Bias in density estimates when the speed of whale sampling vessels is reduced in high density areas

S. TANAKA (Institute of Cetacean Research)

Abstract

In order to cover the entire research area within an allocated time period, the Protocol of JARPA fixes a distance sampling vessel should proceed each day. Due to this procedure, some part of the track lines would be left unobserved because the proceeding speed of vessel is inevitably reduced in a high whale density area. It is considered that the unobserved area causes a bias in the estimate of whale density. In this paper, effects of size of high density area and reduced speed of vessel on the estimate of whale density are examined on the basis of a simple model.

The space is expressed by a straight line of one dimension, and the length of distance planned to proceed each day is taken as a unit of length. One high density area of length w (<1) is included in the unit space. Whales are distributed homogeneously in each of high and low density area and the density is d+b and b, respectively. The position of high density area x is expressed by the left side end point of the area. The distribution of x is homogeneous within 0 - 1 range. If high density area is not included in the space, sampling vessels can proceed one length unit within one time unit or working hour in a day (speed=1). In a high density area the speed is reduced to v(<1). The length of actual observed area in a day is ℓ(<1) and the number of schools (or whales) observed (or taken) is n.

For calculating expectations E(n/ℓ) and E(n)/E(ℓ), n and ℓ are expressed as functions of x, and n/ℓ, n and ℓ are integrated by x over a range of 0 - 1. These expectations include parameters of w, v, d and b. The estimator proposed by Burt & Borchers (1997) corresponds to E(n/ℓ), while the formula used for JARPA corresponds to E(n)/E(ℓ).

E(n/ℓ) has a positive bias regardless of the values of w and v. On the contrary, E(n)/E(ℓ) gives either positive or negative bias depending on a combination of w and v values. Some data obtained by JARPA in Area V suggest that possible ranges of parameter values may roughly be w=0.1-0.2, v=0.05-0.1 and dw/b=3-10. Within these ranges of parameter value, the bias of E(n/ℓ) could be 25-80%, and E(n)/E(ℓ), -50-+5%. The ratio of these two estimators would be as high as 2.

The bias in E(n)/E(ℓ) is considered to be connected with a distorted sampling ratio between high and low density areas. Possible negative bias in the JARPA estimates suggests that the sampling ratio is lower in high density areas than in low density areas.

It is very important to develop estimating methods free from biases due to unobserved area and uneven sampling ratios. It would also be necessary to formulate a procedure of survey which would not require sampling vessels to cruise in the dark without surveying.

I. Introduction

The fact that the estimates of the population abundance by sighting in JARPA stays about half of the estimates obtained by the IDCR is now at issue (Nishiwaki et al, 1997). One possible cause of this phenomenon may be in the procedure of JARPA that, in order to cover the research area thoroughly within a limited interval of research days, proceeding distance per day is set, and when this distance is not covered in a day, the vessel is obliged to cruise during night-time so that it can start from pre-determined point the next morning (Nishiwaki et al., 1997). When a high density area is encountered, proceeding speed is reduced, and a distance sampling vessel actually proceed each day is shortened. The higher the density of whales the longer the distance of idle cruising and sighting effort (proceeding distance) is reduced. For this reason, underestimation could occur as a whole.

This issue cannot be overlooked also in terms of representativeness of samples. Reduction of effort in high density area means that sampling ratio in low density areas gets relatively higher, and biological parameters in low density areas are emphasized in the samples. If there are differences in biological parameters such as sex ratio and age composition between high density areas and low density areas, there is a possibility that bias is introduced in the estimates of biological parameters.

The standard density estimation method used by the IWC is that, firstly sea areas are stratified and then the total number of whales sighted (in actuality the number of schools) is divided by total distance searched (actual proceeding distance), in each stratum. Denoting daily number of whales sighted by n_i, and actual proceeding distance by l_i,

$$\text{Density} = \sum_i n_i \,/\, \sum_i l_i$$

is obtained (Nishiwaki et al. 1997). This method means to average n_i / l_i using l_i as weight. If n_i / l_i and l_i are correlated, the density estimates will have bias.

As one method to eliminate this bias, the following formula can be considered:

$$\left(\sum_i^M n_i/l_i\right) / M \qquad \text{or} \qquad \left(\sum_i^M L_i n_i/l_i\right) / \sum_i^M L_i$$

Here M stands for the number of research days, and L_i stands for planned daily proceeding distance. If daily L_i is constant, the two formulae are identical. The latter formula was used by Burt & Borchers (1997). When this formula is used, an estimate from JARPA gives rather similar value as that from IDCR.

The formula used by Burt & Borchers is based on the assumption that n_i / l_i represents the density of all the space to be surveyed on the day. In other words, the portion already passed is supposed to be a random sample from the whole space. However, the high density area is a continuos space and the portion of space already passed is naturally continuous. Therefore, there is a possibility that it does not become random sample. In such a case, bias will be introduced in estimates. In this paper, characteristics of n/ℓ is examined by means of a simple model within the scope treatable analytically and compared with $\Sigma n/\Sigma\ell$

II. Model for sampling operation

Space: pre-determined proceeding distance per day is taken as the unit of length and set as 1 in single-dimension space.

Time: time required to pass the unit space in case the schools of whales are distributed uniformly at density b is set as 1.

Searching survey: research vessels start moving from the left end of the space and proceed to the right. Whales are counted on the basis of the number of schools.

