the determination of the means, proportions, variances, and confidence limits important for stock assessment.

2. Demonstrate the application of these equations and compare the results of two phase and single sampling schemes, using data collected during 1984.

THE DEMONSTRATION PROGRAMME

Virtually all herring catches sampled in NAFO Division 4T in 1984 were triple sampled. The usual 50 fish random sample was frozen for laboratory analysis. Subsequently, this is referred to as the "single phase" sample. The lengths of approximately 250 additional fish were measured. A second phase sample composed of one or two fish per half centimetre length interval (a subset of the length frequency sample) was also frozen for laboratory analysis. The length frequency sample and its subset are referred to as the "two phase" sample.

The determination of the overall Division catch matrix for each of the two spawning groups is based on a summation of catches at age in smaller area-time-gear units. Typically these represent catches in unit areas (Figure 1) during one month. The only two important gears are gillnets and purse seines, and the latter are restricted to a brief fall fishery in one or two unit areas. The basic modules used may vary depending upon the availability of samples and the abundance of the catches.

For the purpose of this paper, the ten available samples from unit area 4Tn for the period from August 30 to September 20, 1984 were processed in the usual manner in the laboratory. These were gillnet-caught samples only. For purposes of comparison between the two sampling schemes a catch of 10,000 tonnes was assumed. For each fish, total length, total weight and gonad weight were measured. Otoliths were used for age determination. Sex and gonad maturity stage were judged by experienced technicians.

Spawning group identification is not a directly observed or measured parameter. However, the most important observation for the final assignment is the gonad maturity stage. At particular times of year, fish belonging to different spawning groups have different gonad maturity stages. In the present analysis, the proportion of fish at gonad maturity stage 5 was evaluated. The model has thus been tested using an easily-identifiable parameter, and the more subjective issues of which stages belong to which spawning group have been avoided, since they are irrelevant to the issue at hand. However, it appears that virtually all of the fish sampled were members of the fall spawning group.

In summary, the two sampling schemes were compared for their estimates of mean fish weight, length composition of the total catch, catch at stage 5, weight/length relationship, and age composition of the total catch.

SYMBOLS USED

N_m total number of fish in phase one (length frequency) of sample m

N_{jm} total number of fish in phase one at length j in phase one of sample m

N'_m, N'_{jm} } comparable parameters for detailed samples

W_m weight of the catch from which sample m came

N'_{ijm} number of fish at maturity stage i at length j in m or number of fish at age i at length j in m

EQUATIONS

1. Two Phase Sampling

Most of this section is based on Gavaris and Gavaris (1983) and Cochran (1977).

Total of all length frequencies:

$$N = \sum_m N_m \quad (1)$$

Total detailed observations at length j:

$$N'_j = \sum_m N'_{jm} \quad (2)$$

Proportion at length j in sample m:

$$P_{jm} = N_{jm}/N_m \quad (3)$$

If W_{jm} is the total weight of fish at length j in sample m, then the estimated average weight at length j is:

$$\overline{W}_j = \sum_m W_{jm}/N'_j \quad (4)$$

This equation provides a biased estimate insofar as the ratio between N'_{jm} and the associated catch may vary among fishery samples, m. This bias may be avoided by ensuring a random distribution of samples over the fishery, and by maintaining a constant N'_{jm} for all m.

Average fish weight in sample m:

$$\overline{W}_m = \Sigma_j \overline{W}_{jm} P_{jm} \tag{5}$$

Catch in number of fish by the vessel providing sample m:

$$C_m = W_m / \overline{W}_m \tag{6}$$

Proportion at maturity stage i (or age i) in length group j of sample m:

$$P_{ijm} = N'_{ijm} / N'_{jm} \tag{7}$$

Overall estimated proportion at length j:

$$P_j = \Sigma_m P_{jm} C_m / \Sigma_m C_m \tag{8}$$

Overall estimated proportion with maturity stage i (or age i) at length j:

$$P_{ij} = \Sigma_m N'_{ijm} / N'_j \tag{9}$$

Overall estimated average fish weight:

$$\overline{W} = \Sigma_j \overline{W}_j P_j \tag{10}$$

Overall estimated proportion at maturity stage i (or age i):

$$P_i = \Sigma_j P_{ij} P_j \tag{11}$$

Estimated total number of fish caught in the unit:

$$C = \text{catch weight}/w \tag{12}$$

Estimated total number of fish caught at maturity stage i (or age i):

$$C_i = CP_i \tag{13}$$

If fish from more than one spawning group are present, then the following equations would apply for the estimation of age composition
(where i indexes age
j indexes length group
k indexes spawning group
and m indexes sample numbers):

$$P_{ijk} = \Sigma_m N'_{ijkm} / N'_j \tag{9a}$$

$$P_{ik} = \Sigma_j P_j P_{ijk} \quad \text{(i.e. proportion at age i and spawning group K within the total \underline{mixed} catch)} \tag{11a}$$

$C_{ik} = CxP_{ik}$ (i.e. estimated number caught at age i of spawning group k) (13a)

$P_i(K) = C_{ik}/\Sigma_i C_{ik}$ (i.e. proportion at age i within spawning group K) (14)

2. Single Phase Sampling

Parameter estimation from single phase random sampling is simpler than for the two phase scheme. The important variates for stock assessment are the proportions and numbers at age or spawning group and the average weight per fish. These can all be calculated without reference to lengths. If single fishery sample values are weighted by their associated catch, C_m (as below), then this is technically a form of stratified random sampling.

The average weight per fish in sample m is the observed mean weight:

$$\overline{W}m = \Sigma y/N'_m \qquad (15)$$

where y represents individual fish weights.

The average weight per fish in the fishery is weighted by the proportions of the total fishery catch associated with each sample:

$$\overline{W} = \sum_m \overline{W}_m C_m / \sum_m C_m \qquad (16)$$

The estimate of the proportion in a given spawning group or age class i is a similar weighted average:

$$P_i = \sum_m P_{im} C_m / \sum_m C_m \qquad (17)$$

The total number caught, C, and the total number at age or spawning group i are given in (12) and (13) above.

Variances and Confidence Limits

Since equations 9 and 9a provide no weighting by cluster size, the appropriate equation for the variance of the proportion with characteristic i in length group j is the simple

$$V_{ar}(P_{ij}) = \frac{P_{ij}\,(1-P_{ij})}{N'_j-1} \qquad (18)$$

Gavaris and Gavaris (1983) provide the following equation for the variance of the overall proportion of the catch at a given age, which is applicable to our two phase sampling.

$$v(P_i) = \sum_{j:P_{ij}\neq 0} \frac{P_j^2 P_{ij}(1-P_{ij})}{N'_j-1} + \frac{P_j(P_{ij}-P_i)^2}{N} \qquad (19)$$

For single phase sampling, the variance of the catch proportion at age is estimated by the following equation for the variance of a proportion in stratified random sampling, where each stratum is the catch, C_m, associated with each fishery sample, m.

$$V(P_i) = \sum_m \left(\frac{C_m}{\Sigma C_m}\right)^2 \frac{P_{im}(1-P_{im})}{N'_m - 1} \qquad (19a)$$

Using Y_j for individual fish weights at length j, the following equation describes the variance of the estimated mean weight at length.

$$v(\overline{w}_j) = \Sigma(y_j - \overline{w}_j)^2 / N'_j(N'_j - 1) \qquad (20)$$

This may also be written:

$$v(\overline{w}_j) = s_j^2 / N'_j \qquad (20a)$$

Cochran (1977) provides the following equation for the variance of a mean in double sampling (his equation 12.24', page 333):

$$v(\overline{Y}st) = \sum_h^L \frac{Wh^2Sh^2}{N_h} - \sum_h^L \frac{WhSh^2}{N} + \frac{g'}{n'} \sum_h^L W_h (\overline{Y}h - \overline{Y}st)^2 \qquad (21)$$

