Northwest Atlantic Fisheries Organization

Serial No. N278 NAFO SCR DOC. 81/II/14

SPECIAL MEETING OF SCIENTIFIC COUNCIL - FEBRUARY 1981

Sampling Variation and Survey Design for Capelin (*Mallotus villosus*) Densities from an Acoustic Survey in Divisions 3LNO, 1980

by

B. S. Nakashima
Department of Fisheries and Oceans, Research and Resource Services
P. O. Box 5567, St. John's, Newfoundland, Canada A1C 5X1

Introduction

Abundance and biomass estimates from acoustic surveys may be used as relative indicators of annual trends in stock size (Carscadden and Miller, 1980; Miller and Carscadden, 1980). However, the inability to determine biases associated with each point estimate precludes the interpretation of acoustic data into absolute estimates of stock size. Taylor and Kieser (1980) recognized four sources of error relating to 1) extrapolation, 2) species composition, 3) calibration, and 4) sampling variation. To date only the latter source, i.e. variance related to sampling design, can be measured with any degree of certainty.

Several recent studies have recognized that variation estimates from line transect and related survey designs must account for the non-random distribution of marine mammals and fish populations (Dark et. al. 1980; Kimura and Lemberg, 1981; Quinn, 1977; Shotton and Dowd, 1975; Taylor and Kieser, 1980). Generally, as populations exhibited stronger aggregating tendencies, the variance of the sample mean which assumes random distribution becomes more biased from the true population variance (Shotton and Dowd, 1975). Recognizing that fish populations are rarely randomly distributed, Shotton and Dowd (1975) proposed using the clustering sample method of Hansen, et. al. (1953) to estimate variance. Compared to other variance models they found that the clustering sample method was the best available because 1) it accounted for serial correlations among observations from contagious distributions and 2) the results could be used to allocate sampling effort based on the degree of inter- and intra-transect variation. Some recent acoustic studies appear to have successfully applied the clustering sample method to calculate confidence intervals around biomass estimates of Pacific whiting (Dark, et. al., 1980) and pollock (Taylor and Kieser, 1980). Another variance model developed by Kimura and Lemberg (1981) has limited practical use since it requires detailed information on the shape, size and distribution of fish schools.

The purpose of this report is to provide empirical estimates of variance for capelin (*Mallotus villosus*) densities from Divisions 3LNO, 1980 as an example to illustrate the methodology involved. The behaviour of one model parameter (δ) will be examined to suggest changes in survey design. It must be emphasized that all variance estimated in this report are a function of sampling design only.

Acoustic Data Analysis

The methodology and results of the Div. 3LNO 1980 acoustic survey used in this analysis were described by Miller and Carscadden (1981). Capelin density estimates (fish m^{-2}) per observation (i.e. 6 min. interval) (Miller and Carscadden, 1981) were used to estimate the variance model parameters in equation (1). The original cruise track (see fig. 6, Miller and Carscadden, 1981) was modified to eliminate areas of overlap and short deviations from the main pattern. Thus the results of our study are based upon the data obtained employing the transects depicted in survey 1 (Fig. 1). Since midwater trawl catches indicated that the Div. 3LNO capelin stock consisted of four groups (Miller and Carscadden, 1981), survey 1 was subdivided *a posteriori* into four strata based upon size differences. To facilitate computations, the four strata in this report are termed surveys 1a, 1b, 1c and 1d which encompass transects 1 to 9, 10 to 16, 17 to 21, and 22 to 31, respectively.

The influence of survey pattern on sampling variation was investigated by examining the density estimates from a systematic zigzag (survey 2, Fig. 2) and systematic parallel (survey 3, Fig. 3) pattern. A fourth survey (survey 4, Fig. 4) was completed involving a systematic zigzag track over an area larger than that covered by surveys 2 and 3. These surveys were run consecutively assuming that the capelin distribution did not change over time between July 1 and 6.

