NOT TO BE CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

International Commission for the Northwest Atlantic Fisheries

Serial No. 5262

ICNAF Res. Doc. 78/VI/76

ANNUAL MEETING - JUNE 1978

On estimating the accuracy of silver hake abundance indices derived from the Soviet research vessel trawling surveys of the Emerald Deep area in 1972-1976

by

P. S. Gasjukov and R. S. Dorovskikh
AtlantNIRO
Kaliningrad, USSR

Publ. in ICNAF Sel. Papers No. 5: 19-24

Introduction

Trawling surveys of the Emerald Deep area were conducted by the Soviet research vessels in the 1972 to 1976 period. This is the area of primary importance in the studying of the Nova Scotia silver hake (4wvx). According to A.S.Noskov [6], 90% of the species is caught here (see also [4]). Therefore, the estimation of the accuracy of abundance indices derived from the survey data is of great interest.

The survey has a random sampling design where the whole Division is represented by a single stratum. 20 hauls are made there annually in randomly scattered points. The mean silver hake catch per standard haul both in weight and in numbers was used as an abundance index.

Two approaches to the estimation of the trawling survey precision are known, which are discussed by Grosslein [3]. In one of these approaches statistical characteristics of the catch per haul of the studied species are widely used. The second approach is based on the comparison between the abundance indices drawn from the set of survey data for a number of years and the values characteristic of the stock obtained in another, independistic, although, according to the data available, this estimate may also suffer from the low determination accuracy.

In addition to observation time series the second technique suggests the presence of a correlation,approaching to a direct proportional dependence, between the studied characteristics.

In this paper, apart from the approaches mentioned above a method of statistical simulation on computer is suggested to consider the accuracy of trawling survey. This method based on the simulation of probability regularities in the spatial distribution of a commercial species under investigation gives supplementary information on characteristics of the abundance index accuracy.

<u>Estimation of accuracy of silver hake abundance index using sampling statistical characteristics</u>

The mean catch per haul in the stratum and its variance are calculated from the known formulae:

$$\bar{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi} \qquad (1)$$

$$S^2_{\bar{x}} = \frac{1}{n_h (n_h - 1)} \sum_{i=1}^{n_h} (X_{hi} - \bar{X}_h)^2 \qquad (2)$$

where $\bar{X}_h$ is the mean catch per haul in the stratum,

$S^2_{\bar{x}}$ is a variance of the mean catch,

X_{hi} is the catch size per i-haul and

n_h is the number of hauls per stratum.

The calculation results from the formulae (1) and (2) and the silver hake catch data collected during the survey of the Emerald Deep area in 1972-1976 are presented in tables 1 and 2. Similar calculation results, however, subject to preliminary log transformation of the initial data are presented in tables 3 and

4. As is known, a log transformation is used to reduce the influence of larger in absolute size, but relatively occasional observations, and also to normalize the data.

The coefficient of the abundance index variation is 0.20, on the average, and 0.05 for transformed data. From the availability of the 95 percent C.I. intersections it can be suggested that year-to-year variations in the silver hake stock size were insignificant during the studied period.

For comparison, it should be noted that in estimation of accuracy of the Georges Bank haddock, cod and yellowtail abundance index the coefficients of variation derived by Grosslein [3] were on the order of 0.25, and 0.10-0.15 for transformed data. These values hold with the coefficients derived for silver hake with the difference that the former were calculated using a stratified sample of about 60 hauls made in a group of strata.

From this it may be concluded that provided other conditions are equal, the accuracy of the silver hake abundance index for the Emerald Deep area will be higher than that of other commercial fish species abundance indices for Georges Bank.

<u>Comparison between the survey abundance indices and other abundance indices</u>

Grosslein's suggestion [3] that the abundance index variance derived from the trawling survey data is overestimated has been confirmed by the comparison between the long-term results of the survey by species and the commercial indices.

Such a comparison has been also made in this paper using the silver hake stock size estimates for Div. 4 wvx derived from the virtual population analyses (Noskov, 1976).

The curves of the silver hake abundance index based on the Soviet research vessel trawling survey data for 1972-1976 and on the stock size derived from the V.P.A. are shown in Fig. 1. They are given in percent of the means for a series of years under observations. The analyses of the curves has revealed a difference between the directions during the transition period from 1974 to 1975. However, because of the lack of such observations (for 5 years only) it is hardly possible to discuss on the presence or absence of a correlation between the above estimates.

This is not a single case of difference between the sequence of the abundance indices gained from the survey data and other indices of the stock state. Thus, Grosslein [3] indicates similar differences between the index curves for Georges Bank cod and haddock from the western Nova Scotia.