Whale density: density in ordinary area is b and is constant. (The number of schools sighted in the unit time and distance is set as b unit)

High density area: high density areas of the length w are arranged at certain intervals (1-w) in the infinite space. The starting point of sighting vessels is determined arbitrarily and is taken as the origin of x axis. The position at the left end of the high density area is set as x in the space 0-1.When part of the two adjoining high density areas is located on each of the two ends of this unit space, the position of the left end of the right side area is taken as x. The proportion of high density area in the unit space is w. Distribution density in high density area is constant at (b+d).

Proceeding speed: the speed in the high density area is v times that in the ordinary area ($0 \leqq v \leqq 1$).

Average density for the entire area: $b(1-w)+(b+d)w=b+dw\equiv dm$. When whales are distributed only in high density area, b=0, and the density in the high density area is d.

Discovery process: all the schools on the sighting track are discovered, and there is no missing.

III. Actual proceeding distance per day ℓ, number of schools sighted n and average density n/ℓ

3.1 In case where high density area is covered completely when the survey in a day is ended

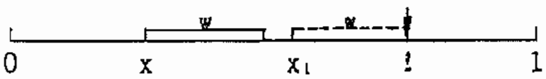

Fig. 1

Assuming that the survey is ended just at the right end of the high density area, and denoting x in this case as x_1,

$$x_1 = \ell - w$$

Neither n nor ℓ changes within the range of $0 \leqq x \leqq x_1$.

$$\left.\begin{aligned} n_1 &= b(1-w/v) + (b+d)w = b(1-w/v+w) + dw \\ \ell_1 &= 1 - w/v + w \end{aligned}\right\} \quad (1)$$

Substituting ℓ_1 into formula of x_1,

$$x_1 = 1 - w/v \qquad (2)$$

If $w/v>1$, this case does not occur because sighting vessels cannot pass the high density area even if $x=0$.

3.2 In case a part of the high density area is left unsurveyed

Two cases are considered: (1) where the high density area is continuous; and (2) where it is separated into two parts at each end of the space.

(1) Where one continuous high density area is exist

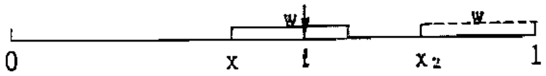

Fig. 2

Let's denote x by x_2 when the right end of the high density area touches the right end of the unit space.

$$x_2 = 1 - w \qquad (3)$$

In the section of $x_1<x<x_2$,

$$\left.\begin{aligned} n &= bx + (b+d)v(1-x) = b(x+v(1-x)) + dv(1-x) \\ \ell &= x + v(1-x) = v + (1-v)x \end{aligned}\right\} \quad (4)$$

Formulae for n and ℓ when $x=x_2$ are obtained by putting $x=(1-w)$ in (4), and

$$\left.\begin{aligned} n_2 &= b(1-w) + (b+d)wv = b(1-w+wv) + dwv \\ \ell_2 &= 1 - w + wv \end{aligned}\right\} \quad (5)$$

When x takes other value than x_2, it always becomes $\ell < \ell_2$. ℓ_2 is the maximum possible proceeding distance.

This case could occur in either $w/v \leqq 1$ or $w/v>1$.

(2) In case where the high density area is separated to the right and left ends of the space

There are three possible cases when the high density area is separated to the right and left ends: ① the case where the vessel is in the right side section of the high density areas,② the case where the vessel is in the ordinary area between the two high density areas, and ③ the case where the vessel is in the left side section of the high density areas.

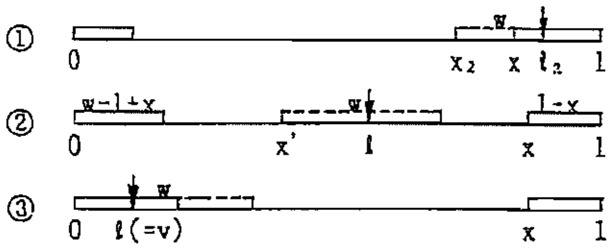

Fig. 3

As seen in 3.1, shift of the passed portion of the high density area will not modify n and ℓ as far as it is in the left side of the arrow mark in Fig. 3 . Similarly, unpassed portion of the high density area will not modify n and ℓ even if it is moved within the space right of the arrow mark. When this principle is applied to each case, it is evident that each of the cases is identical with the case shown by dotted line in Fig. 3.

① is identical with the case when $x=x_2$. n and ℓ become n_2 and ℓ_2 in (5). This case occurs when $x_2<x\leqq\ell_2$.

② occurs when $x>\ell_2$ and is identical with the case of $x=x'$ in Fig. 3 and (4) can be applied. Here x' is

$$x' = \ell - \{w - (1 - x)\}$$

From(4), ℓ at this time is

$$\ell = v + (1 - v)x'$$

Deleting ℓ from two formulae above, we get

$$x' = 1 + (1-w)/v - x/v \tag{6}$$

In case $w/v\leqq 1$, it is possible to pass through the high density area of the width of w even when x=1 (the same as x=0). Therefore, the range of x is $\ell_2 \leqq x<1$. In case $w/v>1$, it is not possible to pass through the high density area at the left end when $1-x<w-v$. Therefore, the range of x where this case applicable is $\ell_2<x\leqq 1-w+v$.