Clustering Sample Model

The clustering sample model used to estimate the relative variance (V_r^2) associated with the mean sample density was

$$V_r^2 = \frac{1-f}{t\bar{n}} \hat{V}^2 [1 + \delta (\bar{n}-1)] \tag{1}$$

where f is the sampling fraction, t is the number of transects in the sample, $\bar{n}$ represents the mean number of observations (i.e. intervals) per transect, and δ indicates the degree of intra-transect variance. The parameter $\hat{V}^2$ is calculated using

$$\hat{V}^2 = \frac{t-1}{t} B^2 + \frac{\bar{n}-1}{\bar{n}} W^2 \tag{2}$$

where

$$B^2 \sum_{i=1}^{t} \frac{(x_i-\bar{x}_t)^2}{(t-1)\bar{x}_t^2} + \sum_{i=1}^{t} \frac{(n_i-\bar{n})^2}{(t-1)\bar{n}^2} - 2 \sum_{i=1}^{t} \frac{(x_i-\bar{x}_t)(n_i-\bar{n})}{(t-1)\ \bar{x}_t\bar{n}} \tag{3}$$

and

$$W^2 = \frac{1}{\bar{X}^2} \frac{1}{N} \sum_{i=1}^{t} \frac{n_i}{n_i-1} \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 \tag{4}$$

For both (3) and (4) x_{ij} is the jth density estimate in the ith transect, x_i is the total density of the ith transect, $\bar{x}_i$ is the mean density of the ith transect, $\bar{x}_t$ is the mean density for all transects in the sample, $\bar{X}$ is the mean sample density, n_i is the number of observations in the ith transect, and N is the total number of observations in the sample.

Within transect variation is estimated by

$$\delta = \frac{\frac{t-1}{t} B^2 - \frac{\hat{V}^2}{\bar{n}}}{(\bar{n}-1)\ \frac{\hat{V}^2}{\bar{n}}} \tag{5}$$

The parameter δ can be employed as an indicator of variance heterogeneity.

The relative variance estimated by equation 1 is analogous to the clustering sample variance and is defined as the coefficient of variation squared (Hansen et. al., 1953). To apply the variance model (equation 1) to acoustic data Shotton and Dowd (1975) assumed that transects were equivalent to clusters and intervals or observations were primary sampling units. For a detailed account of the method, the reader is referred to Hansen, et. al. (1953)

To compare the results of this study to published studies, normality of the mean was assumed and 95% confidence limits were constructed by multiplying the coefficient of variation by 1.96.

Results

The relative variance for each survey area was calculated assuming that f was negligible. This can be observed by comparing the cruise track lengths (Tables 1, 2) to the total area surveyed (Fig. 1, 2, 3, 4). For surveys 1d, 2, 3, and 4, W^2 was approximated by assuming $W^2 = 0$ since the $\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2$'s per transect were small.

The mean densities for each survey and their relative variances and 95% confidence intervals were summarized in Table 3. These results demonstrated that a large proportion of the variance observed in survey 1 can be attributed to the sampling variance of survey 1b. The n_i's and $\bar{x}_i$'s also tended to be more variable in survey 1b than in surveys 1a and 1d (Table 1). The difference among the four surveys was probably due to the presence of four size groups occurring in the survey 1 area. The post-stratification of the large initial survey into four strata on biological grounds (Miller and Carscadden, 1981) resulted in separate density and variance estimates per strata. The cluster sampling analysis can then be applied to consider a survey design to reduce variation where possible.

The comparison of the effect of survey pattern on sampling variation suggested that the parallel and zigzag patterns were equally applicable since the mean densities, confidence intervals, and coefficients of variation were similar between surveys 2 and 3 (Table 3). Further, these two surveys were similar to the results from survey 1d which was over a significantly larger area. Survey 4 density estimate was 50%

of the density estimates for the 1d, 2 and 3 surveys while the coefficient of variation was similar among all four surveys. This lower mean density may be attributed to a reduction in the spawning population with time since survey 4 was conducted after surveys 1d, 2 and 3.

