NAFO Sci. Coun. Studies, 16: 123–128

Uncertainty in Biomass Assessments and Survey Planning

Z. I. Kizner
All-Union Research Institute of Marine Fisheries and Oceanography (VNIRO)
17 Verkhnaya Krasnoselskaya, Moscow, 107140, USSR

Abstract

Simple relationships expressing dependence of the sampling related error in biomass estimates (by means of a survey) on statistical characteristics of fish concentration density fields under examination and on parameters of survey itself, have been derived with the help of mathematical statistics. As for hydroacoustic surveys, the anisotropy index, correlation radius along the transects and the variation coefficient, serves as field characteristics on which the error depends, and the direction of survey with respect to the axis of the correlation ellipse and frequency of transects serve as survey parameters. Dependencies offered are applicable to surveys over large regions and can be used in practise both for a *posteriori* estimation of the error made when evaluating biomass assessment, and for survey planning on the basis of *apriori* information about statistical characteristics of concentration density fields. They might make a basis for the procedure of survey operative control.

Introduction

A biomass estimate of fishing resources is the main result of a hydroacoustic or a trawl survey. Since estimates of such a kind cause a certain effect on making decisions which very often have a considerable economic and ecological meaning, it is necessary to supply them with confidence intervals indicating the limits of possible errors with desired probability. Thus, it is important to find out which survey parameters and statistical characteristics of fish concentration fields, the error of the obtained biomass estimate might depend on and how this dependence can be expressed mathematically, while accounting for the probability nature of the estiamte in question.

If such a dependence has been revealed, it would be possible to solve an inverse (quite important from the practical point of view) problem of determining parameters of an optimal survey allowing to find a biomass estimate, the error of which does not exceed the defined level with desired probability.

This paper is devoted to all these problems.

Method

The first consideration is confined to the easiest and most widespread method of estimating biomass B in a region under consideration: the relationship $B = \bar{\rho}S$ is used, where S is the region area, and $\bar{\rho}$ is the average surface density of concentrations in the region, evaluated using information obtained through an 'instant' (i.e. rather short in time) survey. The error in acoustic survey data can be regarded as an additive one (distributed normally), so the arithmetic mean should be used as the average density. The error structure of the data of a trawl survey is usually more complicated, and therefore an appropriate transformation of the data should be performed first, then the arithmetic mean must be calculated, and finally, the mean must be transformed inversely.

If it is not to consider the measurement errors (which can be assumed as known) one should regard the relative error δ of estimate B as a sampling relative error of the estimate of average density $\bar{\rho}$. As for hydroacoustic surveys, neither B nor δ depend on integration interval (if an echointegrator is used). Thus, in this context echo-surveys can be considered similarly to trawl surveys, assuming that the information obtained from a survey of any of these types correspond to knots of a regular rectangular grid covering the region under examination, with the steps h_x and h_y, along coordinate axes x and y, connected with the direction of a survey (Fig. 1); these points will be called knots for short. The difference between them lies in types of the error in the initial data (additive or, say, multiplicative) and in fact that for a trawl survey usually $h_x \sim h_y$ (i.e. the steps are close in their orders of magnitudes), while in the case of an echo-survey one of the steps (further on h_x), corresponding to the distance covered by a vessel between two successive echo pulses, is much less than the other one (h_y) which represents the distance between transects: $h_x \ll h_y$. For simplicity of the exposition we shall

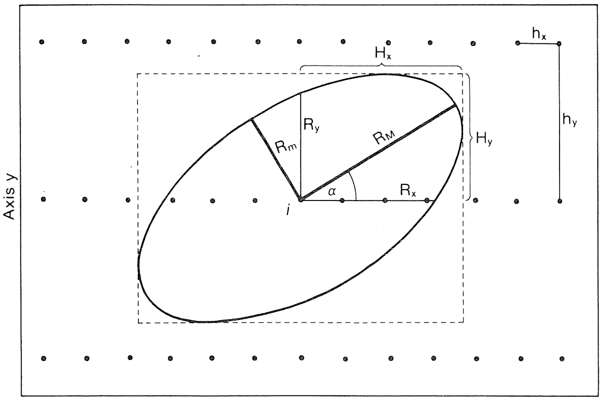

Fig.1. Main notations.

assume, that in the case of a trawl survey the necessary transformation is already made and ρ represents the transformed density (nevertheless, we shall call it density).