③ is the same as the case of x=0. Here it is not possible to pass through the high density area, and this case is confined to when $w/v>1$. Putting x=0 in (4), following formula is obtained

$$n_3 = (b+d)v = bv + dv; \qquad \ell_3 = v \tag{7}$$

3.3 General relations between n and ℓ

Generally, there exist the following relations between n and ℓ. Denoting the length of high density area actually passed by w_0,

$$w_0/v + (\ell - w_0) = 1 \tag{8}$$

is obtained because the total time span is 1. Consequently, there exists a relation as

$$n = bl + dw_o = bl + d(1-l)v/(1-v) \qquad (9)$$

3.4 Average density n/l corresponding to various values of w/v and x

A list of formulae of n/l is presented in the following for all the cases given in 3.1 and 3.2.

(1) $w/v \leqq 1$

i) $0 \leq x < 1-w/v$: $\quad n/l = n_1/l_1 = b + dw/(1-w/v+w)$

ii) $1-w/v \leq x < 1-w$: $\quad n/l = b + dv(1-x)/\{v+(1-v)x\}$

$$= b + d\frac{v}{1-v}\left[\frac{1/(1-v)}{x+v/(1-v)} - 1\right]$$

iii) $1-w \leq x < 1-w+wv$: $\quad n/l = n_2/l_2 = b + dwv/(1-w+wv)$

iv) $1-w+wv \leq x < 1$: $\quad n/l = b + dv(1-x')/\{v+(1-v)x'\}$;

$$x' = 1 + (1-w)/v - x/v$$

(2) $w/v > 1$

ii) $0 \leq x < 1-w$: same as ii) in (1) $w/v \leqq 1$

iii) $1-w \leqq x < 1-w+wv$: same as iii) in (1) $w/v \leqq 1$

iv) $1-w+wv \leqq x < 1-w+v$: same as iv) in (1) $w/v \leqq 1$

v) $1-w+v \leqq x < 1$: $\quad n/l = b + d$

IV. Expected value, relative bias and variation range of daily observed average density n/l, and ratio of expected values of n and l when x is distributed evenly over 0-1

Expected value of n/l is obtained by expressing n/l as function of x, and integrating it regarding x for the range of 0 to 1, i.e.

$$E(n/l) = \int_0^1 n/l \; dx \qquad (10)$$

The relative bias is given by

$$RB = \{E(n/l) - d_m\}/d_m = \{E(n/l) - (b + dw)\}/(b + dw) \qquad (11)$$

The case of $w/v \leqq 1$ and the case of $w/v > 1$ are considered separately.

4.1 $w/v \leqq 1$

The range of x, 0~1, for integration is divided into four seetings as in 3.4 (1). When integration is done, the following formula is obtained (for derivation of formula, see Appendix 1 (1)).

$$\int_0^1 n/\ell \, dx = b + d[-wv/\{(1-w/v+w)(1-w+wv)\} + v(1+v)/(1-v)^2 \ln\{(1-w+wv)/(1-w/v+w)\}] \quad (12)$$

It should be noted in this formula that the term with b and that with d are completely separated. The value of b is simply added to the other term, and no bias is included in this part. Expressing the portion inside the [ ] relating to d in (12) by $F_I(w, v)$,

$$E(n/\ell) = b + d\, F_I(w, v) \quad (13)$$

is obtained. The relative bias is then

$$RB = d(F_I(w, v) - w)/(b + dw) = (dw/b)/(1 + dw/b)\cdot(F_I(w, v)/w - 1) \quad (14)$$

RB is determined by dw/b, w and v. These values will be obtained from observed data (Appendix 2). Here, the maximum value of n/ℓ is n_1/ℓ_1 and the minimum value is n_2/ℓ_2.)

4.2 w/v>1

Integration range is divided into four sections as in 3.4 (2). The result of integration is given by the following formula (See Appendix 1(2))

$$\int_0^1 n/\ell \, dx = b + d[(w-v)/(1-v) - v(1-w)/\{(1-w+wv)(1-v)\} + v(1+v)/(1-v)^2 \ln\{(1-w+wv)/v\}] \quad (15)$$

Similarly to (12), b and the term with d are completely separated. In the same way as in (13), it is expressed

$$E(n/\ell) = b + d\, F_{II}(w, v) \quad (16)$$

and the relative bias is

$$RB = (dw/b)/(1 + dw/b)\cdot(F_{II}(w, v)/w - 1) \quad (17)$$

The parameters included here are also dw/b, w and v. The maximum value of n/ℓ is b+d and the minimum value is n_2/ℓ_2.

4.3 E(n) and E(ℓ)

Next, bias in the conventional estimation formula $\Sigma n/\Sigma \ell$ is considered. Here, it can be assumed that expected value of this estimate gradually approach E(n)/E(ℓ) as the number of observations increases. And hence the characteristics of E(n)/E(ℓ) is examined here. The expected value of n or ℓ is given by the area surrounded by the line for n or ℓ and the x axis in the range of 0-1. As we see in (1), (4), (5) and (7), the line for n or ℓ against x is either constant or straight line for each interval. Therefore the area is obtained as follows,

i) $w/v \leqq 1$

Since $a_1 = 1 - w/v + w$ and $a_2 = 1 - w + wv$

$$E(\ell) = (1-w/v+w)(1-w/v) + (1-w+wv)wv + \{1-w/2\cdot(1-v)/v\}(w/v-wv)$$

$$= 1 + w(1-v)/v\cdot\{w/2\cdot(1/v+v)(1-v) - 1\} \quad (18)$$

And from(9)

$$E(n) = bE(\ell) + dv/(1-v)\cdot(1-E(\ell)) \quad (19)$$

ii) $w/v > 1$

Since $a_1 = 1 - w + wv$ and $a_2 = v$

$$E(\ell) = (1-w+wv)wv + v(w-v) + (1-w+v+wv)/2\cdot(1-w+v-wv)$$

$$= 1 - w(1-v)\{(1-v)(1/w-w)/2 + 1\} \quad (20)$$

As for n, (19) is applied as it is.