The δ parameter in the variance model (equation 5) described the contribution of inter- or intra-transect variance to the sample variance. If δ approaches 1.0, inter-transect variance accounts for most of the observed variances, whereas if δ approaches $\frac{-1}{\bar{n}-1}$ then the observed variance is attributed to intra-transect variance (Shotton and Dowd, 1975). Knowing the source and magnitude of variation allows the investigator to plan an appropriate survey design. Using the above criteria, survey 1a variation is predominantly due to within transect variance and variation observed in surveys 1d, 2, 3 and 4 were due to between transect variance. The remaining surveys (1b, 1c and 1) had intermediate δ values. After a close examination of the $\bar{x}_t$'s (Tables 1, 2) and x_{ij}'s (Miller and Carscadden, 1981) along with the δ parameter for surveys 1d, 2, 3 and 4, we concluded that both the intra- and inter-transect observations were fairly uniform. Thus sampling intensity in these areas could be reduced without any significant loss in information. The low δ for survey 1a would suggest high within transect variation thus indicating a need to increase observations within transects.

Discussion

This study has demonstrated the application of the clustering sample method to estimate variance for acoustic estimates of density and to use the results to plan future surveys. The model appears to fit the biological observations (Shotten and Dowd, 1975; Taylor and Kieser, 1980). The 95% confidence intervals for Div. 3LNO capelin densities were similar to those reported in other studies (Table 4).

While some suggestions for future survey design can be made from this analysis, the overall results are strongly influenced by the low densities encountered. Either the zigzag or parallel survey design can be employed without any reduction in precision according to this study. However, Kimura and Lemberg (1981) have demonstrated through simulation analysis that the choice of survey pattern is dependent on sampling intensity. According to their study, the zigzag pattern would have resulted in lower confidence intervals than for the parallel design. The choice of transect cannot be made on the basis of one comparison since the estimated densities were small ($\leqslant$0.002 fish m^{-2}) and the fish appeared to have a uniform distribution in the area. The behaviour of δ's would probably change if the survey were conducted at higher population levels where serial correlations among observations would become meaningful. The decline in density estimates between surveys 1d, 2 and 3 and survey 4 may have indicated that these transect comparisons were performed near the end of the spawning period. Thus the aggregated behaviour of spawning schools would not have been observed and the assumption of unchanging distribution during the survey comparisons was violated.

The effect of post-stratifying according to biological characteristics of the population allowed us to dissect the variation observed for survey 1 into smaller components. However, the influence of post-stratifying on variance estimation is unknown. The results of this analysis may be employed to conduct a future survey of the Div. 3LNO capelin stock using an a priori survey design which would partition the sampling effort according to the expected distribution of capelin in the area.

Based on the variance analysis and the low densities encountered, especially for the spawning adults, we conclude that sampling variation was relatively small. However, our earlier caution still applies. Sampling variation represents only one source of the total variation associated with acoustic surveys and its immediate application as demonstrated here is to help in designing acoustic surveys. Furthermore, functional aspects of the acoustic equipment prevented sampling to a depth of 20 meters below the transducer and also in the zone immediately adjacent to bottom. Presence of fish concentrations in these zones could substantially bias the biomass estimate downward from the true value.

References

Carscadden, J. E., and D. S. Miller. 1980. Analytical and acoustic assessments of the capelin stock in Subarea 2 and Division 3K, 1979. NAFO SCR Doc. 80/II/13, Serial No. N045.

Dark, T. A., M. O. Nelson, J. J. Traynor, and E. P. Nunnallee. 1980. The distribution, abundance and biological characteristics of Pacific whiting, Merluccius productus, in the California-British Columbia region during July-September, 1977. Mar. Fish. Rev. 42: 17-33.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow. 1953. Sample survey methods and theory. Volume I Methods and Applications. J. Wiley & Sons, Inc., New York, 638 p.

Kimura, D. K., and N. A. Lemberg. 1981. Simulation study of the variance of line intercept density estimates (an investigation into the variability of hydroacoustic biomass estimates). (Manuscript reviewed for publication.) 44 p.

Miller, D. S., and J. E. Carscadden. 1980. An acoustic survey of capelin (Mallotus villosus) in Divisions 3LNO, 1979. NAFO SCR Doc. 80/II/14, Serial No. N046.

1981. Acoustic survey results for capelin (Mallotus villosus) in Divisions 2J3K and 3LNO, 1980. NAFO SCR Doc. 81/II/5, Serial No. N269.

Quinn, T. J. 1977. The effects of aggregation on line transect estimators of population abundance with application to marine mammal populations. M. Sc. Thesis. Univ. Wash., Seattle. 116 p.