In fig.1, besides the above mentioned curves, the sequence of the silver hake abundance indices for a number of years is also shown from the data of the Canadian research vessel trawling survey of 4wvx in 1971-1975 [4]. The comparison to the stock size estimates calculated by means of the V.P.A. also resulted in a considerable discrepancy. Besides, a discrepancy between the above values and the abundance indices presented by the Soviet research vessels was revealed.

On Regularities in Spatial Distribution of Silver Hake

In using the method of statistical simulation the probability regularities of the commercial fish species spatial distribution should be set. These are considered in the paper by Taylor [7]. The author, proceeding from theoretical grounds and summarizing the results of treatment of the data from the trawling survey of Georges Bank in 1948-1950, comes to a conclusion that the amount of fish per haul fits the negative binomial distribution.

Such a distribution may be derived from different models, however, the assumption that spatial distribution of individuals clustered in groupings or aggregations has random numbers (the Poisson distribution), and the number of individuals in a grouping fits the logarithmic series distribution is a most common biological interpretation.

The probability of occurrence of r-individuals at the given distribution law is described by a formula:

$$P_r = \frac{\Gamma(r - k) \cdot P^r}{r! \, \Gamma(k) \cdot Q^{k-r}}$$

where Γ is a gamma-function, $Q = 1 - P$, $P = \frac{m}{K}$, m is the distribution mean and K is parameter associating the variation D^2 of the distribution with the mean through the following relation:

$$D^2 = m + \frac{m^2}{K}$$

The hypothesis on fitness the observation data to the specific distribution law may be checked, if standard statistical methods are applied. However, as far as silver hake from the Emerald Deep area is concerned, the complications arise because of limited observation data.

The hypothesis that the silver hake distribution is the negative binomial was checked by means of the data from the comparative gear tests during the joint USSR-USA surveys of groundfish in 1973-1975.

The results of using χ^2 - criterion are presented in tables 6-11. As is evident from the analysis, the observation data hold with the hypothesis of the negative binomial distribution.

A direct check of the hypothesis on fitting the negative binomial distribution to the data of trawling survey of the Emerald Deep area was made using the method of moments [1], [2] based on the consideration of conformity between two third moments obtained theoretically and empirically. The following statistics was applied:

$$T = \frac{[x^3]}{n} - D^2 \left\{ \frac{D^2}{m} - 1 \right\}$$

where $\frac{[x^3]}{n}$ is the third sampled moment

D^2 and m are variances and the mean of the negative binomial distribution.

To describe the statistics variance, T, the following formula is used:

$$D^2 (T) = 2 \cdot m \cdot (k + 1) \cdot p^2 \cdot Q^2 \left[2 \cdot (3 + 5P) + 3KQ \right] \cdot \frac{1}{n}$$

The results of computation of T, of its variance and mean square deviation from the trawling survey data are presented in table 5. The table shows that the hypothesis of the negative binomial distribution is consistent with the observation data.

It should be noted that this hypothesis fits diagrams of the relation between the mean catch and variance in the stratum given by Hennemuth [5], which may also surve as an indirect proof of its acceptability.

The study on accuracy of silver hake abundance indices using the method of statistical simulation

The method of statistical simulation for estimating the accuracy of trawling survey is one of the promising approaches, which may be applied in case of lacking a priori information. Prior to using this method, the probability law of spatial distribution of a commercial population should be set. In simulating the realizations of a random value with the appropriate probability characteristics at a given number of hauls, the statistical characteristics of accuracy of the method may be simulated, the lack of a priory information being compensated with the multy-version calculations using the parameters of the distribution law and possible range of their variation.

In simulation experiments the parameters of the silver hake negative binomial distribution derived from the data collected in the Emerald Deep area in 1972-1976 were used. Table 5 shows that during the observation period the mean varied from 800 to 2500 individuals per haul, and the coefficient, K, ranged from 0.8 to 2.6. Therefore, the calculations were made at m = 500, 1500, 2500 and at K = 0.5, 1.25, 2.00; 20 hauls were taken as the number generally recorded in practice.

Results of the experiments presented as histograms of the abundance index, which were drawn based on 500 realizations of a model are shown in figs. 2-4. Some conclusions resulting from their analysis are given below.

The distribution of the mean catch per haul is of clearly asymmetrical nature, the asymmetry decreasing with the increase of the parameter K and extending with the increase of m.

The range of positive deviations from the modal distribution exceeds significantly that of negative deviations.

The interval between the possible estimates of the "observed" abundance index is widened with the increase of m and reduced with the increase of the parameter K. This is easily explained by the fact that lower values of the parameter correspond to higher concentration level of the population and vice versa, higher estimates imply more uniform spatial distribution.

