This report not to be quoted without prior reference to the Council*

International Council for the
Exploration of the Sea

C.M.1992/D:6
Ref.: G, H, J

REPORT OF THE WORKSHOP ON THE ANALYSIS OF TRAWL SURVEY DATA

Woods Hole, 4-9 June 1992

---

*General Secretary
ICES
Palægade 2-4
DK-1261 Copenhagen K
DENMARK

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Participants

van Beek, F.	Netherlands
Bergh, M.	South Africa
Björnsson, H.	Iceland
Conser, R.	USA
Degel, H.	Denmark
Fogarty, M.	USA
Foote, K.	Norway
Forrester, J.	USA
Fréchet, A.	Canada
Fryer, R.	U.K. (Scotland)
Gagnon, P.	Canada
Gavaris, S. (Chairman)	Canada
Godø, O.	Norway
Heessen, H.	Netherlands
Korsbrekke, K.	Norway
Moriyasu, M.	Canada
Munro, P.	USA
Murawski, S.	USA
Myers, R. *	Canada
Nichols, S.	USA
Nicholson, M.	UK (England)
Pennington, M.	USA
Poulard, J.-C.	France
Rivoirard, J.	France
Sanchez, F.	Spain
Shepherd, J.	UK (England)
Sinclair, A.	Canada
Smith, S.	Canada
Somerton, D.	USA
Sparholt, H.	ICES
Strømme, T.	Norway
Steinarsson, B.	Iceland
Vølstad, J. H.	Norway
Warren, W.	Canada

* part time

1.2 Terms of Reference

The terms of reference (C.Res.1991/2:10) are:

a) extend the statistical analysis of travel survey data, with respect to estimating population abundance trends;

b) analyze the differences and changes in the fishing power of research vessels;

c) study temporal changes in spatial distributions;

d) compare the usefulness of random versus fixed-station survey design;

e) recommend any improvements in survey design based on the results of the above analyses.

1.3 Overview

The subject of the analysis of trawl survey data was previously addressed by the Methods WG at its 1989 meeting in Nantes (Anon 1990), under the somewhat uninformative heading of "Preprocessing".

A number of general issues were treated there, and that report should be read as background for this one.

The main conclusions reached in 1989 were:

- Most global abundance indices reduce to some form of average over the data, often not much different to a (possibly weighted) arithmetic mean.

- The arithmetic mean (and simple standard indices based on it) performs surprisingly well as an abundance index, at least on well-behaved data. No procedure tested performed significantly better than the standard indices available.

- Both random stratified and fixed station survey designs have advantages and disadvantages. The latter perform best when the spatial distribution is persistent (i.e., similar from one year to another), whereas the former may be preferred when the spatial distribution is volatile (variable from year to year). It was not, however, possible to arrive at a quantitative treatment of this balance of advantage at that time.

- GLM (general (ized) linear models) methods are particularly useful when auxiliary information is available and/or needs to be allowed for (e.g., allowance for ship effects in multiple-ship surveys).

- Methods based on random sampling theory are unbiased, but may have relatively high variance. Conversely model-based methods (interpolation, geostatistics, GLM's, etc.) depend on the assumptions made (form of model, distributional assumptions, etc.) and may be biassed.

- Consistent bias is not a serious problem if the index will only be used to calibrate sequential population analysis, and not for direct estimator of absolute population size.

- If interactions between time and any other factors (e.g., spatial/station effect) are included in a model, the year effects are likely to be meaningless. The index should be based on integration of the fitted values instead.

The 1989 report includes the following summary of the applicability of various classes of methods.

At this Workshop it has been possible to clarify some of the outstanding questions, especially over the relative merits of fixed and random station designs and the dependence on the persistence or volatility of the spatial distribution .

There remains a fundamental difference of opinion over the applicability of sampling theoretic and model-based methods: the former assume that the observations are a (precisely measured) sample of those possible, whereas the latter generally regard observations as error-prone observations of the truth (usually of variable precision). Which of these views of the observational process is nearer to the truth is primarily a metaphysical question, which the WG was (not surprisingly) unable to resolve. The Workshop participants agreed to focus on the following objective:

To study problems associated with obtaining a global index of abundance for a defined geographic region to be used:

a) in calibration of VPA
b) as a direct estimate of stock size

with the aim to make recommendations regarding:

a) analysis of existing data
b) design and estimation for subsequent surveys

1.4 Acknowledgements

Workshop participants extend their appreciation to the Chairman, who was nominated and elected - - in his absence - - a few hours prior to convening of the meeting. The workshop extends its thanks to the management and staff of the Northeast Fisheries Science Center for logistical support. In particular, efforts of the computing staff including David Hiltz, William Kramer and Edgar Kleindinst were critical to the completion of the analytical tasks. The report was prepared by Ms. Joanne McDonald and Ms. Elizabeth Holmes.

2 ANALYSIS OF EXISTING DATA

2.1 Introduction

2.1.1 The International Young Fish Survey

The IYFS surveys in the North Sea, Skagerrak and Kattegat conducted in February each year, began in the years 1960-1961. The first surveys were aimed exclusively at juvenile herring and only a part of the North Sea was covered. Over the years the objectives of the survey were broadened to include sampling of young gadoids. This meant that the survey area had to be extended to cover the distribution of all species and the northern North Sea and the Skagerrak/Kattegat were included. Since 1980 the whole North Sea and Skagerrak/Kattegat are covered. About 400 hauls are made each year. Up to 8 countries participated in the survey. A typical example of area allocation is given in Figure 2.1.1.1. This allocation has changed over the years. In each rectangle at least two hauls, by research vessels from different countries, are made. Trawling positions are usually chosen at random, although some vessels use fixed fishing positions in order to reduce possible gear damage. In 1976 a standard gear, the French GOV bottom trawl, was proposed and the introduction was completed in 1978. Haul duration is 30 minutes and trawling is mainly carried out during daytime. The primary objective of the survey is to provide annual indices of recruitment for herring, sprat, cod, haddock, whiting, Norway pout and mackerel.

Standard indices are calculated by taking the arithmetic mean per hour for all hauls within a rectangle, then the arithmetic mean for all rectangles within the species-specific standard area is calculated.

The standard area for herring consists of 57 rectangles where during a 10 year period the highest catches of 1-group herring were made. When the herring standard index is calculated night-hauls are excluded.

The 1-group herring in the North Sea, Skagerrak and Kattegat may be considered to belong to one stock. The 2- and 3-group herring in Skagerrak and Kattegat belong to another stock.

A large amount of effort has been put into standardizing the gear and survey design, especially promoted by the Working Group on the International North Sea, Skagerrak and Kattegat Bottom Trawl Survey (see e.g. Anon. 1992).

Introduction

The International Young Fish Survey has during several decades been conducted each year as a coordinated survey between laboratories in England, France, Scotland, Germany, Holland, Sweden, Norway and Denmark, with the purpose of monitoring fish stocks in the North Sea and producing data for calculating indices for the same stocks.

For various reasons, but often connected to processing of the data, it is useful to make a stratification of the area based on similar catch rates (abundance).

Method

The total estimated biomass (B) for an area, can be expressed as:

$$\hat{B} = \sum_l \overline{CPUE} * A_l \tag{1}$$

Where l is a subdivision of the whole area and A_l is the area of subarea l.

The variance of B is:

$$V(\hat{B}) = \sum_l A_l^2 * V_l\ (\overline{CPUE}) = \sum_l \frac{A_l^2 * V_l(CPUE)}{n_l} \tag{2}$$

where n_l is the number of hauls in stratum l and V_l (CPUE) is the variance of catch per unit area in subdivision l.

According to the Neymann allocation for optimal sampling (Cochran 1977), the minimum variance of the total biomass estimated is achieved if the samples are distributed between strata as follows:

$$n_l = N \frac{A_l \sqrt{V_l}(CPUE)}{\sum_k A_k \sqrt{V_k}(CPUE)} \tag{3}$$

N = the total number of hauls available and k is a summation constant over all subareas.

Setting:

$$D_l = A_l \sqrt{V_l}(CPUE) / (\sum_k A_k \sqrt{V_k}(CPUE) \tag{4}$$

and substituting (2) into (1) gives:

$$V_{min}(\hat{B}) = \sum_l \frac{A_l^2 * V_l(CPUE)}{N * D_l} \tag{5}$$

Rearranging the equation, the minimum expected variance for an area divided in subareas can be expressed as:

Looking at the North Sea, taking the CPUE for a certain square as the mean of the CPUE of all hauls in this square and setting the total numbers of squares in the North Sea equal to N in equation 3, it is possible to find the minimum expected variance for the whole North Sea as the sum of the minimum expected variance for each subdivisions by using eq. 3. Because of the high unbalance in the data set, the mean of catches in each square has been used. Consequently, no weighting of data within squares has taken place.

Looking at the IYFS data base on year class I of herring in the period from 1981 to 1992, it is possible to identify six strata. Initial inspection of the data suggested that a reasonable stratification of the North Sea would be into five subareas and the remainders. Each of the five areas is geographically concatenated. A concatenated area was defined so at least the corners of the squares would be joined. By moving border squares from one subarea to an adjacent subarea, each time calculating the variance of the total biomass estimated using eq. 3, it was possible to define a final area partition which gave the lowest variance. The result is shown in figure 2.1.1.2.

2.1.2 Icelandic Groundfish Survey

The Icelandic groundfish survey started in 1985. The area of investigation covers the Icelandic shelf down to the 500 m depth contour. 600 stations were considered a reasonable effort to reach an acceptable level of coefficient of variation of cod indices. In order to work the 600 stations within a reasonable time limit, five commercial, standardized, stern trawlers are leased.

The allocation of trawling stations is based on the stratified random sampling theory. The stratification scheme is based on pre-estimated cod density patterns derived from commercial as well as research vessel catch data, which were summarized by statistical squares. The statistical square basis allows flexibility in post-stratification with respect to different species.

Based on biological and hydrographical considerations, the survey area is divided into two areas, a northern area and a southern area.

The allocation of statistical squares to strata is based on the estimated density of cod in each square. Information on cod density was derived from three different sources. The trawler captains and their advisors graded each square with respect to their experience of fishing in March. Commercial fisheries data yielded additional information on cod density, as did results from previous research surveys.

Ten strata were constructed from the statistical squares, four in the southern area and six in the northern one. Statistical squares in each stratum are not necessarily adjacent, which allows more flexibility in constructing homogeneous strata with regard to fish density.

Stations were divided between strata in direct proportion to the product of the area of each stratum and its estimated cod density. Finally, the trawl stations of a stratum were allocated to each square within the stratum in direct proportion to the area of the square.

Stations within each statistical square were divided equally between fishermen and project members from the Marine Research Institute. Project members selected random positions for their stations. Fishermen were asked to fix their stations in accordance with their knowledge and experience of fishing and fishing grounds. Trawling is done both day and night, and sampling is distributed uniformly over the 24 hours.

This sampling method may be classified as "semi-random stratified" since only half of the stations are randomly selected.

2.2 GLM & Multiplicative Models

2.2.1 Diagnostics and model considerations

The standardized index used for herring abundance from the IYFS is based on a general linear model fitted to survey data from a combination of many different vessels fishing over different subareas of the survey area for a varying number of years. Typically, such models have severely unbalanced designs which makes it difficult to uniquely estimate main effects and often result in dubious, significant interaction terms.

The Working Group evaluated the model for the herring data by applying a number of diagnostic tools to the results of the general linear model analyses based on using years, ships, subareas and day/night as main effects with multiple levels. Catches of age one herring were log transformed with a constant of 0.5 added to all of the catches as per the original analyses (Sparholt 1990). The Group's analyses differed from Sparholt's in that we did not include depth in the analyses and hour of day was only readily available as either daytime or nighttime. The pattern of log catch as a function of depth is given in Figure 2.2.1.1. Almost all of the observations were in depths less than 150 m. The Group excluded depth from our analyses because it appeared that any relationship between depth and catch was more likely driven by the very few observations made at the deeper depths than by any underlying biological processes. Future surveys may have to include more hauls in deeper water to resolve this issue. Also, depth is partially aliased with subarea and will therefore could introduce further problems into the analyses.

The raw residuals from the model for data from 1981 to 1992 are plotted in Figure 2.2.1.2. Scaling the residuals by the standard error would reduce the scatter in this plot, but the pattern of mainly positive residuals for fitted values less than 4.0 would remain. The band of residuals across the bottom the graph was due to the substitution of log(0.5) for zero catches.

The influence of individual observations of catch of herring on the fitted values can be assessed from the leverage plot in Figure 2.2.1.3. The horizontal line across the plot marks the cutoff between high and low leverage points. The majority of the high leverage points represent cases where only a few observations provided all of the information for the estimate of an individual level of the subarea or vessel effect.

The half normal plot in Figure 2.2.1.4 identifies the major components of the linear model. Basically, subarea 2 (and 5) and four vessels (AND2, DAN2, GOS, TRI) account for most of the variation identified as being due to the main effects. Closer examination of the data revealed the extent of the imbalance in the 'design'. Subarea 4 was only sampled by vessel 2 (ARG) over the whole time period and therefore any contribution from this level of the subarea effect will be confounded by vessel effects. Only two or three vessels sample subarea 5 and 6 but coverage by these vessels was not consistent over the 12 years of the survey in our analyses. Night tows made up only 15% of the total number of tows over the 12 years and most of these night tows were made in subarea 3 where they represented 29% of the total number of tows for this subarea. Any contributions to the model from subareas 4, 5 and 6 and from the inclusion of day/night differences were probably confounded with vessel and year effects.

The Working Group re-analyzed the data using only observations from daytime tows, subareas 1-3 and from six vessels (ARG, CIR, DAN2, SCO2, THA, TRI) which fished these subareas more or less consistently over the 12 years. The main effects of year, ship and subarea were still significant (Type I and III sums of squares) for this reduced data set. The Group's attempt at constructing a more 'balanced' data set resulted in somewhat fewer high leverage points than observed for the full data set (Figure 2.2.1.5). However, those that still exist need to be investigated further. The half normal plot of mean effects in Figure 2.2.1.6 indicates that the model is still being driven by one subarea and 3 vessel effects. Note year effects do not figure prominently in the model.

The Group predicted 'standardized' catch rates for the six vessels used in the analysis. These catch rates were obtained by averaging the predicted values for each vessel over the three subareas (Figure 2.2.1.7). Trends were fairly similar for the six vessels with the estimates from SCO2 and THA being consistently lower than those from the other four vessels. The dominance of vessel and subarea effects is of concern for this model, especially when standardized catch rates are estimated from it to track annual changes in herring abundance. The standardized catch rates for the English vessel Cirolana (CIR) are plotted for each of the three subareas in Figure 2.2.1.8. Note the difference in catch rates between the subareas. Vessels do not appear to consistently sample each subarea at the same rate over time. Cirolana fished between 16 and 38% of its total survey sets in subarea 2 over the 10 years that it participated in the survey. It is possible that annual trends in the survey catches of herring may fully or partially reflect effects due to changes in what subareas are being fished by which vessel in any one year, instead of only annual changes in herring abundance. The highly unbalanced design makes interpretation of any of the main effects, especially year, extremely difficult. This problem will be further exacerbated if interaction terms are included in the model.