Where h indexes strata and W_h is the relative weight of statum h, and where $g' = (N-n')/(N-1)$. Cochran's n' is the total number in the first phase, i.e. equivalent to our N. Assuming the population is very large, and using our symbols, equations 20a and 21 lead to the following equation for the variance of the overall mean weight:

$$v(\overline{w}) = \Sigma_j P_j^2 v(\overline{w}_j) + \frac{\Sigma_j P_j(\overline{w}_j - \overline{w})^2}{N} \qquad (22)$$

For single phase sampling, the variance of the overall mean weight is estimated by the following equation for the variance of a mean in stratified random sampling, where each stratum is the catch, C_m, associated with each fishery sample, m.

$$V(\overline{w}) = \Sigma \left(\frac{C_m}{\Sigma C_m}\right)^2 \frac{\Sigma(y - \overline{W}_m)^2}{N'_m (N'_m - 1)} \qquad (22a)$$

where y represents individual fish weights, and C_m is substantially larger than N'_m.

The following equation describes the variance of the total catch in numbers, where W is the total catch weight (Gavaris and Gavaris, 1983):

$$Var(C) = \frac{W^2 Var(\overline{w})}{w^4} \qquad (23)$$

The same authors provide this equation for the variance of the catch at age i:

$$v(C_i) = C^2 v(P_i) + P_i^2 v(C) - v(P_i)v(C) \tag{24}$$

Approximate 95% confidence limits for the overall mean weight, the total number of fish caught, and the total caught at stage 5 were calculated using the normal approximation. That is, the confidence interval was defined by 1.96 times the standard error of the mean estimate. Confidence limits for catches at age were only computed for those ages where $P_i > 0.05$, since the normal approximation becomes invalid for the calculation of confidence limits on proportions when the latter are very low, unless samples are very large (Cochran, 1977).

RESULTS AND DISCUSSION

Relative catches at length calculated on the basis of the two sampling schemes were quite similar (Table 1). The 30 to 32 cm length intervals were dominant in both cases. However larger fish were more prominent in the two phase results. Since the calculation of mean fish weight is weighted by the length composition, this results in a higher mean weight from the two phase regime (Table 2).

The lower variance for the mean weight from the two phase sampling also results in a low variance for the total number caught (Table 2). However, the total catch at a given maturity stage (Table 2) or at age (Table 3) tends to have higher variance and broader confidence limits in both cases.

Age-length keys computed from the two schemes do not seem to differ substantially (Table 4). Catches at age (Table 3 and Figure 2) are also similar, although the two phase scheme results in increased estimates at older ages, a direct result of the larger length frequency samples there.

Mean weights at length from both sampling schemes have broad fluctuations in variance (Table 5). This is primarily a result of unequal sample sizes at length. The more stable sample size from the two phase scheme results in more stable variances. However, neither data set meets the basic requirement of regression analysis for constant variance of the ordinate as the abscissa varies. A larger data set, permitting subsampling of an adequate number of weight values per length interval, would be required in order to compute the coefficients of a weight/length relationship.

The choice between a single and a two phase sampling scheme depends primarily upon the costs associated with the various stages in the sampling process. In the case of 4T herring investigations, the only significant cost is associated with the processing of fish for the determination of age, maturity stage, sex, and weight. The first phase length sample is relatively "free" of cost. So for a given level of precision, two phase sampling is least costly. Independently of cost factors, neither scheme offers inherently superior accuracy or precision which could not be emulated by the other. For example, if a single phase scheme appears to be inferior in its coverage of the total length range, then this apparent deficiency could be compensated by taking larger individual samples (N_m). If a two phase scheme produces unacceptably high variance in parameter estimates at length, then an increase in the number of second phase observations per length group could rectify this situation.