V. Bias in population density estimates

5.1 Characteristics of n/ℓ

With respect of range 0-1 of the values of w and v, ($F_I(w,v)/w$-1) included in (14) and (17)areshown in Table 1. $F_I(w,v)$ is applied for the case of $w/v \leqq 1$, and F_{II} (w,v), for w/v>1. If distribution density b and d are given, $E(n/\ell)$and RB are obtained through (13) and (14) or (16) and (17). Further, when b=0, i.e. whales are distributed only within the high density area, (F_I (w,v)/w - 1) gives RB directly. When $b \neq 0$, the coefficient (dw/b)/{1+ (dw/b) } get smaller when dw gets smaller against b, that is heterogeneity in distribution is smaller, and then the values of RB also become smaller. As one example, Table 2 shows $E(n/\ell)$,the range of n/ℓ and RB for dw/b=3 and b=1 and hence b+dw=4, where dw is relatively small. $E(n/\ell)$ has positive bias in all cases.

Table 1 and Table 2 show that there is no bias when v=1.0 or 0, or w=1.0 or 0, but this does not have any significant meaning. RB tends to be high in the vicinity of w=v, and it can exceed 1 when both w and v become extremely small. In the range of w>0.5 or v>0.5, RB is relatively small, and in the case of dw/b=3 (Table 2), it seldom exceeds 10%.

Table 2 shows the range of n/ℓ together with $E(n/\ell)$. The observed value for each sample can vary within this range. In case v>0.5, the minimum value is above 60% of the true value, and is higher as the w is larger. The maximum value is below 175%, and becomes the largest at w=v. In case w>0.5, the minimum value at v>0.2 is over 50% but is lower when v is smaller, while the maximum value reaches 175%

where v is small. In the case of $w/v \leqq 1$, the maximum value is applied to the entire region of x of $0 \leqq x \leqq 1 - w/v$. When w=0.1 and v=0.5, for example, all the n/ℓ takes the value of 4.333 between 0 and 0.8 of x. If x is given arbitrarily, this value is obtained eight out of 10 times.

In Table 2 dw/b=3 is applied, and hence d=3/w if b=1. Sicnce v is considered to be reduced as d becomes higher (Appendix 2 (2)), care is needed when observing this Table. For example, v becomes small when w is small, and in the extreme, when $w \rightarrow 0$, it becomes $d \rightarrow \infty$, $v \rightarrow 0$.

Appendix 2(4) shows some examples of the observed values for w, v, dw/b and others. According to these data, it may be assumed that w=0.1-0.2, v=0.05-0.1 and dw/b=3～10. Looking at this range of w and v for dw/b=3 in Table2, bias of $E(n/\ell)$ could reach almost 70%.

5.2 Bias in $E(n)/E(\ell)$ and comparison with $E(n/\ell)$

E(n) and $E(\ell)$ are calculated from (18), (19) and (20), and their ratios are shown in Table 3. Here, it is assumed dw/b=3 and b+dw=4, same as in Table 2. The value of $E(n)/E(\ell)$ is slightly larger than the true value in case $w/v \leqq 1$. On the other hand, when w/v>1, slight positive or negative bias is introduced for v>0.3. If w>0.1, then considerable negative bias occurs in the range of v<0.1, and when v=0.05, the bias is about 50%. If dw/b becomes larger, the degree of bias gets higher but the direction of bias, positive or negative, will be more or less unchanged. The ratio of $E(n)/E(\ell)$ to $E(n/\ell)$ is also shown in Table 3. All of the values of this ratio is smaller than 1. If v>0.3, this ratio is mostly over 90%, but sharply declines as v drops, to around 50% at v=0.05. When considering that the value of v in JARPA is around 0.05-0.1, the fact that the population abundance estimate in JARPA stands about half that of the IDCR could be explained.

VI. Discussion and conclusion

6.1 About models

The model used here assumes that there exist one high density area as a rectangular distribution with density (b+d) and length w. Distribution of actual whale schools is more irregular. However, as we have already seen in Fig. 3, when the proceeding distance of certain ℓ is given, the form of distribution of actual schools within already-passed area, or not-passed area is not relevant. Therefore, generality would not be lost by having the high density area represented by rectangular distribution when certain n and ℓ is observed. It is assumed here that, in obtaining expected values, the left end of high density area x is distributed uniformly between 0 and 1. But in an arbitrary distribution, the distribution of left end of corresponding rectangular area to n and ℓ will not necessarily satisfy this condition. Therefore, there is a possibility that expected values are distorted. However, in case where relatively clear high density area are distributed randomly as Kishino &

Kasamatsu (1987) used in the simulation, it may be considered that this model can approximate actual situation fairly well.

In obtaining expected values, it is assumed that the values of w, d, b and v are the same for every high density areas. In actuality, they change according to areas and this models will not fit the reality. However, if these values are limited in a relatively narrow range and are distributed randomly, it may be safe to think that the conclusion obtained here is reasonable, at least qualitatively.