One of the tasks to be solved by the method of statistical simulation was that of determination of the frequency of cases in which the estimate of accuracy of the abundance index derived by means of $2 \cdot \sigma$ does not contain the actual mean distribution. As is evident from the calculations, the amount of such cases is ranged between 6 and 10%. The practical implication of this conclusion is that approximately in 1 case out of 10 the estimate of the accuracy of the abundance index by means of sampling variance is erroneous. This should be kept in mind when the sequence of the abundance indices for a number of years is compared, as it is done in fig. 1.

Fig. 5 shows the results of computer experiments aimed at the studying the variations of the accuracy characteristics relative to the abundance index, with the change of the haul number per survey in the Emerald Deep area. The parameters of the possible silver hake distribution have been accepted from the observation data for 1975. The surveys with 20, 50, 100 and 200 hauls were simulated.

The asymmetrical distribution is still maintained with 50 hauls, however, with the increase of their number the histograms of the abundance index approximate to the normal curve.

If with 20 hauls per survey the accuracy of the abundance index is on the order of 40%, with 50 hauls per survey it will be decreased to 25%. With further increase of the haul number the accuracy of the abundance index is increased very slowly. Thus to gain the estimate of the abundance index on the order of 10% of accuracy 200 hauls should be made.

So it is evident that more efficient accuracy of trawling survey cannot be gained with the increase of the haul number, which is consistent with Grosslein's conclusion [3].

Summary

Various approaches have been tested to estimate the accuracy of the silver hake abundance index from the trawling surveys of the Emerald Deep area. Statistical characteristics of the catch per haul, comparisons to other independent stock size estimates and the method of statistical simulation were used. In realization of the latter the analysis of probability regularities in the spatial distribution of the studied species has been made.

The accuracy of abundance indices is relatively low, on the order of 40%, which does not allow to discover the year-to-year variation in the stock size.

The incompatibility of the sequence of abundance indices for a number of years derived from the Soviet and Canadian research vessel trawling surveys and the stock size estimates obtained from the virtual population analysis was revealed. As shown by the method of statistical simulation, the estimate of accuracy of the abundance index derived with the use of statistical characteristics corresponds to actual values. In some cases, however, a confidence interval derived from the double mean square deviation does not contain the true estimate of the index.

To improve the trawling survey method, a priori information on the distribution of the commercial fish species should be given due attention.

References

1. ANSCOMBE F.J. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika, vol.37, p. 358-382.
2. BLISS C.I. Fitting the negative binomial distribution to biological data. Biometrics, vol. 9, N 2, 1953, p. 176-200.
3. GROSSLEIN M.D. Some observations on accuracy of abundance indices derived from research vessel surveys. Int.Comm.Northw.Atlant.Fish.Res.Doc. 71/59.
4. DOUBLEDAY W.G., HUNT J.J. and HALLIDAY R.G. The 4wvx silver hake fishery. Int.Comm.Northw.Atlant.Fish.Res. Doc. 76/VI/59.
5. HENNEMUTH R.C. Variability of Albatross IV Catch per tow. Int. Comm.Northw,Atlant.Fish.Res.Doc. 76/VI/104.
6. NOSKOV A.S. The assessment of the silver hake stocks at Nova Scotia in Div. 4 vwx. ICNAF.Res.Doc. 76/XII/157.
7. TAYLOR C.C. Nature of variability in trawl catches. Fish.Bull. U.S. 58, 1953, 145-166.
8. SISSENWINE M.P. Using the USA Research Vessel Spring Bottom Trawl Survey as an Index of Atlantic Mackerel Abundance. Int.Comm.Northw.Atlant.Fish.Res.Doc. 76/XII/144.

Table 1. Mean silver hake catch per haul in numbers in Emerald Deep area for 1972-1976 and the accuracy estimates.

Years	Mean catch per haul x_h	Variance of mean catch per haul σ_x^2	Mean square deviation σ_x	$\frac{\sigma_x}{x_h}$	Confidence interval $x_h - 2\cdot\sigma_x$	$x_h + 2\sigma_x$
1972	963	43681	209	0.22	545	1381
1973	866	34225	185	0.21	496	1236
1974	1291	38025	195	0.15	901	1681
1975	2635	311364	558	0.21	1519	3751
1976	1508	108900	330	0.22	848	2168

Table 2. Mean silver hake catch per haul (in kg) in Emerald Deep area for 1972-1976 and the accuracy estimates.

