Ann. Inst. Statist. Math.
Vol. 43, No. 3, 435–453 (1991)

DETECTION OF HETEROGENEITY AND ESTIMATION OF POPULATION CHARACTERISTICS FROM THE FIELD SURVEY DATA: 1987/88 JAPANESE FEASIBILITY STUDY OF THE SOUTHERN HEMISPHERE MINKE WHALES

HIROHISA KISHINO[1]*, HIDEHIRO KATO[2], FUJIO KASAMATSU[3] AND YOSHIHIRO FUJISE[3]

[1] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*
[2] *National Research Institute of Far Seas Fisheries, 5-7-1 Orido, Shimizu, Shizuoka 424, Japan*
[3] *Institute of Cetacean Research, 3-32-11 Ohjima, Koto-ku, Tokyo 136, Japan*

(Received November 24, 1988; revised April 3, 1991)

Abstract. To get reliable information of the age structure of whale population, Japan conducted a feasibility study of scientific research in the Antarctic in 1987/88. Though the sample was not large enough, it was the first data free from the problem of selectivity and whaling ground bias. From the analysis, it was found that the biological characteristics are highly heterogeneous spatially or other ways. Considering this, we recognize that the survey should be designed to collect the sample from the whole research area uniformly to obtain unbiased estimates of population characteristics. However, in an actual biological field survey, it is difficult to keep the sampling fractions precisely the same for each sampling units. Therefore, it is important to detect the heterogeneity in the sample, and poststratify the data corresponding to the heterogeneity. The methodology of the estimation and model evaluation presented here will be useful for the development of biological field survey in general.

Key words and phrases: Line transect, population abundance, heterogeneity of biological parameters, post stratification, bootstrap resampling, model selection, AIC.

1. Introduction

In 1974, the International Whaling Commission (IWC) adopted the New Managenent Procedure (NMP) to utilize the whale stocks effectively and safely. In general, a population size of animals grows exponentially when it is far below the carrying capacity of the environment. But as the population size comes close to

* Now at Ocean Research Institute, University of Tokyo, 1-15-1 Minami-dai, Nakano-ku, Tokyo 164, Japan.

the carrying capacity, the growth rate decreases. Therefore, if we plot the surplus production against the population size, the resultant curve is concave. The population size which attains the maximal of this curve is called the maximum sustainable yield (MSY) level. The NMP aimed at maintaining the population level close to the MSY level. It protected stocks more than 10% below the MSY level, and allowed the catch of 90% of MSY from stocks above the MSY level.

In this same year, IWC started the worldwide survey called the International Decade of Cetacean Research (IDCR). Since 1978, the Southern Hemisphere minke whales have been monitored by sighting survey. They migrate north to the tropical region for breeding in winter and south to the Antarctic for feeding in summer. The Antarctic Ocean north of 60°S was divided into six areas by longitude, and the IDCR has surveyed the feeding population, an area to area every year. Sighting surveys have been conducted by the line transect procedure, where the sighting is made by observers on a vessel running along the trackline designed to cover the whole observation area with equal probability. The population abundance is estimated by the sighting data. The results of the analysis and the methodology of estimation are discussed at the IWC Scientific Committee, which is held in May or June every year. Presently, an agreement is going to be established in regard to the method of estimation of the abundance.

Although the concept of NMP has been appreciated, the IWC Scientific Committee recognized the difficulty in getting the reliable estimates of MSY and the relative position of the current population abundance to the MSY level, which are required for the implementation of the NMP. This problem urged IWC to suspend the NMP, and set the tentative catch limit of 0, effective from 1986, pending a reassessment of the situation by 1990 (IWC (1983)).

Since then, five adaptive management procedures have been proposed. Three of them incorporated the estimation procedure of the above values into the management scheme, by fitting the models of dynamic equation to the time series of monitored abundance index (Cooke (1989), De la Mare (1989), Punt and Butterworth (1989)). The others made more of the stability against possible misspecification of the model, and set the yearly increase or decrease of the catch limit from the information of the current trend and/or the level of the monitored abundance relative to the targeted one (Magnusson and Stefansson (1989), Sakuramoto and Tanaka (1989)). The more intensively the stock is monitored, the more stability the management can achieve. Hence, there is a trade-off between the investment to the monitoring and the yield which can be taken. And there is also a trade-off between the average level of the catch and the stability of the catch (Cooke (1989)). The efficiency and the stability of those procedures are being tested from various points of view by extensive simulation studies.

The stability and efficiency of the management of whale population could be considered to increase further, if the information of the mortality rate and the recruitment (or birth) rate is available, as this will make it possible to predict the future state of the population. These rates can be estimated most directly from the time series of the abundance index in conjunction with that of the age distribution (Tanaka (1990)). The age of a minke whale is determined by counting the number of lamina on the ear plugs.

The age data has already been accumulated from the whales caught by the commercial whaling. However, several problems have been pointed out about the data. The most serious one is the problem of selectivity; whalers tend to select larger whales. Also, the fishing grounds have been confined to the high density area near the ice edge, at most several dozens of n.miles (nautical miles) from the ice edge.

To get reliable information of age structure, Japan conducted a feasibility study of scientific research in the Antarctic in 1987/88 (i.e., winter season between 1987 and 1988). This paper discusses the theoretical aspects of the statistical procedures applied to the data from the Japanese feasibility study of the Southern Hemisphere minke whales in 1988. Detailed results and their explanations are found in the companion papers (Kasamatsu *et al.* (1990), Kato *et al.* (1990)). It was considered that the heterogeneity of whale distribution and the biological characteristics should be larger in lattitude than longitude. Therefore, the research area was extended north to 55°S line, although it was restricted rather narrowly (105°E to 115°E) in longitude. The line transect sampling procedure was adopted. The trackline was designed so as to cover the whole area uniformly and all the schools sighted were collected. Although animals are highly heterogeneously distributed, the line transect procedure samples the schools proportionally according to the local densities encountered. This enables us to estimate the unbiased population characteristics. The detailed design of the procedure is discussed in the following section.

In Section 3, we develop statistical methodology for the estimation of the population abundance and age distribution. The age distribution obtained by this study was found to be quite different from that obtained by the commercial whaling. Smaller and immature whales were rare in the data from the commercial whaling. However, it became clear that there were in fact many young whales in the Antarctic Ocean. Bootstrap resampling technique (Efron (1979), Hall (1988)) and the theory of two stage random sampling enabled us to calculate the standard errors of the estimates.

