Spatial and temporal structure of the western North Pacific minke whale distribution inferred from JARPN sightings data

Hiroshi Okamura*, Koji Matsuoka+, Takashi Hakamada+, Makoto Okazaki* and Tomio Miyashita*

Contact e-mail: okamura@enyo.affrc.go.jp

ABSTRACT

The density of minke whales (*Balaenoptera acutorostrata*) in the western North Pacific was examined using a generalized additive model in order to investigate the spatial and temporal distribution patterns. The data used were a subset of JARPN sightings data collected from 1994 to 1999. The process for estimating the density was divided into two parts: the detection process for the estimation of the effective search half-width; and the encounter process for the estimation of the encounter rate. Model selection was carried out using information criteria. The selected model for the detection process included 'sightability', a synthetic index of detectability, as a covariate, and for the encounter process included the interaction between latitude and longitude and the interaction between month and latitude. The trend surface of the transformed density predicted by each month revealed no clear gaps. The monthly transition of the density distribution also suggested the northward seasonal feeding migration of the minke whales.

KEYWORDS: COMMON MINKE WHALE; NORTH PACIFIC; INDEX OF ABUNDANCE; STOCK IDENTITY; MIGRATION

INTRODUCTION

The JARPN (Japanese Whale Research Program under Special Permit in the Western Part of North Pacific) programme was carried out from 1994-99 with the primary aim of determining the stock structure of the common minke whale (*Balaenoptera acutorostrata*) in the western North Pacific. The results of the programme were reviewed in February 2000 by the Scientific Committee of the International Whaling Commission (IWC, 2001). Although several papers presented to the review meeting suggested that there was no explicit evidence for multiple stocks of minke whales off Japan, it was pointed out that this conclusion might be due to the inappropriate pre-stratification of the western North Pacific (Martien and Taylor, 2000; Taylor, 2000; Taylor and Chivers, 2000). It was suggested that the examination of density distribution patterns might provide valuable information to determine the appropriate partition of the area. The IWC Scientific Committee recommended that the sightings data should be analysed using a multiple regression model such as a generalized linear model (GLM; McCullagh and Nelder, 1989) that includes the covariates of year, month, Beaufort Sea state and sea temperature (IWC, 2001). This paper examines the available sightings data using a multiple linear regression and a generalized additive model (GAM; Hastie and Tibshirani, 1990) to include several important covariates influencing the detection and encounters. It also examines the monthly distribution patterns.

MATERIALS AND METHODS

The data used were a subset of the JARPN sightings data presented in Matsuoka *et al*. (2000). Fig. 1 shows the sub-areas of the western North Pacific used by the IWC. This paper concentrates on the data available for sub-areas 7, 8 and 9, as the primary issue raised in IWC (2001) was whether more than one population exists in these three sub-areas or not. The pooled effort by 1° square is shown in Fig. 2. Most effort occurred in the northern part of the sub-areas whilst effort in the southern parts was sparse. The monthly plots of the tracklines surveyed are provided in Matsuoka *et al*. (2000).

The sightings data have been divided into two panels, one for detection (perpendicular distance, environmental conditions at detection such as the sea state) and the other for searching activities (effort, year, month, day, averaged environmental conditions such as the sea surface temperature for on-effort portions of the day). The density index was calculated through these two processes, one to estimate the effective search half-width and the other to estimate the encounter rate.

The detection process

The effective search half-width (including the effects of several covariates) is estimated by the following method of Beaver and Ramsey (1998) as described below.

(1) The perpendicular distance from the transect to the *i*th detected pod of whales is the detection distance, y_i. When a set of covariates, $x_i = (x_{i1},\ldots,x_{ip})$, is associated with the *i*th detected pod, it is assumed that the effective search half-width surveyed under condition x_i is w_i, where

$$\log(w_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

The ordinary least squares regression of $\log(y_i)$ on the covariates provides unbiased estimates of the parameters $(\hat{\beta}_1,\ldots,\hat{\beta}_p)$;

(2) Determine average detectability conditions for the covariates, $\bar{x}_j$;

(3) Adjust all detection distances to the average conditions according to

$$\tilde{y}_i = y_i \times \exp\left(\sum_{j=1}^{p} \hat{\beta}_j (\bar{x}_j - x_{ij})\right)$$

(4) Use the adjusted detection distances to select a semi-parametric estimator of the effective half-width,

* National Research Institute of Far Seas Fisheries, 5-7-1 Orido, Shimizu, Shizuoka 424-8633, Japan.
+ The Institute of Cetacean Research, 4-18, Toyomi, Chuo, Tokyo 104-0055, Japan.

