SOME COMMENTS ON THE PROCEDURE FOR TESTING ESTIMATORS OF KRILL ABUNDANCE WHICH UTILISE SURVEY DATA

D.S. Butterworth*, D.L. Borchers* and D.G.M. Miller**

Abstract

Since krill distributional data do show evidence of spatial correlation, estimators of abundance which attempt to model such effects (such as those based on Kriging techniques) may provide improved abundance estimates from survey data. However, computer simulation studies are first required to test whether such estimators, and alternative possible survey designs, are indeed likely to provide improved performance in practice. Such studies require a simple method for computer generation of krill distribution patterns, which are compatible with existing information on the distributions from surveys. "Two-level" models of krill distribution are considered. These achieve overall spatial correlation by placing krill swarms at random within larger aggregation features termed concentrations; these concentrations are then located at random within the survey area. These "two-level" models provide an encouraging improvement in fits to the distribution of inter-swarm distances observed on the 1981 FIBEX survey by MV *SA Agulhas*. However, evidence of model misspecification remains. Further work is needed before such models can be used as the basis for the simulation studies required - some suggestions are made in this regard.

Résumé

Vu que les données sur la distribution mettent bien en évidence une corrélation spatiale, les paramètres d'estimation d'abondance tentant de modéliser ces effets (tels que ceux fondés sur les techniques de Kriging) pourraient fournir de meilleures estimations d'abondance à partir des données des campagnes d'évaluation. Cependant, en premier lieu, des études par simulation informatisée sont nécessaires pour vérifier si ces paramètres d'estimation, et d'autres modèles de campagne possibles, sont bien susceptibles, sur le plan pratique, d'offrir de meilleurs résultats. Ces études nécessitent une méthode simple de production informatisée de schémas de répartition du krill qui soient compatibles avec les informations actuelles, fournies par les campagnes d'évaluation, sur les répartitions. Des modèles "à deux niveaux" de répartition du krill sont considérés. Ils réalisent la corrélation spatiale globale en plaçant les essaims de krill au hasard, à l'intérieur de grands rassemblements de krill nommés concentrations; ces concentrations sont ensuite situées au hasard dans l'aire étudiée. Ces modèles "à deux niveaux" sont prometteurs quant à une amélioration des ajustements à la distribution des distances entre les essaims observées par la campagne FIBEX 1981 du MV *SA Agulhas*. Quelques traces d'erreurs de spécification du modèle subsistent toutefois. Avant de pouvoir utiliser ces modèles comme bases des études par simulation, il importe de poursuivre les travaux - quelques suggestions sont avancées à ce sujet.

---

* Department of Applied Mathematics, University of Cape Town, Rondebosch 7700, South Africa
** Sea Fisheries Research Institute, Private Bag X2, Roggebaai 8012, South Africa

Резюме

В связи с тем, что данные по распределению криля действительно указывают на пространственную взаимосвязь, показатели численности (например основанные на методе Крайга), направленные на моделирование подобных эффектов могут представить более точные оценки численности на основе съемочных данных. Тем не менее, первой целью исследований по компьютерному моделированию является проверка вероятности действительного улучшения эффективности на практике подобных показателей и возможных альтернативных схем съемок. Такие исследования требуют внедрения простого метода компьютерного производства моделей распределения криля, которые совместимы с существующими съемочными данными по распределению. Рассматриваются "двухступенчатые" модели распределения криля. С помощью этих моделей можно получить общее пространственное соотношение, произвольно поместив скопления криля в пределах бльших агрегаций, называемых концентрациями; затем эти концентрации произвольно помещаются в пределах района съемки. Использование этих "двухступенчатых" моделей обеспечвает получение более точных подгонок к распределению расстояний между скоплениями, зарегистрированными в ходе съемки FIBEX, проведенной судном *SA Agulhas* в 1981 г. Тем не менее, имеется доказательство неправильного построения модели. Прежде чем эти модели могут быть использованы в качестве основы для требуемых исследований по моделированию, необходимо продолжить работу в этой области - в этой связи в данном труде приводится несколько предложений.