Usually fish concentration density fields have a typical patch-like structure. Certain patches have irregular shapes and are located in disorder (they can gather to create a big aggregation or drift apart to distances which considerably exceed their own sizes); density within one patch, as well as its shape are subject to random perturbations. Thus, when speaking about an 'instant' survey over a large region (macro-scale survey), one can (to the first approximation) consider the density ρ as a stationary homogeneous random field (if the field is not homogeneous in the whole region the latter can be divided into homogeneous strata).

Isotropic fields

Let us first assume that the field $\rho(x,y)$ is an isotropic one. This assumption is usually valid for fields in open ocean regions, far from shelves, jet currents, equator and other physio-geographical phenomena which can give rise to the existing specified directions. The homogeneous isotropic field ρ has got one and the same variance D_ρ for all its points, and its normalized autocorrelation function A, characterizing statistical interdependency of ρ values in any two points, depends itself only upon the distance r between them: $A = A(r)$. In the general case, correlation radius R, the minimum distance at which correlation between density values becomes negligible, depends on direction. But for an isotropic field R = const, so all points in which density correlates with the density in a fixed point, are in fact concentrated within the circle of radius R (correlation circle) with its centre in the fixed point.

We shall need two more values besides R to characterize the function A:

$$\alpha_1 = \frac{1}{\pi R^2}\iint_\omega A(r)\,dx\,dy = \frac{2}{R^2}\int_0^R rA(r)\,dr \qquad \dots (1)$$

$$\alpha_2 = \frac{1}{R}\int_0^R A(r)\,dr \qquad \dots (2)$$

These values are the integral average values of A over correlation circle ω and correlation radius respectively. Finally, we shall assume values h_x, h_y and R to be small enough compared to the region sizes: in reality, it is exactly the situation. This will allow us to consider all knots as 'equal in rights', neglecting differences between internal points and those belonging to the boundary strip.

According to a relation, well known in mathematical statistics, the variance $D_{\bar\rho}$ of the estimate of the average density, $\bar\rho = \sum_{i=1}^{N}\rho_i/N$, calculated for all N knots (N is the total number of knots), is presented by the sum:

$$D_{\bar\rho} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} K_{ij} \qquad \dots (3)$$

where $K_{ij} = D_\rho A(r_{ij})$ are correlation moments of the field ρ for the i-th and j-th knots, r_{ij} is the distance between these knots.

Regular surveys (transects are parallel to each other, steps are constant). One can replace N by S/h_xh_y in (3) (the relative error of this approximation is as less as smaller the steps h_x and h_y are). Taking into account the 'equality in rights' of all knots one can fix an internal knot with any number, i for example. Then, using the fact that the field ρ is an isotropic one, the formula (3) can be rewritten in a different form:

$$D_{\bar\rho} = \frac{D_\rho}{S} h_x h_y \sum_{j=1}^{N} A(r_{ij}) \qquad \dots (4)$$

Stations of a **trawl survey** are usually installed far-between, thus, one can write for the orders of magnitudes of the steps: $h_x \gtrsim R$, $h_y \gtrsim R$. In this case $A(r_{ii}) = 1$ (as $r_{ii} = 0$), and for $i \neq j$ all values of $A(r_{ij})$ are practically equal to zero, so from (4) we obtain:

$$D_{\bar\rho} = D_\rho h_x h_y/S \qquad \dots (5)$$

This relationship, written in the form $D_{\bar\rho} = D_\rho/N$, is well-known and often used when processing independent experimental data.

In the other limiting case which seems to be rarely realized in practice and which, however, is very important for understanding the main results of the paper, when $h_x \ll R$ and $h_y \ll R$, the expression $h_xh_y\sum_{j=1}^{N}A(r_{ij})$, being a part of (4), can be replaced with high accuracy by the intergral of A over correlation circle. In other words, using (1), the equality (4) can be rewritten in the form:

$$D_{\bar\rho} = \pi\alpha_1 D_\rho R^2/S \qquad \dots (6)$$

Hydroacoustic survey occupies an intermediate position between the two above described cases: now $h_x \ll R$, $h_y \gtrsim R$, and thus, density in the i-th knot correlates only with the density in knots located on the same transect at a distance no more than R is. That is why, when replacing expression $h_x \sum_{j=1}^{N} A(r_{ij})/2$ by the integral of A over correlation radius, we obtain out of (1) and (4):