This model does not specifically taken into consideration the relation between v and d. It is because a constant d value is assumed for all high density areas. However, in actuality , the relation as (2. 3) in Appendix 2 could be postulated. Although it is possible to introduce this relation into the formula, it will make things more complex. In case where v and d are related, the idea of having the arbitrary distribution represented by rectangular distribution as mentioned before would become invalid.

It is assumed here that all whale schools will be sighted. In actuality, however, missing rate is relatively high. If the sighting rate is not related to w, d or b, it could be expected that this impact is not substantial. Analytical examination by means of a model will make it easy to grasp features of entire picture. In the results of calculation here general characteristics of $E(n/\ell)$ and $E(n)/E(\ell)$ were shown. It is expected that these characteristics will be preserved for more complex system, but it is necessary to execute a simulation through more realistic models in order to discuss it in more detail.

6.2 Bias and variation of population density and sampling ratio.

Assuming that actual situation can be seen by means of simple models, let's have look at characteristics of population density estimates. $E(n/\ell)$ has positive bias of 25-67% for the actually observed range of w=0.1-0.2 and v=0.05-0.1. Therefore, the effect of night-time idle cruising cannot be eliminated even by formula of Burt & Borchers. On the other hand, in the method used in JARPA, positive or negative biases are generated. There is no bias when w is close to or a little larger than v but, when v becomes small and the frequency of idle cruising increases, negative bias will increase sharply. It is possibile that it becomes about half of the true value in the actually observed range of w and v.

The formula used for JARPA data is the same as the formula to calculate CPUE. Often, catcher searching hour (CSW) is used for effort. CSW is considered to decrease in a high density area because the total hours for handling whales would increases within a fixed working hours of a day as more whales are taken. Therefore CPUE for entire area tends to underestimate whale density (Zahl, 1982, 1983, Cooke, 1985). The bias is increased as handling time per whale is increased, and this situation is just the same as the case where v becomes small in sampling vessels.

Variance of the estimates is not calculated here, but the range of n/ℓ in Table 2 should be noted. When v becomes small, the range of n/ℓ extends from 1/4 to 8-fold of the true value. As the variability of density between unit spaces is not taken into consideration, variation of n/ℓ here is a kind of measurement error. In evaluating the error in actual estimates, coefficient of variation affected by both spatial variability and measurement error is obtained and so there may be no problem for evaluating error. Nevertheless, the precision of population abundance estimates must be deteriorated considerably.

Biological data from individual whales taken in different days are lumped together within the same stratum. Therefore, estimation of biological parametevs is considered to follow the same manner as used in JARPA formula. The fact that $E(n)/E(\ell)$ could have negative bias means that sampling ratio in the high density area could be lower than that of the ordinary area. If there is no difference in biological parameters between the high and low density areas, no problem occurs about difference in sampling ratio. However, if there exist differences, problem will arise for representativeness of samples.

6.3 Future steps to be taken

The procedure of establishing a proceeding distance per day and carrying out night-time cruising when this distance is not covered is very advantageous for covering the extensive sea area thoroughly and implementing the survey as planned. But bias will be introduced into population density estimates and further problem will be brought about regarding representativeness of samples. By formulae proposed to date, bias in density estimates is not eliminated. Furthermore, no consideration has yet been made regarding correction of distorted sampling ratio. In order to resolve these problems, the following measures will be advised.

1) examine whether differences in biological parameters exist between the high density area and ordinary areas. If it is shown that there exist differences, high density area and ordinary area should be treated as separate strata (Zahl, 1983).

2) develop methods to correct bias in population size estimates. Simulation method would be effective (Clarke & Borchers;1997). As $E(n/\ell)$ has positive bias, geometric mean of n/ℓ may be a candidate for better estimates. Usually, variation of the value of n is assumed to follow log-normal distribution, and it may be reasonable to apply geometric mean.

3) develop a sampling procedure not requiring night-time cruising. Schweder (1998) proposed a method of covering the entire area without night-time cruising, by modifying sampling ratio of schools depending on school size. Although the practicability of this method is unknown, this idea deserves thorough examination.

References

Burt, M.L. and Borchers, D.L. 1997. Minke whale abundance estimated from the

1991/92 and 1992/93 JARPA sighting surveys. Paper SC/M97/23 presented at IWC/SC Intersessional Working Group held in Tokyo in May 1997.

Clarke, E.D. and Borchers, D.L. 1997. Proposal for the development of unbiased abundance estimators for JARPA sightings survey data. Paper SC/49/SH30 presented at IWC/SC meeting at Bournemouth, U.K in September 1997.

Cooke, J.G. 1985. On the relationship between catch per unit effort and whale abundance. Rep.int. Whal.Commn, 35:511-19.

Kishino, H. and Kasamatsu, F. 1987. Comparison of the closing and passing mode procedures used in sighting surveys. Rep.int.Whal.Commn, 37:253-8.

Nishiwaki, S., Matsuoka, K., Kawasaki, M., Kishino, H, and Kasamatsu, F. 1997. Review of the sighting surveys in JARPA. Paper SC/M97/1 presented at IWC/SC Intersessional Working Group held in Tokyo in May 1997.

Schweder, T. 1998. A proposed modification for the JARPA sighing and sampling protocol. Rep.int.Whal Commn, 48, Report of the Scientific Committee, Annex U, page 2.

