SC-CAMLR-VI/BG/39

THE ERROR OF THE BIOMASS ESTIMATE AS A FUNCTION OF SURVEY PARAMETERS AND THE STATISTICS OF A DENSITY FIELD OF KRILL AGGREGATIONS

Z.I. Kizner
(USSR)

Abstract

Simple relationships expressing the dependence of relative sampling error in a biomass estimate (from a survey) on the statistical characteristics of the fish concentration density field and on parameters of the survey itself, have been derived with the help of mathematical statistics and methods of calculus of probabilities. The biomass estimate is determined as the product of the average density for the region under examination and the area of this region. For a hydroacoustic survey, the anisotropy parameters correlation radius along transects and coefficient of variation serve as field characteristics on which the error depends. The direction of the survey with respect to the axis of the correlation ellipse and the frequency of transects (sections) serve as survey parameters. The relationships described here can be used in practice for both a posteriori estimation of the error made when calculating the biomass, and for survey planning on the basis of a priori estimates of the statistical characteristics of the concentration density field. These might provide a basis for the operative control of surveys.

Résumé

La statistique mathématique et les calculs de probabilités ont permis d'établir des relations simples exprimant la dépendance de l'erreur relative d'échantillonnage dans l'estimation de la biomasse (à partir d'une prospection) sur les caractéristiques statistiques du champ de densité des concentrations de poissons et sur les paramètres de la prospection même. L'estimation de la biomasse est déterminée comme étant le produit de la densité moyenne pour la région examinée et la superficie de cette région. Pour une prospection hydroacoustique, les paramètres d'anisotropie, le rayon de corrélation le long des transects et le coefficient de variation servent de caractéristiques de terrain dont dépend l'erreur. La direction de la prospection quant à l'axe de l'ellipse de corrélation et la fréquence des transects (sections) servent de paramètres de prospection. Les relations ici décrites peuvent être utilisées en pratique à la fois pour une estimation a posteriori de l'erreur faite en calculant la biomasse et pour la préparation de prospections sur la base d'estimations a priori des caractéristiques statistiques du champ de densité des concentrations. Elles peuvent fournir une base pour le contrôle opératoire des prospections.

Resumen

Se han deducido correlaciones simples que expresan la dependencia del error relativo de muestreo en la estimación de biomasa (obtenida de una prospección) de las características estadísticas del campo de densidad de las concentraciones de peces y de los parámetros de de la prospección en sí, utilizándose la estadística matemática y los métodos de cálculo de probabilidades. La estimación de biomasa se determina como el producto de la densidad promedio para la región bajo estudio y el área de esta región. Para una prospección hidroacústica, los parámetros de anisotropía, el radio de correlación a lo largo de secciones transversales y el coeficiente de variación sirven como características de campo de las cuales depende el error. La dirección de la prospección con respecto al eje de la elipse de correlación y la frecuencia de las secciones transversales sirven como parámetros de prospección. Las relaciones aquí descriptas pueden ser usadas en la práctica para una estimación a posteriori del error cometido al calcular la biomasa, así como para la planificación de una prospección sobre la base de estimaciones a priori de las características estadísticas del campo de densidad de las concentraciones. Las mismas podrían proporcionar una base para el control operativo de las prospecciones.

Резюме

Простые соотношения, выражающие зависимость относительной ошибки выборки при оценке биомассы (по данным съемки) от статистических характеристик поля плотности концентраций рыбы и от параметров самой съемки, были выведены с помощью математической статистики и метода расчетов с учетом теории вероятностей. Оценка биомассы определяется как произведение средней плотности по рассматриваемому району и площади этого района. В случае гидроакустической съемки параметры анизотропии, радиус корреляции вдоль разрезов и коэффициенты вариативности служат характеристиками поля, от которых зависит ошибка. Направление съемки по отношению к оси корреляционного эллипса и частота разрезов (секций) служат параметрами съемки. Описанные здесь соотношения могут на практике использоваться как для последующего определения ошибки, сделанной при вычислении величины биомассы, так и для планирования съемок на основе предварительной оценки статистических характеристик поля плотности концентрации. Это может стать основой для оперативного контроля при проведении съемок.

THE ERROR OF THE BIOMASS ESTIMATE AS A FUNCTION OF SURVEY PARAMETERS AND THE STATISTICS OF A DENSITY FIELD OF KRILL AGGREGATIONS

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

1. Introduction

Estimates of fishing objects biomass are the main result of hydroacoustic and trawl surveys. Since these estimates 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 on which parameters of survey and statistical characteristics of fishing objects the error of the obtained biomass estimate might depend and how this dependence can be expressed mathematically with account for the probability nature of the estimate in question.