$$D_{\bar{\rho}} = 2\alpha_2 D_\rho R h_y/S \quad \ldots (7)$$

In practical calculations for large N the estimate of the average density $\bar{\rho}$ can be regarded as a random value distributed according to a normal law. Thus, its absolute error ε depends on the confidence probability β and $\sigma_{\bar{\rho}} = \sqrt{D_{\bar{\rho}}}$ which is a standard error of $\bar{\rho}$:

$$\varepsilon = t_\beta \sigma_{\bar{\rho}} \quad \ldots (8)$$

In (8) $t_\beta = \sqrt{2}\,\phi^{-1}(\beta)$ and $\phi^{-1}(\beta)$ is the Laplace's inverse function of a defined confidence probability β; tables of values of t_β can be found in any book on mathematical statistics.

Thus, passing from ε to the relative error $\delta = \varepsilon/\bar{\rho}$, from (5) – (8) we obtain the following relationships:

for $h_x > R$, $h_y \gtrsim R$ (trawl survey)

$$\delta = t_\beta v \sqrt{h_x h_y/S} \quad \ldots (9)$$

for $h_x \ll R$, $h_y \ll R$ ('superfrequent' survey)

$$\delta = t_\beta b v R/\sqrt{S} \quad \ldots (10)$$

for $h_x \ll R$, $h_y \gtrsim R$ (hydroacoustic survey)

$$\delta = t_\beta c v \sqrt{R h_y/S} \quad \ldots (11)$$

where $b = \sqrt{\pi a_1}$, $c = \sqrt{2a_2}$ and $v = \sigma_\rho/\bar{\rho}$ is the variation coefficient of the density field ρ. Note that, if correlation properties of fields under examination are similar (i.e. when reduced to normalized argument, r/R, their correlation functions coincide), then b and c are universal constants.

From (9) – (10) it is clear that the minimum possible error is practically made in the case of steps (h_x and h_y for a trawl survey or h_y — for a hydroacoustic one) of the order of R (or some less); 'superfrequent' survey is inefficient because in this case (expression (10)) the error does not depend on h_x and h_y, i.e. because of correlation between data it does not decrease with the increase of the number of knots.

Expressions (9) – (11) represent desired dependencies of biomass estimate error upon survey parameters and isotropic field statistical characteristics. Consequently, resolving equality (9) relative to $h_x h_y$, and (11) – relative to h_y, we can obtain a mathematical basis for survey optimal planning. Thus, if it is necessary to assess the stock size in a certain region of the open ocean with the help of a hydroacoustic survey with such accuracy that relative error should not exceed the level $\triangle$ with the probability β, and if *apriori* estimate of v is accurate enough, then distance between transects, h_y, is to be taken from the following condition:

$$h_y < \left(\frac{\triangle}{t_\beta c v}\right)^2 \frac{S}{R} \quad \ldots (12)$$

Irregular surveys. The relationships obtained can be easily generalized to cases of irregular surveys, when distances between stations and transects are not constant (e.g. random) but the orders of magnitudes of the distances satisfy the above-mentioned inequalities, relating them with the correlation radius. In such a case, equation (10) stays valid; for trawl surveys instead of equation (9), one has the well-known equality $\delta = t_\beta c v \sqrt{N}$; and for hydroacoustic surveys (not only with parallel transects) the relationships (11) and (12) should be replaced by $\delta = t_\beta c v \sqrt{R/L}$ and $L \geqslant (t_\beta c v/\triangle)^2 R$, where L is the total length of the survey trajectory (i.e. the sum of the lengths of the transects). It is convenient sometimes to use somewhat different forms of the two latter relationships: $\delta = t_\beta c v \sqrt{R/l_x}/\sqrt{N_y}$ and $N_y \geqslant (t_\beta c v/\triangle)^2 R/l_x$, where l_x is the average length of transects and N_y is the number of transects.