Resumen

Debido a que los datos de distribución del krill presentan evidencia de una correlación espacial, los estimadores de abundancia que intentan modelar tales efectos (como aquellos basados en las técnicas Kriging) podrían proporcionar estimaciones de abundancia mejoradas utilizando los datos de prospección. Sin embargo, se necesita primero hacer estudios de simulación computerizada para investigar si en efecto, tales estimadores, y posibles diseños alternativos de prospección, podrían proporcionar un mejor rendimiento. Tales estudios requieren un método simple para generar patrones de distribución de krill por medios computerizados, que sean compatibles con la información actual obtenida de las prospecciones. En este documento se consideran modelos de distribución de krill de "dos niveles". Estos logran una correlación espacial global mediante la posición aleatoria de las agrupaciones de krill dentro de mayores agregaciones denominadas concentraciones; estas concentraciones a su vez, son localizadas aleatoriamente dentro del área de prospección. Estos modelos de "dos niveles" proporcionan un alentador avance en los ajustes de la distribución de las distancias entre agregaciones observadas durante la prospección FIBEX de 1981, realizada por el buque *SA Agulhas*. Sin embargo, aún existe evidencia de errores de especificación en los modelos. Se necesita llevar a cabo más investigación antes de que se puedan utilizar tales modelos como la base para los estudios de simulación requeridos - se plantean algunas sugerencias a este respecto.

1. INTRODUCTION

A "straightforward" approach to abundance estimation from surveys of marine resources is to place a set of transects at random in the area of interest. The density estimate from each transect is then treated as an independent estimate of the density in the area, and the mean and associated standard error of the set of estimates from each transect (weighted appropriately in relation to transect length if necessary) immediately provide an estimate of overall abundance and of its variance. For the best possible precision (given a constraint on the survey effort available), the area should be pre-stratified on the basis of *a priori* expectations of density variations, with effort allocated so as to minimise the anticipated variance. Further, within any stratum, the directions of the transects should be chosen to lie parallel to any *a priori* expectation of density trend. Substitution of a fully random survey design by a systematic one would be a satisfactory procedure in most circumstances.

An objection that has been raised to this approach is that it is discarding the information on the spatial variation of density along a transect which is obtained during the survey. In the case of krill, it has been suggested (Foote and Stefansson, 1991) that estimation procedures such as Kriging, which model spatial correlation effects, could be used to produce improved estimates of abundance and its variance by making use of these within-transect data.

Incorporating spatial correlation in the estimation procedure requires that the values of additional parameters be estimated from the data. If krill swarms are in fact located at random within the stratum concerned, such an exercise can lead only to a reduction in the precision of the abundance estimate eventually obtained. Accordingly, the first question which needs to be addressed is whether there are data that do indeed show that krill swarm distributions exhibit spatial correlation. An answer is provided by the analysis by Miller and Hampton (1989a) of the distribution patterns of the swarms encountered during the 1981 FIBEX survey of an area of the south-west Indian Ocean by MV *SA Agulhas*. This demonstrates that that distribution was definitely non-random at the scale of the whole area surveyed, although there were indications of randomness at a smaller scale.

Accepting therefore that krill distributions do exhibit spatial correlation, the next question is whether estimators which attempt to model such effects will provide improved estimates of overall abundance from survey data than those, such as the "straightforward" approach above, which do not. This may also have implications for survey design, as different designs may provide improved estimates if a technique such as Kriging, for example, is used for estimation purposes. This question cannot be answered immediately, because the benefits of basing estimation on a model which better reflects the real situation, may nevertheless still be outweighed by the costs arising from the estimation of additional model parameters which may increase the variance of the abundance estimate.

The answer is therefore case dependent, and can only be obtained by computer simulation studies. Essentially these involve four steps:

(i) setting up a "real" krill distribution on the computer;

(ii) carrying out a "survey" of this distribution, to provide data of the kind which would be obtained from an actual survey;

(iii) analysing the data provided using the estimation procedure proposed, to obtain an abundance estimate; and

(iv) comparing that abundance estimate with the abundance of the "real" krill distribution (which is known to the computer), to ascertain how well the estimation procedure has performed.

Since stochastic factors are involved in these steps (both in generating the distribution, and in the sampling process effected by the survey), the procedure has to be repeated many times on the computer to ascertain the statistical properties (such as bias and variance) of the combination of survey design and estimation procedure proposed.

The "real" distribution set up in step (i) should reflect the known characteristics of krill distributions as closely as possible. Generating such distributions on the computer is simple if they are required to be random representations of an underlying average density, which is either constant or follows a given trend in space. However, once spatial correlation effects need to be introduced, this exercise becomes distinctly non-trivial. Ideally one would wish to generate the simulated distributions from empirical observations, but the essential difficulty is that the data are collected along <u>one</u>-dimensional transects, whereas it is a <u>two</u>-dimensional structure which has to be generated. Techniques are available which can generate spatially correlated data using the correlation model underlying the Kriging technique [decomposition of covariance matrix, see Fishman (1973)]. However, a possible objection to using these is that the true potential of a Kriging-based estimation procedure may be inflated if the "real" distributions are generated in this way, because this approach cannot make allowance for model misspecification (i.e., differences between the process actually governing the "real" distribution, and that assumed for the purposes of estimation from the data collected).