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

This paper is devoted to all these problems.

2. Basic relationships

Let us at first confine to the easiest and most widespread method of estimating biomass B in a region under consideration:

$B=\bar{\rho}S$, where S is region area, and $\bar{\rho}$ is average surface density of concentartions in it, determined using materials obtained through "instant" (i.e. rather short in time) regular survey. In this case, if not to consider the so-called errors of measurement (which can be assumed as known) introduced by the method of obtaining initial information, one should regard relative error δ of estimate B as a sampling relative error of the estimate of average density $\bar{\rho}$. In case of hydroacoustic survey, neither B, nor δ do depend on integration interval (if using an echointegrator). That is why in this context echo surveys can be considered similarly to trawl surveys, assuming that in case of both these types of survey the choice of information is performed in knots (intersections) of a regular rectangular grid covering the region under examination, with steps h_x and h_y , along coordinate axes x and y, connected with direction of survey (for short, these points will be called knots). The difference between them lies in the fact that in case of trawl survey steps h_x and h_y are usually close to each other in their values $(h_x \sim h_y)$, while in case of 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 that the other one (h_y) which represents a distance between transects (sections) : $h_x \ll h_y$.

In the majority of cases fishing objects concentration density fields have a typical patch-like structure, at the same time certain patches with a higher density have irregular shapes, are arranged in disorder (they can get rearranged in big aggregations or drift apart at distances considerably exceeding their own sizes);

density within one patch, as well as its shape, are subject to random perturbations. Thus when speaking about "instant" survey of a large aquatorium, in the first approximation one can consider concentration density ρ as a stationary, i.e. independent on time, homogeneous random field.

Isotropic fields

Let us at first assume that field $\rho(x,y)$ is an isotropic one. This is typical of concentrations density fields in open ocean regions, far from shelves, jet currents, equator and other physio-geographical factors which can give rise to the existing specified directions. Homogeneous isotropic field ρ has got one and the same dispersion $\mathcal{D}_\rho$ for all its points, and its normalized correlation function A, characterizing statistical interdependency of ρ values in any two points, depends itself only on distance r between them: $A = A(r)$. Correlation radius R, the minimum distance at which correlation between density values becomes negligible, is a vivid characteristic of field correlation properties (i.e. of the function A). In the general case, this radius depends on direction, but for isotropic field $R = const$, so all points in which density correlates with density in a fixed point, are in fact concentrated within the circle of radius R (called correlation circle) with its centre in this fixed point. We shall need two more values besides R:

$$a_1 = \frac{1}{\pi R^2} \iint_\omega A(r)\,dx\,dy = \frac{2}{R^2} \int_0^R r\,A(r)\,dr, \qquad (1)$$

$$a_2 = \frac{1}{R} \int_0^R A(r)\,dr. \qquad (2)$$

These values represent 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 as compared to the size of region under examination (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 boundary strip.

According to a relation, well-known in mathematical statistics, dispersion $\mathcal{D}_{\bar{\rho}}$ of the estimate of average density $\bar{\rho} = \sum_{i=1}^{N} \rho_i / N$ calculated for all knots (N is the general number of knots), is expressed by the double sum

$$\mathcal{D}_{\bar{\rho}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij} , \tag{3}$$

where $K_{ij} = \mathcal{D}_\rho A(r_{ij})$ are correlation moments of field ρ for the i -th and j -th knots, r_{ij} is the distance between these knots. One can replace $N = S/h_x h_y$ in (3) (note that in case of arbitrary configuration of the region under examination, this relationship is, generally speaking, approximate, though it is as accurate as smaller the steps h_x and h_y are). Taking into account "equality in rights" of all knots one can fix any internal knot with number i ; then using the fact that the field ρ is isotropic, the formula (3) can be rewritten in a different form:

$$\mathcal{D}_{\bar{\rho}} = \frac{\mathcal{D}_\rho}{S} h_x h_y \sum_{j=1}^{N} A(r_{ij}). \tag{4}$$

In case of trawl survey stations are usually installed far-between, thus, by the order of magnitude grid steps are no less than the correlation radius: $h_x \gtrsim R$, $h_y \gtrsim R$. In this case

$A(r_{ii})=1$ (as $r_{ii}=0$), and at $i \neq j$ all values of $A(r_{ij})$ are practically zero, so from (4) we obtain:

$$\mathcal{D}_{\bar{\rho}} = \mathcal{D}_{\rho}\, h_x h_y / S. \tag{5}$$

This relatioship written in the form $\mathcal{D}_{\bar{\rho}} = \mathcal{D}_{\rho}/N$ is well-known and often used when processing independent experiments (see, for example, Venttsel, 1962).