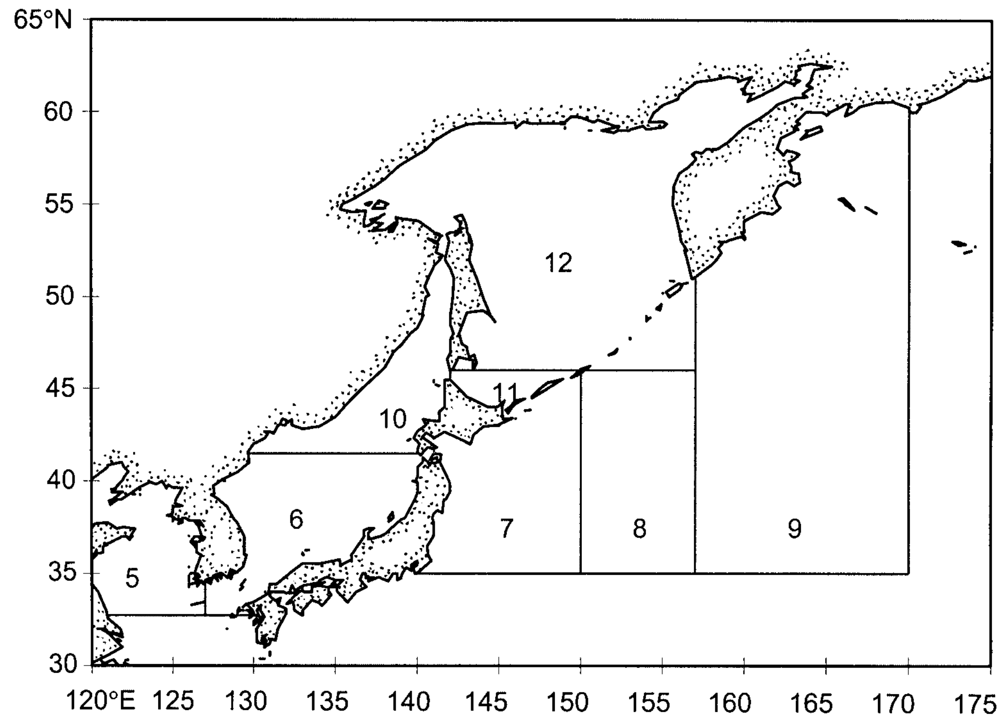


Fig. 1. Sub-areas for the western North Pacific minke whales.

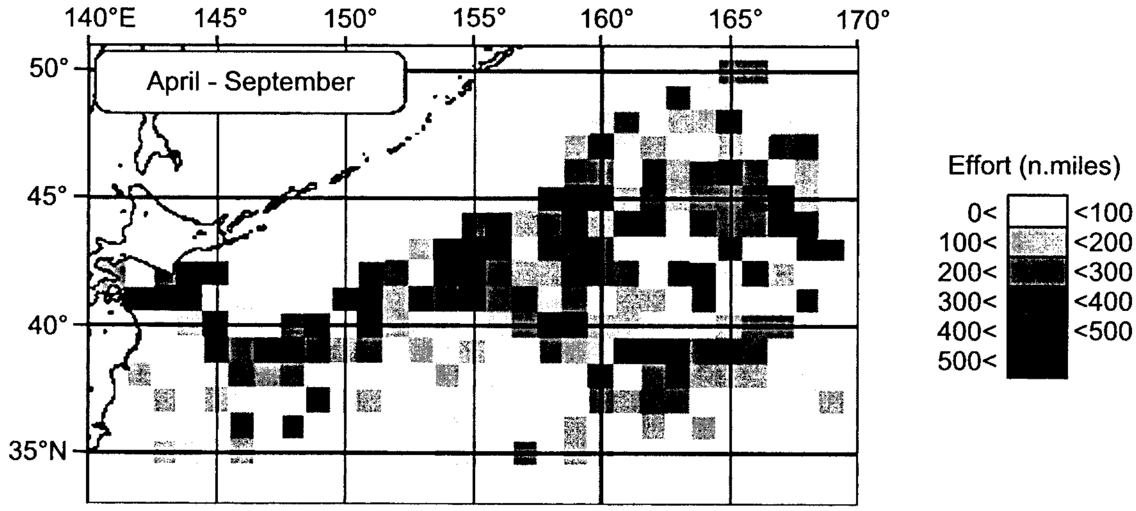


Fig. 2. Search efforts surveyed by JARPN from April to September in 1994-1999. The amount of effort in one degree square cell is divided into six categories with a unit of 100 n.miles.

$\hat{w}_0$, at average conditions. Program DISTANCE (Laake *et al.*, 1993) provides an estimate of effective half-width.

(5) Estimate the constant term with

$$\hat{\beta}_0 = \log(\hat{w}_0) - \sum_{j=1}^{p} \hat{\beta}_j \bar{x}_j$$

The model selection for multiple linear regression is made using the Bayesian Information Criterion (BIC; Swartz, 1978) assisted by Akaike's Information Criterion (AIC; Akaike, 1973). The BIC that approximates the logarithm of the Bayes factor in an approximate manner (Kass and Raftery, 1995) is given by

$$\hat{D} + \log_e(N)p$$

where D is the deviance residual, N is the number of observed pods, p is the number of parameters and 'hat' $(\hat{\cdot})$ denotes an estimate from fitting the model. D is the amount defined from the log-likelihood and is given by

$$D(\mu; y) = 2l\,(\mu^*; y) - 2l\,(\mu; y)$$

where y is the logarithm of the perpendicular detection distance, $\mu = E(y)$, $E(\,)$ is the expectation, l is a log-likelihood function and μ^* is the parameter estimate maximized under no limitation (Chambers and Hastie, 1992). The AIC is obtained by replacing $\log_e(N)$ in the BIC with 2.