One simple approach to generating spatially correlated data is to use a two stage process. Thus, the krill distribution is envisaged to be comprised of larger scale concentrations distributed at random through the area; then, within each concentration, the krill swarms themselves are distributed at random. The simulated krill distributions generated by Butterworth (1989) and Mangel (1989) in their studies of the potential utility of krill fishery CPUE data to index abundance had this two-level structure. Generation of such data again becomes straightforward, because of the random nature of the placement of concentration centres and swarms within concentrations. However, viewed overall, the distribution of the krill <u>swarms</u> generated will exhibit positive spatial correlation.

This is an attractive basis upon which to proceed, but the final question which remains is whether the spatial correlation patterns exhibited by actual krill distributions are compatible with a simple two-level structure. This paper makes an initial attempt to address this question but reanalysing the data considered by Miller and Hampton (1989a) using simple two-level models.

2. DATA AND METHODS

The krill data analysed by Miller and Hampton (1989a) are shown in Figure 1, which is a reproduction of Figure 1 of their paper. There are significant differences between day- and night-time distribution patterns; further, an underlying density trend could be argued from inspection of this plot. However, for the purposes of the simple analyses of this paper, these complications will be overlooked.

If the krill swarms detected were randomly distributed over the area surveyed, the distances between successive swarms encountered should follow an exponential distribution.

Miller and Hampton (1989a) showed that the data exhibit significant deviations from such a distribution, so that the hypothesis that the overall distribution is random can be rejected.

Instead, we model a two-level distribution structure of concentrations and swarms-within-concentrations (as discussed above, and in line with the hierarchy of structure discussed in Miller and Hampton, 1989b) under the assumption of a random distribution at each level. Two models are considered. The first (A) is based on the relationship:

$P_A(y)$ = Prob[encounter swarm]

= Prob[encounter swarm if within concentration] x Prob[within concentration] + Prob[encounter swarm (i.e., next concentration) if outside concentration] x Prob[outside concentration]

$$= \lambda_1 e^{-\lambda_1 y} w + \lambda_2 e^{-l_2 y} (1 - w) \qquad (1)$$

where **y** is the distance travelled since the last swarm was encountered,

λ_1 is proportional to the number of swarms per unit area within a concentration,

λ_2 is proportional to the number of concentrations-per-unit area, and

w is the fraction of the survey area covered by concentrations.

Model (B) is less formally motivated, and has the form:

$P_B(y)$ = Prob[encounter swarm]

$$\begin{aligned} &= \alpha e^{-\lambda_1 y} && y \leq Y \\ &= \alpha e^{-\lambda_1 Y} e^{-\lambda_2 (y-Y)} && y > Y \end{aligned} \qquad (2)$$

where $\alpha = [\lambda_1^{-1}\{1 - e^{-\lambda_1 Y}\} + \lambda_2^{-1} e^{-\lambda_1 Y}]^{-1}$, and

Y is related to the (typical) radius of the concentration.

The basic assumption underlying this approach is that once a distance greater than **Y** has been travelled without encountering a swarm, the vessel has passed outside the concentration boundary so that the probability-per-unit distance (g) of encountering a swarm drops to a lower value. In principle, the constants of proportionality for the two expressions of equation (2) could differ, but in the interests of parsimonious parametrization, continuity (though, of course, not derivative continuity) of probability as a function of y (and therefore continuity at **Y**) is assumed. In reality, concentrations would have different radii, so that the true probability function would be some weighted average of equation (2) over different values of **Y**, which would then smooth out any discontinuity. Equation (2) may thus still remain a reasonable empirical representation of such a function.

Both models have three parameters to be estimated from the data:

(A): λ_1, λ_2, w

(B): λ_1, λ_2, Y

The parameters were estimated by numerical maximization of the respective likelihood functions $\prod_{i=1}^{n} P_A(y_i)$ and $\prod_{i=1}^{n} P_B(y_i)$. The data to which these models were then applied comprises n = 1 566 successive inter swarm distances (y_i), where these distances were measured from centre-to-centre of each intercept of a krill swarm (see Miller and Hampton, 1989a, for the definition of this measure, which they name BSWC). [Note: Not all the krill aggregations encountered were necessarily "swarms" as conventionally defined, but this level of detail is ignored for this analysis, so that the term "swarm" is used throughout for convenience.]