In addition to the diagnostic study reported above participants also looked at a multivariate ANOVA of the full data set by including all ages 1, 2 and 3, in the response matrix. The main effects of year, ship, subareas and day/night were all significant as they were for the univariate models. The application of distributions other than the normal to the problem was also considered but abandoned after looking at the results of the diagnostic study. Both the multivariate model and any non-normal model will also be adversely affected by the unbalanced nature of the data. Therefore, participants suggest that questions concerning redesign of the survey to obtain more balance for the main effects be addressed before any alternative models are investigated. The important elements of such a redesign should include consistent coverage of subareas by the same vessels over time as well as having all vessels cover the same subareas. Wherever possible, the same vessels should be used over time and when old vessels are replaced calibration studies with the new vessels should be undertaken.

North Sea Herring: Fishing power

Several vessels have been used in the IYFS since its inception in 1981. No empirical experiments have been conducted to compare the fishing power of these vessels. Rather, efforts have been made to standardize fishing methods and gears in order to reduce the inter-vessel differences. The potential of using the multiplicative model to compare the power of these vessels was investigated.

The treatment of zero observations is an important aspect of multiplicative models. The herring working group has designated six subareas for North Sea herring (Figure 2.1.1.2). The following table indicates that there was considerable variation in the percentage of non-zero observations among these areas and at different ages.

Percentage non-zero observations at age for North Sea herring

	Area					
	1	2	3	4	5	6
Age 1	89	97	42	100	98	58
Age 2	77	69	67	100	91	65
Age 3	44	21	66	94	76	55
N	1716	796	1764	289	58	231

The temporal and spatial distribution of fishing hauls by the vessels involved in the survey is also important in such analyses. Very few vessels have been used in every year of the survey. Several have participated only for short periods. Furthermore, most vessels cover only a limited area of the North Sea (Figure 2.2.1.9). An attempt was made to compare the fishing power of some vessels by identifying those vessels and areas in which there exists sufficient temporal and spatial overlap of fishing as well as few zero hauls to allow GLM analyses. The analysis used areas 1 and 2 and the vessels CIR, DAN2, SCO2, THA, and TRI. The basic model is

$$\ln(C_{v,s,y} + 0.5) = \mu + V + S + Y + V*Y$$

where v, s, y indicate vessel, subarea, and year. The V*Y interaction allowed us to look at differences in catchability among vessels. The resulting year effects for the different vessels are given in Figure 2.2.1.10

The results indicate that vessel THA had relatively low catchability in 1985-88. Vessel SCO2 catchability looks low for 1990-92. The rest of the vessels appear to have similar catchabilities

2.2.2 Fishing Power

Eleven years of herring data from the IYFS was used to examine if such data could be used to test for differences in vessel fishing power, using both a paired and a General Linear Model (GLM) approach. The ln(catch age 1) was the dependent variable and only non-zero tows from subarea 2 was used in the analysis. Subarea 2 is a relatively homogenous area with moderately high abundances and few zero tows (Figure 2.1.1.2).

2.2.2.1 Pairwise comparisons

Tows from three vessels, CIROLANA (cir), DANA (dan2) and TRIDENS (tri) were paired on the basis of square and year. When two vessels each did multiple tows in the same square/year combination, an effort was made to pair them by shoot time and date. It was not possible to determine the actual distance between the members of a pair. There were 48 cir-dan2 pairs, 10 tri-dan2 pairs and 7 tri-cir pairs. Plots of the transformed catch for the cir-dan2, the tri-dan2 and the tri-cir pairs are given in Figures 2.2.2.1a-c. The lines shown are not regression lines, but are lines with slope = 1.0. The plots give some heuristic evidence of the following relationships: dan2 > cir, tri > dan2 and tri > cir.

The paired differences were formed and a paired t-test was used to test the hypotheses: cir effect - dan2 effect = 0, tri effect- dan2 effect = 0 and cir effect - tri effect = 0. Results are given in Table 2.2.2.1.

No difference was detected between tri and cir or between tri and dan2, even though some evidence was seen in Figure 2. The small number of observations is undoubtedly a contributing factor. A significant difference was detected between cir and dan2, with dan2 catching more herring. Conversion coefficients for cir-dan2 were calculated using Bradu and Mundlak (1970), and are given in Table 2.2.2.1.2.

2.2.2.2 GLMs

Data were not paired for the GLM analysis and included all non-zero tows of tri, dan2 and cir. No quadratic relationship was seen on plots of depth vs transformed catch by year, by vessel and by day/night, so depth was included in the model as a covariate. A model including all 2 factor interactions was run, and non significant effects were excluded. the resulting model is shown in Table 2.2.2.2.1. The significant depth*year interaction was investigated by plotting transformed catch vs depth by year. There appears to be a negative relationship between catch and depth in 1981 (Figure 3), while the other years either have positive or no apparent relationships. No linear regressions were fit to catch and depth by year.

The 17 df for the significant ship*year interaction term (Table 2.2.2.2.1) means that the F test is testing 17 independent hypotheses simultaneously. The estimable function (available in PROC GLM) for the ship*year effect was used to determine what these 17 hypotheses are:

1) cir81 - tri81 = cir90 - tri90
2) cir82 - tri82 = cir90 - tri90
3) cir83 - tri83 = cir90 - tri90
4) cir84 - tri84 = cir90 - tri90
5) cir85 tri85 = cir90 - tri90
6) cir86 - tri86 = cir90 - tri90
7) cir87 - tri87 = cir90 - tri90
8) cir88 - tri88 = cir90 - tri90
9) cir89 - tri89 = cir90 - tri90
10) dan2 82 - tri82 = dan2 90 - tri90
11) dan2 83 - tri83 = dan2 90 - tri90
12) dan2 84 - tri84 = dan2 90 - tri90
13) dan2 85 - tri85 = dan2 90 - tri90
14) dan2 86 - tri86 = dan2 90 - tri90
15) dan2 87 - tri87 = dan2 90 - tri90
16) dan2 88 - tri88 = dan2 90 - tri90
17) dan2 89 - tri89 = dan2 90 - tri90

Hypothesis 1 is testing if the difference between cir and tri in 1981 is equal to the difference between cir and tri in 1990. A similar interpretation is made for the other hypotheses. The ESTIMATE statement in PROC GLM was used to estimate and test each hypothesis. Only hypothesis 2 was significant (Table 2.2.2.2.2), so it appears to be a major contributor to the significant ship*year interaction. The interpretation is that the difference between cir and tri in 1982 is significantly different from the difference between cir and tri in 1990. The MEANS statement in PROC GLM could be used to generate ship*year means, and the significant contrast could be plotted, allowing the scientist to better interpret the contrast, and to decide if it is really of importance.

The 1982 data was then deleted, and the ship*year interaction term was removed from the model. The glm is shown in Table 2.2.2.2.3. The vessel differences tri-cir, tri-dan2 and cir-dan2 were estimated to see how they compared with the differences from the paired analysis. Unbiased estimates of these vessel differences were possible because no vessel interaction term was in the model. The tri-cir and tri-dan2 differences were significant (Pr > |T| = 0.0001 and 0.0155), but the cir-dan2 difference was not. The vessel tri showed greater fishing power in both cases (positive estimates), while the non-significant cir-dan2 contrast was negative, indicating a possible higher fishing power for dan2 over cir. These results agree with the observations from the pairs plots (ie tri > dan2, tri > cir and dan2 > cir). The non-significance of the cir-dan2 difference probably reflects the lack of control of variability inherent in observational data of this type.

2.2.2.3 Comparisons between pairwise and GLM estimates

A paired and a GLM analysis were done on non-zero age 1 herring data from the IYFS to explore the possibility of using these methods to detect vessel differences. Significant differences in vessel fishing power were seen using both methods, but the conclusions drawn from the statistical tests were different. The same trends were seen in both analyses, however.

The vessels in the paired data set were paired up on the basis of square and year, so it is possible that there is a considerable temporal and spatial separation within a pair. This can introduce significant variation, and is a disadvantage of pairing data in this manner.

The GLM analysis yielded some useful results here, and was fairly straightforward to do. However, the data set was restricted to non-zero tows from three vessels in subarea 2. A GLM analysis would become more complex (i. e, very unbalanced) if more areas and vessels were included in the analysis. The increased unbalance and greater variability in the data would make interpretation and detection of differences more difficult. Given the disadvantages of the two data sets considered here, a series of paired tow experiments designed to reduce variability between the tows would give more information.

2.2.3 Adjusted standard index

2.2.3.1 Estimates of fishing power

In order to estimate the fishing power of the various vessels participating in the IYFS the catch rates of 1-ring herring were analyzed in GLM models. The models used include year, day/night, depth, ship and rectangle as main effects. No interaction effects were considered. Zero-catches were excluded from the model which was based on log transformed data. Thus, it was implicitly assumed that the frequencies of zero-catches do not differ by vessel and correction factors are only necessary for non-zero catches. As shown later on there seems to be a significant year-area interaction as the proportion of herring in Division IIIa in one year (1988) was very high. It might, therefore, have been more reasonable to exclude data from Division IIIa, at least for that year. The estimated fishing power by year and as a mean over all years are given in Table 2.2.3.1. The values by year are obtained by separate GLMs where the ship in question in the given year has been renamed in the GLM SAS program and kept unchanged in all other years.

Thus, all other factors were almost unchanged as only a small proportion of the data were changed. Actual inspection of the parameter estimates in each run confirmed this. The fishing power estimates were fairly constant over the years for Anton Dorhn (AND2), Argos (ARG), Cirolana (CIR), Eldjärn (ELD), Scotia (SCO2), Tridens (TRI), and Walter Herwig (WAH2). The fishing power of Dana (DAN2) was low in 1982 and high in 1985, of Explorer (EXP) high in 1981 and low in 1989, of Iris (IRIS) low in 1984 and high in 1991 and of Thalassa (THA) very low in 1985 and 1988 and low in 1986 and 1987. In 1992 the fishing power was on an average level (4.13) for Thalassa and in this year Thalassa used a Scan-mar Echo Sounder System to check the gear performance during hauling for the first time. As observed by Sparholt (1990) the average fishing power of Eldjärn and Thalassa is low.

A GLM model similar to the above ones was applied to the 2-ringer catch rates excluding data from Division IIIa as the main part of the 2-ring herring in this area is belonging to another stock. The results of this GLM are also given in Table 2.2.3.1 and shows that Eldjärn and Thalassa have also a low fishing power for 2-ring herring.

The day/night effect was also estimated by the GLMs and the catch rates during day-time is 1.68 times those during night-time for 1-ring herring and 1.82 for 2-ring herring.

2.2.3.2 Standardization of index

In order to get a simple improvement of the standard IYFS 1-ring herring index it is in the following attempted to get a corrected standard index where the catch rates by haul are corrected for the difference in fishing power between vessels and difference between day and night catches. In order to take into account that only parts of rectangles in Division IIIa have depths (between 10 and 200 m) where 1-ring herring appears the catch rates in rectangles in Division IIIa have been down weighted by the fraction in each rectangle with depths between 10 and 200 m. These fractions are given in Table 2.2.3.2.1.

Based on the fishing power estimates in Table 2.2.3.1, correction (or conversions) factors were calculated and given in Table 2.2.3.2.2. A fishing power value of 4.00 was the standardizing value and the correction factors were exp(4.00-fishing power). All vessels were considered as having constant fishing power over the years except Thalassa. Four time periods were considered for this vessel. Especially the period 1985-1988 is problematic with a correction factor of 11.70. It might be better to delete these data from the index calculation. However, they were included here and the catch in no/hr for each haul in the database were corrected by multiplying by the correction factor for both ship and time of day effects. After correcting the individual haul data the means of the means over rectangles were calculated for the entire North Sea and Division IIIa. The standard index is only based on a restricted number of rectangles in the so-called standard area for 1-ring herring. However, when taking the arithmetic means there is no need to restrict the data to a limited area and especially the inclusion of Division IIIa seems important as big proportions of 1-ring herring in some years are found in this area (Table 2.2.3.3). The comparison of the obtained indices with the VPA is given in Section 2.5.1.

2.2.4 Year-class model

Working papers # 6 and 7 presented a multiplicative analysis of RV data for two Canadian cod stocks. The model included year class, age, and spatial effects, and demonstrated the spatial segregation of age groups in both stocks. This treatment of the data takes advantage of the fact that the surveys estimate the same year classes at successive ages. However, changes in the exploitation pattern or distribution of the fish can have important effects on the parameter estimates. The working papers presented methods of detecting and treating such difficulties.

Analysis was performed to develop an index of year-class strength using all ages, vessels and areas. In this case, only non-zero catches were used. The basic model was;

$$\ln(C_{s,a,yc}) = \mu + S + A + YC + S*A$$

where YC represents year class.

The trend in estimated year class strength is given in Figure 2.2.4.1. The trend is for increasing yearclass strength from 1978-1986, followed by a decline to values similar to those from the late 1970's-early 1980's. The S*A interaction term was included to account for important differences in the relative abundance of the age groups among areas (Figure 2.2.4.2).

2.3 Spatially Oriented Methods

2.3.1 Geostatistical Methods

a) Position of the Problem

The Working Group address here the problem of estimating the mean value of a variable over a domain from the known values of this variable (here catch) at sample locations.

The unknown mean value considered here would be the arithmetic mean of values at every location within the domain. For this mean value the Group will consider a linear estimator of the data values. In the case of irregularly spaced data, a weighted average should be better used. Otherwise, the arithmetic mean of data values is likely to be sufficient. If the data locations are very dense throughout the domain, this simple arithmetic mean would be close to the mean value. In general, however, this is not the case and the essential point is to know how precise the estimator is, which will be characterized by the variance of the error between true and estimated values (which will be referred to as the global estimation variance).

In the geostatistical approach, the first thing is to look at the data and at their spatial structure. The spatial structure is often described by the variogram, which measures the half mean variability between two points x and x+h as a function of their distance h. This variogram is first computed on the data. Then a variogram model is fitted to this observed variogram. This model is afterwards to compute variances and perform kriging for instance. These methods are model based and results depend on how realistic the model is. If one wants to use the location of data through the spatial structure, one has to make more assumptions.

There is a general geostatistical formula for the global estimation variance which depends on:

- the geometry of the domain;
- the location of data; and
- the spatial structure of the variable.

This formula is (with the arithmetic mean as estimator):

Variance =
2 gamma (V,I) - gamma (V,V) - gamma (I,I)

where

gamma (V,V) is the average variogram over the domain (i.e., the average of the variogram value gamma (y-x) between two points x and y which sweep independently over the domain)

gamma (I,I) is the average variogram over the data points

gamma (V,I) is the average variogram between the data points and the domain

In practice, the variogram is estimated from the data values and what we get is in fact an estimation of the variance.

Taking into account the locations of the data and the spatial structure through this formula gives a variance which is not necessarily smaller than the classical sigma**2/n. For this point, see Anon. (1991). It can be larger, or, in the case of nugget variogram (no spatial correlation) equal (in this case there is no matter about fixed or random locations).