Shotton, R., and R. G. Dowd. 1975. Current research in acoustic fish stock assessment at the Marine Ecology Laboratory. ICNAF Res. Doc. 75/16, Serial No. 3468.

Taylor, F. H. C., and R. Kieser. 1980. Hydroacoustic and fishing surveys for walleye pollock in Dixon Entrance and nearby areas, July 5-23, 1978. Can. MS Rep. Fish. Aquat. Sci. 1572. 73 p.

Table 1. Model parameters for survey 1, June 14-29.

Transect	Transect Length (km)	n_i	$\bar{x}_i$
1	124.1	66	0.006
2	179.6	103	0.006
3	109.3	59	0.009
4	100.0	59	0.005
5	103.7	59	0.015
6	81.5	47	0.008
7	177.8	94	0.008
8	198.2	118	0.006
9	270.4	129	0.004
10	277.8	154	0.026
11	276.0	155	0.139
12	287.1	164	0.119
13	290.8	155	0.503
14	279.7	156	0.212
15	283.4	151	0.331
16	211.1	112	0.251
17	187.1	95	0.091
18	151.9	89	0.078
19	138.9	74	0.025
20	103.7	56	0.022
21	101.9	60	0.025
22	94.5	56	0.002
23	96.3	50	0.003
24	94.5	56	0.002
25	113.0	59	0.002
26	77.8	50	0.003
27	81.5	49	0.003
28	74.1	44	0.002
29	77.8	39	0.003
30	74.1	42	0.002
31	75.9	41	0.002

Table 2. Model parameters for surveys 2, 3 and 4.

	Transect	Transect length (km)	n_i	$\bar{x}_i$
Survey 2 July 1-2	1	29.6	15	0.002
	2	51.9	40	0.002
	3	51.9	26	0.002
	4	51.9	25	0.003
	5	55.6	29	0.001
	6	46.3	26	0.002
	7	50.0	30	0.002
	8	50.0	30	0.002
	9	55.6	31	0.001
	10	51.9	26	0.002
	11	55.6	32	0.002
	12	51.9	27	0.002
Survey 3 July 3-4	1	55.6	29	0.002
	2	54.2	35	0.002
	3	52.3	31	0.002
	4	64.8	38	0.002
	5	61.1	33	0.002
	6	57.4	36	0.001
	7	48.2	24	0.002
Survey 4 July 4-6	1	74.1	38	0.001
	2	79.6	43	0.001
	3	74.1	39	0.002
	4	77.8	42	0.001
	5	79.6	42	0.001
	6	79.6	43	0.001
	7	79.6	45	0.001
	8	55.6	29	0.001

Table 3. Relative variance (V_r^2), intra-transect homogeneity (δ), coefficient of variation ($V_{\bar{x}}$), and 95% confidence intervals (C.I.) for estimated densities ($\bar{x}$).

Survey	t	N	$\bar{x}$	V_r^2	δ	$V_{\bar{x}}$	C. I.
1a	9	734	0.007	0.017	0.047	12.8%	± 25.1%
1b	7	1047	0.224	0.301	0.506	54.7%	±107.2%
1c	5	374	0.054	0.067	0.423	26.2%	± 51.4%
1d	10	486	0.002	o.004	1.000	6.5%	± 12.7%
1	31	2641	0.099	0.100	0.422	31.6%	± 61.9%
2	12	337	0.002	0.007	1.000	8.4%	± 16.5%
3	7	226	0.002	0.008	1.000	8.9%	± 17.4%
4	8	321	0.001	0.009	0.995	9.5%	± 18.6%

Table 4. Comparison of 95% confidence intervals for the Div. 3LNO 1980 capelin surveys to previously published acoustic estimates.