The covariates considered are Air Temperature in degrees Celsius (*AT*), Sightability (*SA*) and Sea State (*SS*). The interaction between covariates is not considered for simplicity. *SA* is based on synthetic impression of average detectability reported by navigation officers, whilst *SS* is based on the height of waves. Thus, *SA* includes more information than *SS*, although it can be criticised for being subjective. In this paper, *SA* and *SS* are treated as categorical variables, whilst *AT* is treated as a continuous variable. The effect of school size is considered in the DISTANCE program (see Discussion).

The encounter process

The expected number of encounters, $E(n_k)$, on day k is first modelled as

$$E(n_k) = L_k \hat{w}_k \exp(f(Year) + f(Month) + lo(LAT, LONG, 1/h) + s(SST))$$

where n_k has a Poisson distribution and

- L_k effort on day k as an offset,
- $\hat{w}_k$ effective search half-width for day k estimated from the above detection process as an offset,
- *Year* 1994 to 1999,
- *Month* April to September,
- *LAT* latitude averaged from the period spent on-effort during a day,
- *LONG* longitude as above,
- *SST* sea surface temperature,
- $f(\bullet)$ factor,
- $lo(\bullet, \bullet, 1/h)$ locally-weighted running-line smoother with the span of $1/h$, i.e. the smoothing parameter for a loess fit. The span is the percentage of total data used to fit the local polynomial at each point,
- $s(\bullet)$ spline smoother with the degree of freedom of 4.

The form of the above model is similar to that of Cumberworth *et al.* (1996). Covariates *Year* and *Month* enter the model as categorical variables, whilst other covariates enter the model as smoothed functions such that the single terms are fitted by spline functions and the pairwise terms by loess functions (Hastie and Tibshirani, 1990). The possibility of an annual trend is examined by treating *Year* as a linear term. The span ($1/h$) of loess functions for the first model is selected by using the minimisation of the BIC assisted by the AIC with the change of h shifted by 1. The BIC for the GAM is given by

$$\hat{D} + \log_e(M) p_e$$

where D is the deviance residual, M is the number of days, p_e is the effective number of parameters that is the sum of the degrees of freedom for parametric parts of the model and the equivalent degrees of freedom for non-parametric parts, and 'hat' $(\hat{\cdot})$ denotes an estimate from fitting the model (Hastie and Tibshirani, 1990; Chambers and Hastie, 1992). The number of degrees of freedom of spline functions is fixed at 4.

After the determination of the span ($1/h$), the existence of over-dispersion and the needed covariates are examined under a fixed value of h. The existence of over-dispersion is investigated by using hypothesis testing and the bootstrap method. The details for hypothesis testing are given below. The stepwise model selection based on the information criteria such as the AIC and the BIC for the GAM is also employed for the variable selection. If there is over-dispersion, the information criteria are modified based on principles of quasi-likelihood. The details are also given below.

After the final model selection, the sensitivity of h is examined by changing values of h and calculating the information criteria. All the analyses in this paper are carried out using the *S-Plus* program.

Examination of an overdispersed Poisson distribution

In order to examine whether the sampling variance exceeds the theoretical variance ($var(n) = E(n)$ for the Poisson model), the overdispersion parameter (c) was estimated from the Pearson chi-square statistics of the global model and its degrees of freedom,

$$\hat{c} = \chi^2 / df$$

where $\chi^2 = \sum_{i=1}^{n} \frac{(n_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$, $\mu_i = E(n_i)$ and $df = M - p_e$ (Burnham and Anderson, 1988). The global model is defined as the first model plus all the pairwise interactions:

$$\text{first model} + \sum lo(X_i, X_j)$$

where X_i is each covariate such as *Year* and *SST*.

To test whether $c = 1$ or $c > 1$, the bootstrap approach with resampling of the residuals can be used (Efron and Tibshirani, 1993). In this case, we use the Pearson residuals,

$$r_i^P = \frac{n_i - \hat{\mu}_i}{V(\hat{\mu}_i)^{1/2}}$$

as the resampling unit.

The overdispersion parameters for B generated bootstrap data are calculated respectively (c^{*b}). Then an approximate significance level (P) is calculated by

$$\hat{P} = \#\{c^{*b} < 1\} / B$$

The hypothesis $c = 1$ is considered rejected if $P < 0.05$ (one-sided test with 5% significance level) following the standard statistical convention.

If $c > 1$, model selection is carried out by using QAIC (Burnham and Anderson, 1988) and QBIC again. QAIC and the QBIC are given by

$$\text{QAIC} = \hat{D} / \hat{c} + 2p_e$$

$$\text{QBIC} = \hat{D} / \hat{c} + \log_e(M) p_e$$

RESULTS

The final covariate selected by both the BIC and the AIC was *SA* in the detection process. The estimated parameters and the estimated mean effective search half-width are shown in Table 1. The hazard rate key function with no adjustment parameters was used to model the detection function since this had the lowest AIC value of the available functions in the DISTANCE program. The influence of school size was not significant. The increase in sightability increased the effective search half-width.

Table 1

The estimates and the standard errors of the parameters for the detection process.