In Section 4, we investigate the heterogeneity of the mean ages with regard to school sizes and local areas, using a model selection procedure based on the information criterion AIC (Akaike (1973, 1985), Sakamoto *et al.* (1986)). It appears that whales tend to form schools after reaching sexual maturity. Further, the proportion of female was significantly larger in the area near ice edge than in the area off shore.

The results obtained here are not conclusive due to the limitation of the research area which is narrow in longitude and the sample size of 249 animals which is not sufficiently large. However, the results of the succeeding surveys are confirming the above findings (Fujise *et al.* (1990), Kato *et al.* (1991)). The effective utilization of the biological informations obtained by these surveys to the management scheme is still controversial (Barlow (1990), Nakamura (1990), Tanaka (1990)). Further studies are required.

2. Survey procedure

Two vessels, Kyōmaru 1 (K01) and Toshimaru 25 (T25) were engaged in the line transect sampling. They surveyed in parallel on the main trackline which had been systematically allocated to cover the whole research area uniformly. The distance between the ships was 12 n.miles and the sighting and sampling were made independently by the two ships.

The area was stratified into the northern and southern strata separated by the boundary of 60°S line. In the northern stratum, the trackline was running in the direction of north-south. In the southern stratum, the trackline was chosen like the trace of a billiard ball with fixed angle of reflection at the ice edge or the boundary of the region (Kato *et al.* (1989)). This is the modification of the zig-zag trackline adopted by the IDCR survey, which covered the wide area in lattitude compared with this feasibility study. According to the ergodic theory, a billiard ball covers the whole area uniformly in the long term. But from simulation study, it was found that more survey effort tends to be laid on the boundary region in the expected time scale of the survey. This problem was solved by fixing the reflection angle to 70°. The cruise track is given in Fig. 1. Since the size of the sample from the northern stratum was only seven, we analyzed only the sample from the southern stratum.

Observers usually identify a school of whales as a unit. Therefore, by regarding schools as primary sampling units and whales as secondary units, the following two-stage sampling procedure was adopted. All schools sighted were included in the sample. Thus, all solitary animals detected were sampled. Only two whales were sampled at random from a school of more than two animals, i.e., generating two random numbers i and j, the i-th and the j-th whales from the left were chased. The detailed procedures including the way of pursuits are described in Kato *et al.* (1989).

Accordingly, the sampling fractions of whales are decomposed into two factors;

1. sampling fractions of the first stage sampling, i.e., sighting probability given by $2wL/A$, where w is the half of the effective search width, L is the length of the trackline and A is the area of the survey, and

2. sampling fractions of the second stage samplings, i.e., the proportions of whales sampled from the schools sighted.

The sampling fractions were different among schools of different school sizes. Larger schools are easier to be found at the sea, i.e., the sampling fractions of larger schools are larger in the first stage sampling. On the other hand, smaller proportions of whales were sampled from larger schools due to the sampling limitation. To compensate for the effect of this variation, the data was post-stratified by school size.

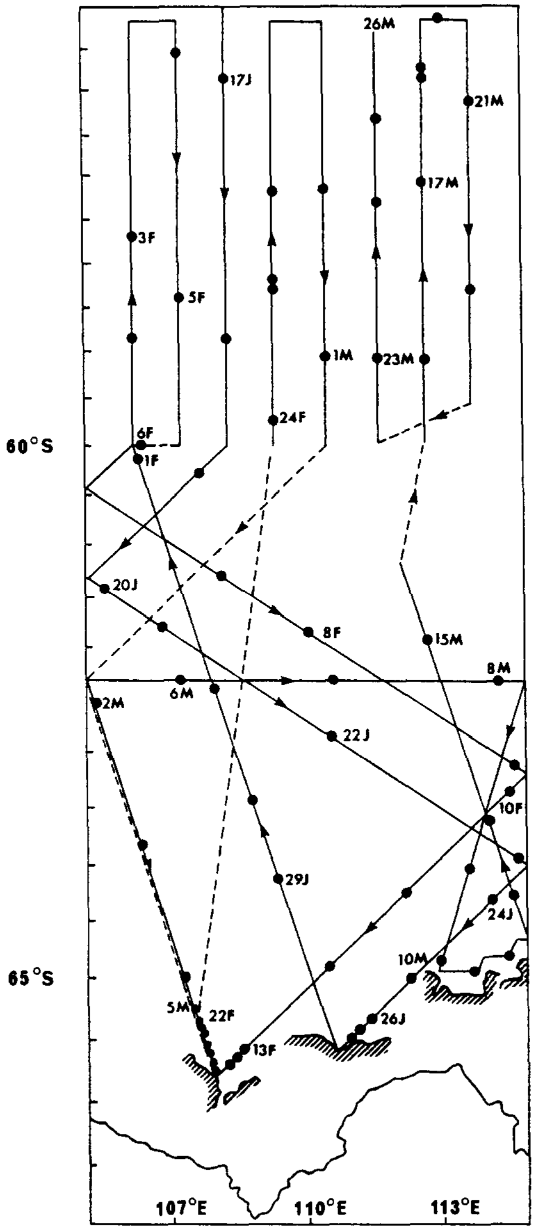

Fig. 1. The trackline of the 1987/88 feasibility study with the marks of dates attached (cited from Kato *et al.* (1989)). The north-south trackline was designed for the northern stratum. In the southern stratum, the trackline was chosen like the trace of a billiard ball with fixed angle (70°) of reflection. The broken lines indicate that the vessels were off-effort. As it is seen from the figure, the vessels proceeded little in the ice-edge area encountered in the mid February. This is because of extraordinary high density of minke whales. Since the total sample size to be collected was limited to 300, it was decided to leave the area and survey the northern stratum on 22nd February, to avoid too much effort from this particular area. After finishing one round in the northern stratum, the vessels returned to the point where the extrapolated line from the trace in the ice-edge area and the western boundary crossed. Then, they surveyed towards south-east. Arriving at the position of 22nd February, they returned along the current line without sampling effort to the boundary. Then, they continued the reflection procedure. Details are explained in Kato *et al.* (1989). From this experience, the length of one day trip have been fixed beforehand in the succeeding surveys.