Years	Mean catch per haul x_h	Variance of mean catch per haul σ_x^2	Mean square deviation σ_x	$\frac{\sigma_x}{x_h}$	Confidence interval $x_h - 2\cdot\sigma_x$	$x_h + 2\sigma_x$
1972	101	361	19	0.19	63	139
1973	116	484	22	0.19	72	160
1974	205	1089	33	0.16	139	271
1975	340	4761	69	0.20	202	478
1976	227	2704	52	0.23	123	331

Table 3. Mean silver hake catch per haul in numbers (log transformation) in Emerald Deep area for 1972-1976 and the accuracy estimates.

Years	Mean catch per haul x_h	Variance of mean catch per haul σ_x^2	Mean square deviation σ_x	$\frac{\sigma_x}{x_h}$	Confidence interval $x_h \pm 2\sigma_x$	
1972	6.25	0.0784	0.28	0.04	5.69	6.81
1973	6.31	0.0576	0.24	0.04	5.83	6.79
1974	6.84	0.0484	0.22	0.03	6.40	7.28
1975	7.44	0.0529	0.23	0.03	6.98	7.90
1976	6.79	0.0625	0.25	0.04	6.29	7.29

Table 4. Mean silver hake catch per haul in numbers (log transformation) in Emerald Deep area for 1972-1976 and the accuracy estimates.

Years	Mean catch per haul x_h	Variance of mean catch per haul S_x^2	Mean square deviation S_x	$\frac{S_x}{x_h}$	Confidence interval $x_h \pm 2S_x$	
1972	4.10	0.0784	0.28	0.07	3.54	4.66
1973	4.10	0.0441	0.21	0.05	3.98	4.82
1974	4.97	0.0529	0.23	0.05	4.51	5.43
1975	5.40	0.0529	0.23	0.04	4.94	5.86
1976	4.91	0.0576	0.24	0.05	4.43	5.39

Table 5. Estimates of parameters of the negative binomial distribution from trawling surveys of Emerald Deep area in 1972-1976.

Years	Mean catch per haul (m)	Error of the mean (σ_m)	Parameter (k)	Error of parameter (σ_k)	Statistics $T \times 10^9$	Mean square deviation $\sigma_T \times 10^9$
1972	885	205	0.98	0.45	- 0.407	1.2
1973	858	190	1.02	0.45	- 0.005	0.98
1974	1292	195	2.32	0.90	- 0.510	0.64
1975	2635	558	1.11	0.48	- 4.022	23.46
1976	1507	330	1.04	0.46	- 1.270	5.05

Table 6. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 41 Yankee trawl, comparative tests, 1973).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	40	49	1.68
16 - 50	13	9	1.65
51 - 100	8	5	0.73
101 - 200	8	6	0.49
201 - 500	11	8	0.98
501 - 1000	4	5	0.30
1001 - 2000	1	3	1.80
2001 - 5000	3	1	0.95
TOTAL:	88	86	8.58

A 12

Table 7. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 36 Yankee trawl, comparative gear tests, 1973).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	53	61	1.11
16 - 50	21	11	7.60
51 - 100	9	6	0.78
101 - 200	2	5	2.47
201 - 500	4	4	0.07
501 - 1000	1	1	0.01
TOTAL:	90	88	12.04

Table 8. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 41 Yankee trawl, comparative gear tests, 1974).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	41	38	0.12
16 - 50	6	13	3.98
51 - 99	9	8	0.003
100 - 200	12	9	0.80
201 - 500	10	9	0.001
501 - 1000	4	3	0.006
1001 - 2000	1	1	0.005
TOTAL:	83	81	4.91

Table 9. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 36 Yankee trawl, comparative gear tests, 1974).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	45	41	0.36
16 - 50	13	17	1.20
51 - 100	11	11	0.002
101 - 500	17	16	0.04
501 - 1000	1	0	0.005
TOTAL:	87	85	1.61

Table 10. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 41 Yankee trawl, comparative gear tests, 1975).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	20	39	9.36
16 - 50	45	24	16.92
51 - 100	16	17	0.11
101 - 200	15	16	0.08
201 - 1000	12	13	0.21
TOTAL:	108	109	26.68

Table 11. Observed and expected silver hake catches at hypothetical negative binomial distribution (No. 36 Yankee trawl, comparative gear tests, 1975).

