This paper not to be cited without prior reference to the authors

INTERNATIONAL COUNCIL FOR THE
EXPLORATION OF THE SEA

C.M. 1974/H : 19
Pelagic Fish (Northern) Committee

THE PRECISION OF ABUNDANCE ESTIMATES FROM YOUNG HERRING SURVEYS IN THE NORTH SEA

by

H.B. Becker and A. Corten
Netherlands Institute for Fishery Investigations,
Haringkade 1, Ymuiden, the Netherlands.

International Council for the Exploration of the Sea

C.M. 1974/H:19
Pelagic Fish (Northern) Committee

THE PRECISION OF ABUNDANCE ESTIMATES FROM YOUNG HERRING SURVEYS IN THE NORTH SEA.

by

H.B. Becker and A. Corten.

Netherlands Institute for Fishery Investigations, Haringkade 1, Ymuiden, the Netherlands.

1. Introduction

In recent years the abundance of I-group herring in the North Sea has been estimated by extensive international trawling surveys. For details on these surveys the reader is referred to the report of the Working Group on North Sea Young Herring Surveys (Anon. 1974a).

One of the main problems in the interpretation of the results from these surveys is the level of precision that can be ascribed to the estimated mean abundance. This mean abundance, expressed as mean number of fish per hour, is often to a large extent determined by one or two catches of exceptional size. What one really ought to know is how much precision can be obtained for a given amount of sampling effort (in number of ship days), or rather, how much sampling effort is required to achieve a certain level of precision.

This paper considers some methods of working out the precision of the abundance estimate. The methods are applied to data of the 1973 International Young Herring Survey. A concise treatment of the same problem, using a different approach, is given in the report of the Working Group (Anon. 1974a).

2. The statistical distribution of trawl catches

Distributions of individual fish in space may be classified conveniently into one of the following types (Anon. 1974b).

A. random and independent of each other
B. uniform
C. heterogeneous (aggregated, contagious).

It can be shown that the numbers of fish in random hauls from an (A)-type distribution belong to a Poisson distribution, with the variance equal to the mean. Samples from a (B)-type distribution would yield more or less constant numbers per haul.

Analysis of data from Young Herring Surveys has shown that the variance of the number per haul is even larger than the mean number per hour, which means that neither distribution (A) nor (B) apply in this case.

Postulating heterogeneity for the distribution of individuals, it has been shown by various authors (e.g. Taylor, 1953) that the negative binomial distribution gives a satisfactory description of the frequency distribution of numbers per haul.

The negative binomial distribution is characterised by the fact that its variance is larger than the mean. It is a mathematical counterpart of the positive binomial distribution. The probability of observing a sample of x individuals is

$$P(x) = \frac{(k + x - 1)!}{(k - 1)!x!} \quad \frac{p^x}{(1 + p)^{k + x}} \qquad (1)$$

$$x = 0,1,2,\ldots\ldots$$

where $p = \frac{m}{k}$

The distribution is determined by two parameters, the mean (m) and a positive exponent (k). The variance of the distribution is given by

$$m + \frac{m^2}{k}$$

As $k \longrightarrow$ and $p \longrightarrow 0$, it can be shown that the distribution converges to Poisson. That means the variance approaches the mean.

3. Fitting a negative binomial distribution to observed data

The fitting of the negative binomial distribution to an observed frequency distribution will be illustrated with data on year class 1971 as measured during the 1973 Young Herring Survey. Fig. 1 shows the spatial distribution of all hauls made during the 1973 Young Herring Survey in the North Sea south of 58° 00 N, excluding Moray Firth, and fig. 2 gives the frequency distribution of numbers per haul. The parameters (m) and (k) of the negative binomial distribution are estimated from the observed data by $\bar{x}$ and $\hat{k}$.

The mean (m) is estimated efficiently by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

where

x_i = number in i^{th} haul

N = total number of hauls in sampling area.