3. Estimation of population characteristic

3.1 *Population abundance and composition of school size*

The total number of schools was estimated by the usual line transect methodology (Burnham *et al.* (1980), Butterworth (1982), Butterworth *et al.* (1984)). Let $g(y)$ denote the probability that a school of whales at y n.miles in perpendicular distance from the trackline is detected. $g(y)$ is called a detection curve. The half of the effective search width w is obtained by $w = \int_0^c g(y)dy$, where $c > 0$ indicates the limiting perpendicular distance of the schools sighted and adopted for the analysis. The use of the following hazard rate model (Hayes and Buckland (1983), Buckland (1985, 1987)) given by $g(y) = 1 - \exp[-(y/a)^{1-b}]$ is considered for the detection curve. Here, a and b are the free parameters to be estimated.

The sightings data were grouped by perpendicular distance into subintervals $[y_{j-1}, y_j)$ $(j = 1, \dots, J)$. $c = y_J$ was set to 2.0 n.miles, where $J = 20$. The multinomial distribution defined by the probabilities of the sub-intervals obtained from the normalized detection function $f(y) = g(y)/\int_0^{2.0} g(y')dy'$ was fitted to the distribution of the numbers of sightings n_j $(j = 1, \dots, 20)$. Under the assumption that all whales on the trackline should be detected, i.e., $g(0) = 1$, an estimate of w is obtained by $\hat{w} = 1/\hat{f}(0)$, where $\hat{f}(0)$ denotes the fitted value of $f(0)$. If the area is A, the total number of schools in the area is estimated as

$$\hat{P} = \frac{nA}{2\hat{w}L} = \frac{n\hat{f}(0)A}{2L}, \tag{3.1}$$

where n is the number of schools sighted. The mean density of schools is estimated by

$$\hat{D} = \frac{n}{2\hat{w}L} = \frac{n\hat{f}(0)}{2L}. \tag{3.2}$$

Evaluation of sampling variabilities by bootstrap resampling

Some assumptions are required for the evaluation of the variance of the estimates. The density of whales is spatio-temporally inhomogenious and the sighting probability varies due to the conditions of the weather and observers.

From equation (3.2), $\hat{D}$, the estimate of the mean density of schools, is approximated as follows;

$$\hat{D} \sim \frac{E[\hat{f}(0)]}{2}\left(\frac{n}{L} - E\left[\frac{n}{L}\right]\right) + \frac{1}{2}E\left[\frac{n}{L}\right]\left(\hat{f}(0) - E[\hat{f}(0)]\right).$$

Thus, if we can assume that the distribution of the perpendicular distances and the number of the schools sighted are independent to each other, the square of the coefficient of variation of $\hat{D}$ can be expressed as the sum of that of sightings per unit effort n/L and that of the estimate of $f(0)$;

$$CV\{\hat{D}\}^2 = CV\left\{\frac{n}{L}\right\}^2 + CV\left\{\hat{f}(0)\right\}^2. \tag{3.3}$$

We note here that L is also a random variable whose randomness is determined mainly by the weather condition of the day, while the line transect theory usually treats it as constant. Let n_i and L_i be the number of sightings and the length of the trackline of the i-th day respectively ($i = 1, \ldots, k$). Since n/L is a ratio estimator (Cochran (1963), Konijn (1973)), the first term of equation (3.3) is estimated by

$$\begin{aligned}\widehat{CV}\left\{\frac{n}{L}\right\}^2 &= \left(\frac{k}{n}\right)^2 \frac{1}{k(k-1)} \sum_{i=1}^{k} \left(n_i - \frac{n}{L} L_i\right)^2 \\ &= \frac{1}{k}\left(\widehat{CV}\{n_i\}^2 + \widehat{CV}\{L_i\}^2 - 2\widehat{CV}\{n_i, L_i\}\right).\end{aligned}$$

Here $CV\{n_i, L_i\}$ is defined by

$$CV\{n_i, L_i\} = \frac{\mathrm{Cov}\{n_i, L_i\}}{E[n_i]E[L_i]}.$$

As to the second term $CV\{\hat{f}(0)\}^2$ of equation (3.3), the variance estimate of $\hat{f}(0)$ under the assumption of the multinomial distribution for n_j is expected to be negatively biased due to the daily variation of the detection probability.

The bootstrap resampling procedure (Efron (1979), Hall (1988)) is applicable to the evaluation of the variance of $\hat{f}(0)$. Taking into account the daily fluctuation of weather and whale density, we can regard the sighting as a two-stage sampling where the survey area of each day defines the primary unit and the whales (or schools) define the secondary units. The daily data were resampled with replacement, assuming that the variation of the effective search width among different days is much larger than that within a day.

The bootstrap procedure was replicated 100 times. The resultant data were post-stratified into three strata, schools of size 1, 2 and 3, and over 3. The effective search widths were estimated separately and the population size in each stratum of the school size was estimated by (3.1).

Table 1 compares the estimated $f(0)$'s with the bootstrap means. We see that the effective search width of large schools is larger than that of small schools. More precisely, it seems that the detection probabilities of schools of size 1, 2 and 3 are similar but they are significantly less than those of schools of size over 3. In fact, the 95% bootstrap confidence interval of $\hat{w}_{2+3} - \hat{w}_1$ for K01 is $(-0.605, 1.489)$ and contains the point 0, whereas that of $\hat{w}_{>3} - \hat{w}_1$ is $(0.339, 1.750)$. The corresponding confidence intervals for T25 are $(-0.538, 0.461)$ and $(0.478, 1.522)$, respectively.

The estimates are approximately equal to the bootstrap means, i.e., the estimates obtained under the assumption of the multinomial distributions are robust with respect to the error distribution. On the other hand, the estimated standard errors under the assumption of multinomial distributions are consistently smaller than the bootstrap standard deviations and the ratios of the two estimates are roughly between 1 and 2.

The numbers of whales in the research area are obtained by summing up those of various school sizes (Kasamatsu *et al.* (1990)). Table 2 lists the results. The bootstrap means and standard deviations were 8623 and 5170 for the ship K01 and 9463 and 2436 for the ship T25, respectively. It is worth noting that, while the search widths and the numbers of sightings are very different, K01 gives similar estimate of the number of whales as T25.

Table 1. Estimated $f(0)$'s with standard errors. The upper figures were obtained by assuming the multinomial distributions for the distributions of perpendicular distance and the lower figures from the 100 bootstrap samples of the survey days.