In the other limiting case which seems to be not ever realized in practice and which, however, is very important for understanding the main results of the work, when $h_x \ll R$ and $h_y \ll R$, expression $h_x h_y \sum_{i=1}^{N}\sum_{j=1}^{N} A(r_{ij})$, being a part of (4), can be replaced with high accuracy by integral of A over correlation circle. In other words, using (1), the equality (4) can be rewritten in the form:

$$\mathcal{D}_{\bar{\rho}} = \pi a_1 \mathcal{D}_{\rho} R^2/S. \tag{6}$$

Hydroacoustic survey occupies an intermediate position between the two abovedescribed cases: now $h_x \ll R$, $h_y \gtrsim R$, and thus, density in the i-th knot correlates only with density in knots located on the same track at a distance no more than R. 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) :

$$\mathcal{D}_{\bar{\rho}} = 2 a_2 \mathcal{D}_{\rho} R h_y / S. \tag{7}$$

Since estimate of average density $\bar{\rho}$ is a random value distributed according to a normal law (N is large), its absolute error ε depends on confidence probability β and

$\sigma_{\bar{\rho}} = \sqrt{D_{\bar{\rho}}}$ which is a standard deviation of $\bar{\rho}$:

$$\varepsilon = t_\beta \sigma_{\bar{\rho}} . \tag{8}$$

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

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

at $h_x \gtrsim R$, $h_y \gtrsim R$ (trawl survey)

$$\delta = t_\beta v \sqrt{h_x h_y / S}, \tag{9}$$

at $h_x \ll R$, $h_y \ll R$ ("superfrequent" survey)

$$\delta = t_\beta b v R / \sqrt{S}, \tag{10}$$

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

$$\delta = t_\beta c v \sqrt{R h_y / S}, \tag{11}$$

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

From (9) - (11) it is clear that the minimum possible error is practically obtained in case of steps (h_x and h_y for trawl survey or h_y - for hydroacoustic survey) 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. it does not decrease with the growth of the number of knots.

Expressions (9) - (11) represent desired dependences of biomass estimate error on survey parameters and isotropic field statistical characteristics. Consequently, when resolving equality (9) relative to $h_x h_y$, and (11) - relative to h_y , we can obtain mathematical basis for survey optimal planning. Thus, if it is necessary to estimate biomass of fishing objects in a certain region of the open ocean with the help of hydroacoustic survey with such accuracy that relative error should not exceed defined level Δ with probability β , and if a priori estimate of $\mathcal{V}$ is accurate enough, then distance between transects h_y is to be taken from the following condition:

$$h_y \leqslant \left(\frac{\Delta}{t_\beta c \mathcal{V}}\right)^2 \frac{S}{R}. \qquad (12)$$

Anisotropic fields

Up to now we have been discussing isotropic density fields. However, quite simple geometric considerations allow us to apply results obtained to anisotropic fields in which one can specify two mutually perpendicular directions in such a way that along one of them correlation radius is maximum, and along the other one - it is minimum. Such a situation occurs, for example, when survey is being carried out in shelf waters where medium, and, correspondingly, concentration characteristics change insignificantly along the shelf, and change considerably - in perpendicular direction. In this case, correlation circle gets deformed and becomes an ellipse, large and small radii of which we shall denote as R_M and R_m. It is clear that now error δ can depend on anisotropy index $k = R_M/R_m \geqslant 1$ and survey direction which we shall define by angle α between axis x and the large axis of correlation ellipse. Introducing 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 sizes of a rectangle with sides being parallel to axes x and y and embracing this ellipse, similarly to the abovedescribed we obtain three variants of survey: $h_x \gtrsim H_x$ and $h_y \gtrsim H_y$; $h_x \ll H_x$ and $h_y \ll H_y$, $h_x \ll H_x$ and $h_y \gtrsim H_y$. At the same time relationship (9) does not change, constants b and c remain the same, and R in (10) is replaced by $\sqrt{R_M R_m}$. For hydroacoustic survey (the third variant) we have instead of (11) and (12):

$$\delta = t_\beta c v \sqrt{R_x h_y / S}, \qquad h_y \leq (\Delta / t_\beta c v)^2 S/R,$$

Thus when determining the accuracy of biomass estimate one has to know correlation radius only along transects Since it depends on k and α, error δ and allowed distance between transects h_y also depend on this parameters. For example, when h_y is defined, the minimum error can be obtained if $\alpha = 90°$; correspondingly, the maximum distance between transects providing for desired accuracy with defined probability, corresponds to such a direction of survey when $\alpha = 90°$.