In the case of strata, or zones sampled with a different location density according to their presumable abundance, the formula above can be used for each zone (with its proper variogram), and variances are to be weighted as usual by the squared surfaces to give the overall variance.

The above formula can be extended to the case of a weighted average as an estimator. When these weights minimize this variance we have the kriging estimator (it is generally used for local estimation i.e., mapping). For the global estimation, however, weighting is not necessary when the data are uniformly spread throughout the field or when there is no spatial correlation. In such cases, it may happen that weighting differently a large sample value would change the estimated mean, but this change would likely be small in comparison to the magnitude of the global estimation variance.

b) North Sea Herring, age 1, years 1981-1992

The Working Group studied the spatial distribution of catch values, year by year, and without distinguishing between vessels. The location of data with a proportional representation of catch values is given year by year in Figure 2.2.1a-k. There is a large zone in the North where there are only small values. Otherwise, small values are present in the vicinity of the large ones. The distributions are skewed. The arithmetic means range from 642 (year 1981) to 5,466 (year 1988; see below). The coefficients of variation range from 2.7 up to 6.2 (year 1990), generally around 3-3.5. The largest value for each year contributes for around 10% of the mean, except 27% for year 1990 (the maximum value is 150,588 for this year).

Year	Coef Var	Arith mean
1981	2.7	642
1982	2.6	1019
1983	3.1	1337
1984	3.2	1564
1985	3.2	2331
1986	3.5	3717
1987	3.3	4354
1988	3.9	5466
1989	4.9	3058
1990	6.2	1454
1991	3.4	1376
1992	2.9	1405

The variograms have been computed year by year, and also averaged, with a distance lag of 5 nm. Most of them are practically pure nugget, showing no evidence of structure for distances between five or more than 100 nm (Figure 2.3.1.2).

The reason for this is the largest values, which would mask the underlying structure if any. But the reality is probably so, even if large values are subject to larger uncertainties. Knowing these nugget variograms does, however, provide some information. A consequence is that the exact locations of the data do not matter. The variance when estimating the mean by the arithmetic average of the data is then sigma**2/n. This would give an error of about 15-20% of the mean. However, there is at least one large distance structure (distinction of the poor North zone) which would suggest a separate treatment of the two zones for the estimation.

c) Iceland Cod, age 4, 1985-91

The large catch values are located in the North (Figure 2.3.1.3 a-g). The arithmetic mean per year ranges from three (1990) to 25 (1988; see below). The distributions of values are skewed, though less than for the North Sea herring. The coefficients of variation range from 2.4 to 4.8.

Year	Coef. Var.	Arith	Krig.
1985	2.6	10.91	9.68
1986	2.7	5.04	4.48
1987	2.7	18.82	16.69
1988	3.4	25.25	25.71
1989	4.8	16.41	14.26
1990	2.7	2.97	2.34
1991	2.4	6.74	5.55
1992		5.28	4.59

The variograms per year show a repeated structure, with a range of 50-70 km and a nugget component which is about half of the sill (Figure 2.3.1.4). The range of 50-70 km addresses the distribution of the large values, but despite this structure, there is also quite an important short distance variability. This nugget component is the witness of either a very short range variability or sampling errors. It is not possible to separate these two effects from this single observation. As a general rule, the larger the nugget proportion, the less representative of their neighbourhood of the values.

The variograms have been used to make a kriging estimation. Comparison between the kriged mean and the arithmetic mean shows that the kriging mean is systematically lower than the arithmetic mean (except for year 1988 where the kriging estimate is slightly larger). This is due to the smaller density of data in the zones with small values.

2.3.2 Empirical Interpolation Methods

Most interpolation methods can be expressed as weighted moving summations over the data, where the sum of the weights is unity. Integration under the interpolated surface also reduces approximately to the weighted summation of the data, and this is why the arithmetic mean multiplied by the area is usually a useful approximation to the integral.

In general the result of the summation involved in the integral is essentially independent of the precise form of the weights involved in the interpolation, and the value obtained for the integral is therefore only weakly dependent on the precise form of interpolation used. For the calculation of global abundance indices, therefore, it is not particularly important to use an optimal interpolation procedure, and simpler sub optimal procedures may be adequate. This is in contrast to the situation when mapping is the primary goal when the method of interpolation is crucial.

Precise interpolation (providing an exact match at the data points) is also not necessary, and interpolating using smoothing splines and similar procedures may be envisaged. This may be advantageous because a smoothing spline can be fitted in a way which takes account of the precision of the observations, giving less weight to imprecise data, for example, and the residuals from the fitted surface can be used to generate an estimate of the variance of the result. Methods of locally weighted least squares (LOWESS and similar acronyms) are suitable for this purpose.

Such methods involve empirical (rather than optimal) interpolation, and may be particularly useful because they

a) can be used equally well with fixed or random survey designs

b) can allow for the variable precision of the observations

c) can cope easily with missing observations, and possibly also can identify and deal intelligently with outliers (including zeroes).

(a) Contouring-based Interpolation

The Methods WG report from the Nantes meeting (Anon. 1990) gives some results from the application of one such method (actually involving a quite sophisticated interpolation procedure), and more general discussion of these issues (sections 2.4.3 and 2.4.4 of Anon. 1990). Rather surprisingly this work indicated that the results were not improved by increasing the degree of smoothing used, probably because, with a log transformation, this resulted in excessive weight being given to the low (and zero) observations.

This same procedure (hyper-gridding using program HYPGRD) has been applied to data sets analyzed at this meeting, and the results (series " cont ") are discussed below.

(b) Locally Weighted Robust Estimators

In addition, the results obtained using a much simpler locally weighted robust estimating procedure have been evaluated. This was considered worthwhile because it might help to reduce the problems associated with extreme observations and zeroes, without the complication of an iteratively re-weighted least squares procedure which uses the fitted values and a variance/mean relationship to allow for the precision of the data (especially the very low precision attached to the zero observations).

The procedure adopted involves a simple locally weighted mean, with a Gaussian spatial weighting function with a range of 1 grid spacing. The rational for this is that if one value at a grid point is to be used to represent the observations within the grid square, which surrounds it, then all these should be given approximately equal weight. Values in adjacent squares should have some influence, and those further away should have very little. A strictly local (e.g., bi-square or tri-cube) weighting is not used in this context, because these prevent interpolation across unsampled areas.

The mean of the spatially weighted observations may be defined in many ways (see e.g., Mosteller & Tukey 1977), including the simple arithmetic mean. In this case the estimation reduces to the arithmetic mean of the data taking account of any unevenness in the spatial distribution of the observations. These results are given as the "lw-arith" series of results, for comparative purposes.

The robust estimator is a robustly re-weighted mean with a bounded hyperbolic weighting function. This approximates to a trimmed mean, and with very scattered observations tends towards the median. Again, a strictly local weighting is avoided since in this context it seems to be unreasonable to ignore outliers altogether: a median type estimator takes note of their presence, but not their actual magnitude. The estimator is discussed in a little more detail in Appendix S. With this median-type estimator no logarithmic transformation is required, and zeroes present no difficulty.

The variance of such estimates at any point can be determined from the residual sum of square (or absolute deviations) of the data from the fitted surface. Provided that the errors of adjacent estimates are uncorrelated, an approximate variance for the integral may be obtained by summing these variances. This will be approximately true if (as here) the range of the local weighting is not more than the grid spacing (otherwise it will be an under-estimate). These estimates are given in Tables 2.5.1.2 and 2.5.2.2. Such estimates are only approximate, and may be unreliable if there is much interpolation across unsampled areas. This is not the situation for the data sets considered here.

The results obtained with this procedure (series "lw-rob") are also discussed below.

2.4 Other Methods

2.4.1 Time series

Abundance indices from trawl surveys are generally much more variable than estimates from VPA. In particular, variance estimates of such indices based on the within survey variability appears to be underestimates. A potential source of variability is that the catchability varies both within and between years, e. g., due to changes in behaviour and as a result of variation in gears and survey routines. Working papers #3 and #12 present techniques that significantly reduce the true variance of abundance indices by making use of the entire time series of survey data. In principle, the "best" current indices is estimated from a weighted average of the predictions from previous years and the present year's survey indices. The method applied in WP-12 assumed on a priori choice of model structure (specifically an integrated moving average model) because of the limitations imposed by the short time series of data available (generally <30 points). Methods from time-series analysis appears to provide estimates from trawl surveys that is more closely related to VPA estimates. In paper #3 it is also demonstrated that ARIMA models (see Box and Jenkins 1976) can provide estimates of yearly changes in catchability, both by using the VPA time series, and from the trawl indices alone. The method using VPA assume that the indices from the VPA's and the survey are independent. For tuning VPA this assumption may not be entirely justified, but it was generally agreed that by deleting the last five years of the series in the analysis, this effect is likely to be negligible. The methods provide reasonable results for a large number of species.

2.4.2 Post stratification

Working Paper #5 indicates that the precision of abundance estimates from trawl surveys can be improved by post-stratification. The method involved constructing new strata by partitioning the initial geographic-strata. The variance estimate for the post-stratified mean includes a component due to the stratification and a component due to deviation from the initial allocation scheme. It should be stressed that the choice of post-stratification boundaries should be based entirely on information which is independent of the survey data to be post-stratified.

Post-stratification can be efficient in multi-species surveys since different post-stratification schemes could be employed for different species, provided that there is a clear basis or post-stratification using either historical information of an auxiliary variable (e.g., environmental data, hydroacoustic estimates etc.).

2.4.3 Resampling procedures

Bootstrap procedures for complex survey schemes were discussed (Working Paper #4). The importance of resampling in accordance with the original sampling scheme was demonstrated. To obtain unbiased variance estimates of stratified means modifications to the number of observations resampled from each stratum were required. The problem involved in obtaining symmetric confidence intervals were also discussed.

2.4.4 Use of covariates

Variation in the spatial distributions of stocks, as measured by research vessel surveys, can have important ramifications for the computation of consistent time series measures for VPA tuning. Changes in geographic distribution may be random, or related to a variety of biotic and abiotic factors. By understanding factors influencing the pdfs of survey catches, then it is envisaged that these covariates can be used to reduce variance and bias of tuning indices.

Two working papers considered the distribution, as indexed by research vessel surveys, in relation to environmental covariates. Two separate methodologies were used: a non-parametric analysis of the distributions of stocks against the cumulative frequency of temperature, depth and salinity (Smith and Nicholson WP #10), and a GLM approach investigation of changes in the centroids and range extensions of stocks in relation to variation in stock abundance and water temperatures (Murawski WP #11). Because of the lack of complete information on these covariates in the North Sea herring data set, it was not possible to apply the methods to the herring problem. Likewise, the use of auxiliary variables in kriging or co-kriging was not possible because a synoptic data set for potential covariates could not be obtained in time to allow analysis.

The two methods proposed for application to covariate are outlined by example below.

2.4.4.1 Assessing the environmental preferences of stocks.

Many researchers have commented upon the fact that estimates of abundance from groundfish trawl surveys often appear to indicate unrealistic large interannual changes in abundance of groundfish species. The possibility of interannual changes in the availability of the fish to the survey gear has been suggested as alternate explanation to the magnitude of these changes. Recent research on the Scotian Shelf (Nova Scotia, Canada) indicates that different species exhibit very different apparent "associations" for hydrographic conditions such as bottom temperature, bottom salinity and depth. It has been hypothesized that interannual changes in the bottom water characteristics may be a factor in changing availability. That is, if the "associations" type of water is not on the bottom in a specific locale, the fish will not be there to be sampled by bottom trawl gear.

Preliminary investigations of possible environmental "associations" of cod caught in the English Groundfish Survey of the North Sea, were undertaken. The distribution of environmental conditions (depth, temperature, salinity) within the survey area is characterized by an empirical cumulative frequency (pdf) curve. Commonly the probability associated with each observation in a pdf is $1/n$, however for more complex random survey designs (e.g., stratified random) this probability may vary between observations. Therefore the pdf is constructed with general probability, π_i

$$f(t) = \sum_i \pi_i \, I(x_i), \tag{6}$$

where,

$$I(x_i) = \begin{cases} 1, \; if \; x_i \le t; \\ o, \; otherwise \end{cases} \tag{7}$$

The 'association' of an animal population with particular environmental condition is measured on an accumulation of the stock along the gradients of available conditions. Fish catches (numbers) are associated with environmental conditions on a trawl set-by-set basis:

$$g(t) = \sum_i \pi_i \frac{Y_i}{\bar{y}} I\,(x_i) \tag{8}$$

Catches larger than the mean, $\bar{y}$, would indicate hydrographic conditions for which fish have a higher association than conditions where catches were smaller than the mean. The difference between *g(t)* and *f(t)* is tested by a method similar to Kolmogorov-Smirnov tests.

Results of these analyses (Figure 2.4.4.1) indicated several main points: (1) there was a distinct age-effect in environmental associations, age 3 cod were consistently distributed differentially from age 1, and 2, and showed distinct association with cooler than average temperatures. Examination of annual cumulation curves indicates that 1985 was an anomalous year. Inter-annual changes in the availability of suitable bottom water condition confuse abundance changes with changes in availability. Thus, the spatial 'persistence' assumption of fixed-station designs may not necessarily apply.

2.4.4.2 GLM approaches to distribution changes

A GLM approach to evaluating factors influencing distribution changes was also proposed (Murawski WP #11). In this approach, geographically weighted average surface and bottom water temperatures, τs_j, τb_j, and average abundance (δ_{ij}) was related to the distribution centroids and the maximum latitude of occurrence for a large number of stocks:

$$\lambda_{ij} = \alpha + \beta_1 \cdot \delta_{ij} + \beta_2 \cdot \tau s_j + \beta_3 \cdot \tau b_j + \epsilon \tag{9}$$

$$\lambda max_{ij} = \alpha + \beta_1 \cdot \delta_{ij} + \beta_2 \cdot \tau s_j + \beta_3 \cdot \tau b_j + \epsilon \tag{10}$$

Where: λ_{ij} = the mean latitude (weighted by log $n + 1$) of occurrence of species i, taken in year (survey) j, λmax_{ij} = the maximum latitude of occurrence of species i in year (survey) j, δ_{ij} = the mean abundance ($\log_e$ [n+1]) of species i taken in year (survey) j, τs_j, τb_j = stratified mean surface and bottom water temperatures computed for year (survey) j, α, β = computed regression coefficients, ϵ = normally distributed random error term.

Analyses of the effects of temperature and abundance were conducted for 36 species found in USA bottom trawl surveys.

Significant ($p < 0.05$) regression models were fitted for 17 of 36 species from spring and autumn survey data (Figure 2.4.4.2). Variations in water temperature were significant in explaining changes in mean latitude of occurrence for 12 of 36 species in both seasons. Maximum distribution response to inter-annual differences in water temperatures occurred for pelagic species, including Atlantic mackerel and Atlantic herring: centers of abundance of these populations shifted by 0.5-0.8° latitude for each 1°C increase in average water temperature. Significant latitudinal range extensions (as measured by the maximum latitude of occurrence) occurred for 5 species in spring and 4 in autumn surveys, associated with warmer water temperatures.