Species	95% CI (%)	Source
Capelin	±12.7 to ± 107.2 ±16.5 to ± 18.6	survey 1; this study surveys 2, 3, 4; this study
Walleye pollock	±22.0 to ± 87.2	Taylor and Kieser (1980)
Pacific hake	±16.0 to ± 97.0	Dark and Nelson (1977) cited in Taylor and Kieser (1980)
Pacific hake	±12.8 to ± 66.7	Dark et. al. (1980)

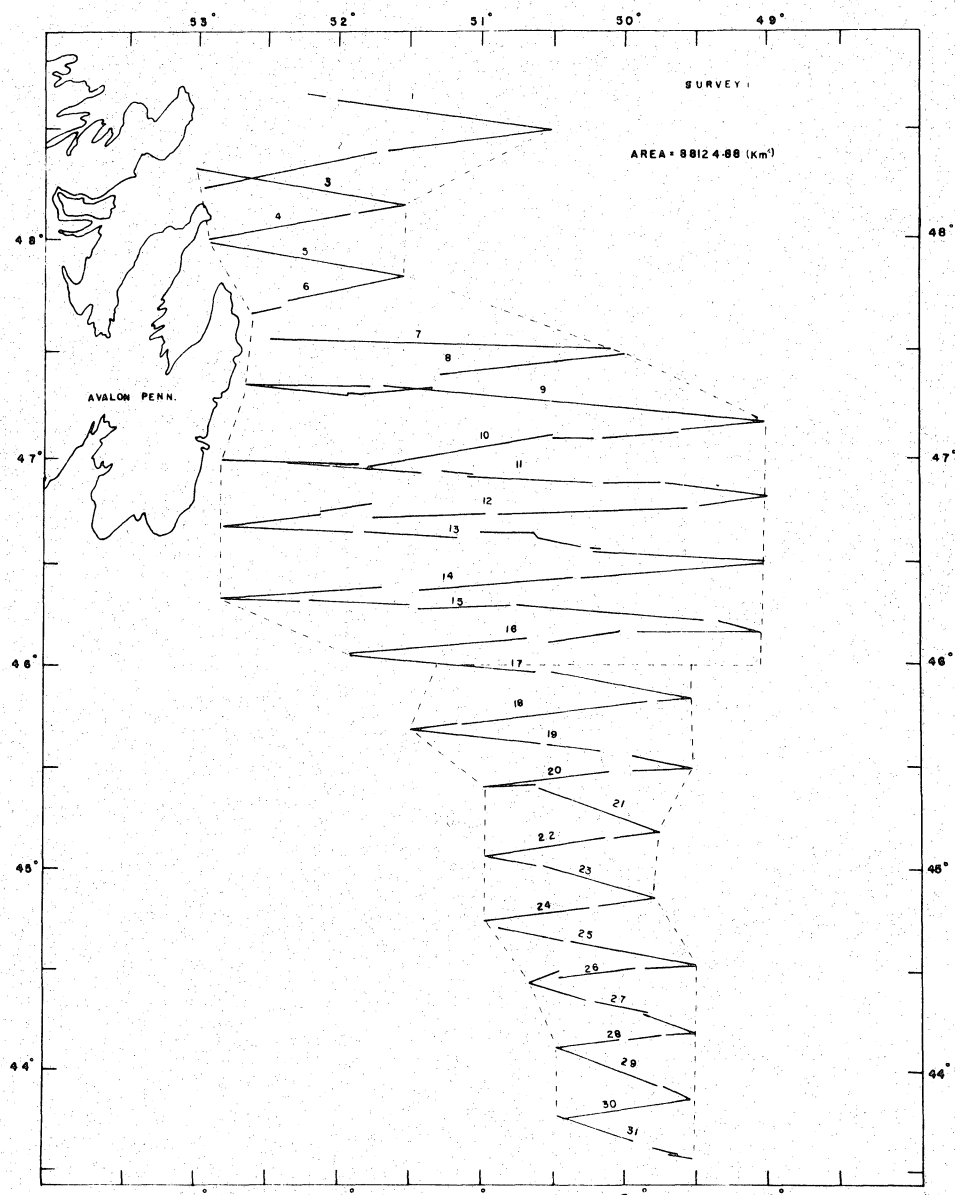

Fig. 1. Transects of Survey 1.