Catch per haul in nos.	Observed number of hauls	Expected number of hauls	χ^2
0 - 15	80	82	0.05
16 - 50	19	16	0.41
51 - 100	6	8	0.72
101 - 200	5	5	0.15
201 - 1000	3	3	0.01
TOTAL:	113	114	1.34

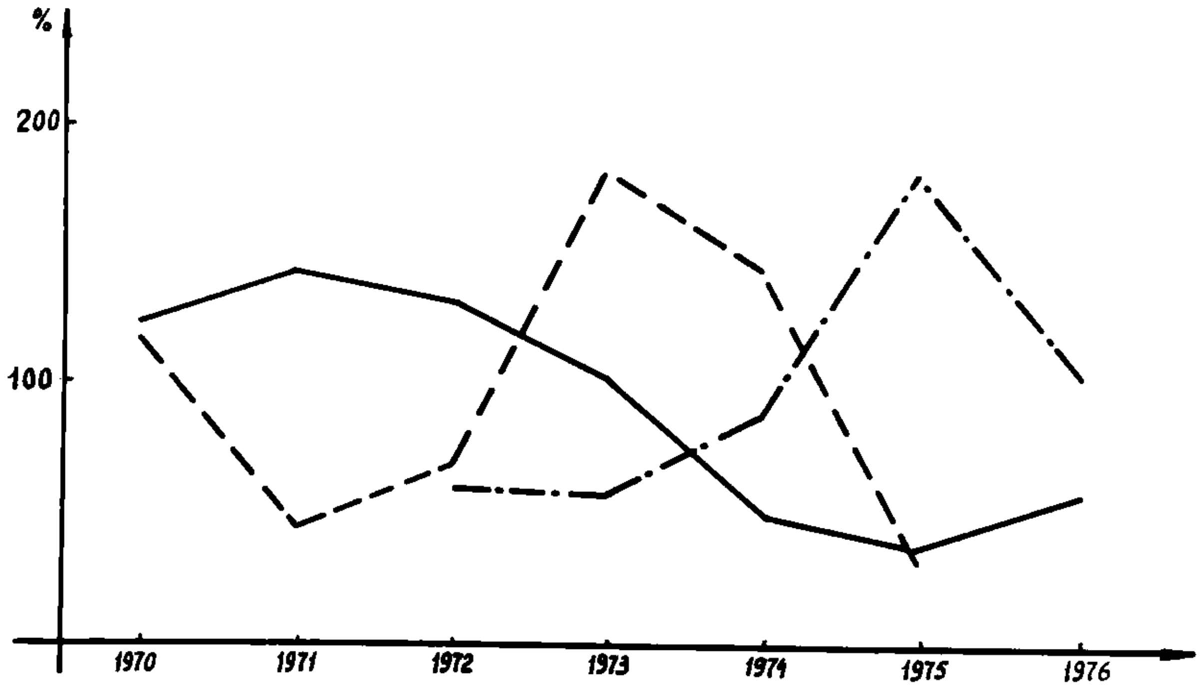

Fig. 1. Comparison between the silver hake abundance indices from the Soviet and Canadian research vessel trawling surveys and the estimates of the stocks size derived from the virtual population analyses.

—— estimate of the stock size in Div. 4VWX derived from the V.P.A. according to Noskov (6);

--- silver hake abundance index from the Canadian research vessel trawling surveys;

—·— silver hake abundance index from the Soviet research vessel trawling surveys.