3. RESULTS AND DISCUSSION

The results of the fits of models (A) and (B) to the data are given in Table 1 and Figure 2. This table also includes the fit for an assumed random distribution, i.e. the negative exponential:

$$P_{NE}(y) = \lambda_1 e^{-\lambda_1 y} \tag{3}$$

Though this is the same form as fitted by Miller and Hampton (1989a) to these data, the results are different. This is because the maximum likelihood estimator:

$$\hat{\lambda}_1 = n \Big/ \left(\sum_{i=1}^{n} y_i \right) \tag{4}$$

was used here, while Miller and Hampton used the estimator:

$$\hat{\lambda}_1 = (\ell n\ 2)/y_{median} \tag{5}$$

These two estimators are deterministically identical (if the model assumed is correct). The estimator of equation (5) has the advantage of being more robust to occasional outliers at large y, and is more appropriate if fitting the data at smaller y is of the greatest concern. However, in the context of modelling the distribution at <u>all</u> scales (not just the smaller scale), the numerous large inter-swarm distances cannot all be dismissed as outliers, so that use of this estimator becomes questionable given the obvious model misspecification. The maximum likelihood estimator of equation (4) has therefore been used instead, and provides comparability with the fitting procedures used for the other models.

The x^2 values in Table 1 and Figure 2 show that the two-level models (A) and (B) achieve fits to the data which are much improved to that of the negative exponential. The fits of the two-level models are very similar, both indicating a change in the pattern of the distribution (to a much slower decline with y) at a scale of some 5 to 7 km. However, both fits show the same systematic deviations from the data, with predictions that are somewhat too low between [0, 0.5] and [2, 20] km, and too high in the complementary ranges. The fact that the x^2 values still indicate statistically significant lack of fit in both cases, means that model misspecification does remain to some extent.

It is appropriate to consider whether the estimates of the model parameter values are meaningful, given other data available from the survey. Attention is confined to model (A), which (on a minimum x^2 basis) provides a better fit than model (B) for the same number of estimable parameters. Consider a simple situation where all swarms are circular with the same radius **r**, and all concentrations also circular with identical radii **R**. If there are d swarms per unit area within a concentration, and the density of concentrations in the survey region is **D** per-unit area, then the following relationships hold:

$$\begin{aligned} \lambda_1 &= 2rd \\ \lambda_2 &= 2RD \\ w &= \pi R^2 D \end{aligned} \tag{6}$$

The mean intercepted swarm length (ℓ) in such a model is given by:

$$\ell = \pi r/2 \tag{7}$$

Data from the survey gives a value for **P** of (26.786/1567), which yields:

$$r = 0.0109 \text{ km} \tag{8}$$

Substituting this and the parameter estimates of Table 1 into equations (6) gives the values:

$$\begin{aligned} d &= 54.6 \quad \text{swarms/km}^2 \\ R &= 13.7 \quad \text{km} \\ D &= 0.00146 \text{ concentrations/km}^2 \end{aligned} \tag{9}$$

This simple model makes a prediction for the fraction (**f**) of the survey area covered by krill, viz:

$$\begin{aligned} \hat{f} &= d\pi r^2 \,.\, D\pi R^2 \\ &= 0.0175 \end{aligned} \tag{10}$$

The corresponding observed value is:

$$\begin{aligned} f_{obs} &= 26.786 \text{ km covered by krill} / 6579.654 \text{ km surveyed} \\ &= 0.00407 \end{aligned} \tag{11}$$

i.e. a some four-fold discrepancy.

This discrepancy is not that discouraging, given the oversimplification of calculations which were performed on the assumptions that all swarms and concentrations are circular and have fixed sizes, and further the model misspecification evident for model (A) [and (B)]. The "problem" seems to be the "high" estimate for **w**, which implies that concentrations covered as much of 86% as the survey area - seemingly an unrealistically high proportion?

Model (A) can be constrained to be consistent with the observed value for **f**. This requires that:

$$\pi r \lambda_1 w/2 = f_{obs}$$

$$\text{i.e.} \quad w = 2\, f_{obs}/(\pi r \lambda_1) = 0.238/\lambda_1 \qquad (12)$$

so that model (A) becomes:

$$P^*_A(y) = 0.238\, e^{-\lambda_1 y} + (1 - 0.238/\lambda_1)\, \lambda_2\, e^{-\lambda_2 y} \qquad (13)$$

The results for this two-parameter version of the model can be found in the column headed: ' Model (A) "constrained" ' in Table 1. Clearly the quality of fit has deteriorated substantially, indicating that the discrepancy between the values of $\hat{f}$ and $\mathbf{f_{obs}}$ in equations (10) and (11) above is not trivially resolvable.