School size	K01	T25
1	3.035 (1.012)	1.873 (0.584)
	2.800 (0.886)	1.755 (0.805)
2 or 3	3.420 (1.995)	1.630 (0.344)
	3.690 (3.254)	1.580 (0.353)
over 3	0.554 (0.088)	0.606 (0.082)
	0.638 (0.185)	0.623 (0.093)
combined	1.918 (0.489)	1.190 (0.188)
	1.999 (0.710)	1.246 (0.264)

Table 2. Estimated number of whales. Figures are the bootstrap means and standard deviations (in parentheses).

School size	K01	T25
1	2068 (782)	1853 (983)
2	2266 (2352)	1747 (558)
3	2410 (2738)	2214 (883)
⋮	⋮	⋮
Total	8623 (5170)	9463 (2436)

Bootstrap resampling applied here may be useful for the analysis of correlation structures, which cannot be estimated analytically, among estimates such as the estimated search width, population abundance of school size 1, that of school size 2 and the correlation between the estimated abundances from the data of K01 and of T25.

Bias due to the daily variability of search area

The density estimate (3.2) is given in the form of a ratio estimator which usually has some bias (Cochran (1963), Konijn (1973)). Let $\bar{D}$ be defined by $E[n] \,/\, E[wL]$. Then the relative bias of $\hat{D}$ with respect to $\bar{D}$ can be expressed as

$$\frac{E[\hat{D} - \bar{D}]}{\bar{D}} = CV\left\{wL\right\}^2 - CV\left\{n, wL\right\}.$$

This bias will be reduced as the number of days k is increased.

Now, we consider the bias of $\bar{D}$ with respect to the average of the mean density D in the area A. Let w_i be the half effective search width of the i-th day ($i = 1, \ldots, k$). The number of sightings and the length of trackline of the i-th day are denoted by n_i and L_i, respectively. We use the term "average" to denote the expectation in the superpopulation. w_i, n_i, L_i ($i = 1, \ldots, k$) are random variables mainly affected by the weather conditions. We assume that (w_i, n_i, L_i) ($i = 1, \ldots, k$) are mutually independently and identically distributed. Then, $\bar{D}$ can be expressed as

$$\bar{D} = \frac{E[n]}{E[wL]} = \frac{E[n_i]}{E[w_i L_i]}.$$

Denoting the local density in the research area of the i-th day by D_i, we get $\bar{D} = E[D_i w_i L_i]/E[w_i L_i]$. Since it holds that $E[D_i w_i L_i] = E[D_i]E[w_i L_i] + \mathrm{Cov}\{D_i, w_i L_i\}$, we have

$$\bar{D} = E[D] + \frac{\mathrm{Cov}\{D_i, w_i L_i\}}{E[w_i L_i]}.$$

When there is a positive correlation between $w_i L_i$ and D_i, i.e., when the search area is larger in high density area than in low density area, $\bar{D}$ is larger than $E[D]$, and vice versa.

Table 3. The correlation coefficients of the estimated number of schools $\hat{P}$, estimated half strip width $\hat{w}$ and length of the trackline L. The values were obtained by 100 bootstrap replications of the resamplings from the daily data.

a) K10

	$\hat{P}$	$\hat{w}$	L
$\hat{P}$		−0.695	−0.220
$\hat{w}$			−0.229
L			

b) T25

	$\hat{P}$	$\hat{w}$	L
$\hat{P}$		−0.662	−0.323
$\hat{w}$			−0.278
L			

To investigate the correlations of the estimated number of schools ($\hat{P}$), the estimated half effective strip width ($\hat{w}$) and the length of the trackline (L), we resample the data daily and calculate the three statistics from the resampled data. Correlation coefficients are obtained from a set of 100 bootstrap replications. The fluctuation of the estimated $\hat{P}$ can be regarded as that of the local density near the trackline. The results show that both $\hat{w}$ and L are negatively correlated to $\hat{P}$, and the correlation coefficients are similar for the two vessels (Table 3). It was observed that this was due to the fact that the observer's search efforts were concentrated in closer range in high density area (Kasamatsu *et al.* (1990)). The bootstrap expectation and standard deviation of estimated number of schools are 3671 and 1391 for K01 and 3708 and 1079 for T25, whereas the estimates from the original data are 3443 for K01 and 3380 for T25. The slightly larger values of bootstrap expectations, in spite of the negative correlation between $\hat{P}$ and $\hat{w}L$, may be explained by the possible difference of $g(0)$ between the normal search condition and that in the high density area.

3.2 *Age distribution*

We estimate here the age distribution in the population. In the sampling procedure adopted in the present study, larger schools tend to be sighted easier than smaller ones. On the other hand, smaller proportion of whales are sampled from larger schools sighted. To control the effect of the unequal sampling probabilities of whales, the data were post-stratified into three classes by school size, i.e., into the strata of schools of sizes 1, 2 and over 2.

The proportion of a particular age class $\bar{X}$ in the population is estimated as the weighted average of the estimates $\bar{x}_i$ $(i = 1, 2, 3)$ in the three strata:

$$\bar{x} = \sum_{i=1}^{3} u_i \bar{x}_i.$$

The sizes of the strata give the weights:

$$u_i = i\hat{P}_i \Big/ \sum_{j\geq 1} j\hat{P}_j \qquad (i = 1, 2),$$

$$u_3 = \sum_{i\geq 3} i\hat{P}_i \Big/ \sum_{j\geq 1} j\hat{P}_j.$$

Here $\hat{P}_i$ is the weighted average of the two estimates of the number of schools in the area A obtained from the data of K01 and T25, with the weights respectively proportional to the inverses of their variances. Table 4 lists the bootstrap means and the variance-covariance matrix of u_i's. $\bar{x}_i$ is estimated as the weighted average of the sample proportions within the schools collected $\bar{x}_{is}$:

$$\bar{x}_i = \frac{\sum_{s=1}^{S_i} M_{is}\bar{x}_{is}}{\sum_{s=1}^{S_i} M_{is}}$$

where M_{is} is the school size of the s-th school and S_i denotes the number of schools in the sample from the i-th stratum.

Table 4. Bootstrap means (left) and the variance-covariance matrix of the weights u_i $(i = 1, 2, 3)$ of the strata of school size 1, 2 and over 3. The values were obtained by resampling 100 times from the daily data.