These results were generally consistent with those in Section 2.4.4.1: distributions of some stocks were significantly influenced by variation in environmental conditions. Use of auxiliary information on correlated variables could potentially be used to reduce the variance and bias imparted to abundance measured from such data. Smith (1990) has suggested methodology for incorporating this information in mean estimates from stratified random designs.

2.5 Comparison of Results

2.5.1 Indices for IYFS herring

As many methods as possible were applied to the IYFS herring data, selected because the ICES WG on the International Bottom Trawl Surveys had identified problems with this data set, and it was, therefore, of interest to know whether any alternatives to the standard index might reduce these.

The various indices available, (see Table 2.5.1.1) from the kriging analysis (essentially arithmetic means), the empirical interpolation methods, and various GLM methods, as well as the standard index were plotted for age 1 (Figure 2.5.1.1). Results were available for some other age groups for some methods, but were not analyzed due to lack of time. The indices were also regressed and plotted (after log transformation) against the standard VPA results, and the results are given in Figure 2.5.1.2 and Table 2.5.1.2.

It can be seen that all indices (except the SW Sector Kriging results, which are based on very few data points) are quite well correlated with each other and with VPA. The most useful indicator of the utility of the index is the residual mean square error (effectively the C.V.) given as the bottom row in the Table. It can be seen that several indices perform slightly, but probably not significantly, better than the standard index. The CV for the locally weighted arithmetic mean index seems to perform better, for reasons which are not clear since this is computationally essentially the same as the arithmetic mean and the standard index (except for a scale change). The reason may be that this index was computed for the whole North Sea, not just the standard area used for the standard index, but excluding division IIIa (unlike the kriging/mean index). It is unlikely that this is generally a preferable method. The results from the locally-weighted robust estimator, the various GLM estimates and the standard index corrected for fishing power are all disappointing, since none seems to be superior to the standard index.

2.5.2 Indices for Icelandic cod

Similar calculations were carried out for the Icelandic Cod age 4 data. The indices available, including several from Stefánsson (1991), are listed in Table 2.5.2.1. The results are plotted as time series in Figure 2.5.2.1. and the regression results are given in Figure 2.5.2.2 and Table 2.5.2.2.

The results are similar, in that no index performs significantly better than the arithmetic mean. The best performer is the Gamma-Bernoulli index of Stefánsson (1991), in agreement with his results.

2.6 Discussion

The results from this exercise were, as in 1989, disappointing. The variation among the indices for IYFS herring suggested that the geographical area included in the index is important, and suggests that further work to identify these areas which <u>are</u> correlated with eventual recruitment might be useful, as a basis for redefining the standard index.

The results for both stocks also suggest that methods which are sensitive to zero values (including anything involving a log transform) perform least well, whilst methods which are weakly sensitive to zeroes (arithmetic mean and variants thereof) perform better. The Gamma-Bernoulli method of Stefánsson (1991), which explicitly treats the zeroes also performs well for the one stock for which results were available. This suggests that future work should also be aimed at the resolution of the (difficult) problem of how best to cope with data including a large proportion of zeroes. To be consistent with the models used in calibration, the comparison of CVs should have been done with the slopes constrained to be 1.

It should be noted that analyzing individual age-groups separately, as here, may be a bit misleading if results from several age groups are to be used in a joint analysis (e.g., VPA calibration). The extent of this problem is not known (Gavaris, pers. comm.), but a possible solution is to analyse several age groups together with a year-class effect. This is preferable to analyzing proportions-at-age, which confuses year and yearclass effects.

Estimates of within survey variance should be compared with those obtained retrospectively, from VPA calibration. If they are not similar, the reasons for this should be explored. It may be appropriate to modify the inverse variance weights estimated from calibration by the exogenous estimates if those are larger, and by a minimum level of variance if those estimated are unreasonably small.

It was noted that a more general comparison may have been obtained by calibrating the respective VPA's with these alternative indices and then comparing the mean square residual around the indices. This was not possible due to time constraints.

3 DESIGN CONSIDERATIONS

3.1 Introduction

A wide range of survey designs can be used for estimating relative fish abundance. The efficiency of these designs will depend to some extent on the interaction between the spatial distribution of the fish and the spatial distribution of the samples.

Although Simple Random Sampling will provide estimates which are unbiased in the long-term, with large between-station variability alternative designs could result in a reduction of the sampling error of the relative abundance estimate.

The precision of the estimates provided by random sampling in heterogenous environments can be improved if the study area can be divided into persistent strata so that catches are more similar within strata than between strata. However, it may be difficult to find an optimal allocation of sampling effort, particularly if there are several survey objectives. One approach is described in Section 3.2. A method of analyzing data using post-stratification is described in Section 2.4.2. Working Paper #14 presents a double-sampling scheme for trawl surveys. Results from the Barents Sea suggest that using acoustic measurements (the first stage) for determining the probability for trawling at a station (the second stage) may increase precision of abundance indices.

Keeping the set of sampled stations fixed from year to year has been advocated for removing the contribution of spatial variability to the standard error of abundance estimates. Although these estimates will generally be biased, if the spatial distribution is persistent from year to year, trends in abundance will be unbiased. A compromise between random and fixed-station sampling is to sample with partial replacement i.e., keeping a subset of stations fixed, and allowing the remainder to be chosen at random. The efficiency of this method will depend upon the persistence of the fixed subset, the amount of spatial variability and the allocation of sampling effort to the fixed and random stations. These aspects are discussed in Section 3.3.

3.2 Constrained Optimal Stratified Sampling

A technique of constrained optimal sampling was proposed as a way to accommodate the multiple objectives of modern marine sampling surveys and as a way to improve the precision of the abundance estimate of a target species. This technique involves the optimal grouping of homogeneous habitats into a small number of allocation strata and the allocation of the available sampling stations to these strata in an optimal manner within specified constraints. The stations are then distributed within the strata proportionally to the region areas. This technique thus allows the secondary objectives of a survey to be met by the imposition of constraints and the primary objective to be optimized within these constraints.

Case study

The technique is implemented by two computer programs: REGROUPE and PARTS; they are documented in Gagnon (1991). In Canada, it has been used to optimize the allocation of the 3Pn4RS winter cod survey and the 4RST summer redfish and shrimp survey. For the winter survey, the gain in efficiency due to the stratification was around 10% and the gain due to the allocation was around 65% .

The REGROUPE program was used on the age 1 herring data from the IYFS survey. The data from all the 12 years available was used together so that the within-year and the between-year variabilities were taken into account. The 176 longitude-latitude (1 deg. by 0.5 deg.) rectangles were used as the regions to be grouped into 6 strata of different sampling intensity. The purpose of this exercise was to calculate what would be the optimal sampling intensity for age 1 herring in every long-lat rectangle of the North Sea from the results of the previous surveys. The IYFS survey is multi-specific, it is not directed for herring but the results from this analysis could be used to implement such a directed survey.

Practical problems with the application of the method.

The number of regions to be considered for this case greatly exceeds that of the previous applications of the method. The REGROUPE program uses a heuristic search method that reduces the number of group configurations to be searched for this case from more than 1.3E36 to around 1.3E9. The computer time required to perform this search is nevertheless considerable (more than one day on a VAX mini).

Usefulness

The success of a survey is the consequence of many factors. Among those, the appropriateness of the survey plan to the objectives of the survey is crucial. The technique of constrained optimal sample allocation can be used to improve the precision of directed surveys without abandoning their secondary objectives. The method presented does not disrupt the stratification of existing surveys and thus allows for the historical continuity of precious time series.

3.3 Fixed and Random Stations

3.3.1 Review

Bottom trawl survey data are commonly used to calibrate VPA. This requires a relative index of abundance for a time series. With sampling over a time series we can consider

- the change from one occasion [for example, year] to the next
- the average over all occasions
- the average for any one occasion

It is commonly supposed that observations on the same sampling units are positively correlated from one occasion to the next. Efficient estimators for this situation, with partial replacement of sampling units, were considered by Patterson (1950).

When estimating change from one occasion to the next, we cannot assume that catchability is constant over the time series. We must then calibrate the VPA so as to match its change from one occasion to the next to the change observed for the survey. This permits the catchability to vary for every pair of occasions considered.

For the generally assumed model of constant catchability, we need to estimate the average over all occasions as well as the change from one occasion to the next. This is equivalent to requiring the average for each occasion. Therefore, we are interested in the relative efficiency of estimators of the average for each occasion from fixed versus random sampling.

The efficient estimator of the average for an occasion can be written as the weighted average of the matched units and the unmatched units where the average for the unmatched units is a simple average and the average for the matched units is a regression estimator. The efficiency of this estimator can be evaluated as a function of correlation between sampling units over occasions and proportion of sampling units replaced from one occasion to the next.

Note that it is possible that the variation in catchability could be reduced by sampling at the same stations from year to year, thereby reducing the model error component in calibration of VPA. Using the recent assessment for Gulf of St. Lawrence cod, the residuals between the time periods when the survey was conducted with fixed stations were compared with those when random stations were used. There was no discernable difference in the magnitude of the residuals between these periods.

Definitions

The relative merits of different survey designs hinge to some extent on the idea of PERSISTENCE. This corresponds to the condition that changes in relative abundance (expressed on a log scale if necessary) are the same at every station in the area of interest. Following the notation of Warren (WP #2), then writing

$$\mu_{iy} = \mu + \phi_i + \Psi_y + \zeta_{iy} \quad i=1...N, \; y=1...Y \tag{11}$$

where μ_{iy} is the mean relative abundance in the y 'th year at station i, and

$$\sum_{i+1}^{N} \phi_i = \sum_{j=1}^{Y} \Psi_y = \sum_{i=1}^{N} \zeta_{iy} = \sum_{y=1}^{Y} \zeta_{iy} = 0 \tag{12}$$

we see that persistence corresponds to

$$\zeta_{iy} = 0 \qquad \forall_{i,y}. \tag{13}$$

if n stations are sampled in the y 'th year, then

$$\epsilon\left[\bar{X}_y\right] = \mu + \Psi y \tag{14}$$

if the n stations are selected at random, and

$$\epsilon\left[\bar{X}_y\right] = \mu + \Psi y + \sum_{i=1}^{\eta} \phi_i + \sum_{i=1}^{\eta} \zeta_{iy} \tag{15}$$

if the stations are fixed, where X is the mean relative abundance for year y. Thus, we see that although the mean relative abundance obtained within a year with a fixed-station survey will generally be biased, differences between years will be unbiased if there is persistence, since e.g.

$$\epsilon\left[\bar{X}_2 - \bar{X}_1\right] = \Psi_2 - \Psi_1 + \sum_{i=1} \left(\zeta_{i2} - \zeta_{i1}\right)/\eta = \Psi \tag{16}$$

The problem with random sampling is that although mean relative abundance will be unbiased in the long term, its standard error will increase with the variance between stations, and in a given year could be further from the true abundance index than the biased estimate obtained by a fixed-station design. Section 3.3 considers ways of assessing survey designs ranging from completely random to completely fixed, with sampling with partial replacement in the middle. Several indices of departure from persistence are considered and used to quantify the relative merits of the different designs.

3.3.2 Identifying persistent stations

It was suggested that a limited number of fixed stations could be selected to predict the overall abundance estimate for surveys that had been conducted for a number of years. This could be useful in a operational context as few fixed stations could be sampled and considered as an adequate index of abundance for a number of years. A more extensive coverage could be undertaken on occasion to verify the usefulness of these stations. Results from the Icelandic cod surveys were analyzed; these were conducted over a period of seven years with a total of 400 fixed stations visited each year.

A stepwise selection of fixed station in a multiple regression allowed to select five stations that could explain all of the yearly estimates of abundance over the period of the first six years. However, when the derived coefficients were applied to the catches of these selected fixed stations in the most recent year (1991), the predicted abundance for the 1991 survey was over estimated by 9 times. This may be explained by two factors.

1) The 1991 estimate for the two most important fixed stations (they explain 99% of the historical abundance estimates) were the highest observed. The 1991 survey abundance estimate is the second lowest in the time series.

2) The time series is too short to adequately select fixed stations to explain the yearly variability in these surveys. The selected stations were chosen as they could predict adequately a few years, but cannot be used in a predictive manner.

Selection of fixed stations from a large number of candidate stations with few years data will inevitably lead to a good explanation of the data without necessarily providing good stable predictions. The use of multiple regression to identify a subset of stations is not a sensible method because; 1) it could produce a silly weighted average of the observations with both positive and negative weights, 2) a pre-specific weighted function might be more appropriate, possibly reflecting the variance of the observations, 3) a subset of contiguous stations or of stations with some other spatial pattern may be desirable, 4) only a small number of stations can be incorporated, due to the small number of years and, hence, degrees of freedom for the regression.

3.3.3 Measures of departure from persistence

Two Working Papers (#2, #13) were presented which considered the effectiveness of random, fixed and partial replacement surveys. The relative merits of these survey strategies depend on how far the spatial/temporal structure departs from persistence. Four measures of this departure were defined:

1) sigma_csi_squared - the mean interaction sums of squares (in a 2-factor ANOVA on stations and years),

2) w = mean interaction sums of squares/mean station sums of squares,

3) rho - the between-station correlation over years (pearson correlation coefficient),

4) rho_sp - the between-station correlation over years (spearman correlation coefficient). This measure is non-parametric, so would investigate persistence on a non-linear scale.

The w and rho measures are related by

w = (1 - rho) / (1 + rho)
rho = (1 - w) / (1 + w)

If there is persistence, the interaction terms are identically zero so

sigma_csi_squared = 0
w = 0
rho = 1
rho_sp = 1.

Progressive departures from persistence correspond to:

- increasing sigma_csi_squared (theoretically to infinity)
- increasing w (theoretically to infinity)
- decreasing rho (to -1)
- decreasing rho_sp (to -1).

When there are only 'small' departures from persistence (ie sigma_csi_squared close to 0, w close to 0, rho close to 1, rho_sp close to 1) fixed station sampling is likely to do well relative to random sampling. As the departures increase, the benefits of fixed station sampling decrease and the advantages of fixed vs random will be case specific.

Working Paper #2 extended Nicholson et al. (1991) to the case of sampling with partial replacement (SPR). Fixed and random sampling are the two extreme special cases of SPR. The bias, variance and mean square error were examined.

The Working Paper concentrated on the bias, variance and mean square error of an estimate of change in abundance. However, results were also presented which would allow the calculation of these quantities for an estimate of relative abundance. Relevant to this was a suggestion that extensive surveys be carried out at, perhaps, 10 year intervals, with fixed stations or SPR applied in the intervening years; i. e., a reasonably accurate base line would be established, abundance in the intervening years could be based on estimates of change, with a check made and if necessary the base line reestablished at appropriate intervals.

A second major point concerned the magnitude of the improvement obtained under fixed stations or SPR. The working paper presented tables of:

1) the probability of fixed stations being more accurate (i. e., lower mean square error) for estimating change than random sampling for selected values of the index w,

2) the probability that the confidence interval under fixed stations would be less than specified fractions of that obtained random sampling.