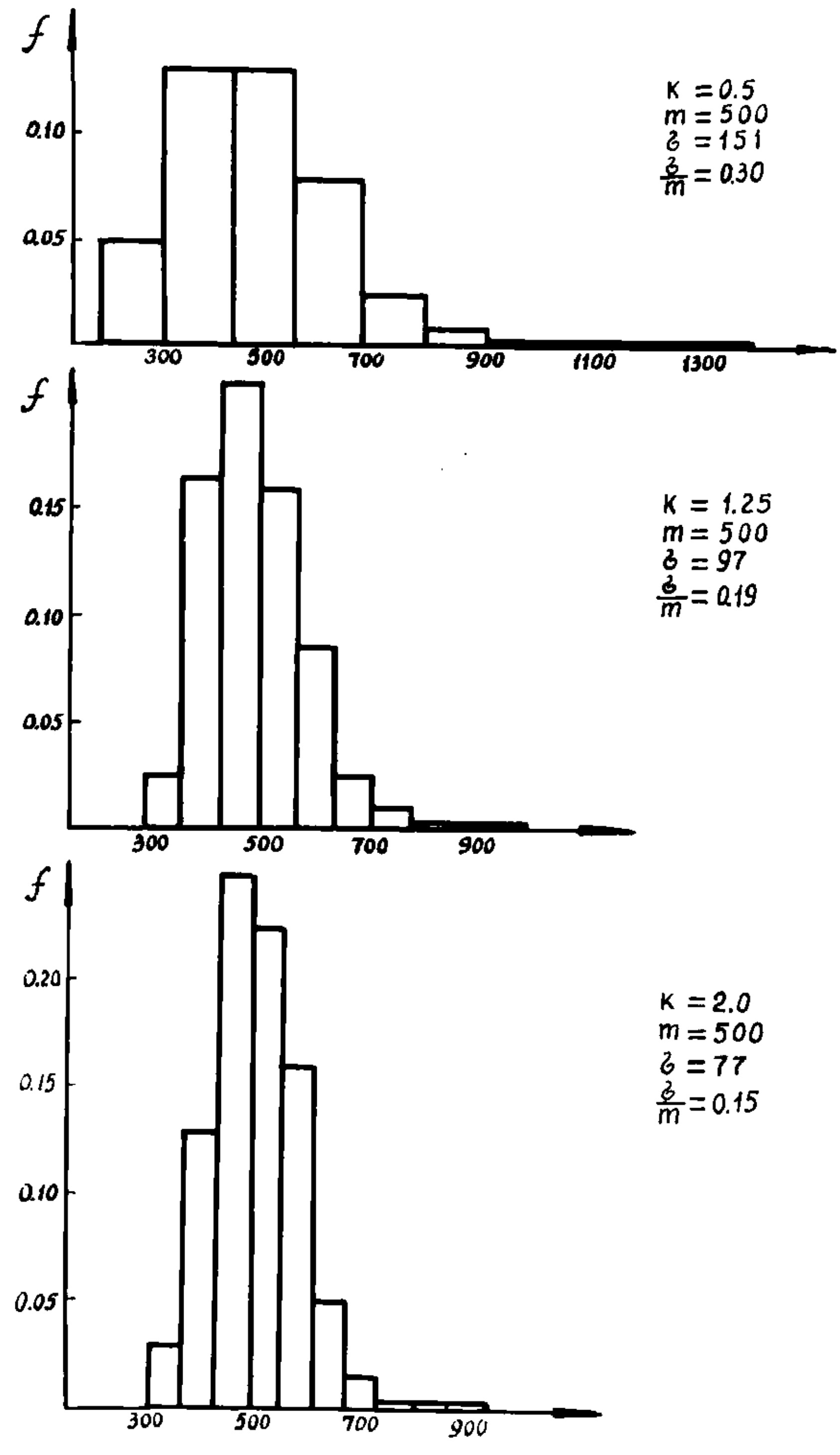

Fig. 2. Histograms of the abundance index from the results of the negative binomial distribution simulation (M = 500, n = 20).