Anisotropic fields

Up to now we have been considering isotropic density fields. However, quite simple geometric considerations allow one to apply results obtained to anisotropic fields in which one can specify two perpendicular directions in such a way that along one of them the correlation radius is maximum, while along the other one it is minimum. Such a situation occurs, for example, when a survey is being carried out in shelf waters where the medium and, correspondingly, the concentration characteristics change insignificantly along the shelf, and change considerably in the perpendicular direction. In this case, the correlation circle gets deformed and becomes an ellipse, and here the large and small radii shall be denoted as R_M and R_m.

It is clear that now the error δ can depend upon the anisotropy index $k = R_M/R_m \geqslant 1$ and the survey direction, which we shall define as the angle α between axis x and the large axis of the correlation ellipse. Let us introduce the symbols $R_x = R_M/\sqrt{k^2\sin^2\alpha + \cos^2\alpha}$ and $R_y = R_M/\sqrt{k^2\cos^2\alpha + \sin^2\alpha}$ for correlation radii in directions of x and y axes, as well as $H_x = R_M R_m/R_y$ and $H_y = R_M R_m/R_x$ for the half-sizes of a rectangle with sides being parallel to the axes x and y and embracing this ellipse (Fig. 1). Similar to the above described, we obtained three variants of a survey: $h_x \gtrsim H_x$ and $h_y \gtrsim H_y$, then, $h_x \ll H_x$ and $h_y \ll H_y$, and finally, $h_x \ll H_x$ and $h_y \gtrsim H_y$. The relationship (9) stays valid, the constants b and c stay the same, R in (10) is replaced by $\sqrt{R_M R_m}$. For a **regular** hydroacoustic survey (the third variant) we have instead of (11) and (12): $\delta = t_\beta cv\sqrt{R_x h_y/S}$, $h_y \leqslant (\triangle/t_\beta cv)^2 S/R_x$, while for an **irregular** hydroacoustic survey with parallel transects we use $\delta = t_\beta cv\sqrt{R_x/l_x}/\sqrt{N_y}$, $N_y \geqslant (t_\beta cv/\triangle)^2 R_x/l_x$. Thus, when estimating the accuracy of biomass assessment, one has to know the correlation radius only along transects. Since it depends on k and α, the error δ and the allowed distance between transects, h_y, (or the number of transects, N_y) also depend on these parameters. For example, when h_y is given, the minimum error can be obtained if $\alpha = 90°$; correspondingly, the maximum distance between transects, providing for desired accuracy with a defined probability, can be achieved when $\alpha = 90°$.

Discussion

One can generalize the approach developed for non-homogeneous fields which can be expanded into sum of a deterministic component (we shall call it a 'trend' for short) and a homogenous random component (noise). However, such a generalization demands more complex mathematical constructions, since formally, in this case the sampling error which has already been discussed, it is connected only with the random component and consequently does not characterize the whole error in biomass assessment completely enough. If, for example, the noise was small compared with the trend, the first place would be occupied by the error caused by the approximate method of assessing biomass (by substituting a finite sum for double integral of trend, in this case). Fortunately, as it has been already mentioned, in practice (at least for large regions) there usually takes place an inverse relation between the deterministic and random components.

Results of numerical experiments, with a computer model developed by Kizner *et al.* (MS 1982) on simulating surveys of various isotropic fields of one and the same geometrical mean size $l = \sqrt{S}$ = const and one and the same step along transects h_x = const (Kalikhman *et al.*, 1986, Fig. 2), allow us to realize that the dependencies derived here are quite universal.