REFERENCES

Cochran, W.G. 1977. Sampling Techniques, 3rd Edition. John Wiley and Sons, Toronto.

Gavaris, S. and C.A. Gavaris. 1983. Estimation of catch at age and its variance for groundfish stocks in the Newfoundland region, pp. 178-182 in W.G. Doubleday and D. Rivard, eds. Sampling Commercial Catches of Marine Fish and Invertebrates. Can. Spec. Publ. Fish. Aquat. Sci. 66.

Kendall, M., A. Stuart, and J.K. Ord. 1983. The Advanced Theory of Statistics, Vol. 3, 4th ed. Macmillan, N.Y.

TABLE 1: Relative catches at length from the two sampling schemes.

LENGTH (cm)	SINGLE PHASE	TWO PHASE
25	.010	.000
25.5	.000	.000
26	.000	.000
26.5	.000	.000
27	.000	.000
27.5	.000	.001
28	.004	.000
28.5	.004	.001
29	.029	.009
29.5	.043	.029
30	.082	.077
30.5	.094	.110
31	.140	.141
31.5	.116	.102
32	.093	.110
32.5	.059	.065
33	.088	.077
33.5	.071	.067
34	.057	.069
34.5	.045	.049
35	.030	.044
35.5	.025	.029
36	.010	.021

TABLE 2: Parameter estimates and associated variances from the two sampling schemes. Estimates at the lower and upper 95% confidence limits are shown in parentheses.

	SINGLE PHASE	TWO PHASE
MEAN WEIGHT(g)	276 (268, 284)	295 (290, 299)
VARIANCE	18.579	5.223
TOTAL CATCH $(\times 10^{-7})$	3.619 (3.508, 3.730)	3.392 (3.340, 3.443)
VARIANCE $(\times 10^{-10})$	31.884	6.912
TOTAL STAGE 5 CATCH $(\times 10^{-6})$	3.801 (2.615, 4.987)	3.854 (2.294, 5.415)
VARIANCE $(\times 10^{-11})$	3.660	6.340

TABLE 3: Estimated catches at age ($x10^{-3}$) from the two sampling schemes with estimates at upper and lower 95% confidence limits.

	SINGLE PHASE			TWO PHASE		
AGE	C	C	C_u	C	C	C_u
3	-	0	-	-	124	-
4	12984	15138	17292	10657	12651	14646
5	8486	10379	12272	8618	10746	12872
6	3166	4624	6082	4135	5595	7058
7	3804	5339	6874	2846	3909	4976
8	-	560	-	-	647	-
9	-	154	-	-	91	-
10	-	0	-	-	78	-
11	-	0	-	-	26	-
12	-	0	-	-	26	-

Table 4: Age-length keys from the two sampling schemes.

SINGLE SAMPLING:

	3	4	5	6	7	8	9	10	11	12	MEAN AGE
25.0	.000	.200	.000	.600	.200	.000	.000	.000	.000	.000	5.8
25.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
26.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
26.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
27.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
27.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
28.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
28.5	.000	1.00	.000	.000	.000	.000	.000	.000	.000	.000	4.0
29.0	.000	.889	.111	.000	.000	.000	.000	.000	.000	.000	4.1
29.5	.000	.769	.231	.000	.000	.000	.000	.000	.000	.000	4.2
30.0	.000	.879	.121	.000	.000	.000	.000	.000	.000	.000	4.1
30.5	.000	.833	.167	.000	.000	.000	.000	.000	.000	.000	4.2
31.0	.000	.745	.255	.000	.000	.000	.000	.000	.000	.000	4.3
31.5	.000	.478	.391	.087	.043	.000	.000	.000	.000	.000	4.7
32.0	.000	.194	.611	.111	.083	.000	.000	.000	.000	.000	5.1
32.5	.000	.111	.667	.167	.056	.000	.000	.000	.000	.000	5.2
33.0	.000	.028	.639	.139	.194	.000	.000	.000	.000	.000	5.5
33.5	.000	.000	.389	.278	.306	.028	.000	.000	.000	.000	6.0
34.0	.000	.000	.080	.360	.480	.080	.000	.000	.000	.000	6.6
34.5	.000	.063	.063	.188	.625	.063	.000	.000	.000	.000	6.6
35.0	.000	.000	.000	.300	.700	.000	.000	.000	.000	.000	6.7
35.5	.000	.000	.000	.222	.667	.000	.111	.000	.000	.000	7.0
36.0	.000	.200	.000	.000	.200	.600	.000	.000	.000	.000	7.0