3) the probability of fixed stations being more accurate than SPR for selected values of w and the replacement fraction.

Obviously not all possibilities could be presented. The objective was to illustrate the approach and give a general idea of what might be expected, rather than give definitive results.

It should be emphasized that the illustrations in the Working Paper .assume that the fixed stations were chosen at random from the set of available stations and the results were given in terms of probabilities. There may well exist sets of fixed stations for which the estimates would be considerably worse or better than indicated by these probabilities. Further, in application, purposeful rather than random, selection of the fixed stations, might well be considered.

Working Paper #13 considered SPR for estimating a relative index of abundance, following Patterson (1950). The estimate is a weighted average of the means of the fixed and random stations, and involves the correlation between fixed stations over time (rho). The weights are a function of the relative variance of fixed and random samples. Cochran (1977) tabulates the gain in efficiency of this technique relative to random sampling for correlations greater than 0.7 and replacement fractions of 0.5 and 0.25, where the reduction in variance would be less than 20%.

3.3.4 Applications

Working Paper #2 presented values of w based on survey data from NAFO Division 2J for the years 1985-1990. The values ranged from 0.11 to 0.91 with an average of 0.42 with a tendency, albeit not without exceptions, for w to be smaller between adjacent years and to increase with time.

At the workshop, values of w were calculated for the English Groundfish Survey data for the years 1977-1981. These data generally had more than one observation for the same year/station combination (unlike the 2J data) permitting allowance of measurement error. The results are given both with and without such allowance for both 1 and 2 year old stock (Table 3.3.4). It will be noticed that w is much smaller than for the 2J data. Allowance for measurement error considerably decreases w for the 2 year old stock. That the persistence index will decrease on allowing for measurement error follows from theory. However, the reduction may be exaggerated in the present case. In contrast to the 2J data, where stations were within 2.5 nautical miles were regarded as fixed, the EGF survey stations covered 450 nautical miles squared, so that a greater degree of spatial variation would inflate the measurement error. This, along with the usual sampling error in estimating variance components, can account for the zero values of the index! The tabulated values thus provide upper and lower limits of w for these data.

The working group examined values of the correlations rho and rho_sp for cod from the southern Gulf of St. Lawrence survey (total numbers) and the English North Sea survey (age 1). The correlations from one occasion to the next for various lags are displayed in Figures 3.3.4.1 - 3.3.4.4. There were 12 fixed stations available for Gulf of St. Lawrence cod from 1971-88. The correlations were very variable, but seldom greater than 0.7. The correlations for North Sea cod were more consistent but only 1977- 81 data (37 fixed stations) were available for examination. The correlations were generally between 0.5 and 0.75. There was little difference between the values of rho and rho_sp.
Data for 61 fixed stations during 1984-87 were available for the Gulf of St. Lawrence cod stock. Values of rho for these were:

Year	1984	1985	1986
1985	.23		
1986	.31	.10	
1987	.46	.20	.62

Using the formulae given by Cochran (1963), the group computed the efficiency of Patterson's regression estimator (relative to random sampling) for estimating relative abundance for rho = 0.4 ... 0.8 - corresponding to the cod data - and for partial replacement fractions of 0.1 ... 0.5. The results are approximate for time series exceeding 6 occasions:

w	rho	Replacement fraction		
		0.3	0.3	0.5
0.43	0.4	2	4	5
0.25	0.6	5	11	13
0.11	0.8	13	29	33

For 4TVn cod, the gain in efficiency of fixed vs random would appear to be at most 10%. For the English Groundfish Survey, the gain in efficiency is likely to be greater.

It is important to note that the efficiency of SPR relative to random sampling for estimating change in abundance will be different.

3.3.5 Graphical methods

A variety of ways of graphically portraying departures from persistence were considered. One method was as follows. The stations were ranked according to mean catch over time. The yearly catches (adjusted for year effects) were then plotted against rank.

Figures 3.3.5.1 and 3.3.5.2 show these plots for the English Groundfish survey cod age 1 and for 12 fixed stations for Gulf of St. Lawrence cod from 1971-88.

If there is persistence, then the adjusted catches for each year will be coincident. In practice they will be perturbed by measurement error. If there are replicate trawls, e. g., the English Groundfish data, then it is possible to superimpose a confidence envelope as an aid to interpreting the interaction terms.

The 'trend' in the ranked mean catches measures the station effects and indicates the heterogeneity of the spatial distribution of the stock. The variability about the mean catch is a measure of the interaction effects. As w increases, the trend becomes less discernible.

Thus, for the English Groundfish survey, there is a reasonably clear trend. For the Canadian cod survey, the trend is less clear. These graphical results corroborate the values of w and rho given in the previous section.

Another use of these plots would be to identify particularly naughty stations which have high interaction terms.

3.3.6 Alternative assessment of persistence

[Due to scheduling, this material was not discussed by the workshop participants]

An alternative index of departures from persistence is based on a test of specific hypotheses concerning the changes in the spatial pattern of abundance. A technique developed by Myers and Stokes (1989) was applied to determine if the catch at fixed sampling stations from the Iceland groundfish trawl surveys changed proportionally throughout the range. That is, for each site, the log catch at that site was regressed against the log VPA estimates over the years from 1985 to 1992. If the population is responding uniformly throughout the range, then the slope should be approximately one for all stations, and there should be no geographical pattern to the slopes.

Figure 3.3.6.1 shows the distribution of slopes for age 4 cod around Iceland. The areas where there are large mean abundances, i. e., in the north appear to have the largest slopes.

In Figure 3.3.6.2, the median slope and median abundance have been calculated for each statistical square around Iceland. There is some evidence of a relationship between mean abundance and the slope.

It appears that for cod around Iceland there is a greater change in survey catch in regions with high mean abundance. That is, when the population increases it appears to increase proportionally more at stations in which the mean abundance is high. This could be because when population abundance is low, there is greater fishing mortality in the regions of high mean abundance. Myers and Stokes (1989) found that in the North Sea haddock responded uniformly throughout the range, whereas cod and whiting did not. Multiplicative models of research surveys should be robust to such changes if they occur.

3.3.7 Summary/discussion

The various approaches which have been considered in the evaluation of repeated sampling do essentially the same thing in different ways. The identity between the measures of non-persistence allows a unified approach to examination of the impact on precision of various sampling designs. The tools used to examine the degree of non-persistence were helpful in understanding the problems. It is recommended that these measures of non-persistence be examined for other available data to permit evaluation of the consequences of alternative survey designs.

Considering the objective function minimized in calibration of VPA led to the conclusion that an estimate of relative abundance was needed for each year. It seems that simple year-to-year differences (or ratios) in survey indices do not fully capture the essence of the problem, which ideally requires the minimization of the variance of apparent catchability about its mean. A biassed mean is acceptable provided the bias is constant. It would be of interest to determine whether the effect of persistence is the same for variations about the mean as for year-to-year differences.

The results suggested that the North Sea cod data might benefit more from a fixed station survey design than the other data examined. With respect to estimating the mean relative abundance index within a year, the expected maximum gain in efficiency for the above estimator for the optimal partial replacement level of 0.5 would be about 13% . The efficiency would be less if greater than 50% of stations were retained fixed and if a simple average were used instead of the efficient estimator. The gains in efficiency for the other stocks was smaller.

Though a fixed station design is not likely to result in greater variance, the potential bias within a year introduced by lack of persistence is not known to be constant, possibly introducing complications for VPA calibration. For stratified designs, the within strata sample sizes are small, creating difficulties in application of the efficient estimators. It is recommended that when fixed stations are used, the potential for variable bias be evaluated.

The implications for assessing changes in relative abundance from year to year were more straightforward. For 1-group cod from the English groundfish survey, there was general agreement amongst the methods that the degree of non-persistence was small, suggesting that fixed-station sampling would have an advantage over random sampling. This suggestion was less strong for the Gulf of St. Lawrence and area 2J cod data sets.

4 CONCLUSIONS AND RECOMMENDATIONS

The following conclusions and recommendations were made:

Pairwise/parallel trawling should be incorporated in the allocation of vessels by area for multi-vessel surveys, permitting comparison of fishing power as data sets are accumulated over the years.

Whenever possible, the same vessel should be retained over time. Before old vessels are replaced by new ones, calibration studies should be undertaken.

Maximize balance with respect to ship, area, day/night and other factors affecting catch rate when designing surveys. An extreme case of an unbalanced design occurs in the IYFS where only one vessel fished in Div. IIIa.

When applying GLMs, it is suggested that interaction effects be investigated, but be aware that their significance may be an artefact in unbalanced data. However, as pointed out by the Methods WG (Anon. 1989), interactions involving year should not be included if the year term is used to index abundance.

When applying GLMs to highly unbalanced data, it was noted that analyses of a restricted subset of the data, with an acceptable degree of balance, would help greatly in interpretation.

Spatially oriented methods can be applied to data from any survey design.

Spatially oriented methods can usefully be applied to describe (map) survey results and summarize (variogram) spatial structure.

The sophisticated methods examined did not appear to perform generally better than simple arithmetic or weighted averages when compared to VPA results. In the time available, it was not possible to conduct a more definitive investigation by calibrating with the index from each method.

Estimation of variance from the geostatistical methods can take into account the location of the data and the spatial structure. These results, however, are model based and invoke some assumptions.

Additional work is required to evaluate all components of variance for survey results (e. g., changes in catchability and/or availability, sampling error, etc.) and their use in the calibration of VPA should be investigated.

Fixed station designs may benefit more from purposeful selection of locations. Correlation between results from subset of stations and the results from an extensive survey may help to identify key locations.

The degree of persistence will likely vary from stock to stock and it is recommended that the various indicators, which have been identified (rho, omega, interaction sum of squares) should be applied to other cases to permit evaluation of the impact of fixed, partial replacement or random station survey designs.

The statistical properties, especially precision, for estimators of change in the survey abundance index relative to the mean over the whole time series should be investigated with respect to partial replacement strategies in survey design.

5 REFERENCES

Anon. 1990. Report of the Working Group on Methods of Fish Stock Assessment. ICES, Doc. C.M. 1990/Assess:15 95pp.

Anon. 1991. Report of the workshop on the application of spatial statistical techniques to acoustic survey data. ICES, Doc. C.M.1991/D:40.

Anon. 1992. Report of the International North Sea, Skagerrak, and Kattegat Bottom Trawl Survey. ICES, Doc. C.M. 1992/H:3.: 21pp.

Box, G.E.P. and G.M. Jenkins. 1976. Time series analysis, forecasting and control, Holden - Day. 575pp.

Bradu, D. and Y. Mundlak. 1970. Estimation in Log Normal Linear Models. Journal of the American Statistical Association, V 65, No. 329. pp 198-211.

Cochran, W.G. 1963. Sampling techniques. John Wiley.

Cochran, W.G. 1977. Sampling techniques. John Wiley & Sons, New York, NY. 428pp.

Gagnon, P. 1991. Optimisation des Campagnes d'echantillonnage: les programmes REGROUPE et PARTS. Rapp. Tech. Can. Sci. Halieut. Aquat. 18/8: 20pp.

Mosteller, F. and J.W. Tukey. 1977. Data analysis and regression. Addison-Wesley. 588pp.

Myers, R.A. and Stokes, T.K. 1989. Density-dependent habitat utilization of groundfish and the improvement of research surveys. ICES, Doc. C.M.1989/D: 15.

Nicholson, M.D., Stokes, T.K., and Thompson, A.B. 1991. The interaction between fish distribution, survey design and analysis. ICES, Doc. C.M.1991/D:11.

Patterson, H.D. 1950. Sampling on successive occasions with partial replacement on units. J. royal Stat. Soc. series B, 12:241-255.

Smith S.J. 1990. Use of Statistical Models for the Estimation of Abundance from Groundfish Trawl Survey Data. Canadian Journal of Fisheries and Aquatic Sciences. 47:894-903.

Sparholt, H. 1990. Using GLM analysis on the IYFS herring data for the North Sea. ICES, Doc. C.M.1990/H:6.

Stefánsson, G. 1991. Analysis of groundfish survey data: combining the GLM and delta approaches. ICES, Doc. C.M. 1991/D:9.

Table 2.2.2.1.1. Results of paired t-tests on mean logged differences in catch of age 1 herring, IYFS, Subarea 2.

Vessels	n	Mean Difference	Standard Error	t	Pr>|T|
tri-cir	7	-0.28	0.87	-0.33	0.76
tri-dan2	10	0.95	0.75	1.27	0.24
cir-dan2	48	-1.19	0.34	-3.47	0.001

Table 2.2.2.1.2. Conversion coefficients to convert CIROLANA catches to DANA equivelant and to convert DANA catches to CIROLANA equivilant age 1 herring, IYFS, subarea 2.

Conversion Coefficients	
CIROLANA to DANA	DANA to CIROLANA
0.27	3.10

Table 2.2.2.2.1. Results of GLM analysis on IYFS age 1 herring. Model is ln(catch age 1) = ship + year + depth + depth * year + ship * year.

Effect	df	Mean Square	Pr > F
ship	2	18.559	0.0041
year	10	16.595	0.0001
depth	1	85.491	0.0001
depth*year	10	19.372	0.0001
ship*year	17	8.074	0.0014
Error	336	3.324	

Table 2.2.2.2.2. Seventeen ship*year contrasts for IYFS age 1 herring data.

Contrast	Mean Square	Pr > F
c81-c90 vs t81-t90	0.441	0.716
c82-c90 vs t82-t90	15.135	0.034
c83-c90 vs t83-t90	0.214	0.800
c84-c90 vs t84-t90	0.060	0.893
c85-c90 vs t85-t90	1.743	0.469
c86-c90 vs t86-t90	0.026	0.930
c87-c90 vs t87-t90	0.031	0.923
c88-c90 vs t88-t90	0.003	0.977
c89-c90 vs t89-t90	0.531	0.690
d82-d90 vs t82-t90	8.795	0.105
d83-d90 vs t83-t90	0.121	0.849
d84-d90 vs t84-t90	0.107	0.858
d85-d90 vs t85-t90	8.339	0.114
d86-d90 vs t86-t90	6.150	0.175
d87-d90 vs t87-t90	1.204	0.548
d88-d90 vs t88-t90	1.795	0.463
d89-d90 vs t89-t90	3.178	0.329

Table 2.2.2.2.3. Results of GLM analysis excluding 1982 data, and without ship*year interaction terms. Vessel difference estimates are also given.

Effect	df	Mean Square	Pr > F
ship	2	26.238	0.0005
year	9	15.168	0.0001
depth	1	100.571	0.0001
depth*year	9	16.819	0.0001
Error	323	3.344	

Vessel Differences	Estimate	Standard Error	Pr> |T|
Tridens-Cirolana	0.961	0.248	0.0001
Tridens-Dana	0.672	0.276	0.0155
Cirolana-Dana	-0.289	0.278	0.2985

Table 2.2.3.1. Fishing power (arbitrary unit, LSMEANS) by vessel and year for 1 ring herring in the IYFS. The estimates are based on GLM models including year, day/night, depth, ship and rectangle as effects. Data in the GLM log transformed. Each value obtained in a separate GLM, with a vessel in a given year regarded as a separate vessel. Estimated fishing power of 2-ring herring (Division IIIa excluded) is shown for comparison.