Zahl, S. 1982. Bias of the CPUE using search time as effort measure. Rep.int.whal. Commn, 32:809-13.

Zahl, S. 1983. Correcting the bias of the CPUE due to a varying whale density. Rep.int.whal. Commn, 33:307-8.

Appendix 1 Integral calculation to obtain expected values of E(n/ℓ)

(1) Integration when $w/v \leqq 1$

Of the four sections in 3.4 (1),each of (i) and (iii) is constant regardless of x. Therefore, the value of n/ℓ should be multiplied by the length of each section. (iv) becomes the same formula as ii) when it is expressed by x'. Therefore, transforming the variable for integration from x to x', and putting dx= - vdx'

$$\int_{1-w+wv}^{1} n/\ell \, dx = \int_{1-w+wv}^{1} (b + dv(1-x')/\{v+(1-v)x'\})dx$$

$$= \int_{1-w/v}^{1-w} (b + dv(1-x')/\{v+(1-v)x'\})vdx' \qquad (1.1)$$

This is the same as the integral of (ii) as multiplied by v. (ii) becomes,

$$\int_{1-w/v}^{1-w} [b + dv(1-x)/\{v+(1-v)x\}]dx$$

$$= \int_{1-w/v}^{1-w} \{b - dv/(1-v)\}dx + dv/(1-v)^2 \int_{1-w/v}^{1-w} [1/\{x+v/(1-v)\}]dx$$

$$= \{b - dv/(1-v)\}(w/v-w) + dv/(1-v)^2 \ln[\{1-w+v/(1-v)\}/\{1-w/v+v/(1-v)\}]$$

$$= bw(1-v)/v - dw + dv/(1-v)^2 \ln[(1-w+wv)/(1-w/v+w)].$$

Adding up integrals of i) through iv), expectation of E(n/ℓ) will be

$$\int_0^1 n/\ell\ dx = (1-w/v)\{b+dw/(1-w/v+w)\} + wv\{b+dwv/(1-w+wv)\}$$
$$+ (1+v)\{bw(1-v)/v - dw\} + dv(1+v)/(1-v)^2 \ln[(1-w+wv)/(1-w/v+w)]$$

$$= b + d[w(1-w/v)/(1-w/v+w) + w^2v^2/(1-w+wv) - (1+v)w]$$
$$+ dv(1+v)/(1-v)^2 \ln[(1-w+wv)/(1-w/v+w)]$$

Inside of the [ ] of second term is $-wv/\{(1-w/v+w)(1-w+wv)\}$,
and then E(n/ℓ) is

$$\int_0^1 n/\ell\ dx = b + d[- wv/\{(1-w/v+w)(1-w+wv)\}$$
$$+ v(1+v)/(1-v)^2 \ln\{(1-w+wv)/(1-w/v+w)\}] \qquad (1.2)$$

and the formula (12) in the main text is obtained.

(2) Integration when w/v>1

Of the four sections in 3.4 (3), (iii) and (v) are constant. As for the integral of (iv), x is transformed to x' to obtain a similar formula as in (1.1). Here, the range of integration is from 0 to (1 - w). Integral of (ii) is

$$\int_0^{1-w} [b + dv(1-x)/\{v+(1-v)x\}]dx$$

$$= \{b - dv/(1-v)\}(1-w) + dv/(1-v)^2 \ln[(1-w+wv)/v]$$

Adding up (ii) through (v), the expected value E(n/ℓ) is

$$\int_0^1 n/\ell\ dx = b + d[w^2v^2/(1-w+wv) + (w-v) - (1+v)(1-w)v/(1-v)]$$
$$+ dv(1+v)/(1-v)^2 \ln[(1-w+wv)/v]$$

Inside of the [ ] of second term is rewritten and E(n/ℓ) becomes

$$\int_0^1 n/\ell\ dx = b + d[(w-v)/(1-v) - v(1-w)/\{(1-w+wv)(1-v)\}$$
$$+ v(1+v)/(1-v)^2 \ln\{(1-w+wv)/v\}] \qquad (1.3)$$

and the formula (15) in the main text is obtained.

Appendix 2 Estimation of values of parameters contained in the formula

(1) w

Data from dedicated sighting vessels can be used for estimating w. In this case, it is assumed that v=1. w is the size of high density area but it is also the proportion of high density area within the unit space. When more than one high density areas are observed within a unit space, total length of all high density areas is taken as the value of w. In actual surveys, the resulted proceeding distance of a dedicated sighting vessel per day is variable around 100 miles. Denoting this distance by ℓ_0 and the length of j th high density area by wj, w can be estimated as in the following:

$$w = \Sigma w_j / \ell_0 \qquad (2.1)$$

There exists a problem of how to determine the range of the high density area, but generally, boundaries are fairly clear.