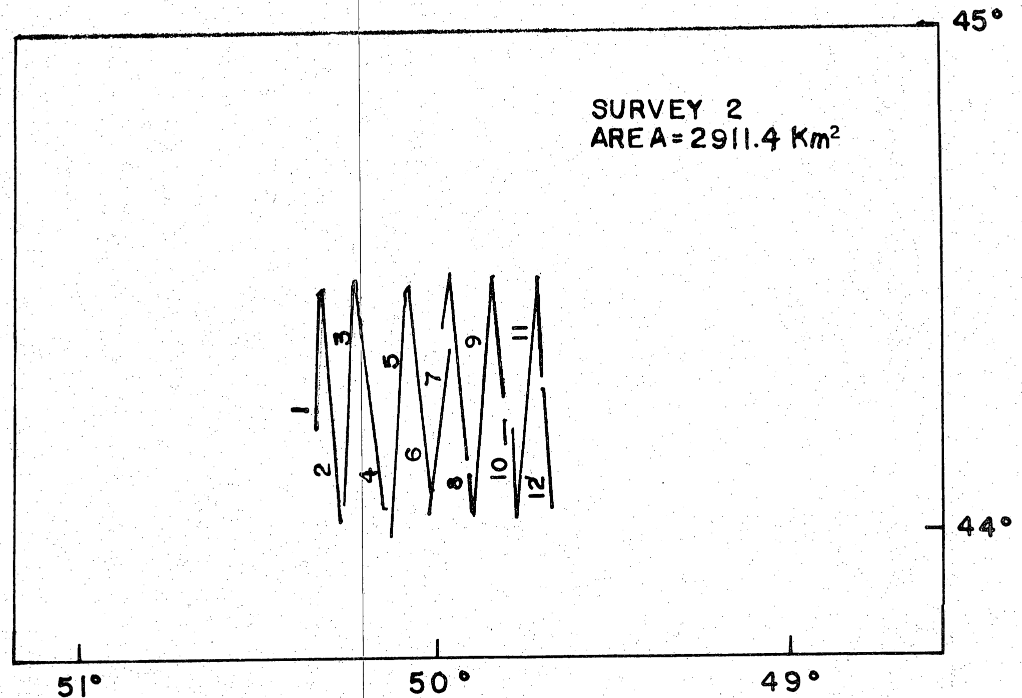

Fig. 2. Transects of Survey 2.

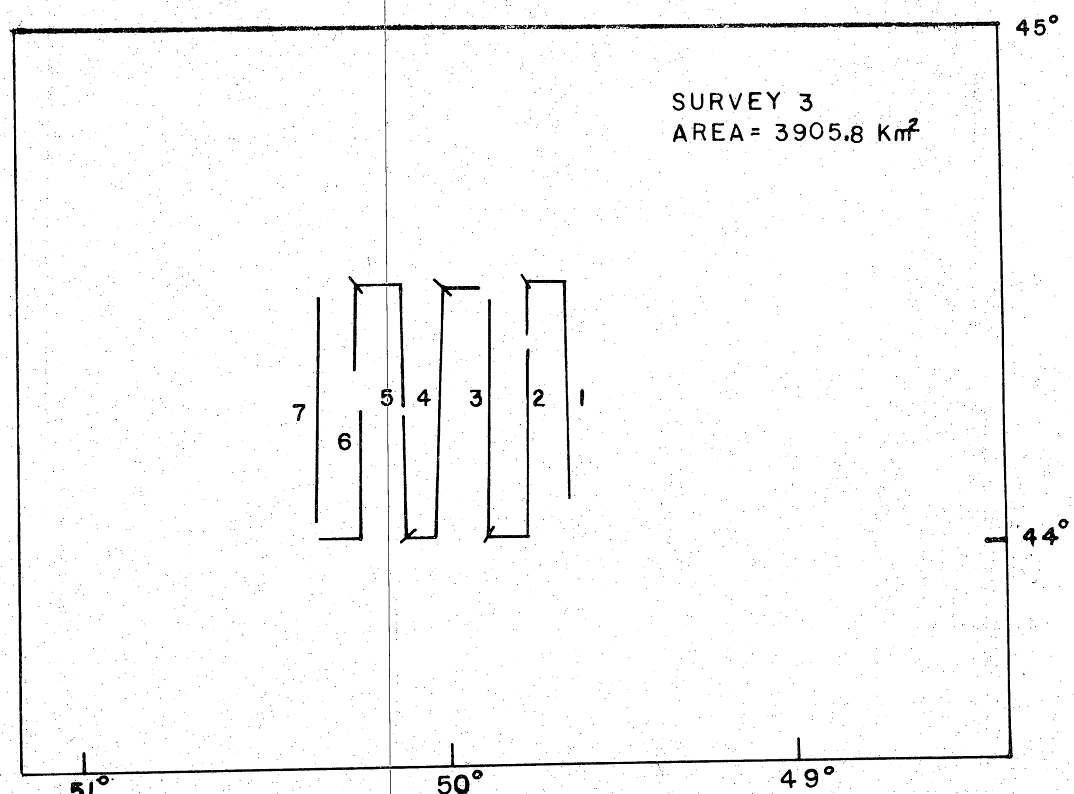

Fig. 3. Transects of Survey 3.