YEAR	VESSEL														
	AND2	ARG	CIR	DAN2	ELD	EXP	GOS	ISIS	JHJ	SCO2	SOL	THA	TRI	TRI2	WAH2
1981	-	-	3.38	-	-	5.27	-	-	-	-	-	-	3.76	-	-
1982	-	-	4.42	3.01	-	3.67	-	-	-	-	-	3.18	4.21	-	-
1983	3.17	3.87	4.62	3.82	2.65	4.80	-	-	-	-	-	3.84	3.77	-	-
1984	3.95	4.38	4.18	4.40	4.08	2.99	-	2.69	-	-	-	3.65	3.89	-	-
1985	4.28	5.24	4.92	6.07	3.00	-	-	4.36	-	3.98	-	0.35	3.70	-	-
1986	3.96	5.58	3.48	4.27	3.80	-	-	4.49	-	4.19	-	2.38	3.93	-	-
1987	-	4.03	4.06	3.44	3.11	-	-	4.86	-	4.45	-	2.16	3.97	-	5.04
1988	-	5.30	3.70	4.58	3.63	-	-	-	-	4.36	-	1.27	4.79	-	3.01
1989	-	4.51	4.00	4.76	3.63	-	-	5.09	-	4.14	-	3.03	2.94	-	3.69
1990	-	4.70	3.49	4.30	2.52	-	-	4.31	-	4.04	-	3.35	4.57	-	4.00
1991	-	3.79	-	4.89	-	-	-	6.04	4.13	3.45	-	3.31	-	4.23	4.19
1992	-	4.06	-	4.32	-	-	3.47	-	-	3.47	3.57	4.13	-	4.38	3.47
All yrs.	3.86	4.43	3.95	4.28	3.32	4.01	3.47	4.47	4.13	3.96	3.57	2.73	3.86	4.24	4.02
All yrs. 2-ring	3.09	3.32	3.11	3.60	2.52	2.78	3.37	3.32	4.06	3.29	3.57	2.88	3.51	3.23	3.64

Table 2.2.3.2.2. Conversion factors by ship to catch rates of 1-ring herring in the IYFS based on Table 2.2.3.1. A fishing power of 4.00 was considered as the reference level. Conversion factor for night hauls is also given.

Ship			Fishing Power	Conversion Factor
AND2	Anton Dohrn		3.86	1.15
ARG	Argos		4.43	0.65
CIR	Cirolana		3.95	1.05
DAN2	Dana		4.28	0.76
ELD	Eldjärn		3.32	1.97
EXP	Explorer		4.01	0.99
GOS	G.O. Sars		3.47	1.70
ISIS	Isis		4.47	0.63
JHJ	Johan Hjort		4.13	0.88
SCO2	Scotia		3.96	1.04
SOL	Solea		3.57	1.53
THA	Thalassa	1982-84	3.56	1.55
		1985-88	1.54	11.70
		1989-91	3.23	2.16
		1992	4.13	0.88
TRI	Tridens	(Old)	3.86	1.15
TRI2	Tridens	(New)	4.24	0.79
WAH2	Walter Herwig		4.02	0.98
Night				1.68

Table 2.2.3.2.1. Fraction of areas of depths between 10-200m in rectangles in ICES Division IIIa.

Rectangle	Area Fraction
46F9	0.1
46G0	0.7
45F9	0.1
45G0	0.3
45G1	0.3
44F7	0.1
44F8	0.4
44F9	0.8
44G0	0.7
44G1	0.6
43F8	0.9
43F9	0.3
43G0	0.3
43G1	0.8
43G2	0.1
42G0	0.3
42G1	0.9
42G2	0.5
41G0	0.3
41G1	1.0
41G2	0.6

Table 2.2.3.3. Means of means by rectangle of catch rates of 1-ring herring in the IYFS. North Sea + Division IIIa. Catch rates corrected for ship effects and day/night effects on an individual haul basis. Catch rates in Division IIIa are multiplied by area fractions given in Table 2.2.3.2.1. In obtaining the index for Division IIIa, the average catch rates have been multiplied by 0.15 in order to make them comparable to North Sea indices.

YEAR	N. Sea + Division IIIa		N. Sea	Division IIIa	Percentage of Total in IIIa
	$\overline{X}$	S.D.	$\overline{X}$	$\overline{X}$	
1981	691	1844	691	-	-
1982	799	1903	806	- *	-
1983	1060	2247	990	269	21
1984	1279	3328	1175	356	23
1985	1797	3602	1657	504	23
1986	2578	6263	2198	942	30
1987	4208	8870	3743	1302	26
1988	3104	8659	1912	2421	56
1989	1781	4770	1635	471	22
1990	956	3727	951	150	14
1991	1261	3377	1304	129	9
1992	953	2113	916	190	17

* Only one hawl

Table 2.5.1.1. IYFS Herring: Indices available

std : Standard index

lw-arith : locally weighted arithmetic mean

cont : interpolation based on contouring algorithm (see Anon 1990)

lw-rob : locally weighted robust estimator

krig : kriging estimate (arithmetic mean of all observations

GLM : Year effects from GLM (see Sec. 2.2.3.2)

kr-sw : kriging (mean) estimate for data in SW sector

kr-se : ditto for SE sector

her-fpc : standard index corrected for fishing power

GLM-D : GLM estimate using Dana data only

GLM-T : ditto, using Thalassa data only

Mult : multiplicative model based on several age groups.

Table 2.5.1.2. Comparisons of IYFS herring indices with age 1 recruitment from VPA. Key to the indices is given in Table 2.5.1.1.

year	VPA	std	lw-arith	dev	cv	cont	dev	cv	lw-rob	dev	cv	krig	se	cv	GLM	kr-sw	kr-se
81	5420	551	24800	6742	0.27	7832			13220	4394	0.33	642	136	0.21	691	786.5	639.1
82	8560	1293	38770	9032	0.23	21730			18560	4998	0.27	1019	185	0.18	799	826.1	1329.1
83	16980	1797	62320	13780	0.22	24790			25520	4786	0.19	1338	197	0.15	1060	1986.9	1822.5
84	15340	2663	74000	21340	0.29	28820			27400	7369	0.27	1565	232	0.15	1279	142.5	2120
85	15820	3416	97850	24820	0.25	29460			39630	9192	0.23	2332	324	0.14	1797	1228.4	2991.4
86	27590	3667	110400	30000	0.27	61320			43100	11110	0.26	3717	563	0.15	2578	1391.4	4049
87	33520	5717	193300	39690	0.21	108300			88650	16370	0.18	4355	618	0.14	4208	3267.5	4859.1
88	27830	4192	94100	22040	0.23	49720			48670	10610	0.22	5466	1050	0.19	3104	104.5	3893.7
89	15290	3468	85100	23280	0.27	26360			29070	5458	0.19	3058	722	0.24	1781	645	3510
90	17680	2146	82930	32900	0.40	11890			18270	5047	0.28	1454	461	0.32	956	442.1	2436.3
91	11940	2433	56190	14450	0.26	23100			28190	6629	0.24	1376	227	0.16	1261	203.4	2302.4
92		2098	65900	15270	0.23	31180			28970	6140	0.21	1406	217	0.15	953	858.9	1729.5

Logarithms & regressions

year	VPA	std	lw-arith		cont		lw-rob		krig		GLM		kr-sw		kr-se	
	8.60	6.31	10.12		8.97		9.49		6.46		6.54		6.67		6.46	
	9.05	7.16	10.57		9.99		9.83		6.93		6.68		6.72		7.19	
	9.74	7.49	11.04		10.12		10.15		7.20		6.97		7.59		7.51	
	9.64	7.89	11.21		10.27		10.22		7.36		7.15		4.96		7.66	
	9.67	8.14	11.49		10.29		10.59		7.75		7.49		7.11		8.00	
	10.23	8.21	11.61		11.02		10.67		8.22		7.85		7.24		8.31	
	10.42	8.65	12.17		11.59		11.39		8.38		8.34		8.09		8.49	
	10.23	8.34	11.45		10.81		10.79		8.61		8.04		4.65		8.27	
	9.63	8.15	11.35		10.18		10.28		8.03		7.48		6.47		8.16	
	9.78	7.67	11.33		9.38		9.81		7.28		6.86		6.09		7.80	
	9.39	7.80	10.94		10.05		10.25		7.23		7.14		5.32		7.74	
slope	1.10	-2.87	0.97	1.86	1.16	-0.91	0.86	2.02	1.12	-3.28	0.95	-1.87	0.24	4.17	1.02	-2.04
se	0.17	1.64	0.12	1.12	0.24	2.36	0.17	1.64	0.18	1.78	0.18	1.70	0.69	6.64	0.14	1.34
r**2	0.82	0.28	0.89	0.19	0.71	0.41	0.74	0.28	0.81	0.31	0.77	0.29	0.01	1.15	0.86	0.23
F	42.25	9.00	69.71	9.00	22.37	9.00	25.62	9.00	37.29	9.00	29.40	9.00	0.12	9.00	53.51	9.00
ss	3.43	0.73	2.63	0.34	3.76	1.61	2.07	0.73	3.55	0.86	2.54	0.78	0.16	11.91	2.90	0.49
cv		0.28		0.19		0.41		0.28		0.31		0.29		1.15		0.23

Table 2.5.1.2. (continued).

her-fpc	GLM-D	GLM-T	mult
691			33
799	5.06	3.61	37
1060	5.29	4.24	64
1279	4.76	3.13	76
1797	4.33	3.27	53
2578	5.46	2.88	102
4208	5.9	3.36	175
3104	5.72	3.12	95
1781	5.7	2.98	75
956	4.7	2.4	37
1261	5.45	4.02	27
953	5.77	4.42	36

her-fpc	GLM-D	GLM-T	mult
6.54			3.50
6.68	1.62	1.28	3.62
6.97	1.67	1.44	4.16
7.15	1.56	1.14	4.33
7.49	1.47	1.18	3.98
7.85	1.70	1.06	4.62
8.34	1.77	1.21	5.16
8.04	1.74	1.14	4.55
7.48	1.74	1.09	4.32
6.86	1.55	0.88	3.61
7.14	1.70	1.39	3.31

0.95	-1.87	0.87	-4.28
0.18	1.70	0.21	1.99
0.77	0.29	0.66	0.35
29.40	9.00	17.72	9.00
2.54	0.78	2.11	1.07
	0.29		0.35

Table 2.5.2.1. Indices Available: Icelandic Cod

AM	:	Arithmetic Mean
SM	:	Stratified arithmetic mean
G-B	:	Gamma-Bernoulli index
lw-arith	:	See Table 2.5.1
lw-rob	:	ditto
cont	:	ditto
kr (mean)	:	kriging (mean) estimate

Table 2.5.2.2. Comparisons of Icelandic cod groundfish survey indices with age 4 recruitment from VPA. Key to the indices is given in Table 2.5.2.1.

Year	VPA	AM	SM	G-B	lw-arith	cont	lw-rob	kr(mean)
85	108	1218	709	2006	1038	527	326	968
86	112	560	277	960	308	153	106	448
87	254	2090	1007	3251	308	153	106	1669
88	233	3081	1346	4306	.1366	466	323	2571
89	137	1532	906	2149	1166	517	274	1426
90	68	289	144	597	362	183	122	234
91	109	640	359	1387	190	82	57	556
92	96	528			294	105	74	469

Logarithms and regressions

Year	VPA	AM		SM		G-B		lw-arith		cont		lw-rob		kr(mean)	
85	4.68	7.10		6.56		7.60		6.95		6.27		5.78		6.88	
86	4.72	6.31		5.62		6.86		5.73		5.03		4.66		6.10	
87	5.54	7.64		6.91		8.09		5.73		5.03		4.66		7.42	
88	5.45	8.03		7.20		8.37		7.22		6.14		5.78		7.85	
89	4.92	7.33		6.81		7.67		7.06		6.25		5.61		7.26	
90	4.22	5.67		4.97		6.39		5.89		5.21		4.80		5.46	
91	4.69	6.46		5.88		7.23		5.25		4.41		4.06		6.32	
92															
		1.62	-1.00	1.61	-1.09	1.35	0.84	0.47	3.94	0.32	3.92	0.35	3.35	1.62	-1.19
		0.33	1.62	0.39	1.90	0.27	1.31	0.73	3.59	0.70	3.42	0.63	3.11	0.35	1.72
		0.83	0.38	0.76	0.44	0.84	0.30	0.08	0.83	0.04	0.79	0.06	0.72	0.81	0.40
		24.15	5.00	15.23	5.00	25.83	5.00	0.42	5.00	0.21	5.00	0.30	5.00	21.39	5.00
		3.42	0.71	2.96	0.97	2.38	0.46	0.29	3.46	0.13	3.15	0.16	2.60	3.42	0.80
			0.38		0.44		0.30		0.83		0.79		0.72		0.40

Table 3.3.4. Values of w (mean interaction sums of squares/mean station sums of squares) for 1 year old cod from the English Groundfish Survey. The upper line ignores measurement error; the lower line allows for measurement error.

	1978	1979	1980	1981
1977	0.35 0.23	0.21 0.10	0.18 0.10	0.50 0.34
1978	- -	0.24 0.11	0.20 0.12	0.32 0.09
1979	- -	- -	0.08 0.01	0.37 0.17
1980	- -	- -	- -	0.42 0.32

Values of w for 2 year old cod from the English Groundfish Survey.