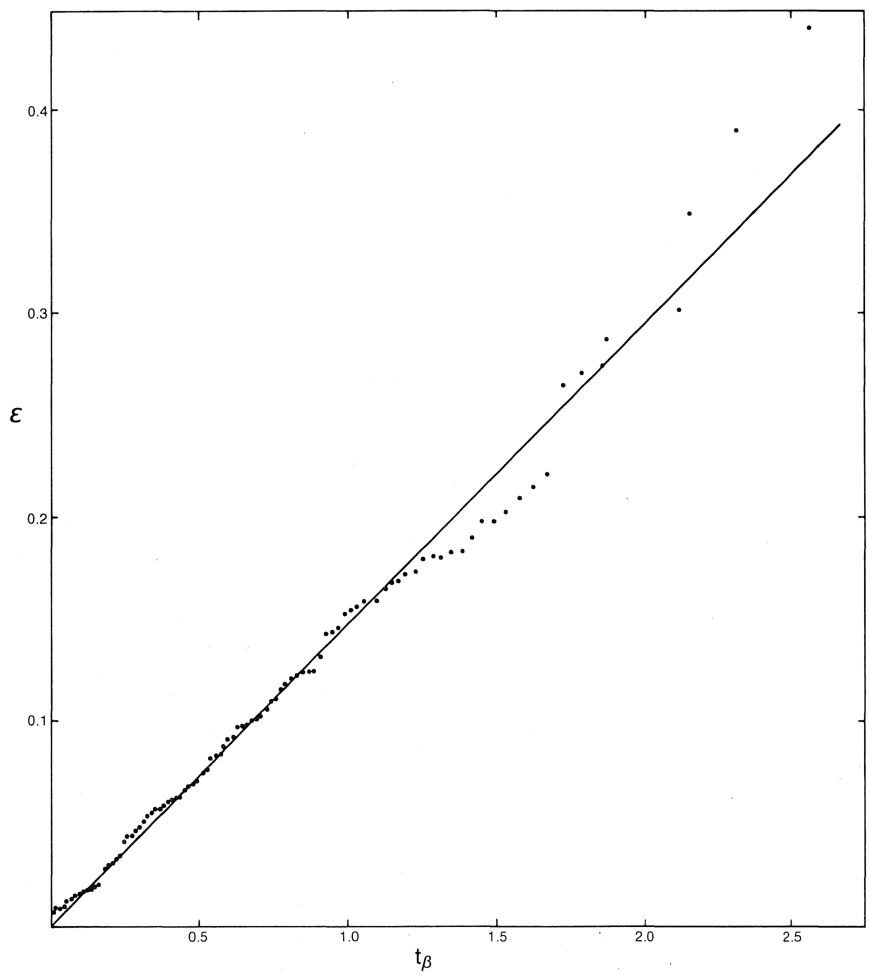

Fig. 2. Dependence of ξ upon t_β: points are experimental data, the straight line is the plot of the function $\xi = 0.15t_\beta$.