	Value	Standard Error	*t* value	Sample size
SA1	-1.660	0.145	-11.436	51
SA2	-1.733	-1.733	-15.046	81
SA3	-1.495	-1.495	-17.782	152
SA4	-1.432	-1.432	-14.220	106
SA5	-0.856	-0.856	-3.600	19
β_0	0.723			
The estimated mean effective search half-width				
w_0		0.452		

For the encounter rate, the span $1/h$ for loess functions selected by minimizing the BIC was 1/2 ($h = 2$), whilst the span for loess functions selected by minimising the AIC was 1/31. The latter value is unreasonably large and thus we used only the result from the BIC ($h = 2$).

The over-dispersion parameter observed for the global model was 1.470. The estimated significance level $\hat{P}$ = 0.0004 was obtained from 10,000 bootstrap samples for the Pearson residuals of the global model. The 95% bootstrap confidence interval of the over-dispersion parameter was [1.195, 1.783]. Because $c > 1$, model selection was carried out using QBIC.

The selected model for the encounter rate with the estimated effective search half-width was given by

$$E(n_k) = L_k \hat{w}_k \exp(lo(LAT, LONG, 1/2) + lo(LAT, LONG, 1/2))$$

which had the lowest QBIC of the models considered (QBIC = 769.89). The covariates of the model that had the second lowest QBIC was *Year* + *lo*(*LAT*,*LONG*,1/2)+*lo*(*Month*, *LAT*,1/2) with *Year* as a linear term (QBIC = 771.59). A sensitivity test for *h* was carried out by changing the value of *h* incrementally and calculating the QBIC for the above final model. The model with $h = 2$ was still selected. The plots of residuals of the final model over each covariate showed no systematic trend.

The plots of smooth terms for the final model are shown in Fig. 3. The top plot is for the loess smoother for the interaction between latitude and longitude and the bottom plot for the loess smoother for the interaction between month and latitude. They indicate that the density in high latitudes is higher and the area with high density moves north as month changes.

The monthly densities in each 1° square were predicted from the final model. The monthly density indices in each cell were calculated by standardisation after the logarithm transformation:

$$\text{Density index} = \frac{\log_e(density(LAT, LONG, Month)) - \overline{\log_e(density(Month))}}{\sqrt{Var(\log_e(density(Month)))}}$$

where:

$$\overline{\log_e(density(Month))} = \frac{1}{L} \sum_{LAT,LONG} \log_e(density(LAT, LONG, Month))$$

$$Var(\log_e(density(Month))) = \frac{1}{L-1} \sum_{LAT,LONG} (\log_e(density(LAT, LONG, Month)) - \overline{\log_e(density(Month))})^2$$

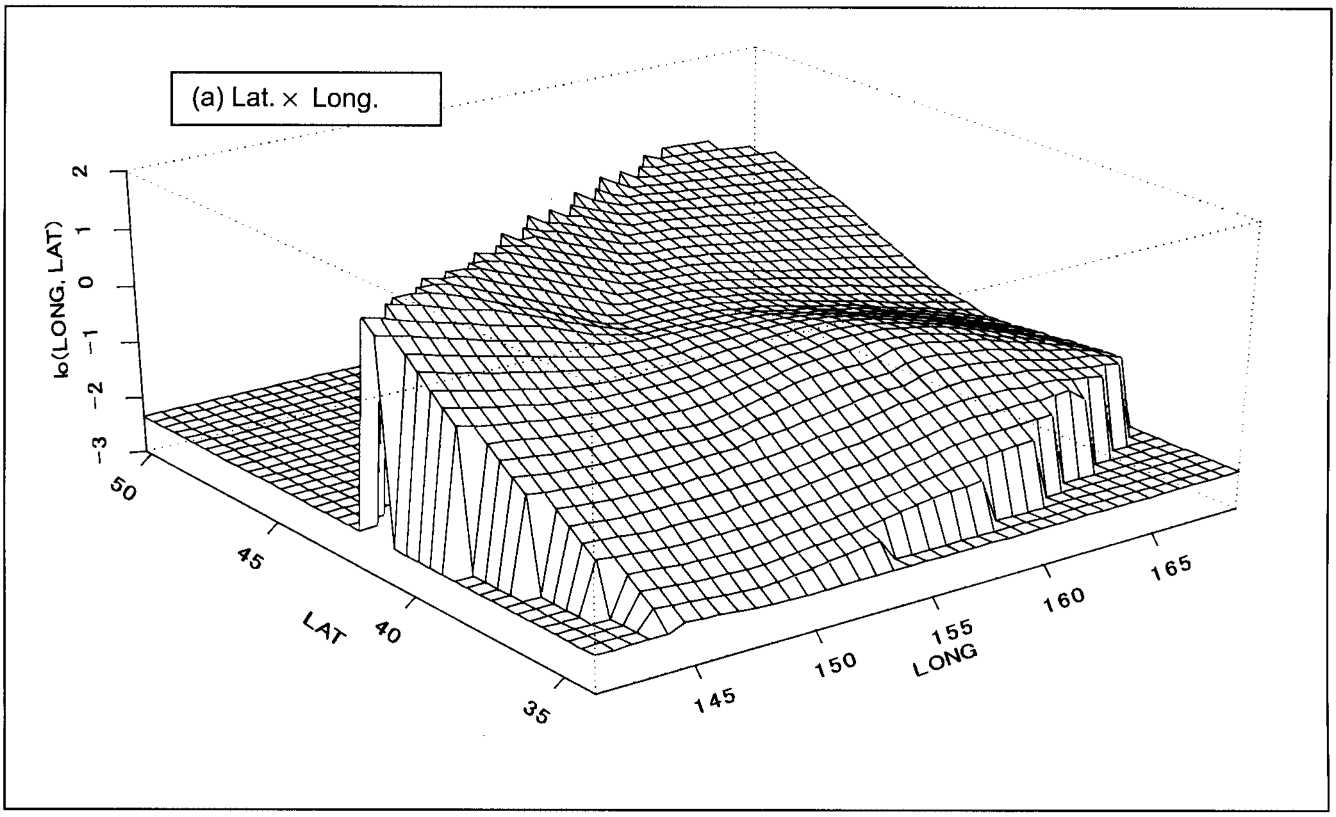
and *L* is the number of combination of *LAT* and *LONG* that is equal to the number of cells in gray- and black-coloured zone in Fig. 4. The trend surfaces of monthly-predicted density indices are shown in Fig. 3.