The upper line ignores measurement error; the lower line allows for measurement error

	1978	1979	1980	1981
1977	0.39 0.09	0.27 0.00	0.34 0.10	0.27 0.05
1978	- -	0.22 0.00	0.44 0.16	0.22 0.00
1979	- -	- -	0.26 0.00	0.28 0.05
1980	- -	- -	- -	0.47 0.30

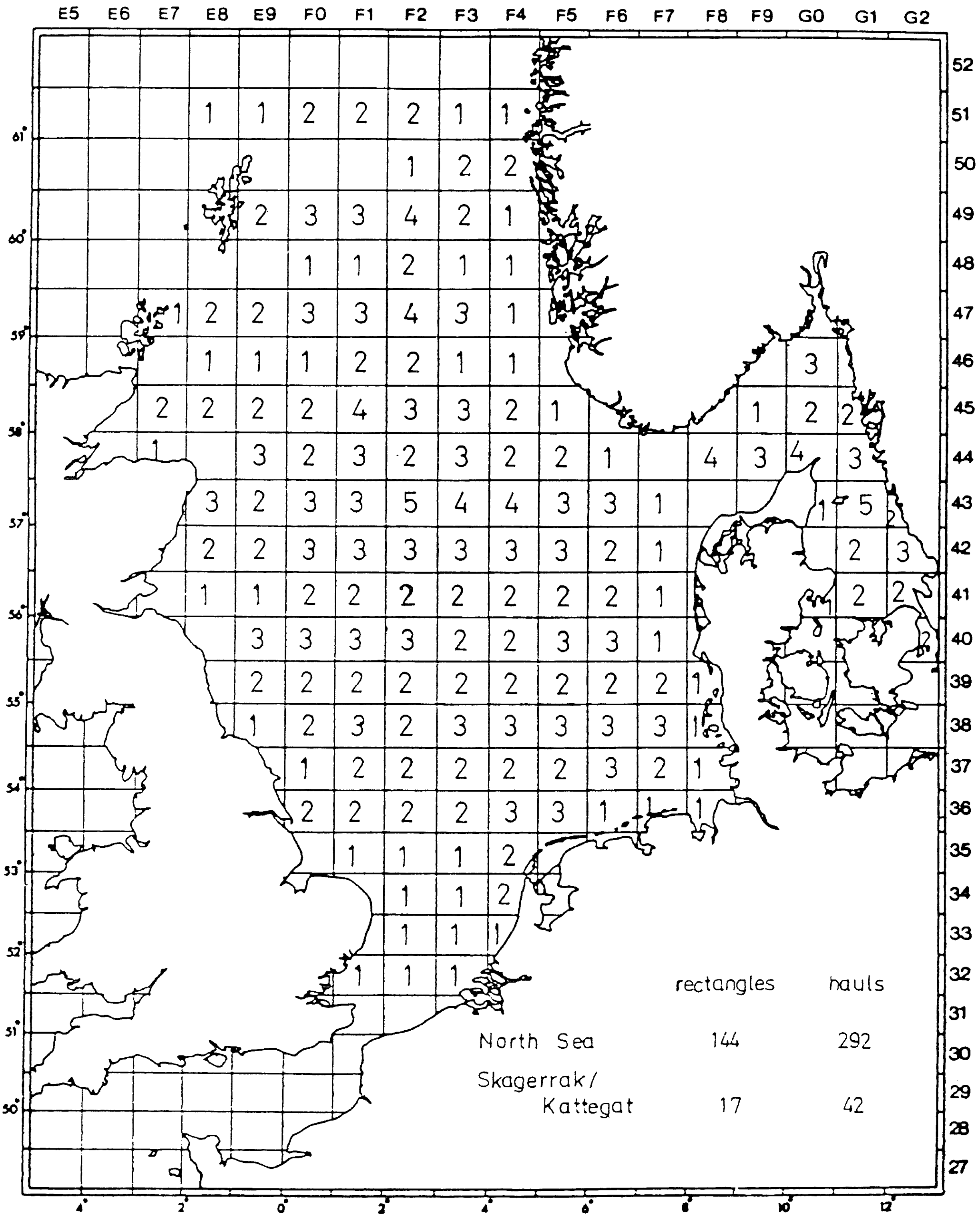
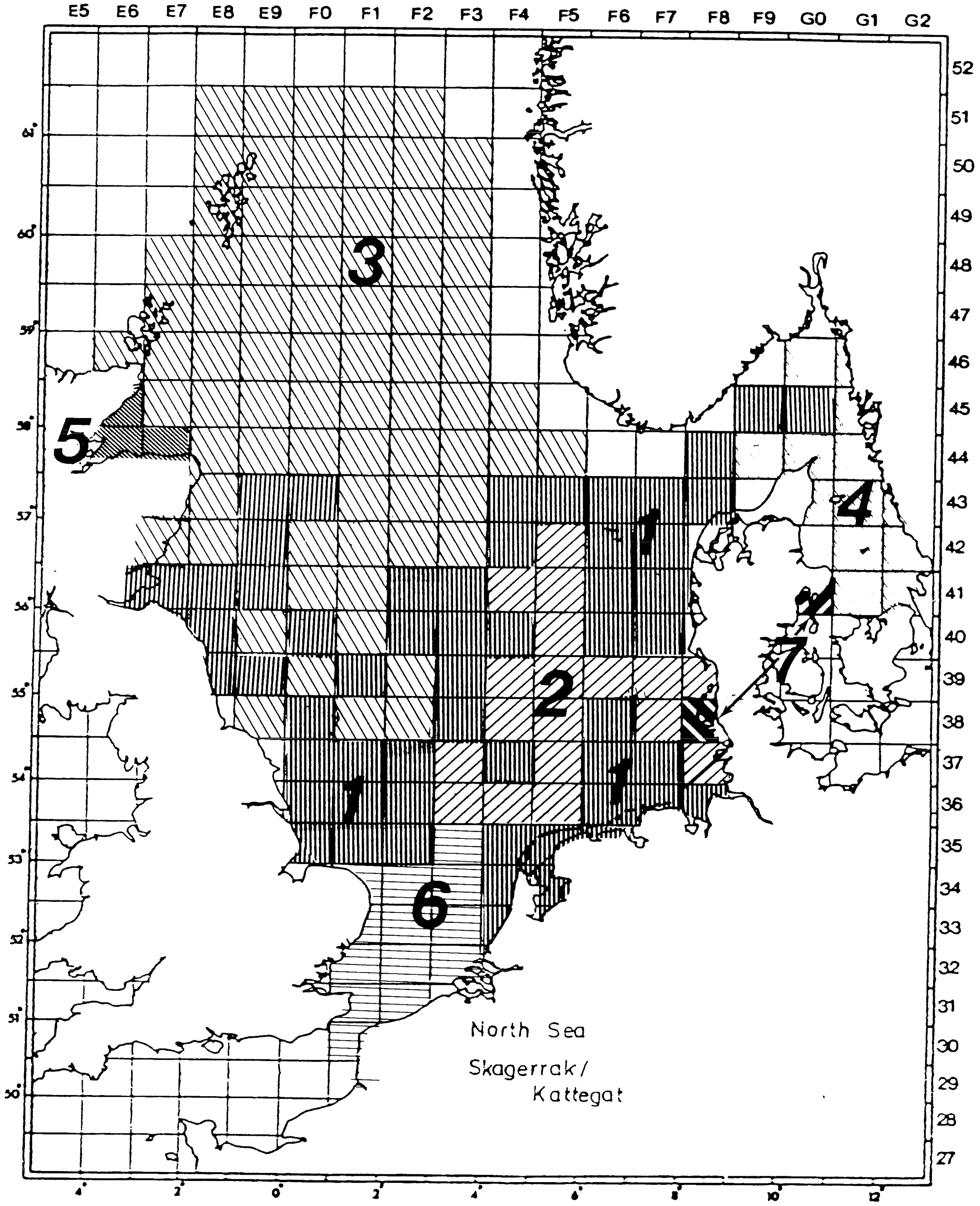


Figure 2.1.1.1. Allocation of haul locations for the International Young Fish Survey (IYFS) in the second quarter of 1991.

Figure 2.1.1.2. Groupings of IYFS rectangles used for herring assessments.



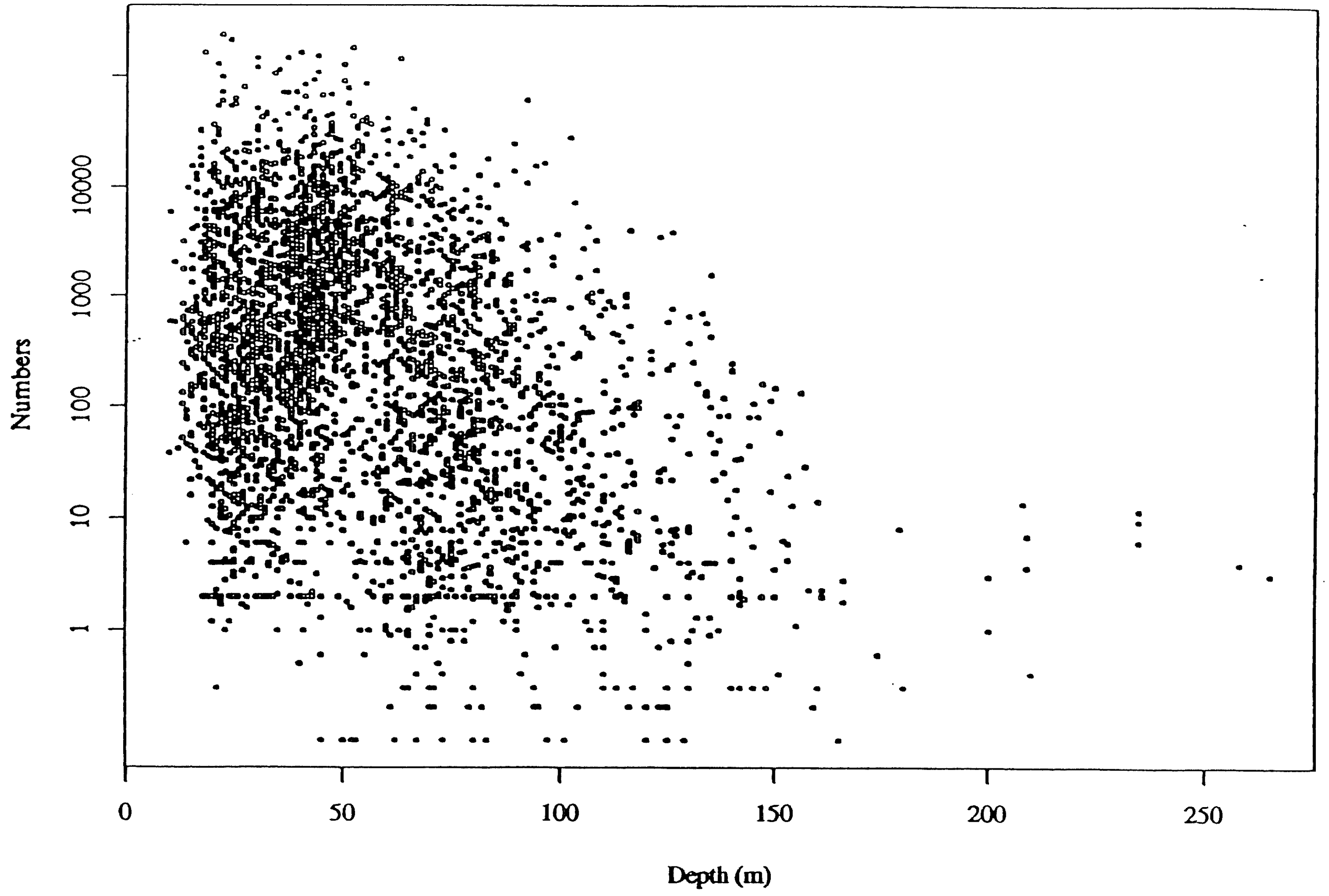


Figure 2.2.1.1. Catch (numbers) of age 1 herring by depth in the IYFS area, 1981-1992.

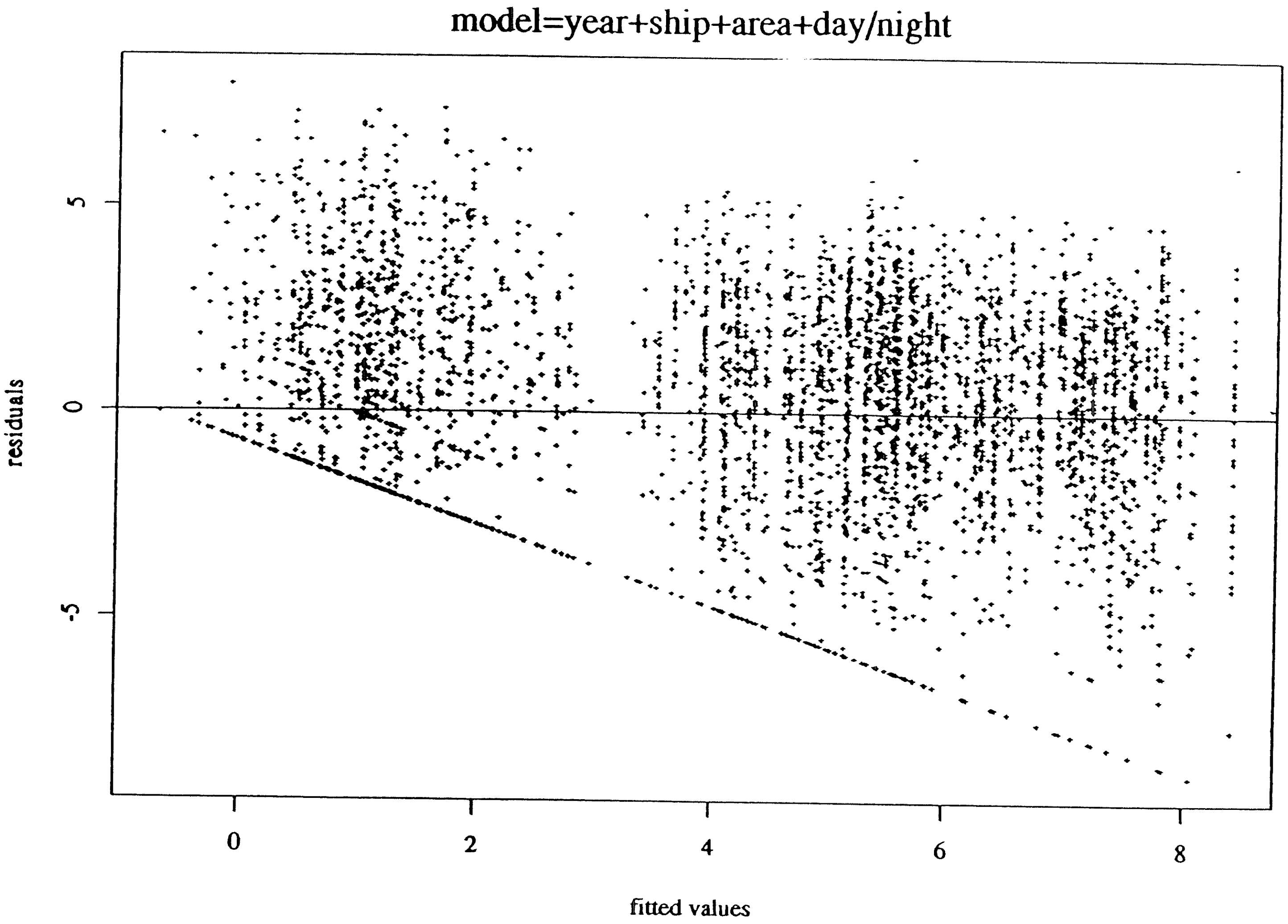

model=year+ship+area+day/night

Figure 2.2.1.2. Residuals about a linear model fit for age 1 herring catch data for the North Sea. The model is specified above the figure.