The value (k) may be estimated from the formula

$$\hat{k}.\ {}^{10}\log\left(1 + \frac{\bar{x}}{\hat{k}}\right) = {}^{10}\log\left(\frac{N}{f_o}\right) \qquad \text{(Bliss 1953)}$$

in which f_o = number of zero hauls.

By substituting $N = 172$
$f_o = 84$

$\bar{x} = 490.63$

we find $\hat{k} = 0.0824$ with 90% efficiency.

χ^2-test for goodness of fit

Using the above values for $\hat{k}$ and $\bar{x}$, the expected frequencies can be calculated for all possible values of (x) by formula (1). These expected frequencies are compared with the observed ones by means of a chi-square test. In order to do this, expected frequencies are first grouped into classes in such a way that the sum of all expectations (e) within a class is more than 5. The small expectations at the tail end of the distribution are all lumped into one class.
Table 1 compares values of expected frequencies (e) with the observed frequencies (f).

Table 1.

x	0	1	2-3	4-8	9-16	17-31	32-53	54-92	93-149	150-238	>238	Total
f	84	5	8	8	6	5	6	5	8	8	29	172
e	84.04	6.92	6.35	7.29	5.84	6.01	5.16	5.55	5.03	5.03	34.78	172

For a total number of (n) classes

$$\chi^2 = \sum \frac{(f-e)^2}{e}$$

has (n-3) degrees of freedom.

We find $\chi^2 = 5.86$ with $p > 0.60$

This indicates a good fit of the negative binomial distribution to the observed data.

4. Precision of the abundance estimate

Two methods are used to calculate the level of precision of the mean number per haul.

1st. method

The catches within a certain area are assumed to belong to a negative binomial distribution. If the number of hauls (N) is large, 95% confidence limits for the population mean (m) can be expressed as

$$m_{1,2} = \bar{x} \pm t \frac{s}{\sqrt{N}} \qquad (2)$$

where (t) is the 97.5% point of Student's t-distribution with (N-1) degrees of freedom.
The variance s^2 of the negative binomial distribution is estimated by

$$s^2 = \bar{x} + \frac{(\bar{x})^2}{\hat{k}} \qquad (3)$$

The precision (d) of the estimate $\bar{x}$ can be expressed as percentage confidence limits of the population mean (Anon. 1973):

$$d = \left| \frac{m_{1,2} - \bar{x}}{\bar{x}} \right| 100 \qquad (4)$$

From (2) and (4):

$$d = \left| \frac{\frac{t.\,s}{\sqrt{N}}}{\bar{x}} \right| .100$$

$$d^2 = \frac{\frac{1}{N}.\ t^2.s^2}{(\bar{x})^2} .\ 10^4$$

Substituting s^2 (formula 3) gives

$$d^2 = \frac{t^2}{N} \left(\frac{1}{\bar{x}} + \frac{1}{\hat{k}}\right).10^4 \qquad (5)$$

This formula enables us to calculate the precentage confidence limits (d) for any value of N (table 2).

Table 2.

N	50	75	100	125	150	175	200	225	250	300	400	500	600	700
d	99	80	69	62	56	52	49	46	43	40	34	31	28	26

For the 1973 Young Herring Survey with N = 172, a precision was obtained of 52%. This means that in 95% of all possible surveys the population mean (m) would fall within the range

$$\bar{x} \pm 0.52\,\bar{x}$$

To get a precision for instance of 31%, (with 95% confidence limits), at least 500 hauls would be necessary.

<u>2nd. method</u>

Another way of calculating confidence limits is to apply normalising transformations to the original data. According to Taylor (1953) the correct transformation to be used for negative binomial distributions with (k) less than 3/4 is the logarithmic transformation.

We use $y = {}^{10}\log\,(x + 1)$ to avoid difficulties with zero hauls.