4. CONCLUDING REMARKS

The ability of the two-level models considered to mimic an observed spatial distribution pattern for krill swarms is encouraging. Nevertheless, evidence of model misspecification remains, and the ability of more complex models to provide improved fits to data needs to be investigated. This should be done before such models are used to provide simulated krill distributions for tests of alternative survey strategies and krill abundance estimators, because realistic distributions are required if the tests are to give reliable results.

One immediate priority for future work would seem to be the application of the models developed here to other krill data sets, to see whether similar estimates of parameter values are obtained. Another would be the extension of the models to allow the λ values to vary with position in a simple way, so as to reflect overall spatial trends in krill density.

ACKNOWLEDGEMENTS

This work was supported in part by the Foundation for Research Development, South Africa.

REFERENCES

BUTTERWORTH, D.S. 1989. A simulation study of krill fishing by an individual Japanese trawler. In: *Selected Scientific Papers, 1988, Part 1 (SC-CAMLR-SSP/5)*. CCAMLR, Hobart, Australia: pp. 1-108.

FISHMAN, G.S. 1973. *Concepts and Methods in Discrete Event Digital Simulation*. John Wiley and Sons. p. 385.

FOOTE, K.G. and G. STEFANSSON. 1991. Definition of the Problem of Estimating Fish Abundance Over an Area from Acoustic Line-Transect Measurements of Density. Document *WG-Krill-91/8*. CCAMLR, Hobart, Australia: p. 19.

MANGEL, M. 1989. Analysis and modelling of the Soviet Southern Ocean krill fleet. In: *Selected Scientific Papers, 1988, Part 1 (SC-CAMLR-SSP/5)*. CCAMLR, Hobart, Australia: 127-235.

MILLER, D.G.M. and I. HAMPTON. 1989a. Krill aggregation characteristics: spatial distribution patterns from hydroacoustic observations. *Polar Biol. 10*: 125-134.

MILLER, D.G.M. and I. HAMPTON. 1989b. Biology and ecology of the Antarctic krill (*Euphausia superba* Dana): A review. *BIOMASS Sci. Ser. 9:* p. 166.

Table 1: Fits to the frequency distribution of inter-swarm distances by the various models considered. Full descriptions of the models and definitions of their parameters are provided in the text.

Distance Interval	Observed	Negative Exponential	Model (A)	Model (B)	Model (A) "constrained"
0 - 0.2	328	73	287	318	79
0.2 - 0.4	277	70	227	254	74
0.4 - 0.6	162	66	179	202	69
0.6 - 0.8	122	63	141	161	64
0.8 - 1.0	86	60	112	128	60
1.0 - 1.2	69	58	88	102	56
1.2 - 1.4	49	55	70	81	52
1.4 - 1.6	42	52	55	65	49
1.6 - 1.8	26	50	44	52	46
1.8 - 2.0	35	48	35	41	43
2.0 - 3.0	101	206	94	108	177
3.0 - 4.0	54	163	34	34	129
4.0 - 5.0	28	128	16	8	96
5.0 +	187	474	184	192	572
x^2	-	2107	69	141	1929
d.f.	-	13	10	10	11
<u>Parameters</u>					
λ_1 (km^{-1})	-	0.238	1.19	1.14	0.39
λ_2 (km^{-1})	-	-	0.04	0.04	0.06
w	-	-	0.86	-	(0.61)
Y (km)	-	-	-	4.57	-