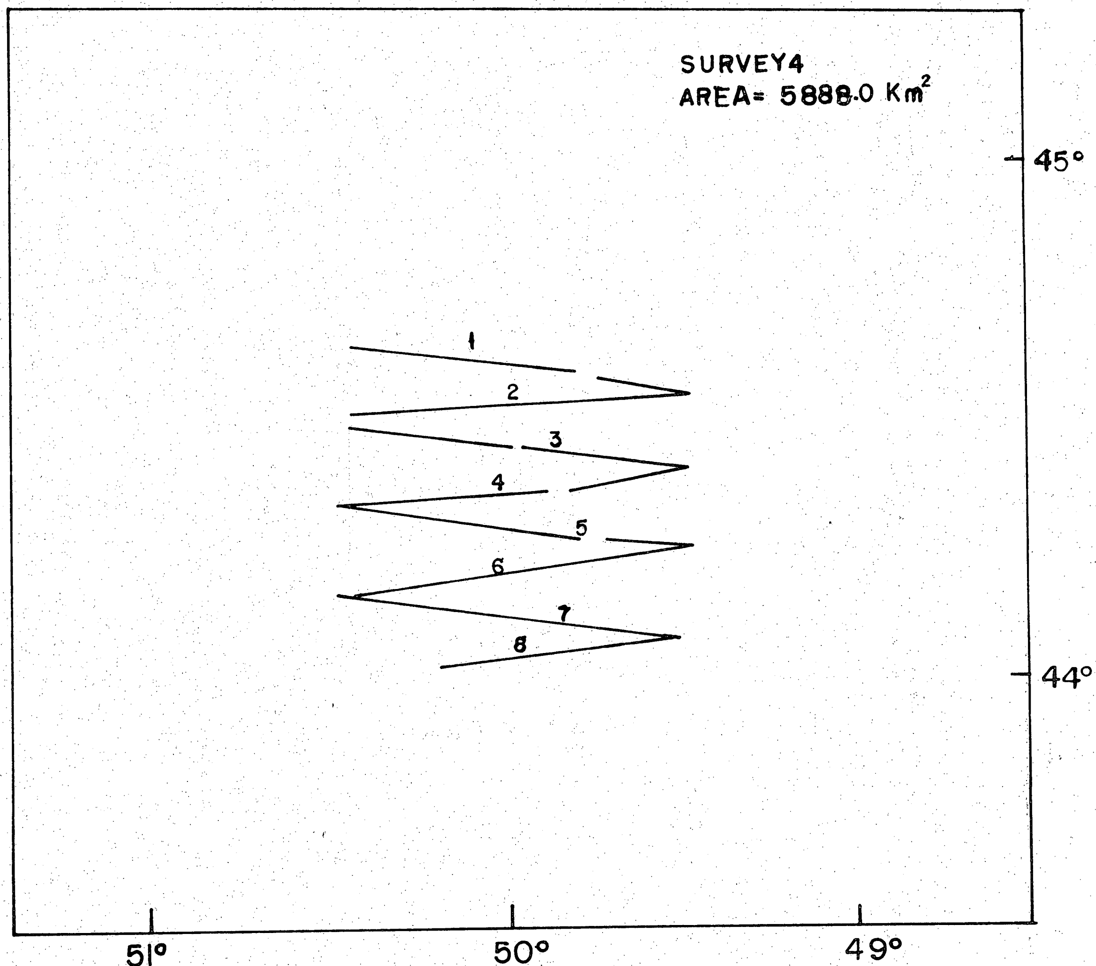

Fig. 4. Transects of Survey 4.