(2) v

The actual proceeding distance of sampling vessels is assumed as ℓ, and total length of the high density areas, w0, as Σwj. The speed in the high density area is v and that in the ordinary area is 1, and then it can be write.

$$\Sigma w_j / v + (\ell - \Sigma w_j) = 1$$

From this formula and data from sampling vessels, v is estimated by

$$v = (\Sigma w_j / \ell_0)/(1 - \ell/\ell_0 + \Sigma w_j/\ell_0) \qquad (2.2)$$

It is considered that v is related to density. Let's assumed that the proceeding is halted at each unit discovery for a time of Δt. In the high density area, there are discoveries of (b+d) Δx within the distance Δx. Assuming that effect of bΔx is already incorporated in the pre-determined proceeding distance, the reduction of the speed is related to the par t o f dΔx. As the time required for passing Δx is {Δx+d $\Delta x \Delta t$} , the reduction of the speed is

$$v = \Delta x/\{\Delta x + d\Delta x\Delta t\} = 1/\{1 + d\Delta t\} \qquad (2.3)$$

v is given by d and Δt. When n and ℓ are given as observed values, Δt is calculated by

$$\Delta t = (1 - \ell/\ell_0) / (n - b\ell/\ell_0) \qquad (2.4)$$

(3) d, b and dw/b

Here data from dedicated sighting vessel are used. Assuming that the total number of discovery in the high density area to be nh, it follows d+b=nh/w, and further

$$b = (n - n_h)/(1 - w), \quad d = (n_h/w - n)/(1 - w) \qquad (2.5)$$

are obtained. It is anticipated that d is fairly larger than b. The proportion of schools distributed in the high density area can be estimated by nh/n. Using these formulae, the following is obtained.

$$dw/b = (n_h - wn)/(n - n_h) \qquad (2.6)$$

(4) Examples of numerical value

Applying a part of the data from the survey in Area V in 1993 (data for January 15-February 13 compiled by Fujise.), values of these parameters are calculated tentatively and folloing values are obtained.

w=0.1~0.24, v=0.02~0.11, Δt=0.11~0.36, d=34~102, b=1.1~5.9, dw/b=2.6~12

As reference values, it may roughly be put w=0.1~0.2, v=0.05~0.1, dw/b=3~10.

Table 1 Value of $F_i(w,v)/w - 1$. This value corresponds to the relative bias (RB) of $E(n/l)$ when whales are distributed only in high density areas (b=0). For other values of b, this value is multiplied by $(dw/b)/\{1+(dw/b)\}$ to give RB.

v \ w	0	0.1	0.2	0.3	0.4	0.5	0.7	1.0
0	0	0	0	0	0	0	0	0
0.02	0	0.4217	0.2091	0.1373	0.1004	0.0772	0.0470	0
0.05	0	.6875	.3396	.2213	.1599	.1207	.0682	0
0.1	0	.9000	.4423	.2855	.2031	.1499	.0775	0
0.2	0	.2406	.5003	.3179	.2208	.1573	.0717	0
0.3	0	.1164	.2473	.3030	.2063	.1429	.0595	0
0.4	0	.0649	.1305	.1840	.1803	.1220	.0474	0
0.5	0	.0372	.0723	.1012	.1159	.0995	.0365	0
0.7	0	.0103	.0192	.0262	.0305	.0315	.0188	0
1.0	0	0	0	0	0	0	0	0

Table 2 Characteristics of n/ℓ. (1) $E(n/\ell)$, (2) range of n/ℓ, (3) relative bias RB of $E(n/\ell)$. b=1, dw=3, true density d_m=b+dw=4.

v \ w		0	0.1	0.2	0.3	0.4	0.5	0.7	1.0
0	(1)	4	4	4	4	4	4	4	4
	(2)	1 - 4or∞	1 - 31	1 - 16	1 - 11	1 - 8.5	1 - 7	1 - 5.286	1or4- 4
	(3)	0	0	0	0	0	0	0	0
0.02	(1)	4	5.265	4.627	4.412	4.301	4.231	4.141	4
	(2)	1.06- 4	1.067- 31	1.075- 16	1.085- 11	1.099- 8.5	1.118- 7	1.191- 5.286	4 - 4
	(3)	0	0.3163	0.1569	0.1030	0.0753	0.0579	0.0352	0
0.05	(1)	4	6.063	5.019	4.664	4.480	4.362	4.205	4
	(2)	1.15- 4	1.166- 31	1.185- 16	1.210- 11	1.242- 8.5	1.286- 7	1.448- 5.286	4 - 4
	(3)	0	0.5156	0.2547	0.1660	0.1199	0.0905	0.0511	0
0.1	(1)	4	6.700	5.327	4.856	4.609	4.450	4.233	4
	(2)	1.3 - 4	1.330- 31	1.366- 16	1.411- 11	1.469- 8.5	1.545- 7	1.811- 5.286	4 - 4
	(3)	0	0.6750	0.3317	0.2141	0.1524	0.1124	0.0581	0
0.2	(1)	4	4.722	5.501	4.954	4.662	4.472	4.215	4
	(2)	1.6 - 4	1.652- 6	1.714- 16	1.789- 11	1.882- 8.5	2 - 7	2.364- 5.286	4 - 4
	(3)	0	0.1804	0.3752	0.2384	0.1656	0.1180	0.0538	0
0.3	(1)	4	4.349	4.742	4.909	4.619	4.429	4.179	4
	(2)	1.9 - 4	1.968- 4.913	2.047- 6.625	2.139- 11	2.25 - 8.5	2.385- 7	2.765- 5.286	4 - 4
	(3)	0	0.0873	0.1855	0.2273	0.1547	0.1072	0.0446	0
0.4	(1)	4	4.195	4.392	4.552	4.541	4.366	4.142	4
	(2)	2.2 - 4	2.277- 4.529	2.364- 5.286	2.463- 6.455	2.579- 8.5	2.714- 7	3.069- 5.286	4 - 4
	(3)	0	0.0487	0.0979	0.1380	0.1352	0.0915	0.0355	0
0.5	(1)	4	4.112	4.217	4.304	4.348	4.298	4.110	4
	(2)	2.5 - 4	2.579- 4.333	2.667- 4.75	2.765- 5.286	2.875- 6	3 - 7	3.308- 5.286	4 - 4
	(3)	0	0.0279	0.0542	0.0759	0.0870	0.0746	0.0274	0
0.7	(1)	4	4.031	4.058	4.078	4.092	4.095	4.057	4
	(2)	3.1 - 4	3.165- 4.134	3.234- 4.281	3.308- 4.444	3.386- 4.621	3.471- 4.818	3.658- 5.286	4 - 4
	(3)	0	0.0078	0.0144	0.0196	0.0229	0.0237	0.0141	0
1.0		For all w values: (1) 4; (2) 4 - 4; (3) 0							