	Means	u_1	u_2	u_3
u_1	0.259	0.00826	−0.00349	−0.00476
u_2	0.225		0.00429	−0.00080
u_3	0.517			0.00557

Figures 2(a), (b) and (c) show the age distributions estimated by this procedure for the three strata. It can be seen that younger whales dominate the stratum of school size 1, and the proportion of older whales increases as the school size increases.

In the estimation of the variances of the estimated parameters in the strata, we regard the sampling of schools as a simple random sampling. Then the characteristic of each sample can be expressed as follows:

$$x_{isj} = \bar{X}_i + \xi_{is} + \eta_{isj},$$
$$i = 1, 2, 3; \quad s = 1, \ldots, S_i; \quad j = 1, \ldots, J_s$$

where the subscripts i, s and j denote the stratum, school in the stratum and whale in the school, respectively, and $\bar{X}_i$ denotes the mean within the stratum. J_s is 1 or 2. ξ_{is} and η_{isj} are independent random variables that represent the fluctuations between schools and between whales. For the purpose of the analysis, the sampling without replacement is treated as a sampling with replacement. When it is expected that there is the variation of biological parameters of the scale of days or so, it should also be incorporated into the model. Under this model, the variance of $\bar{x}_i$ is estimated by

$$\widehat{\mathrm{Var}}\{\bar{x}_i\} = \frac{1}{S_i(S_i-1)} \sum_{s=1}^{S_i} \frac{M_{is}^2}{\bar{M}_i^2} (\bar{x}_{is} - \bar{x}_i)^2,$$

where $\bar{M}_i$ is the sample mean of the school size in the i-th stratum. Assuming the independence between the estimated composition of school size and the estimated biological parameters, the variance of $\bar{x}$ is estimated by

$$\widehat{\mathrm{Var}}\{\bar{x}\} = \sum_{i=1}^{3} \sum_{k=1}^{3} \widehat{\mathrm{Cov}}\{u_i, u_k\} \bar{x}_i \bar{x}_k + \sum_{i=1}^{3} \widehat{\mathrm{Var}}\{\bar{x}_i\} \left(\widehat{\mathrm{Var}}\{u_i\} + u_i^2 \right).$$

Table 5 shows the estimated age composition with standard errors. The age composition of the whales caught by the commercial whaling had a peak around 10 to 15 years old. However, this result also provides a strong evidence for the fact that there are many young whales in the Antarctic in summer. The estimated standard errors are much larger than those obtained under the assumption of simple random sampling of individual whales.

Table 5. The estimated age distribution in the population with standard errors.

Age class	Proportion
1– 5	0.3346 (0.0545)
6–10	0.1594 (0.0293)
11–15	0.1403 (0.0267)
16–20	0.1579 (0.0278)
21–25	0.1003 (0.0228)
26–30	0.0734 (0.0196)
31–35	0.0062 (0.0037)
36–40	0.0217 (0.0117)
41–45	0.0060 (0.0067)

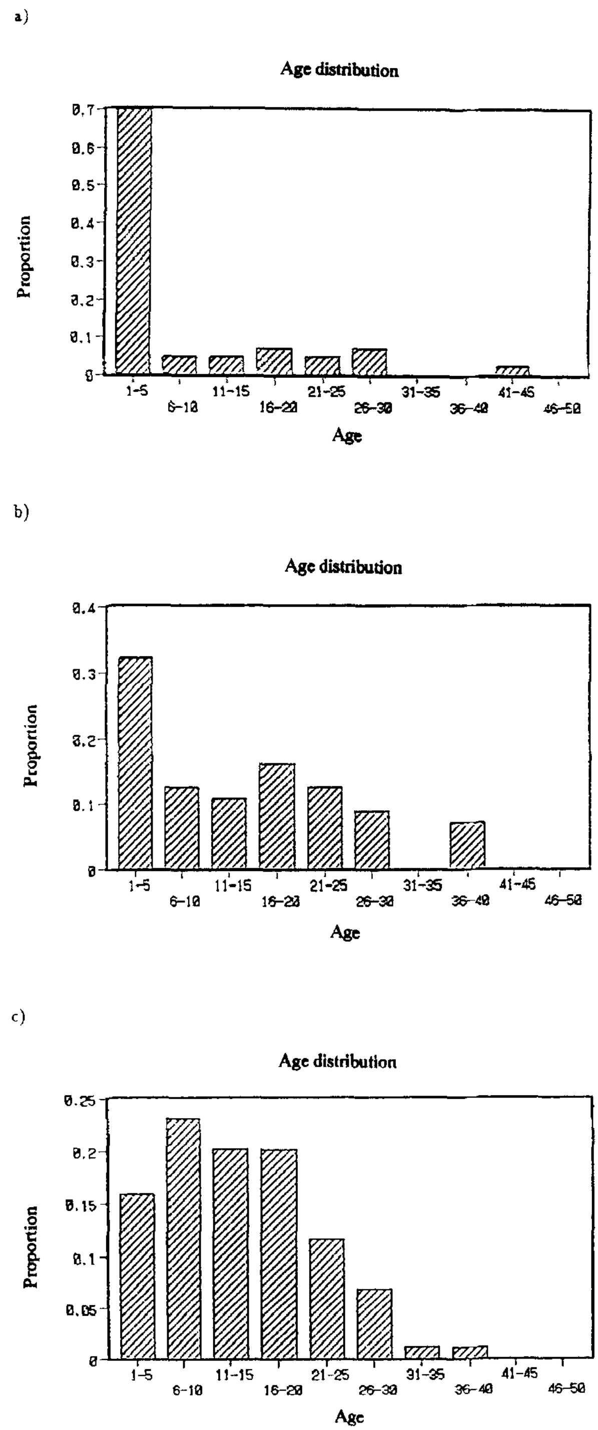

Fig. 2. The age distributions in the strata of school size 1 (a), 2 (b), and over 2 (c)

4. Heterogeneity of age composition among different school sizes and areas

As it was seen in the previous section, the proportion of older whales is larger in the strata of larger school size. On the other hand, it is known that the school size becomes large in the area near the ice-edge. To investigate the factor that mainly controls the age distribution, we further post-stratified each stratum defined by school size by two areas, the area within about 60 n.miles from the ice edge and the area off shore.