After transformation of the original values we find:

$$\bar{y} = 0.9837$$

$$s_y^2 = 1.4583$$

$$s_{\bar{y}} = 0.0921$$

Transformation back to the mean and its confidence limits of the original units is done by the following formula: (Jones 1954)

$${}^{10}\log\,(\bar{x} + 1) = \bar{y} \pm t.s_{\bar{y}} + 1.15\ s_y^2\ \frac{N - 1}{N}$$

The mean number per haul ($\bar{x}$) is found to be 447, with 95% confidence limits 294 and 679. The upper limit is

$$\frac{679 - 447}{447} \ \text{x}\ \ 100\% = 52\% \text{ above the mean.}$$

Thus the precision of the mean, calculated by this method, appears to be the same as the one found by the former method.

5. Possibilities to increase the precision of the estimated mean number per haul

It should now be considered whether the precision can be increased while the amount of sampling effort (number of ship/days) remains the same.

Going back to formula (5), it appears that the precision of the sample mean is determined by three parameters: $\hat{k}$, $\bar{x}$, and N.
The value of $\hat{k}$, according to Taylor, is an intrinsic property of the population being sampled, and does not depend on the sampling method. The only parameters affected by the sampling method are the sample mean ($\bar{x}$) and the number of tows (N). Since both parameters occur in the denominator of the formula, an increase in either of them will result in a smaller (d^2) and thus increase the precision.

The sample mean $\bar{x}$ can be increased either by using a larger gear, or by tows of longer duration. However, one cannot increase both ($\bar{x}$) and (N) at the same time. For a given number of ship days, one has to choose between a large number of short hauls, or a smaller number of long hauls. In formula (5), an increase in N has a larger effect than an increase in ($\bar{x}$). Suppose for instance that we could double the number of hauls by reducing the length of the tow to half its original duration. Then (d) would decrease approximately by a factor $\sqrt{2}$. For the 1973 Young Herring Survey this would mean an increase of the precision from 52% to 37%.

Apart from the duration of the haul, the sample mean ($\bar{x}$) is also determined by the size of the net. Therefore, it is advisable to use the largest possible gear, as long as it doesnot affect the number of hauls that can be made.

Another possibility to narrow the confidence limits of the estimated mean exists when the fish have a tendency to concentrate in certain parts of the sampling area.
If certain strata of high fish density can be delimited, a relatively high proportion of all hauls can be concentrated in these areas. Separate estimates of the mean density and its confidence limits are made for the high density and low density strata, and these figures are then combined into an overall mean density plus confidence limits. The large number of hauls in the high density stratum results in a more precise estimate for this area, which ultimately also increases the precision of the overall estimate.

The stratification of the sampling area should be made before the sampling at sea starts. Strata should be designed using for instance information on hydrographic conditions, depth contours, or the results of a preliminary survey. The Working Group on North Sea Young Herring Surveys has recently decided to introduce for future surveys a stratification based on historical catch rates in the various statistical rectangles (Anon. 1974a).

References

Anon. 1974(a). Report of the Working Group on North Sea Young Herring Surveys. I.C.E.S. CM 1974/H:6.

Anon. 1974(b). Second Report of the Working Group on Standardisation of Scientific Methods for Comparing the Catching Performance of Different Fishing Gear. I.C.E.S. CM 1974/B:

Anon. 1973. Marine Fisheries Research Department, Annual Report 1972. Southeast Asian Fisheries Development Centre, Singapore.

Barnes, H., 1952. The use of transformations in marine biological statistics. J. Conseil. Exp. Mer., 18: 61-71.

Bliss, C.J., and R.A. Fisher., 1953. Fitting the negative binomial distribution to biological data and a note on the efficient fitting of the negative binomial. Biometrics, 9: 176-200.

Gulland, J.A., 1954. A note on the statistical distribution of trawl catches. Rapp. P.-v. Cons. int. Explor. Mer, 140, (1): 30-39.

Jones, R., 1954. The analysis of trawl haul statistics with particular reference to the estimation of survival rates. Rapp. P.-v. Cons. int. Explor. Mer, 140 (1): 28-29.

Taylor, C., 1953. Nature of variability in trawl catches. Fishery Bull. Fish Wildl. Serv. U.S., 54: 145-63.