DISCUSSION

It is perhaps not surprising that only sightability remains in the model of detection process since sightability is the overall judgment for detectability made in the field and will represent a subjective integration of a number of factors including wave height, swell, wind speed, weather conditions etc. In this paper, the influence of school size was not considered as a covariate in the detection model because the resulting model cannot be used to estimate the density in regions where no whales were found if school size is treated as a covariate. Although Beaver and Ramsey (1998) recommended the method of Drummer and McDonald (1987), for simplicity, the adjustment contained in the DISTANCE program was used here. The appropriateness of this should be considered in a future study. However, it should be noted that in the present data set the school size was almost exclusively one (total number of detected schools = 422, total number of detected individuals = 443).

GAMs have been applied to obtain distribution patterns of density in several other areas (e.g. Palka, 1995; Hedley *et al.*, 1999). The GAM analysis in this paper results in a number of density distribution maps for western North Pacific minke whales. These reveal no conspicuous drops in the central part of the western North Pacific (Fig. 3 (a) and Fig. 4) i.e. they do not suggest a need to sub-divide this area of the North Pacific. However, this conclusion requires some qualification.

Model selection was carried out using the BIC (or QBIC), not the AIC values. AIC sometimes results in more parameters than BIC because AIC tends to overestimate the number of parameters needed (Kass and Raftery, 1995). In fact, the sensitivity test of the span of loess functions for the first model presented here showed that the BIC resulted in the model with the span of 0.5 whereas the AIC tended to suggest a much smaller value ($1/h = 0.032$). However, it cannot be ruled out that the result of the BIC is too conservative such that the selected span is too large to detect any true gaps in the study area. Other extended information criteria such as AIC_c (Burnham and Anderson, 1988) and CAIC (Bozdogan, 1987) were examined for selecting an appropriate *h* value. The result was that AIC_c selected $1/h = 0.05$ and CAIC selected $1/h = 0.5$. The former value is very small and would require considerably more data to reasonably apply such a complex model. Therefore, somewhat arbitrarily we produced plots such as those in Fig. 3 and Fig. 4 for the final model with values of $1/h$ of 0.25 and 0.125. These generally were similar to Figs 3 and 4 except that the increase in density from lower to higher latitudes lacked smoothness to some degree. However, we believe that model selection tools other than information criteria should be considered to look for the presence of gaps in distribution in any future study.