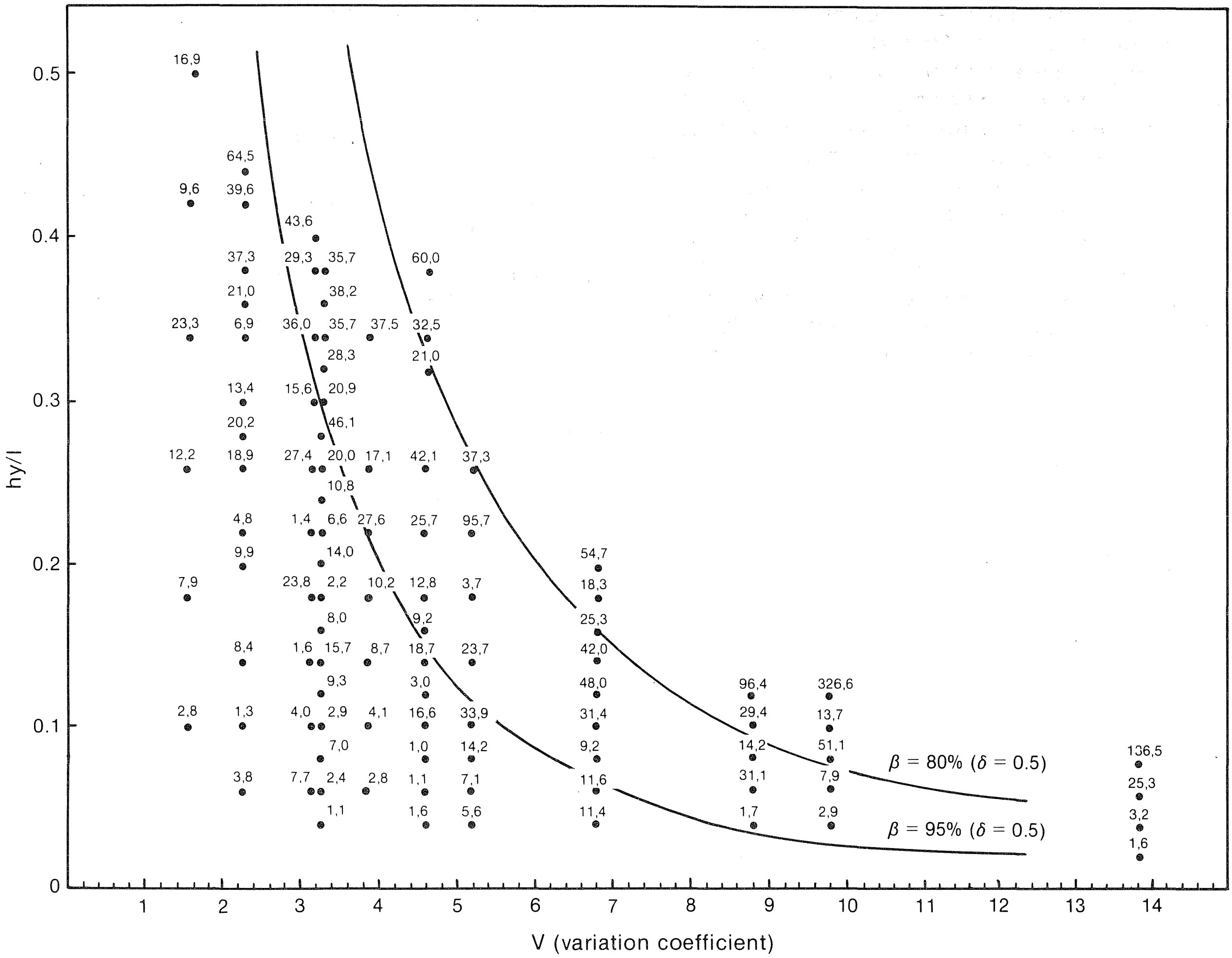

Fig. 3. Dependence of the relative distance between transects on the variation coefficient of the field: points are experimental data, numbers near the points are corresponding values of δ, the curves are theoretical dependencies for $\delta = 0.5$ and $\beta = 80\%$ and 95%.

A survey carried out using a discrete model can be interpreted as hydroacoustic or trawl survey with frequent stations along transects: from the above it must be clear that it depends on correlation properties of fields being simulated. Nevertheless, in any case, if the error δ in biomass assessment obeys the relationship (9) or (11), then the values $\xi = \delta / v \sqrt{h_y / l}$, corresponding to each experiment, should be proportional to t_β with a coefficient of proportion constant for every field (since S and h_x are constant). (Empirical probability β is the ratio n_ξ / n of the number n_ξ of experiments in which ξ does not exceed the given level, and the total number of experiments.)

A checkup with the help of these data show that it is true (see Fig. 2 and 3). Moreover, from the Fig. 2 it is clear that the relationship $\xi = 0.15 t_\beta$ is satisfied with good accuracy (in Fig. 3 the plots of the function $h_y / l = (\delta / 0.15 t_\beta v)^2$ for $\delta = 0.5$ and $\beta = 80\%$ and 95% are given).

Thus, the coefficient of proportion is actually one and the same for all simulated fields, despite the fact that they have different (in some cases rather significant) trends.

Analogous checking shows that similar relationships (practically the same ones) are valid for some other more complicated methods of estimating total biomass (such as local and weighted averaging).

Conclusion

Results presented here can be used in practise both for a *posteriori* estimation of the relative error in biomass assessment by means of survey data and for survey optimal planning, i.e. for determining its parameters by already known estimates of field statistical characteristics, allowed error and desired confidence probability. In the case of planning, if there are not enough *apriori* data of this kind, one can use the method of operative control, when all the necessary field characteristics are being obtained and gradually checked in the course of the survey itself. According to them, on the basis of the relationships offered, the survey parameters are determined and step-by-step they become more and more accurate, and the survey is gradually being transformed into an optimal regime. However, one should take into account the fact that if this process takes too long and the survey is transformed into the optimal regime too late (or does not have time at all to get transformed into it), then the biomass assessment error might turn out to be higher than that desired for given confidence probability.

References

KALIKHMAN, I. L., Z. I. KIZNER, B. R. ZARIPOV, and V. D. TESLER. 1986. Choice of distance between transects of hydroacoustic survey. *Rybn. Khoz.*, **9**: 51-54.

KIZNER, Z. I., V. D. TESLER, and B. R. ZARIPOV. MS 1982. Construction and analysis of a statistical model of fish concentration density field. Contr. for Symposium on Fisher. Acoustics, Bergen, Norway, 1982, No. 65.