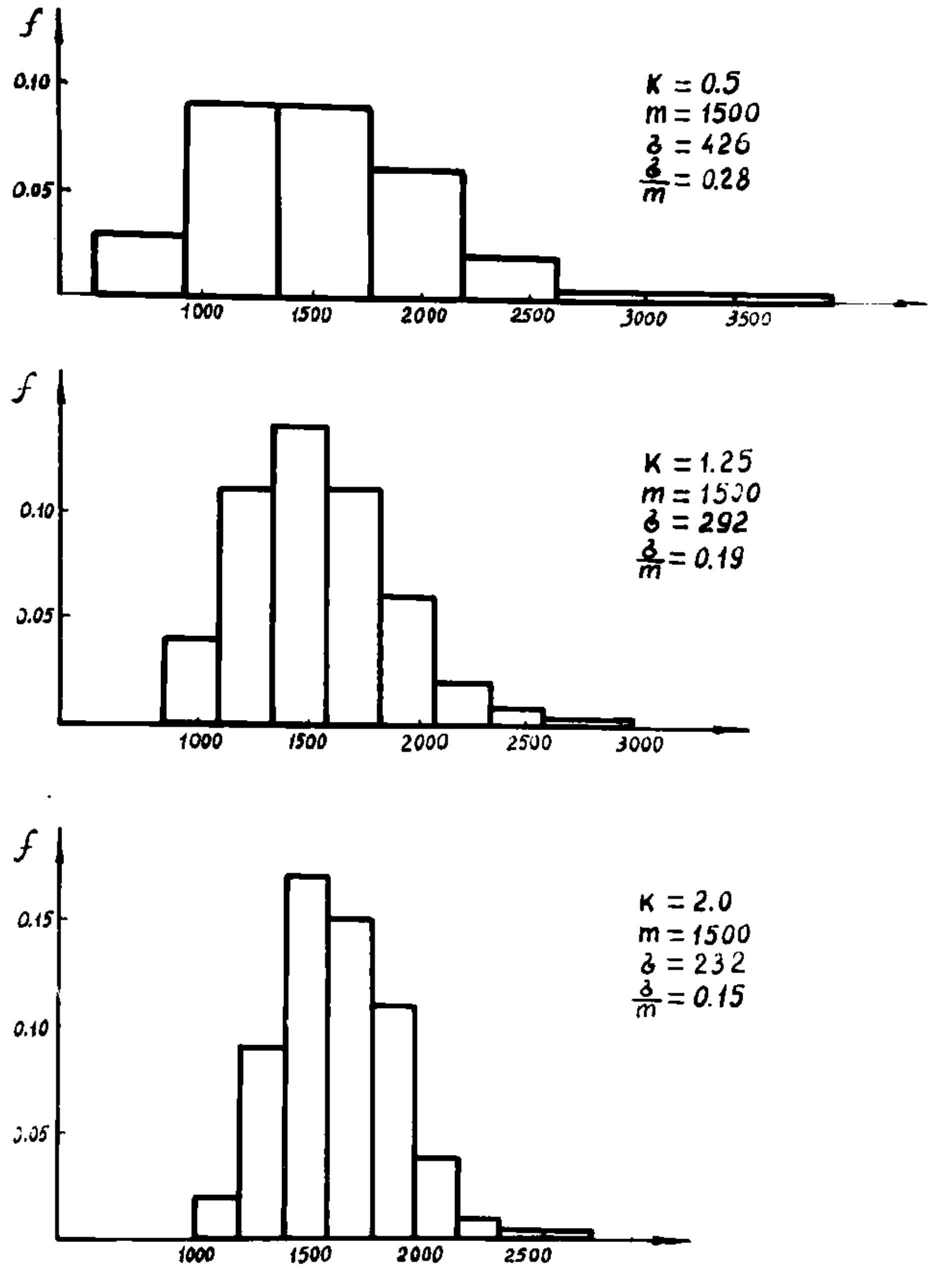

Fig. 3. Histograms of the abundance index from the results of the negative binomial distribution simulation (m = 1500, n = 20).

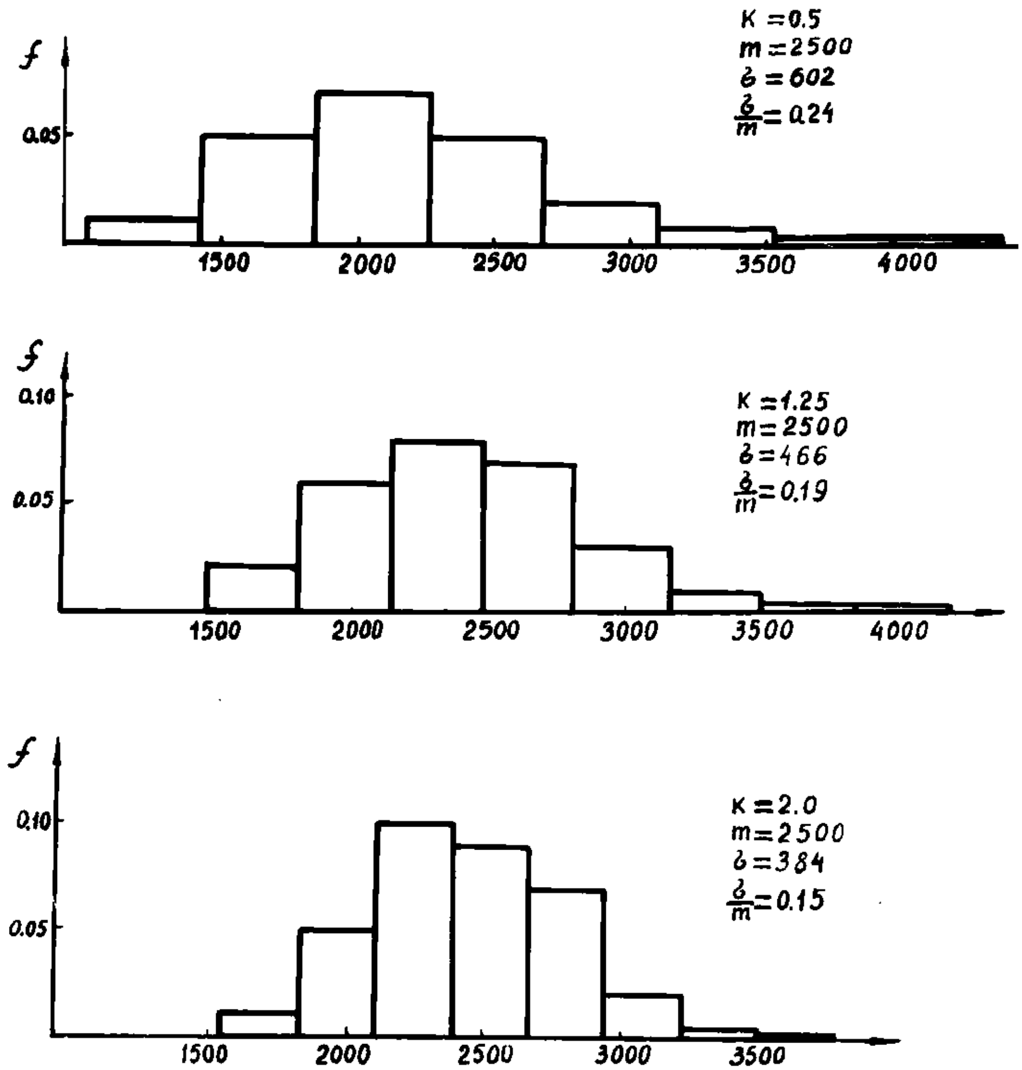

Fig. 4. Histograms of the abundance index from the results of the negative binomial distribution simulation (m = 2500, n = 20).