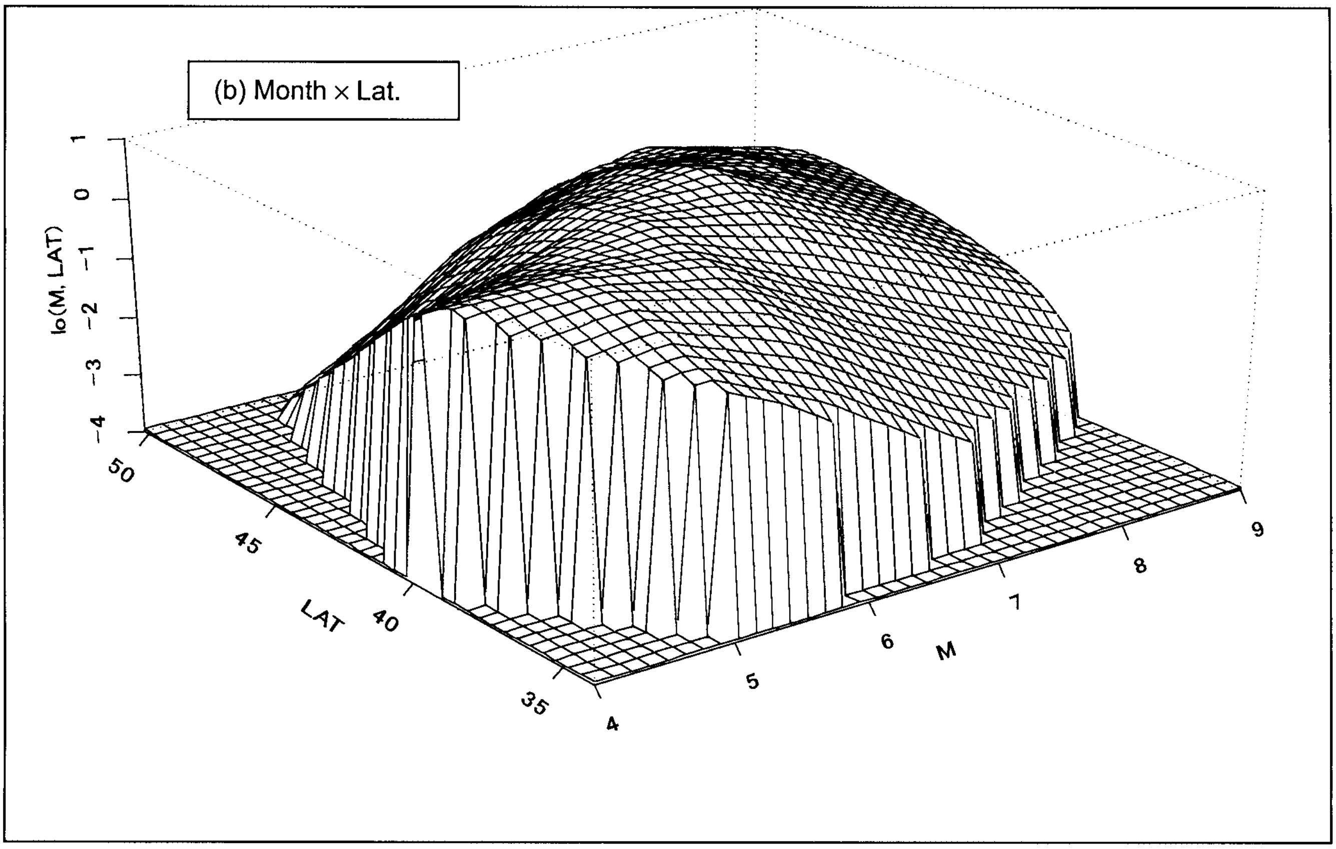


Fig. 3. Plots of fitted spline functions and perspective plots of fitted less smooth functions in the generalized additive model. The top plot is for latitude and longitude, and the bottom for month and latitude.

Fig. 3 (b) reveals some monthly variation in density with indices in high latitudes increasing whilst those in low latitude decreased as time passes. This result agrees with Hatanaka and Miyashita (1997). Fig. 4 shows these monthly changes in density more explicitly. In April, the peak in density indices occurs between 38°N-39°N and 154°E-160°E (Fig. 4 (a)), but this peak is not present in June and July (Fig. 4 (c) and (d)). However, there was clearly more effort in June and July than in April and May. In particular, the data for April are too patchy to be useful. It is