The sizes of sub samples and the estimated mean ages ($\bar{x}_{ij}$; where i $(= 1, 2, 3)$ is the subscript for school size and j $(j = 1, 2)$ for area) in the strata are shown with standard errors ($\hat{\sigma}_{ij}$) in Tables 6 and 7. As for the strata of school size over 2, the $\bar{x}_{ij}$'s were obtained as the simple averages of the sample mean within the schools. Conditional on a given school size, difference in age between areas appears small. However, we see the heterogeneity among different school sizes, especially between the solitary whales and those belonging to schools. We investigate this heterogeneity through the process of statistical model selection.

Table 6. The numbers of the schools (whales) of the sub-samples in the strata classified by school size and area.

School size	Off shore	Near ice edge
1	27 (27)	16 (16)
2	13 (19)	15 (22)
over 2	10 (13)	91 (152)

Table 7. The estimated mean ages in the strata classified by area and school size. The numbers in the parentheses are the standard errors.

School size	Off shore	Near ice edge
1	6.778 (1.896)	10.250 (2.616)
2	14.269 (3.261)	13.733 (1.824)
over 2	11.450 (2.397)	14.319 (0.703)

4.1 *Model selection*

Suppose that the strata divided by school size and area are classified into L groups as;

$$
\begin{array}{ll}
\text{1st group:} & (i_{11}, j_{11}), \ldots, (i_{1m_1}, j_{1m_1}), \\
\text{2nd group:} & (i_{21}, j_{21}), \ldots, (i_{2m_2}, j_{2m_2}), \\
\vdots & \quad \vdots \qquad\qquad\quad \vdots \\
L\text{-th group:} & (i_{L1}, j_{L1}), \ldots, (i_{Lm_L}, j_{Lm_L}),
\end{array}
$$

and assume that the means or the proportions μ_{ij} are the same within each group, that is,

$$\begin{aligned}
\mu_{i_{11},j_{11}} &= \cdots = \mu_{i_{1m_1},j_{1m_1}} = \mu_1,\\
\mu_{i_{21},j_{21}} &= \cdots = \mu_{i_{2m_2},j_{2m_2}} = \mu_2,\\
&\vdots \qquad\quad \vdots\\
\mu_{i_{L1},j_{L1}} &= \cdots = \mu_{i_{Lm_L},j_{Lm_L}} = \mu_L.
\end{aligned}$$

From the central limit theorem, $\bar{x}_{ij}$ $(i = 1,2,3;\ j = 1,2)$ will follow a normal distribution approximately. The maximum log likelihood (MLL) of this model of $\bar{x}_{ij}$ $(i = 1,2,3;\ j = 1,2)$ is expressed as;

$$\begin{aligned}
\text{(4.1)}\qquad \mathrm{MLL} = &-3\log 2\pi\\
&-\frac{1}{2}\sum_{l=1}^{L}\sum_{h=1}^{m_l}\left\{\log\hat{\sigma}^2_{i_{lh}j_{lh}} + \frac{(\bar{x}_{i_{lh}j_{lh}} - \hat{\mu}_l)^2}{\hat{\sigma}^2_{i_{lh}j_{lh}}}\right\},
\end{aligned}$$

where $\hat{\mu}_l$ $(l = 1,\dots,L)$ are obtained as the averages of $\bar{x}_{i_{lh}j_{lh}}$'s within the groups with the weights inversely proportional to $\hat{\sigma}^2_{i_{lh}j_{lh}}$.

When there are several alternative models of the heterogeneity of the age distribution, the model which has the minimum AIC (Akaike's Information Criterion; Akaike (1973, 1985); Sakamoto *et al.* (1986)) defined by

$$\begin{aligned}
\mathrm{AIC} = &-2\times(\text{Maximum log likelihood of the model})\\
&+2\times(\text{Number of the free parameters of the model})
\end{aligned}$$

is selected as the best model.

Our confidence of the selection of the best model depends on the magnitude of the differences among the AIC's of the alternative models. Suppose that there are m alternative models $1,\dots,m$ and their AIC values are $\mathrm{AIC}(k)$ $(k = 1,\dots,m)$, respectively. If the prior probability p_k is given to Model k, the posterior probability of Model i after obtaining the data is defined by

$$\begin{aligned}
\hat{\mathrm{P}}(i|\text{data}) &= \frac{p_i\exp\left(-\frac{1}{2}\mathrm{AIC}(i)\right)}{\sum_{k=1}^{m}p_k\exp\left(-\frac{1}{2}\mathrm{AIC}(k)\right)}\\
&= \frac{1}{1+\sum_{k\neq i}\frac{p_k}{p_i}\exp\left(\frac{1}{2}(\mathrm{AIC}(i)-\mathrm{AIC}(k))\right)},
\end{aligned}$$

where the likelihood of each model is estimated by $\exp(-(1/2)\mathrm{AIC})$.

There are L parameters concerning means. Since the term outside and the first term in the bracket of equation (4.1) are common among the models, it can be neglected when comparing the models. Therefore, it is enough to calculate the following statistics

$$\text{(4.2)}\qquad \mathrm{AIC}^* = \sum_{l=1}^{L}\sum_{h=1}^{m_l}\frac{(\bar{x}_{i_{lh}j_{lh}} - \hat{\mu}_l)^2}{\hat{\sigma}^2_{i_{lh}j_{lh}}} + 2L.$$

When $\bar{\boldsymbol{x}}_{ij}$ $(i = 1,2,3;\ j = 1,2)$ are vectors such as the age distribution, equation (4.2) becomes

$$\mathrm{AIC}^* = \sum_{l=1}^{L}\sum_{h=1}^{m_l}(\bar{\boldsymbol{x}}_{i_{lh}j_{lh}} - \hat{\boldsymbol{\mu}}_l)^T \hat{\Omega}_{i_{lh}j_{lh}}^{-1}(\bar{\boldsymbol{x}}_{i_{lh}j_{lh}} - \hat{\boldsymbol{\mu}}_l) + 2LM,$$

where M is the dimension of the vectors and Ω_{ij} is the variance-covariance matrix of $\bar{\boldsymbol{x}}_{ij}$. It is important in the analysis of such tables that the variance covariance structure should be estimated adequately, particularly when we analyze the data obtained by a field survey (Scott and Rao (1981)).