/ds

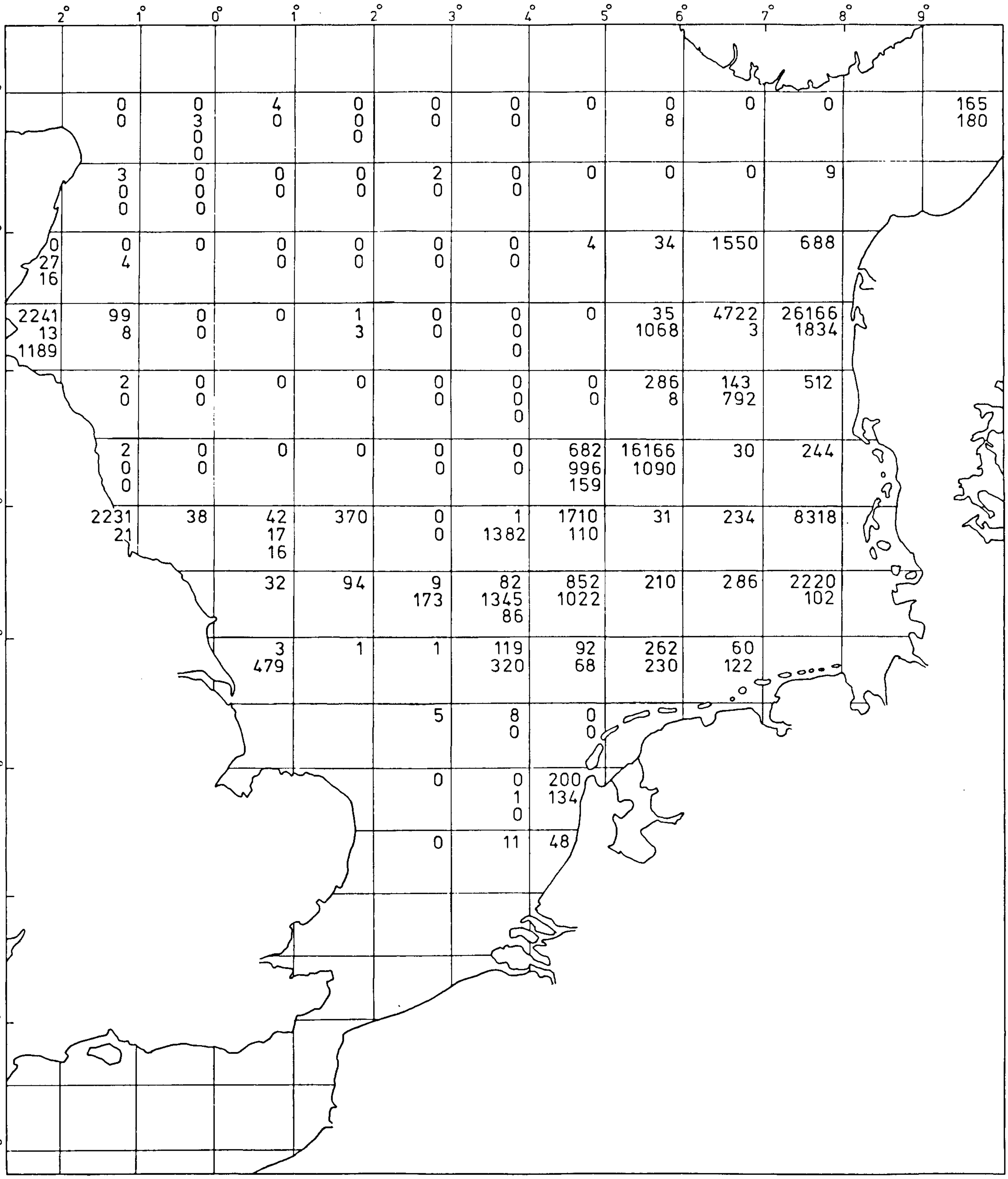

Fig. 1. Catches of I-group herring during the Young Herring Survey 1973 in numbers per hour trawling.