DOUBLE SAMPLING:

	3	4	5	6	7	8	9	10	11	12	MEAN AGE
25.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
25.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
26.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
26.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
27.0	.000	1.00	.000	.000	.000	.000	.000	.000	.000	.000	4.0
27.5	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
28.0	.000	1.00	.000	.000	.000	.000	.000	.000	.000	.000	4.0
28.5	.000	.800	.200	.000	.000	.000	.000	.000	.000	.000	4.2
29.0	.053	.789	.105	.000	.053	.000	.000	.000	.000	.000	4.2
29.5	.000	.933	.000	.067	.000	.000	.000	.000	.000	.000	4.1
30.0	.042	.833	.083	.042	.000	.000	.000	.000	.000	.000	4.1
30.5	.000	.778	.222	.000	.000	.000	.000	.000	.000	.000	4.2
31.0	.000	.632	.316	.053	.000	.000	.000	.000	.000	.000	4.4
31.5	.000	.435	.435	.043	.043	.043	.000	.000	.000	.000	4.8
32.0	.000	.200	.667	.133	.000	.000	.000	.000	.000	.000	4.9
32.5	.000	.133	.733	.067	.067	.000	.000	.000	.000	.000	5.1
33.0	.000	.100	.500	.250	.150	.000	.000	.000	.000	.000	5.5
33.5	.000	.056	.278	.500	.167	.000	.000	.000	.000	.000	5.8
34.0	.000	.000	.167	.583	.250	.000	.000	.000	.000	.000	6.1
34.5	.000	.160	.080	.280	.480	.000	.000	.000	.000	.000	6.1
35.0	.000	.063	.000	.125	.688	.125	.000	.000	.000	.000	6.8
35.5	.000	.000	.067	.400	.333	.133	.067	.000	.000	.000	6.7
36.0	.000	.111	.000	.259	.148	.259	.037	.111	.037	.037	7.4

TABLE 5: Mean Weights at Length and their Variance

LENGTH	ONE PHASE			TWO PHASE		
	n	$\overline{W}$	$V(\overline{W})$	n	$\overline{W}$	$V(\overline{W})$
270				2	174.5	24810.2
275						
280	1	195.0		2	201.5	32943.2
285	2	222.5	40201.2	6	198.3	6408.5
290	9	211.9	4565.6	20	206.1	1819.0
295	14	223.0	30.9	18	218.6	34.2
300	40	228.7	15.9	25	229.0	14.0
305	47	236.4	7.3	11	246.5	24.7
310	59	251.6	7.5	20	256.3	48.1
315	51	269.6	9.9	25	275.6	38.6
320	41	279.3	11.4	17	290.4	36.1
325	41	303.3	20.5	17	305.6	47.8
330	39	316.4	20.9	20	325.6	40.1
335	37	338.3	18.6	19	334.5	95.9
340	27	344.0	30.5	15	329.1	142.1
345	18	358.1	63.6	28	373.9	30.8
350	14	358.0	222.1	16	383.3	51.5
355	10	387.3	79.0	19	394.6	71.2
360	6	372.8	1254.1	31	436.4	66.3