Table 3 Values of $E(n)/E(\ell)$, (line 1) and its ratio to $E(n/\ell)$, (line 2).
b=1, dw=3, true density d_m=b+dw=4.

v \ w		0	0.1	0.2	0.3	0.4	0.5	0.7	1.0
0	(1)	4or1	1	1	1	1	1	1	1or4
	(2)	1or0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25or1
0.02	(1)	4	1.823	1.586	1.555	1.595	1.695	2.178	4
	(2)	1	0.3462	0.3428	0.3524	0.3708	0.4005	5.260	1
0.05	(1)	4	2.867	2.306	2.203	2.244	2.377	2.926	4
	(2)	1	0.4728	0.4594	0.4723	0.5008	0.5449	0.6958	1
0.1	(1)	4	4.215	3.199	2.963	2.946	3.041	3.437	4
	(2)	1	0.6291	0.6004	0.6101	0.6392	0.6833	0.8121	1
0.2	(1)	4	4.478	4.288	3.833	3.685	3.667	3.799	4
	(2)	1	0.9483	0.7796	0.7798	0.7903	0.8199	0.9013	1
0.3	(1)	4	4.288	4.431	4.272	4.037	3.944	3.933	4
	(2)	1	0.9860	0.9344	0.8703	0.8740	0.8905	0.9413	1
0.4	(1)	4	4.174	4.294	4.322	4.213	4.080	3.995	4
	(2)	1	0.9949	0.9778	0.9494	0.9278	0.9344	0.9656	1
0.5	(1)	4	4.103	4.182	4.223	4.214	4.143	4.025	4
	(2)	1	0.9980	0.9917	0.9813	0.9693	0.9638	0.9794	1
0.7	(1)	4	4.030	4.053	4.070	4.077	4.075	4.037	4
	(2)	1	0.9997	0.9990	0.9978	0.9964	0.9952	0.9952	1
1.0		For all w values:	(1) 4:	(2) 1					

Bias in density estimates when the speed of whale sampling vessels is reduced in high density areas

S. TANAKA (Institute of Cetacean Research)

In Table 2 of the main text, it is shown that E(n/ℓ) has positive bias regardless of w and v values. As a method to remove this bias, geometric mean of daily n/ℓ values, instead of arithmetic mean, may be suggested. It is often assumed in the sighting survey data that n values are distributed log-normally. The ranges of n/ℓ given in Table 2 indicate that the distribution of n/ℓ has a long right hand side tail. By taking logarithm of n/ℓ values, a more symmetrical distribution of n/ℓ may be obtained.

To ascertain this , exp[E{ln(n/ℓ)}] are calculated for various values of w and v by a numerical integration method. The results are shown in Addendum Table 1. If v>0.2, the bias in density estimates will be reduced considerably, although a small positive bias still remains. The bias is smaller then 5% when v>0.5. In case of v<0.2, negative bias may be introduced. At v=0.05 and w=0.2, for instance, the negative bias could be as high as 20%. Generally, geometric mean will give a smaller bias but it could be either positive or negative when v value is small.

Addendum Table 1. Comparison of arithmetic mean, E(n/ℓ), and geometric mean, exp[E{ln(n/ℓ)}]. True value is 4 with b=1 and dw=3.

v	\ w	0.1	0.2	0.3	0.4	0.5	0.7
0.05	n/ℓ	6.063	5.019	4.664	4.480	4.362	4.205
	ln(n/ℓ)	3.67	3.24	3.26	3.38	3.52	3.81
0.1	n/ℓ	6.700	5.327	4.856	4.609	4.450	4.233
	ln(n/ℓ)	4.69	3.86	3.71	3.73	3.78	3.94
0.2	n/ℓ	4.722	5.501	4.954	4.662	4.472	4.215
	ln(n/ℓ)	4.38	4.50	4.15	4.04	4.01	4.01
0.3	n/ℓ	4.349	4.742	4.909	4.619	4.429	4.179
	ln(n/ℓ)	4.21	4.38	4.36	4.17	4.09	4.03
0.4	n/ℓ	4.195	4.392	4.552	4.541	4.366	4.142
	ln(n/ℓ)	4.12	4.22	4.28	4.22	4.12	4.03
0.5	n/ℓ	4.112	4.217	4.304	4.348	4.298	4.110
	ln(n/ℓ)	4.07	4.13	4.17	4.18	4.13	4.03
0.7	n/ℓ	4.031	4.058	4.078	4.092	4.095	4.057
	ln(n/ℓ)	4.02	4.04	4.05	4.05	4.05	4.03