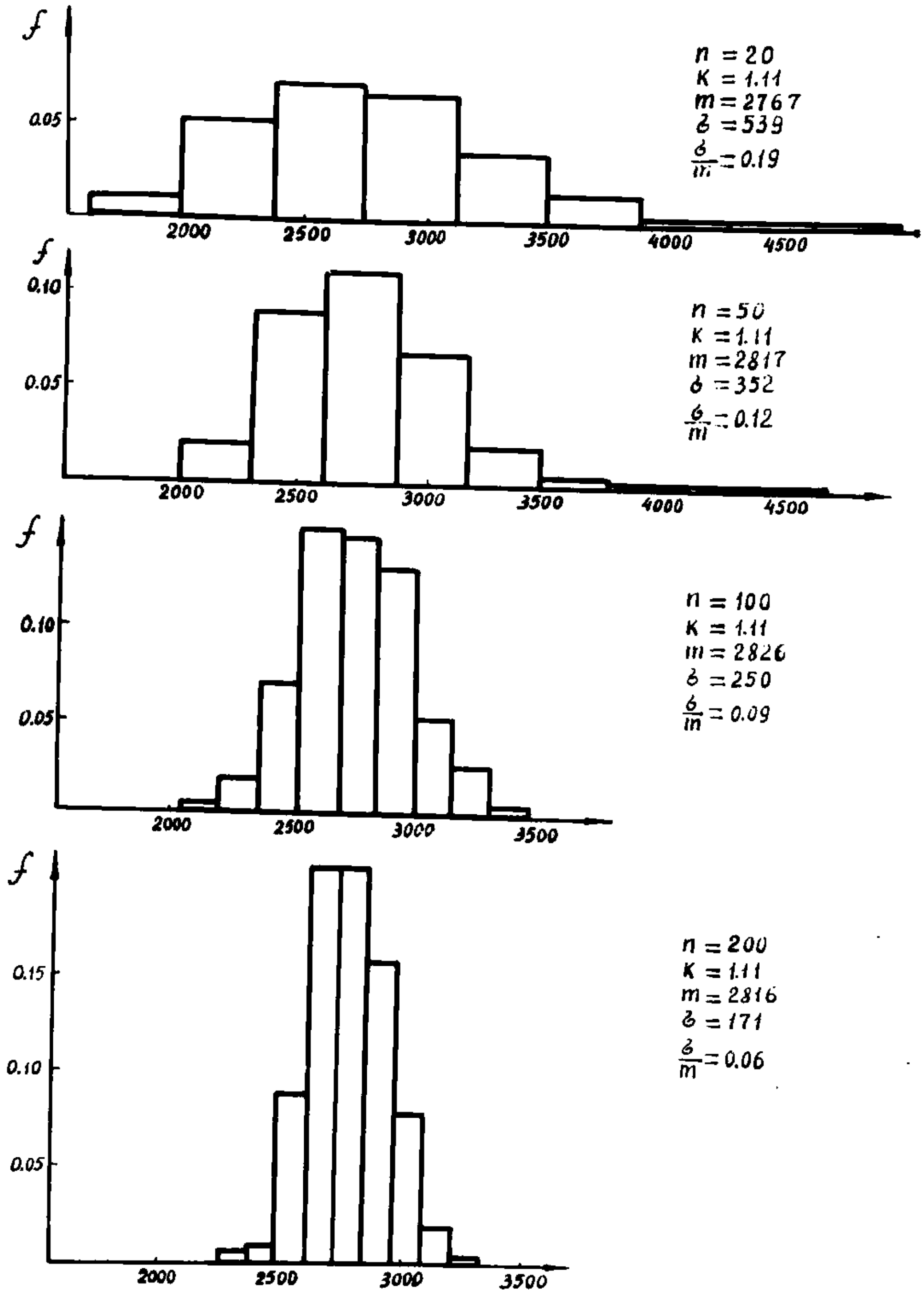

Fig. 5. Histograms of the abundance index from the results of the negative binomial distribution simulation (m = 2500, n = 20, 50, 100, 200).