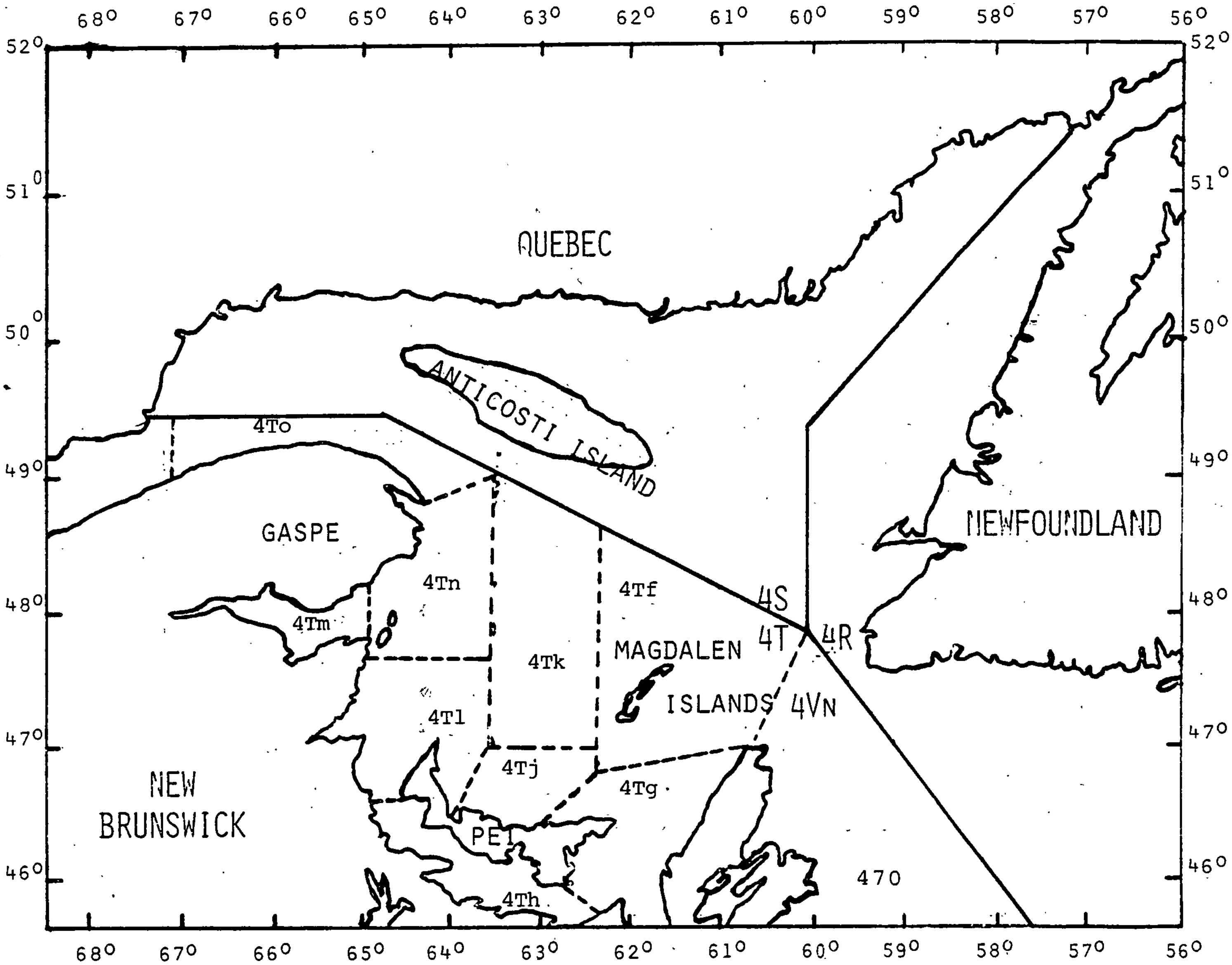

Fig. 1. Map showing statistical unit areas for the Southern Gulf of St. Lawrence.

Figure 2: Catches at age from the two sampling schemes.