3. Discussion

One can also present some other generalizations of the abovedescribed approach, and, in particular, for non-homogeneous fields which can be expanded into sum of deterministic component (for short, we shall call it a trend) and homogeneous random component (noise). However, such a generalization demands more complex mathematical constructions, since, formally, in this case, the sampling error, that has already been mentioned here, is connected only with the random component, and, consequently, does not characterize the whole error in biomass estimate completely enough. Thus, if noise were small as compared to trend, the first place would be occupied by the error caused by the approximate method of estimating biomass (in this case - by substituting a final sum for double integral of trend). Fortunately, as it has already been mentioned, in practice there usually takes place an inverse relation between deterministic and random components.

Results of numerical (computer) experiments on simulating surveys of various isotropic fields of one size $\ell = \sqrt{S} = const$ with one and the same step along transects $h_x = const$ (Kalikhman et al., 1986, fig. 2), allow us to realize that dependences derived here are quite universal; these model fields are represented by 50x50 numerical matrices and in general have weak autocorrelation. When analysing these data, it is convenient to use relative distance between transects h_y/ℓ instead of h_y.

Survey carried out with the help of this discrete model can be interpreted as hydroacoustic or trawl survey with stations installed frequent along transects - from the abovesaid it is clear that it depends on correlation properties of fields being simulated. However, in any case, if error δ in biomass estimate obeys dependence (9) or (11), then (since S and h_x are constant) values $\xi = \delta / v\sqrt{h_y / l}$, corresponding to each experiment , should be proportional to t_β with coefficient of proportion constant for every field; in this case empirical probability β is defined as a ratio n_ξ / n of the number n_ξ of experiments in which ξ does not exceed the given level, and the total number n of experiments. Check-up with the help of these data showed that it is true (see fig.1). Moreover, from the fig.1 it is clear that relationship $\xi = q\, t_\beta$ or(which is the same) $\delta = q\, t_\beta v \sqrt{h_y / l}$, where $q \approx 0.15$, is satisfied with food accuracy. Thus, coefficient of proportion is actually the same for all model fields, despite the fact that they have different (in some cases rather significant) trend.

Also note, that in data used (Kalikhman et al., 1986) value δ represents an average error for several methods of estimating B , including via expression $B = \bar{\rho} S$, and this means that relationships similar to those derived in this paper, are valid for other methods of estimating biomass.

4. Conclusion

Results mentioned here can be used in practice both for a posteriori estimation of relative error in biomass estimate by means of survey matherials and for survey optimal planning, i.e. determining its parameters by already known estimates of field statistical characteristics, allowed error and desired confidence probability. In last case, if there are not enough a priori data of this kind, one can use the method of operative control, when all the necessary field characteristics are being determined and gradually checked in the course of survey itself. According to them, on the basis of dependences offered, survey parameters are chosen and detailed, and survey is gradually being transformed into an optimal regime. In this case, however, one should take into account the fact that if this process takes too long and survey is transformed into optimal regime too late (or does not have time at all to get transformed into it), then biomass estimate error might turn out to be higher than that desired for defined confidence probability.

References

1. Kalikhman, I.L., Kizner, Z.I., Zaripov, B.R., Tesler, V.D. 1986. Choice of distance between tracks of hydroacoustic survey. Rybnoe Khoziajstvo, No 9, p. 51-54 (in Russian).
2. Venttsel E.S. 1962. Calculus of probabilities. State publishing house of physics and mathematics literature, Moscow, USSR, 564 p. (in Russian).