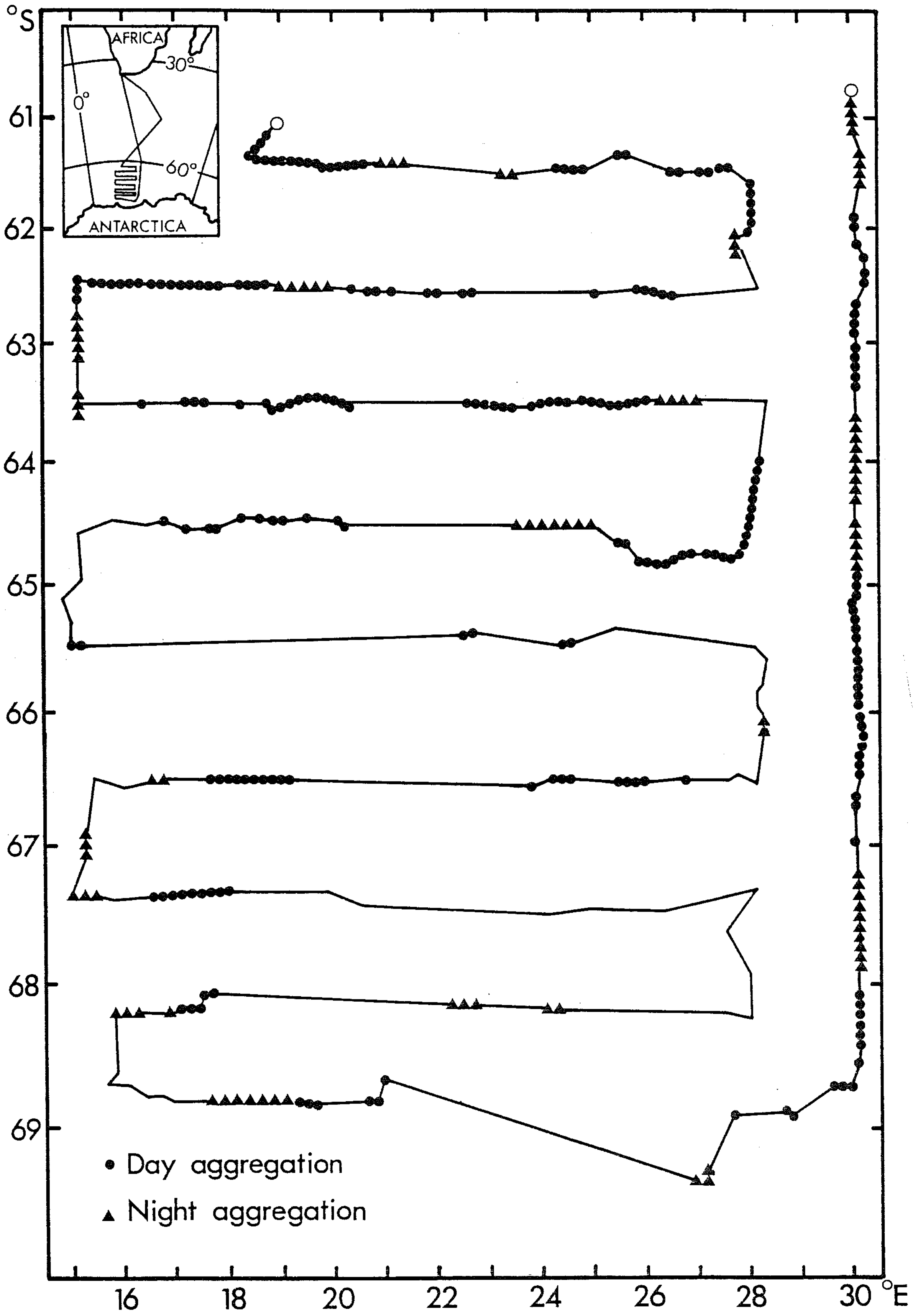


Figure 1 : Krill aggregations detected in the south-west Indian Ocean during FIBEX, February to March, 1981. The cruise track of MV *SA Agulhas* is also shown.