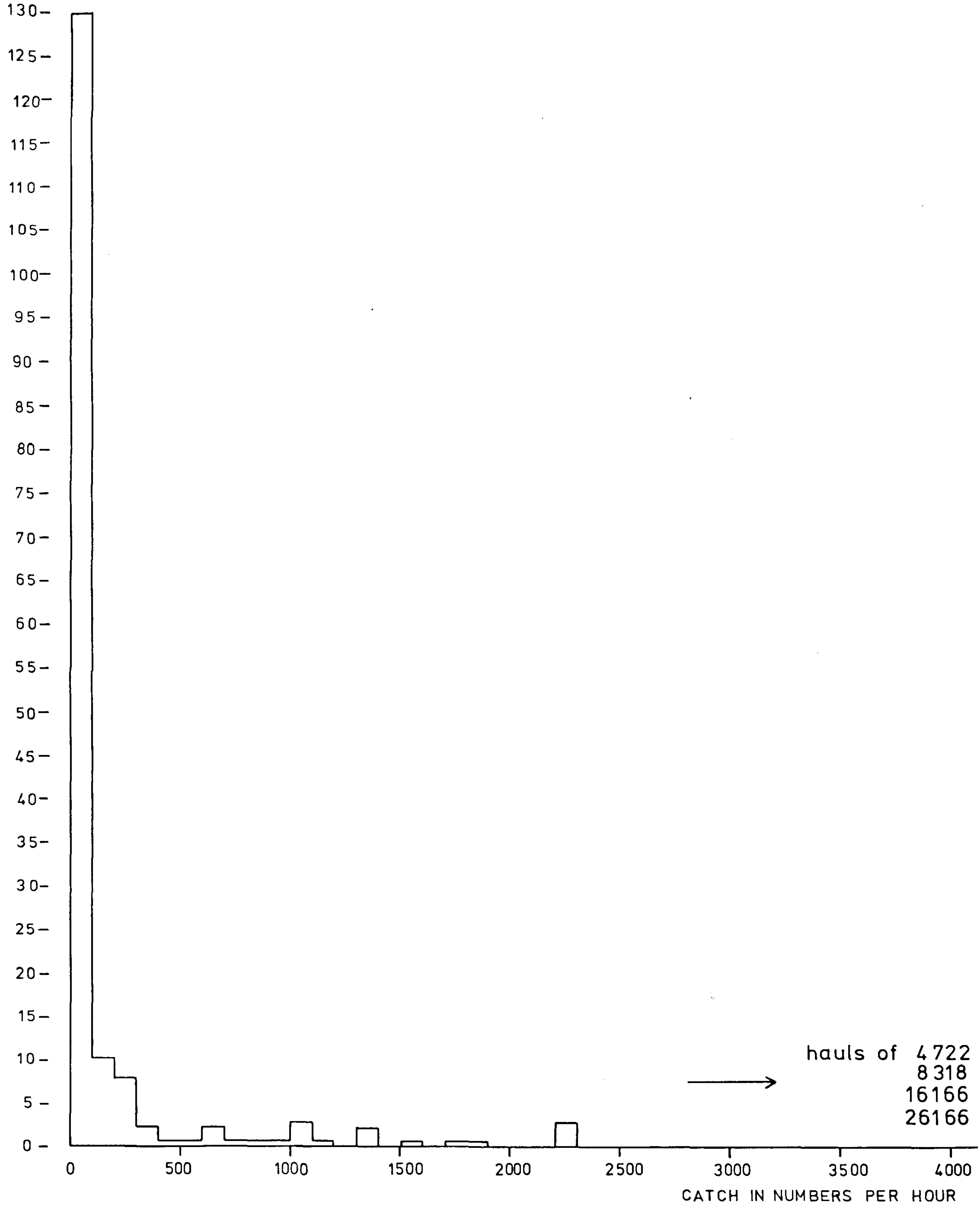

Fig. 2

Frequency distribution of catches I-group herring during Young Herring Survey 1973