List of Symbols

A - normalized correlation function of the field ρ

B - biomass or its estimate

D_ρ - dispersion of the field ρ

$D_{\bar{\rho}}$ - dispersion of $\bar{\rho}$

h_x, h_y - grid steps along axes x and y

k - anisotropy index

K_{ij} - correlation moment

l - size of a model field

N - number of the grid knots (size of a sample)

R - correlation radius of an isotropic field

R_M, R_m - large and small correlation radii

R_x - correlation radius in the direction of tracks

S - area of the region under survey

v - variation coefficient

x, y - coordinate axes (the survey transects are parallel to x-axis)

α - angle between axis x and the large axis of the correlation ellipse

β - confidence probability

δ - relative sampling error of biomass estimate

Δ - desired relative accurancy of a survey

ε - absolute sampling error of biomass estimate

ρ - surface density of concentrations

$\bar{\rho}$ - average density (over the region under survey)

$\sigma_{\bar{\rho}}$ - standard deviation of $\bar{\rho}$

$\Phi^{-1}(\beta)$ - Laplace's inverse function on β

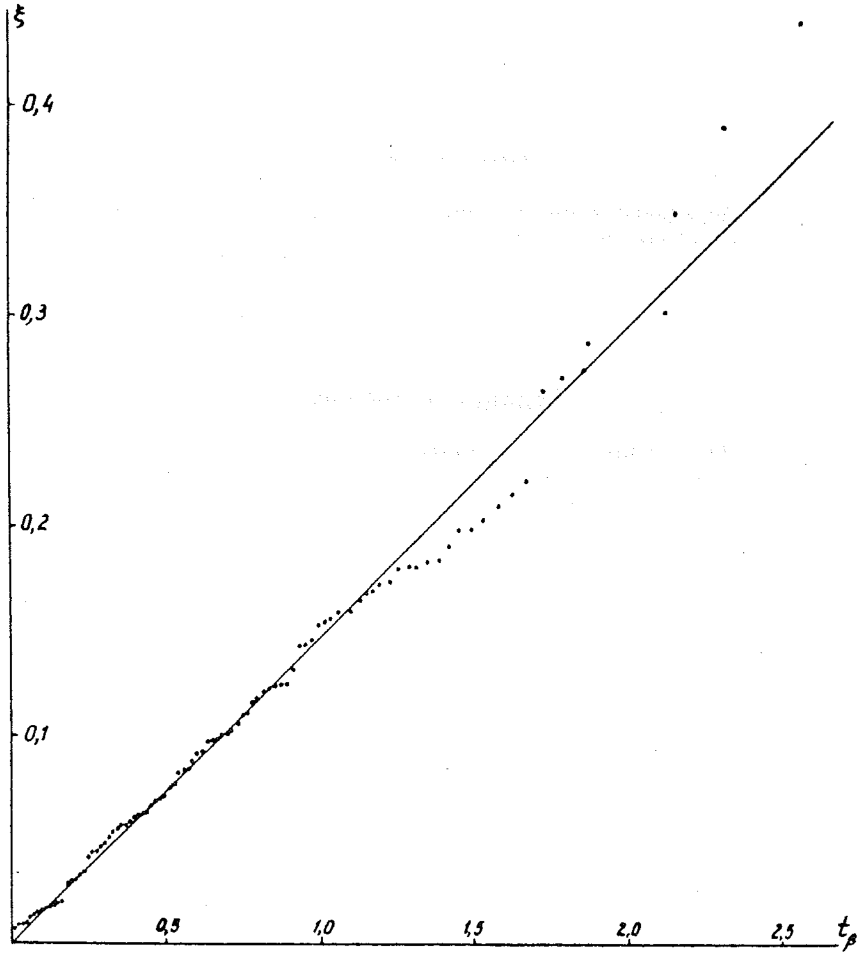

Fig.1. Dependence of value $\xi = \delta / b\sqrt{h_y / \ell}$ on t_β : points are experimental data, straight line is the graph of equation $\xi = 0.15\, t_\beta$

Légende de la figure

Figure 1 Dépendance de la valeur $\xi = \delta / v\sqrt{h_y/l}$ sur t_β : les points correspondent à des données expérimentales, la ligne droite est le graphe de l'équation =0,15t .

Leyenda de la Figura

Figura 1 Dependencia del valor $\xi = \delta / v\sqrt{h_y/l}$ de t_β : los puntos corresponden a datos experimentales, la línea recta es el gráfico de la ecuación =0.15t .

Подпись к рисунку

Рисунок 1 Зависимость величины $\xi = \delta / v\sqrt{h_y/l}$ от t_β : точки - данные экспериментов, прямая линия - график уравнения = 0,15t .