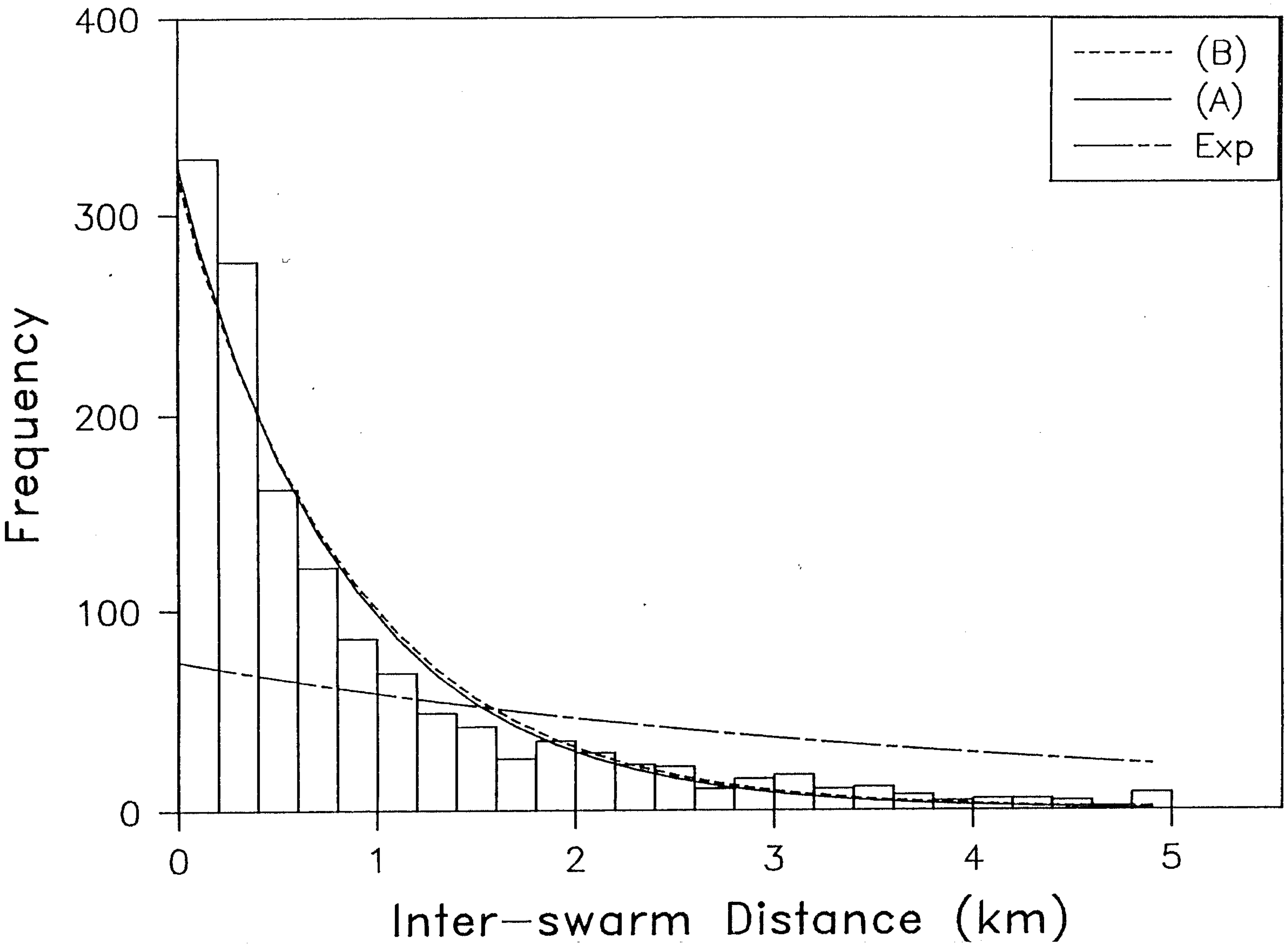
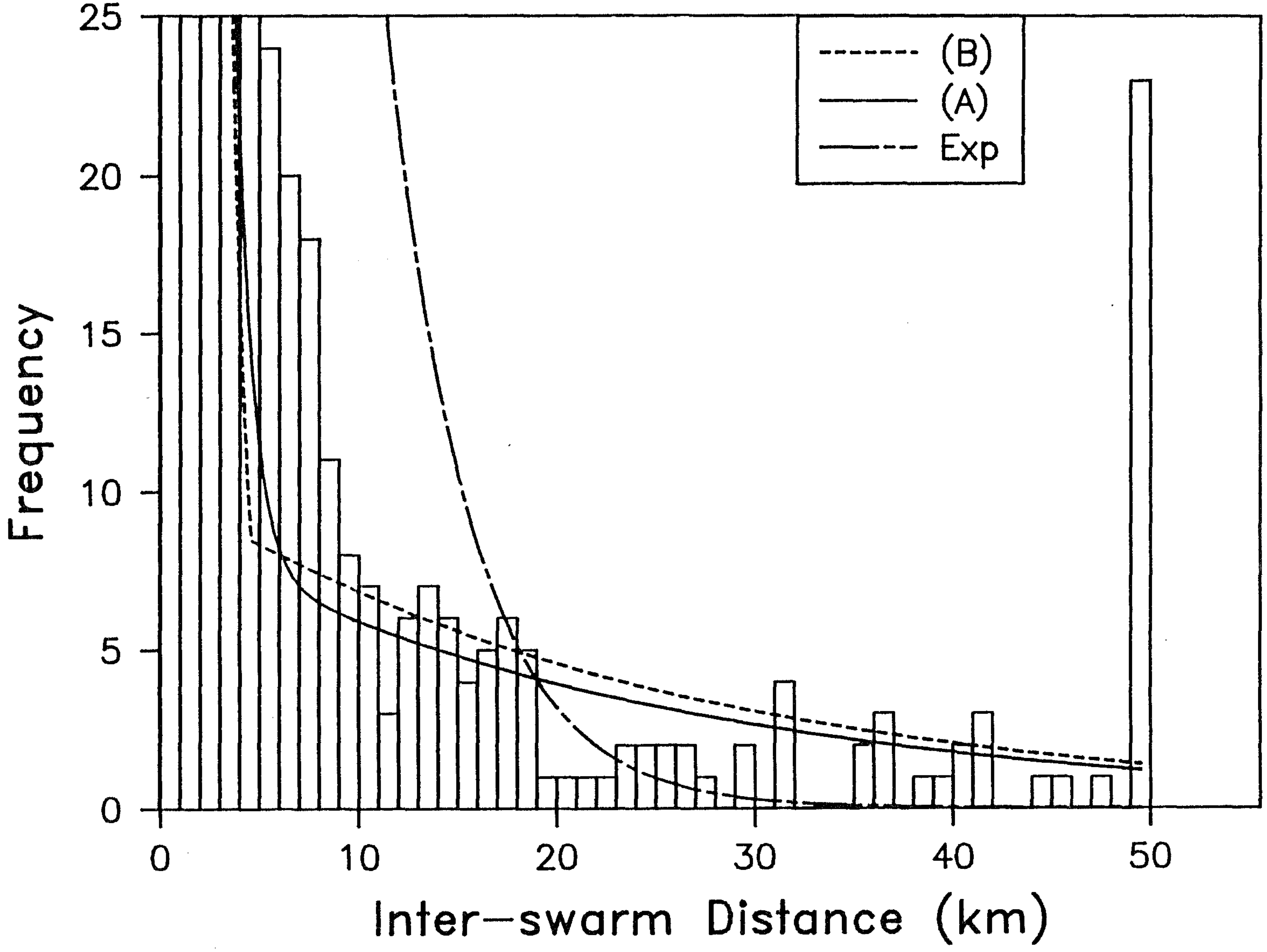