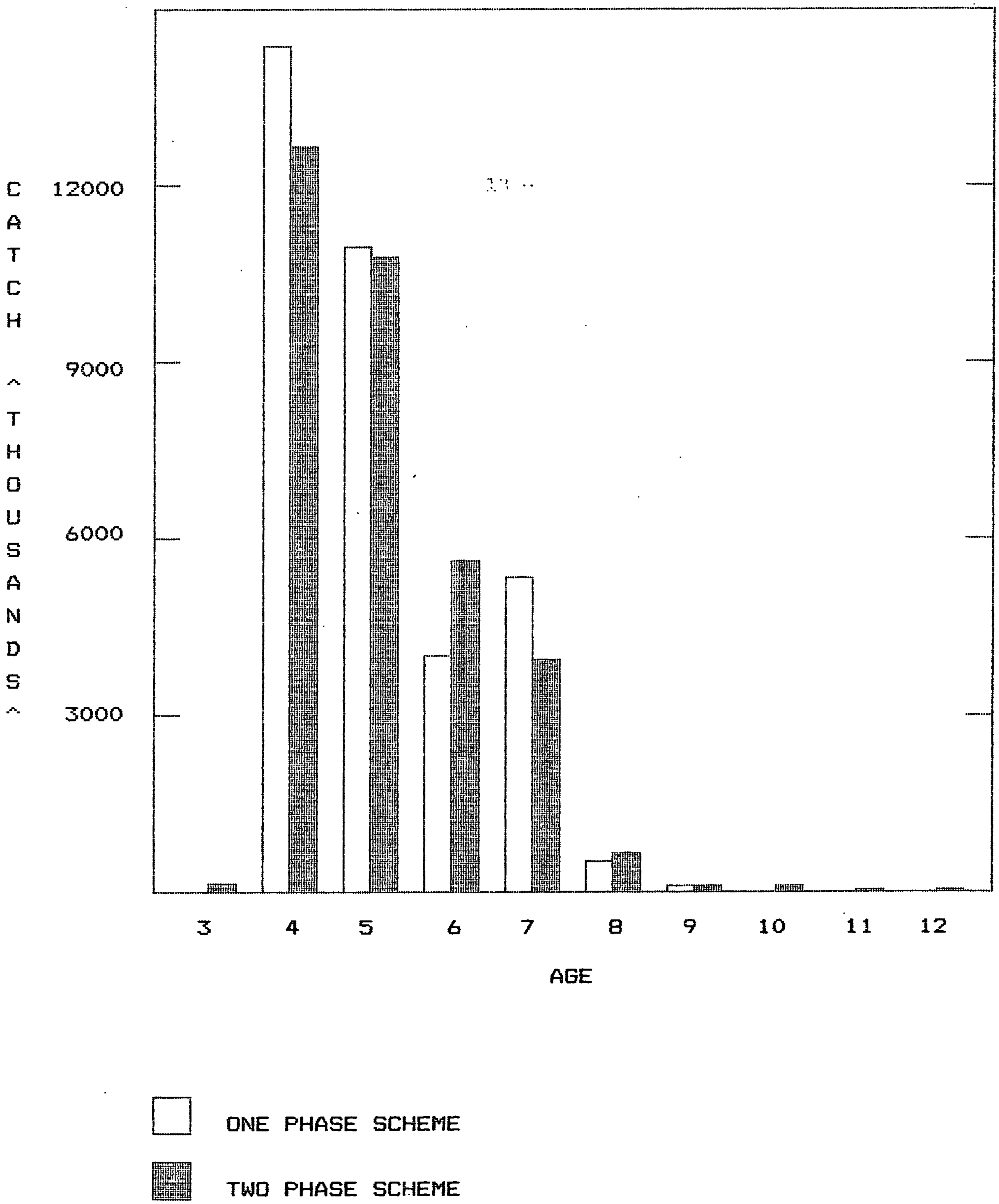

Figure 3: Length-weight relationships based upon the two sampling schemes.