4.2 *Result*

We compare the following four models listed in Table 8. Model 1 assumes that the mean ages are the same among the different school sizes and areas. Model 2 assumes the homogeneity between the two areas, but allows the difference between the solitary whales and those in schools. Model 3 assumes the homogeneity among different school sizes, but allows the difference between the two areas. Model 4 considers, first, the heterogeneity between the solitary whales and those in schools, and second, the difference between the ice edge area and off shore area for the whales in schools.

Table 8. AIC*'s and the posterior probabilities of the models of heterogeneity of the mean age among different school sizes and areas, given the equal prior probabilities. The first and second components in the parentheses denote the school size class and sampling area, respectively.

	Grouping of strata	AIC*	Posterior
Model 1	All strata combined	18.00	0.002
Model 2	1'st group: (1,1) + (1,2) 2'nd group: (2,1) + (2,2) + (3,1) +(3,2)	6.51	0.679
Model 3	1'st group: (1,1) + (2,1) + (3,1) 2'nd group: (1,2) + (2,2) + (3,2)	11.14	0.067
Model 4	1'st group: (1,1) + (1,2) 2'nd group: (2,1) + (2,2) 3'rd group: (3,1) + (3,2)	8.50	0.251

The values of AIC*'s in respective models were 18.00 (Model 1), 6.51 (Model 2), 11.14 (Model 3), and 8.50 (Model 4). Therefore, Model 2 was selected as the best model among the four from this data. The posterior probabilities of the four models were 0.002, 0.679, 0.067 and 0.251 respectively for equal probability prior distribution.

We see from Table 9 that the proportions of female whales are significantly smaller in the area off shore compared with those in the area near ice edge. Although we see a mild tendency of declining of proportion of female in larger school sizes, it will be difficult to judge it as significant from the present data. Analyzing the data in Table 9 in the same way as the mean ages (Table 8), we get AIC* values 10.41 (Model 1), 12.40 (Model 2), 5.96 (Model 3) and 14.39 (Model 4). This result supports strongly, with posterior probability 0.860, Model 3 which takes account of the difference between two areas.

Table 9. The estimated proportions of female and the standard errors in the strata classified by area and school size.

School size	Off shore	Near ice edge
1	0.3548 (0.0874)	0.5294 (0.1248)
2	0.3000 (0.1069)	0.5278 (0.0946)
over 2	0.2000 (0.1106)	0.4516 (0.0419)

The means and the standard errors of age for male and female are shown in Table 10. We see no significant differences of mean ages of male whales among different areas or different school sizes, whereas the mean age of female whales seems to be younger in the area off shore than those in the area near the ice edge.

Table 10. The estimated mean ages of male and female in the strata classified by area and school size. The numbers in the parentheses are the standard errors.

a) Male

School size	Off shore	Near ice edge
1	9.412 (2.836)	13.375 (4.136)
2	15.818 (3.643)	10.800 (2.851)
over 2	12.944 (2.212)	15.285 (0.913)

b) Female

School size	Off shore	Near ice edge
1	2.300 (0.423)	7.125 (3.073)
2	4.800 (2.596)	15.167 (2.841)
over 2	2.333 (0.667)	13.364 (0.998)

The AIC* values of the four models for male were 8.24 (Model 1), 7.63 (Model 2), 8.42 (Model 3) and 8.76 (Model 4). The posterior probabilities of the four models were 0.248, 0.336, 0.226 and 0.191 respectively for the equal prior probabilities. We cannot select one particular model as best confidently. On the other hand, the AIC*'s for female data were 128.58 (Model 1), 101.71 (Model 2), 9.28 (Model 3) and 100.13 (Model 4), and Model 3 could be selected as a best model without any hesitation. It was already noted by Kato *et al.* (1989) that solitary whales seemed to be younger than whales in a school.

The proportion of female whales has been estimated around 0.6 from the data of commercial whaling (Kato (1982)). This figure is somewhat larger than the result obtained from the present study (Table 9). As it was mentioned in Introduction, the whaling grounds have been located near the ice edge. Table 9 shows that the proportion of female seems to be higher in the area near the ice edge than those in the off shore. However, even if we consider only the strata of the area near the ice edge, the figures are somewhat smaller than 0.6. This may be partly due to the fact that female whales are larger than male whales by 1 or 2 feet and as a result females were caught more in the commercial whaling.

5. Discussion

We have seen that the biological characteristics of whales are highly heterogeneous spatially. Therefore, a survey should be designed to collect sample from the whole area uniformly to make unbiased estimation of population characteristics. The line transect procedure will satisfy this.

However, it is difficult to keep the sampling fractions precisely equal in a field survey. In the present study, larger schools were more easily detected than smaller ones and in those schools smaller proportions of whales were sampled. It is important to detect the heterogeneity of the characteristics under study and poststratify the data corresponding to the heterogeneity, particularly when the sampling fractions are correlated to the heterogeneity. The methodology of the estimation and model evaluation presented in this paper will provide the basis of the investigation of biological characteristics from a field survey.

From the analysis, it turned out that the heterogeneity in area was also significant. On the other hand, the sampling fractions were suspected to be higher in the ice edge area, because of the good weather conditions. Since the bootstrap procedure required heavy computation task, we could not re-estimate the population characteristics, post-stratifying by school size and location. However, in the analyses of the data from the succeeding surveys both of these heterogeneities have been taken into account (Fujise *et al.* (1990), Kato *et al.* (1991)).

In the survey, non-negligible number of whales which were selected to be samples were missed during the pursuit. Herein arises the problem of the non-response error. It was reported (Kato *et al.* (1989)) that no significant difference in the observed body length was seen between the missed whales and the collected whales, and no significant difference between the observed body length and the measured one was detected. However, further study and effort to minimize the possibility of the bias from this problem is required.

We see some apparent inconsistencies in the data and results associated with Tables 7–10. For example, average age for the off shore area for schools of size 2 is 14.3 for both sexes, 15.8 for males and only 4.8 for females. The proportion of females is 0.3. Thus $15.3 \times 0.7 + 4.8 \times 0.3$ gives an average age of 12.5, substantially below 14.3. The mean age in the stratum for both sexes, μ is expressed as $\mu = \rho\mu_f + (1-\rho)\mu_m$ where ρ is the proportion of female whales, and μ_f and μ_m are the mean ages of female and male whales in the stratum respectively. We should note, however, that the estimate μ obtained by $\hat{\mu} = \hat{\rho}\hat{\mu}_f + (1-\hat{\rho})\hat{\mu}_m$ could be biased because of the possible correlation between $\hat{\rho}$ and $\hat{\mu}_f$, $\hat{\mu}_m$. This correlation might be related to the school structure. However, the confirmation of this fact is left for future study.