Figure 2 : Comparison of the observed frequency distribution of inter-swarm distance (**y**) with the fits to these data by the negative exponential model, and by the two-level models (A) and (B). The same plot is shown at two different scales to better indicate the behaviour of the fits at small and at large values of **y**.


25
20
15
10
5
0
Frequency
0
10
20
30
40
50
Inter–swarm Distance (km)
(B)
(A)
Exp

Liste des tableaux

Tableau 1: Ajustements de la distribution de fréquences des distances séparant les essaims par les divers modèles considérés. Les descriptions intégrales des modèles et les définitions de leurs paramètres figurent dans le texte.

Liste des figures

Figure 1: Concentrations de krill détectées dans le secteur sud-ouest de l'océan Indien pendant la campagne FIBEX, de février à mars 1981 et trajet du MV *SA Agulhas*.

Figure 2: Comparaison de la distribution observée de fréquences de distances entre les essaims (**y**) avec l'ajustement de ces données par le modèle exponentiel négatif, et par les modèles à deux niveaux (A) et (B). Ce tracé est présenté à deux échelles différentes pour mieux indiquer le comportement des ajustements à des valeurs faibles ou élevées de **y**.

Список таблиц

Таблица 1: Подгонки к частотному распределению расстояний между скоплениями различными рассмотренными моделями. Полные описания моделей и определения их параметров приводятся в тексте.

Список рисунков

Рисунок 1: Агрегации криля, зарегистрированные в юго-западной части Индийского океана в ходе съемки FIBEX в феврале-марте 1981 г. Также указан маршрут плавания *SA Agulhas*.

Рисунок 2: Сравнение зарегистрированного частотного распределения расстояний между скоплениями (**у**) с подгонками к этим данным негативной экспоненциальной моделью, и двухступенчатыми моделями (А) и (В). Та же схема показана в двух разных масштабах в целях лучшей индикации поведения подгонок при больших и малых значениях **у**.

Lista de las tablas

Tabla 1: Ajustes de la distribución de las distancias entre cardúmenes realizados por los diversos modelos considerados. En el texto se proporciona una descripción completa de los modelos y de las definiciones de sus parámetros.

Lista de las figuras

Figura 1: Concentraciones de krill detectadas en el océano Indico sudoccidental durante FIBEX, febrero a marzo de 1981. Se muestra además la derrota del buque *SA Agulhas*.

Figura 2: Comparación de la distribución de frecuencia de la distancia observada entre cardúmenes (**y**) con el ajuste de estos datos, realizada por el modelo exponencial negativo y por los modelos de dos niveles (A) y (B). El mismo gráfico se muestra a dos escalas diferentes para indicar, de mejor manera, el comportamiento de los ajustes al emplear valores altos y bajos de **y**.