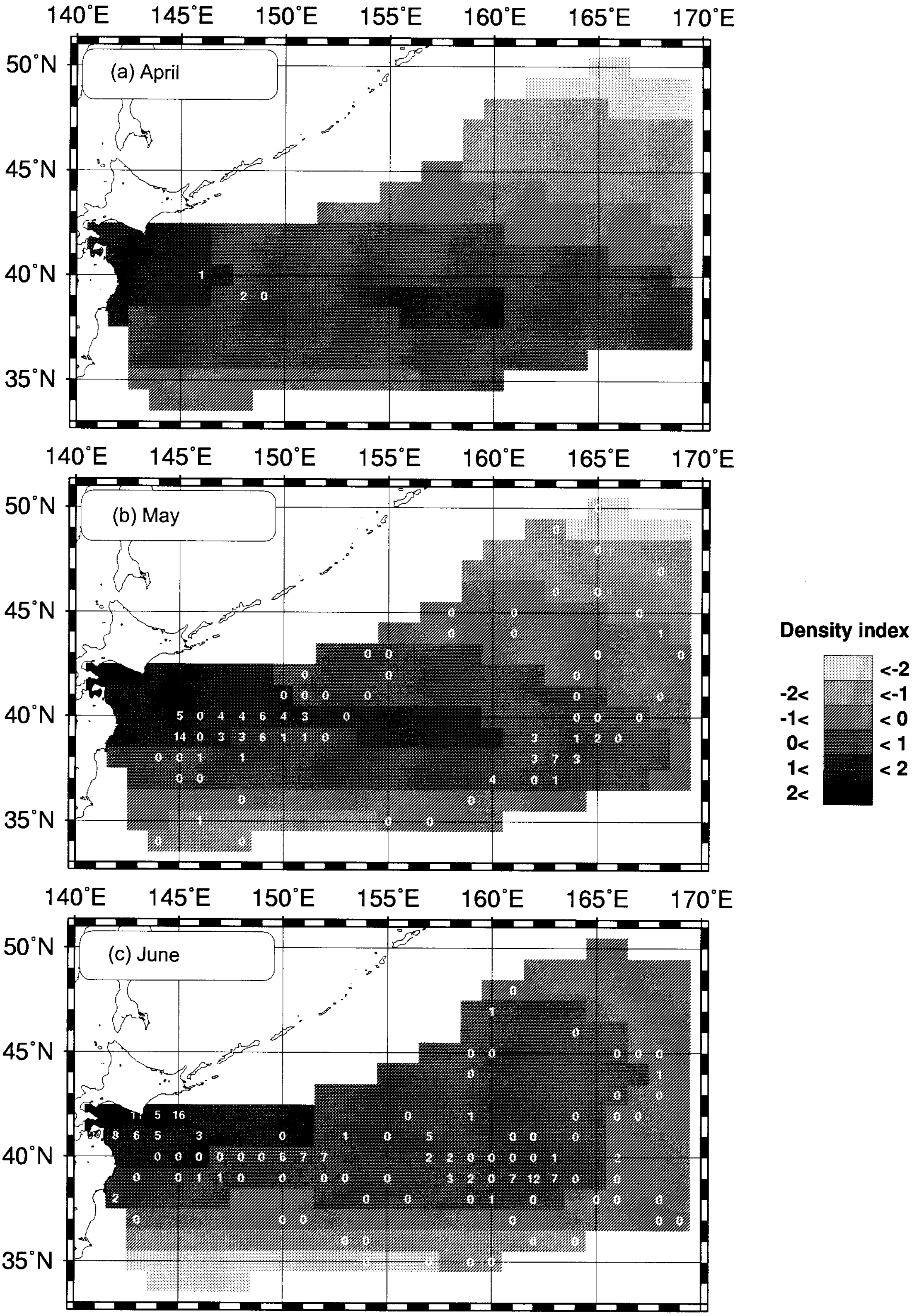


Fig. 4. The predicted density indices of North Pacific minke whales in (a) April, (b) May, (c) June, (d) July, (e) August and (f) September. The density index calculated in one degree square cell is standardised after logarithm transformation. The figure in one degree square cell is the actual sighted number for schools of whales.

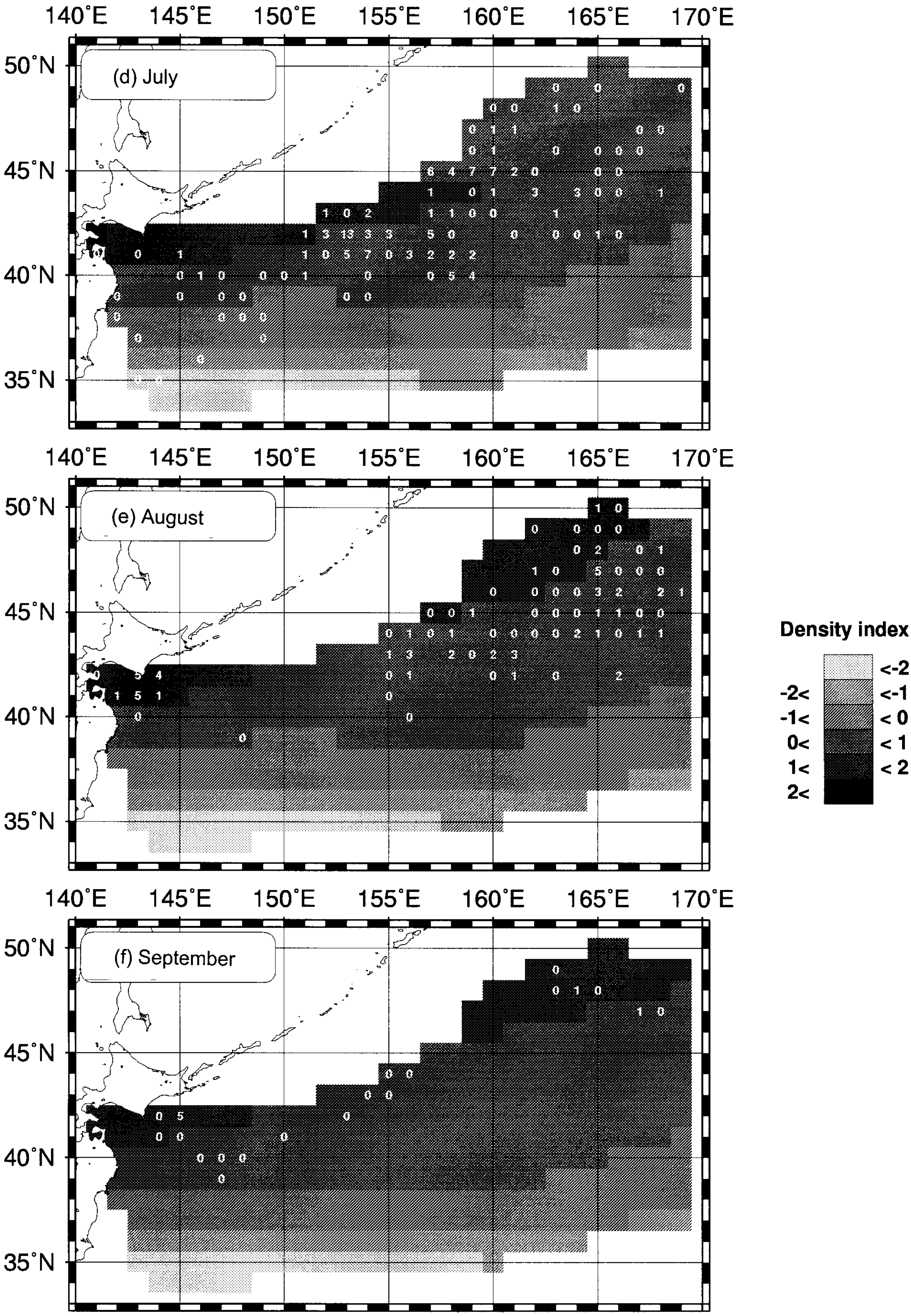


Fig. 4. (Continued).

also plausible that the high-density area in the western part of Fig. 4 reflects many sightings off the coast of Hokkaido which may reflect segregation of young whales (Hatanaka and Miyashita, 1997).