Acknowledgements

Dr. H. Akaike and Dr. T. Nakamura read the original manuscript and gave us useful comments. We also appreciate useful suggestions of Dr. Y. Taga and Dr. N. Kashiwagi. Mr. Y. Iwata and Mr. H. Sugimoto helped our analysis with programming and we thank them. The elaborated comments by the three referees improved the manuscript very much.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csaki), 267–281, Akadémiai Kiadó, Budapest.

Akaike, H. (1985). Prediction and entropy, *A Celebration of Statistics, the ISI Centenary Volume* (eds. A. C. Atkinson and S. E. Fienberg), 1–24, Springer, New York.

Barlow, J. (1990). The utility of demographic models in marine mammal management, Paper SC/42/Mg16 (submitted to the 42nd IWC Scientific Committee, June 1990).

Buckland, S. T. (1985). Perpendicular distance models for the line transect sampling, *Biometrics*, **41**, 177–195.

Buckland, S. T. (1987). An assessment of the performance of line transect models for fitting IWC/IDCR cruise data, 1978/79 to 1984/85, *Report of the International Whaling Commission*, **37**, 277–279.

Burnham, K. P., Anderson, D. R. and Laake, J. L. (1980). Estimation of density from line transect sampling of biological populations, *Wildlife Monograph*, No. 72.

Butterworth, D. S. (1982). A possible basis for choosing a functional form for the distribution of sightings with right-angle distance: some preliminary ideas, *Report of the International Whaling Commission*, **32**, 555–558.

Butterworth, D. S., Best, P. B. and Hembree, D. (1984). Analysis of experiments carried out during the 1981/82 IWC/IDCR Antarctic minke whale assessment cruise in Area II, *Report of the International Whaling Commission*, **34**, 365–392.

Cochran, W. G. (1963). *Sampling Technique*, 2nd ed., Wiley, New York.

Cooke, J. G. (1989). Simulation studies of two whale stock management procedures, *The Comprehensive Assessment of Whale Stocks: The Early Years* (Special Issue **11** of *Report of the International Whaling Commission* (ed. G. P. Donovan)), 147–156.

De la Mare, W. K. (1989). Further simulation studies on management procedures, *The Comprehensive Assessment of Whale Stocks: The Early Years* (Special Issue **11** of *Report of the International Whaling Commission* (ed. G. P. Donovan)), 157–170.

Efron, B. (1979). Bootstrap method: another look at the jackknife, *Amer. Statist.*, **7**, 1–26.

Fujise, Y., Kato, H. and Kishino, H. (1990). Reproductive segregation of the minke whale population in high latitudinal waters with some estimations of pregnancy and sexual maturity rates, data from Japanese research takes in 1987/88 and 1988/89, Paper SC/42/SHMi10 (submitted to the 42nd IWC Scientific Committee, June 1990).

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion), *Ann. Statist.*, **16**, 927–985.

Hayes, R. J. and Buckland, S. T. (1983). Radial distance models for the line transect method, *Biometrics*, **39**, 29–42.

International Whaling Commission. (1983). Chairman's Report of the 34th Annual Meeting, *Report of the International Whaling Commission*, **33**, 20–42.

Kasamatsu, F., Kishino, H. and Hiroyama, H. (1990). Estimations of the number of minke whale schools and individuals based on the 1987/88 Japanese feasibility study data, *Report of the International Whaling Commission*, **40**, 239–248.

Kato, H. (1982). Some biological parameters for the Antarctic minke whale, *Report of the International Whaling Commission*, **32**, 935–945.

Kato, H., Hiroyama, H., Fujise, Y. and Ono, K. (1989). Preliminary report of the feasibility study on southern minke whale under the Japanese Proposal to the special permit, *Report of the International Whaling Commission*, **39**, 235–248.

Kato, H., Kishino, H. and Fujise, Y. (1990). Some analyses on age composition and segregation of southern minke whale using samples obtained from the Japanese Feasibility Study in 1987/88, *Report of the International Whaling Commission*, **40**, 249–256.

Kato, H., Fujise, Y. and Kishino, H. (1991). Preliminary analysis on age and reproductive data of southern minke whales obtained from the Japanese research take in 1988/89, *Report of the International Whaling Commission*, **41** (in press).

Konijn, H. S. (1973). *Statistical Theory of Sample Survey Design and Analysis*, North-Holland, Amsterdam.

Magnusson, K. G. and Stefansson, G. (1989). A feedback strategy to regulate catches from a whale stock, *The Comprehensive Assessment of Whale Stocks: The Early Years* (Special Issue **11** of *Report of the International Whaling Commission* (ed. G. P. Donovan)), 171–190.

Nakamura, T. (1990). A new look at the Bayesian cohort model for time series data obtained from research take of whales, Paper SC/42/SHMi17 (submitted to the 42nd IWC Scientific Committee, June 1990).

Punt, A. E. and Butterworth, D. S. (1989). Results of first stage screening trials for a proposed whale stock management procedure, *The Comprehensive Assessment of Whale Stocks: The Early Years* (Special Issue **11** of *Report of the International Whaling Commission* (ed. G. P. Donovan)), 191–198.

Sakamoto, Y., Ishiguro, M. and Kitagawa,G. (1986). *Akaike Information Criterion Statistics*, Reidel, Dordrecht.

Sakuramoto, K. and Tanaka, S. (1989). A simulation study on management of whale stocks considering feedback systems, *The Comprehensive Assessment of Whale Stocks: The Early Years* (Special Issue **11** of *Report of the International Whaling Commission* (ed. G. P. Donovan)), 199–210.

Scott, A. J. and Rao, J. N. K. (1981). Chi-squared tests for contingency tables with proportions estimated from survey data, *Current Topics in Survey Sampling* (eds. D. Krewski, R. Platek and J. N. K. Rao), 247–266, Academic Press, New York.

Tanaka, S. (1990). Estimation of natural mortality coefficient of whales from the estimates of abundance and age composition data obtained from research catches, *Report of the International Whaling Commission*, **40**, 531–536.