In conclusion, our analysis revealed no evidence to suggest a further division of the western North Pacific. However, in addition to the factors discussed above it is apparent that there was little effort to the south of the central part of the study area (Figs 2 and 4). Thus the density in that area is a result of extrapolation predicted from the model and this is important in reaching the conclusion. Future survey should try to increase sighting effort (and sightings) in this southern area.

ACKNOWLEDGMENTS

We greatly appreciate Y. Fujise, H. Hatanaka, K. Hiramatsu, S. Kawahara, T. Kitakado, H. Shono and Y. Takeuchi and reviewers, M. Bravington and S. Hedley for useful comments on earlier version of this paper. We also thank all captains, crews and researchers who participated in the JARPN surveys from 1994 to 1999. The first author also thanks S. Okamura for her tolerance and deep affections.

REFERENCES

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. pp. 267-81. *In:* B.N. Petran and F. Csaàki (eds.) *International Symposium on Information Theory*. 2nd. Edn. Akadèemiai Kiadi, Budapest, Hungary. 451pp.

Beavers, C.S. and Ramsey, F.L. 1998. Detectability analysis in transect surveys. *J. Wildl. Manage.* 62(3):948-57.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52(3):345-70.

Burnham, K.P. and Anderson, D.R. 1988. *Model Selection and Inference*. Springer Verlag, New York. 353pp.

Chambers, J.M. and Hastie, T.J. (eds.). 1992. *Statistical Models in S.* Wadsworth & Brooks/Cole, Pacific Grove, California. 608pp.

Cumberworth, S.L., Buckland, S.T. and Borchers, D.L. 1996. A spatial modelling approach for the analysis of line transect data. Paper SC/48/O 12 presented to IWC Scientific Committee, June 1996, Aberdeen (unpublished). 11pp. [Paper available from the Office of this Journal].

Drummer, T.D. and McDonald, L.L. 1987. Size-bias in line transect sampling. *Biometrics* 43:13-21.

Efron, B. and Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York. 436pp.

Hastie, T.J. and Tibshirani, R.J. 1990. *Monographs on Statistics and Applied Probability*. 43. *Generalized Additive Models*. Chapman & Hall, London. 335pp.

Hatanaka, H. and Miyashita, T. 1997. On the feeding migration of Okhotsk Sea-West Pacific stock minke whales, estimates based on length composition data. *Rep. int. Whal. Commn* 47:557-64.

Hedley, S., Buckland, S.T. and Borchers, D.L. 1999. Spatial modelling from line transect data. *J. Cetacean Res. Manage.* 1(3):255-64.

International Whaling Commission. 2001. Report of the Workshop to Review the Japanese Whale Research Programme under Special Permit for North Pacific Minke Whales (JARPN), Tokyo, 7-10 February 2000. *J. Cetacean Res. Manage. (Suppl.)* 3:377-477.

Kass, R.E. and Raftery, A.E. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773-95.

Laake, J.L., Buckland, S.T., Anderson, D.R. and Burnham, K.P. 1993. *Distance User's Guide, Version 2.0*. Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO. 72pp.

Martien, K.K. and Taylor, B.L. 2000. The limitations of hypothesis testing as a means of demographically delineating independent units. Paper SC/F2K/J3 presented at the JARPN Review Meeting, Tokyo, Japan, 7-10 February 2000 (unpublished). 32pp. [Paper available from the Office of this Journal].

Matsuoka, K., Hakamada, T., Fujise, Y. and Miyashita, T. 2000. Distribution pattern of minke whales based on sighting data during the JARPN 1994-1999. Paper SC/F2K/J16 presented at the JARPN Review Meeting, 7-10 February 2000 (unpublished). 17pp. [Paper available from the Office of this Journal].

McCullagh, P. and Nelder, J.A. 1989. *Monographs on Statistics and Applied Probability*. 37. *Generalized Linear Models*. 2nd Edn. Chapman & Hall, London. 511pp.

Palka, D. 1995. Influences on spatial patterns of Gulf of Maine harbor porpoises. pp. 69-75. *In:* A.S. Blix, L. Walløe and Ø. Ulltang (eds.) *Developments in Marine Biology*. 4. *Whales, Seals, Fish and Man: proceedings of the International Symposium on the Biology of Marine Mammals in the Northeast Atlantic. Tromso, Norway, 29 November-1 December 1994*. Elsevier Science B.V., The Netherlands. 720pp.

Swartz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461-4.

Taylor, B.L. 2000. Genetic population structure in the western North Pacific minke whale: an analysis of mtDNA data. Paper SC/F2K/J6 presented at the JARPN Review Meeting, Tokyo, Japan, 7-10 February 2000 (unpublished). 4pp. [Paper available from the Office of this Journal].

Taylor, B.L. and Chivers, S.J. 2000. An example of the calculation of the statistical power to detect population sub-division in North Pacific minke whales. Paper SC/F2K/J7 presented at the JARPN Review Meeting, Tokyo, Japan, 7-10 February 2000 (unpublished). 12pp. [Paper available from the Office of this Journal].