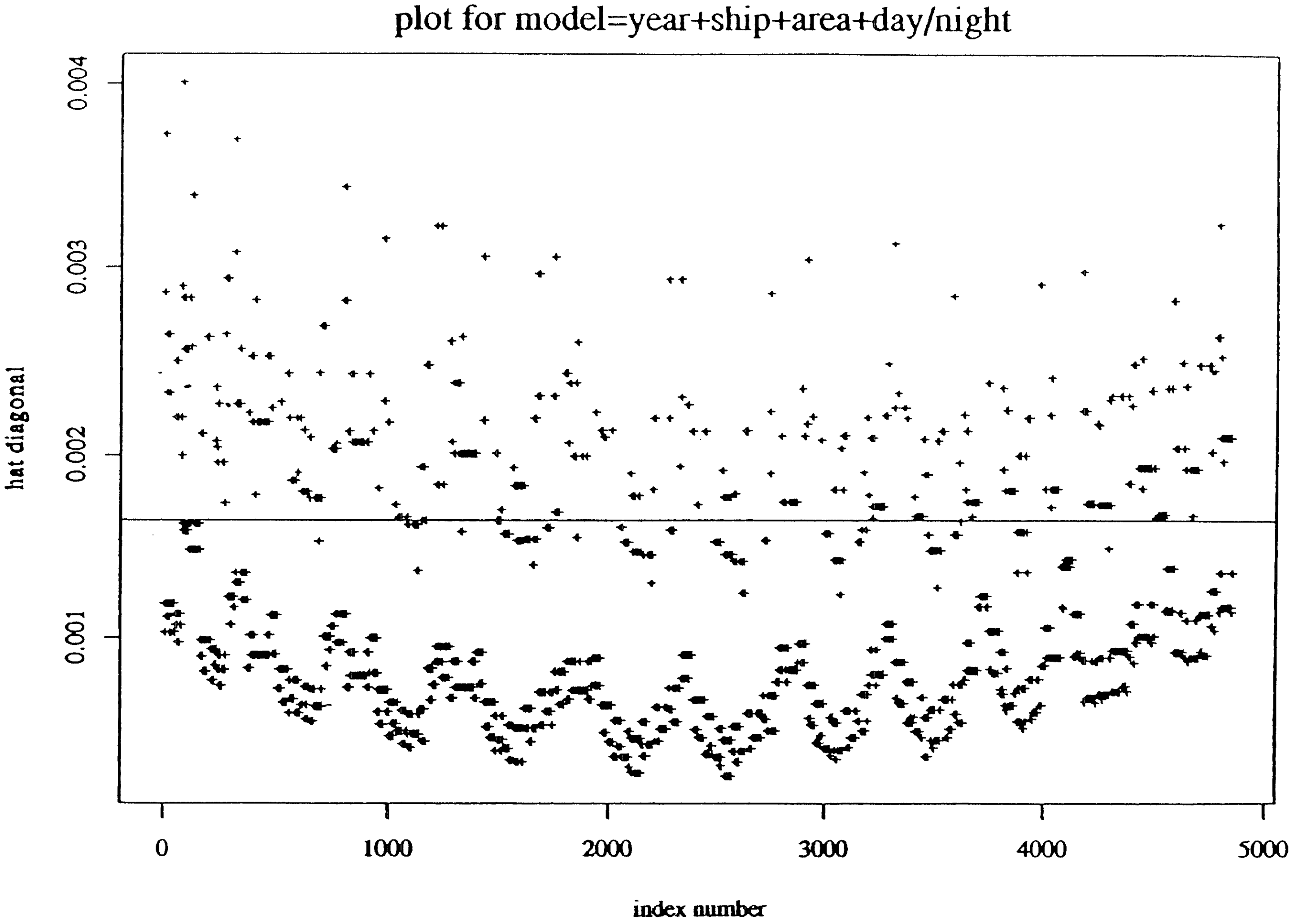


Figure 2.2.1.3. Leverage plot for the influence of extreme data points on the linear model fit of age 1 herring data for the North Sea, given in Figure 2.2.1.2. The horizontal line across the plot marks the cutoff between high and low leverage points.

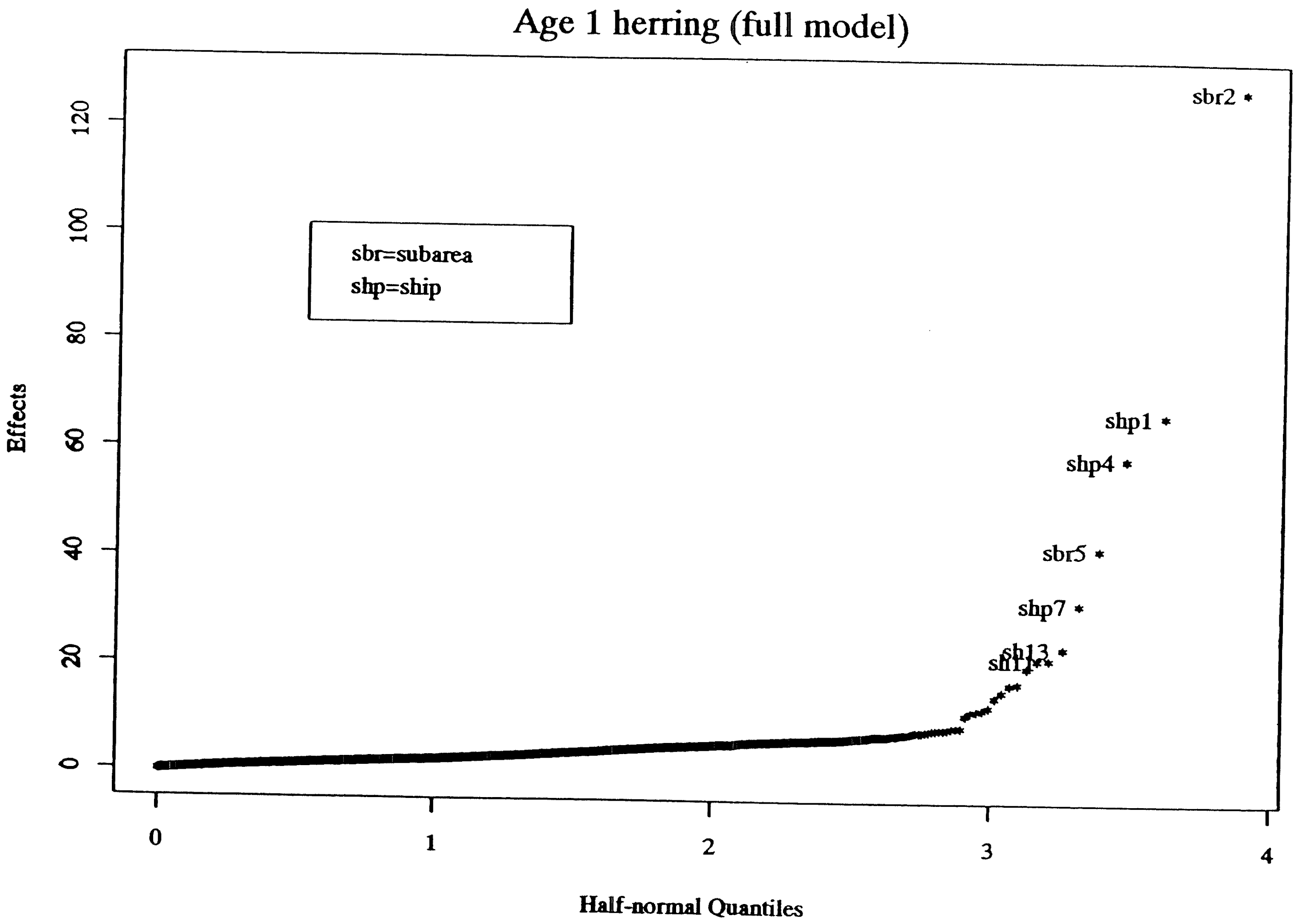


Figure 2.2.1.4. Half-normal plot of mean effects from the linear model of age 1 herring catch for the North Sea.

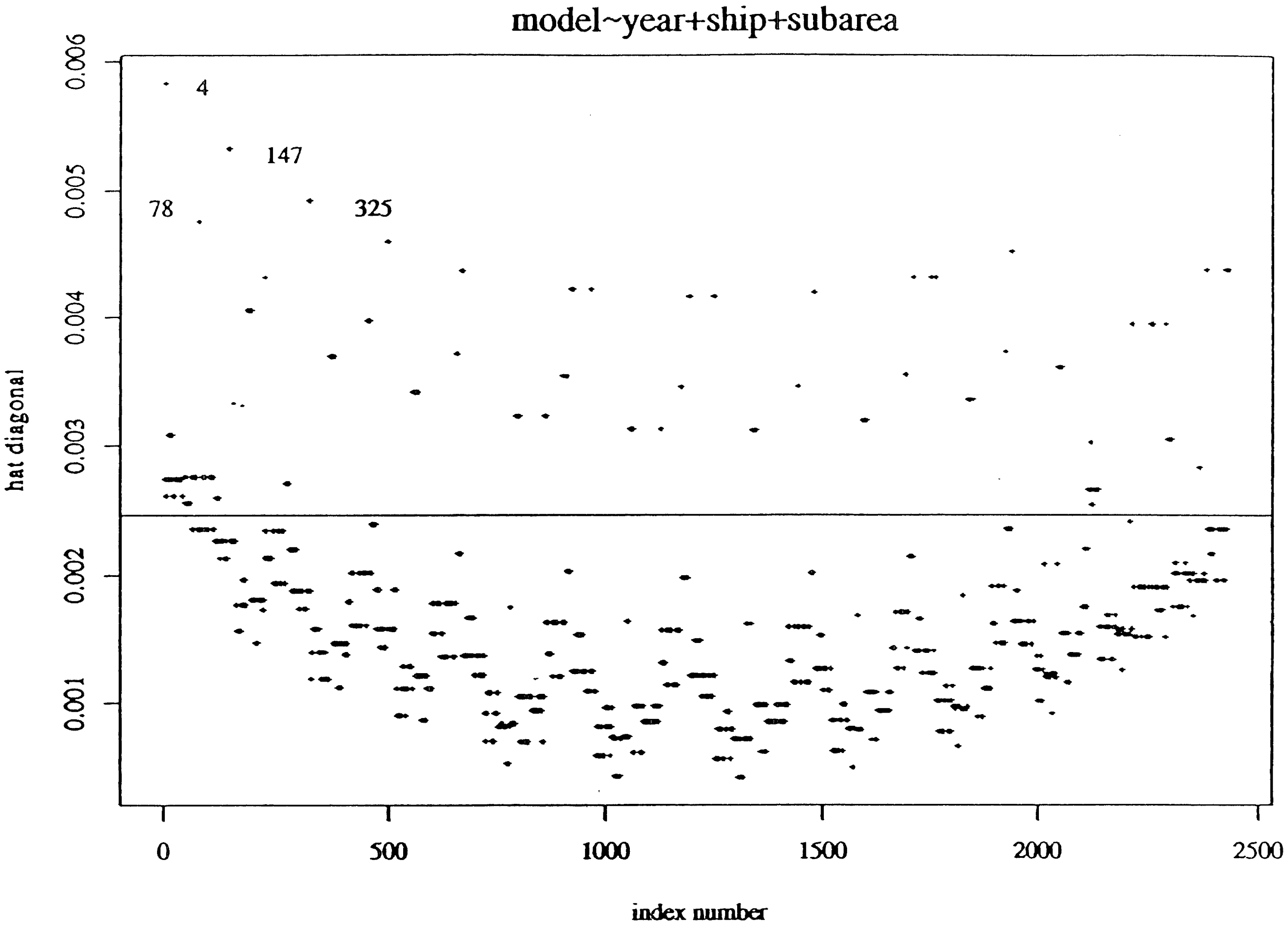


Figure 2.2.1.5. Leverage plot for the influence of extreme data points from a reduced model of age 1 herring catch in the North Sea, using only daytime tows, sub-areas 1-3, and six research vessels.

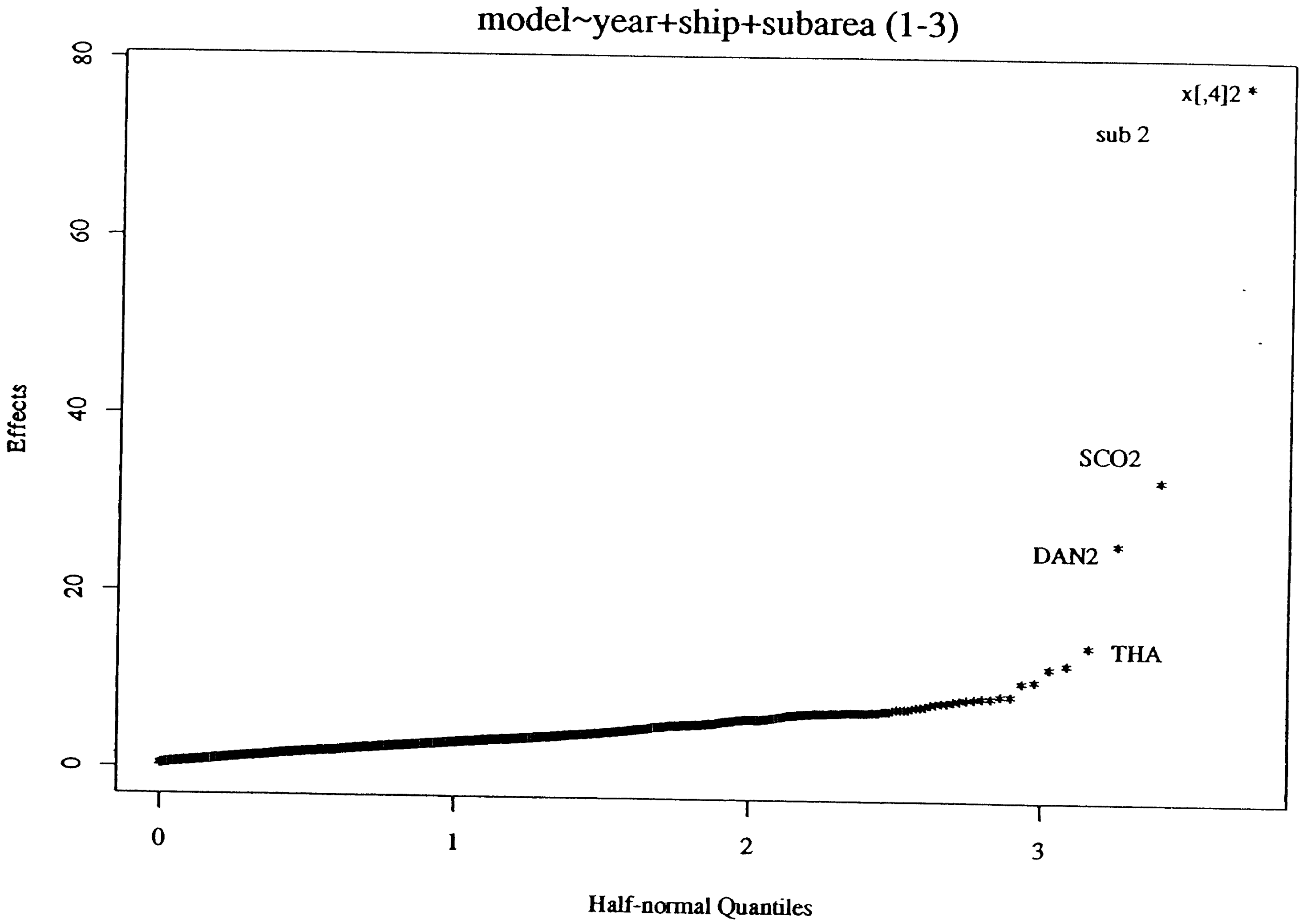


Figure 2.2.1.6. Half-normal plot of mean effects from the reduced linear model of age 1 herring catch in the North Sea given in Figure 2.2.1.5.

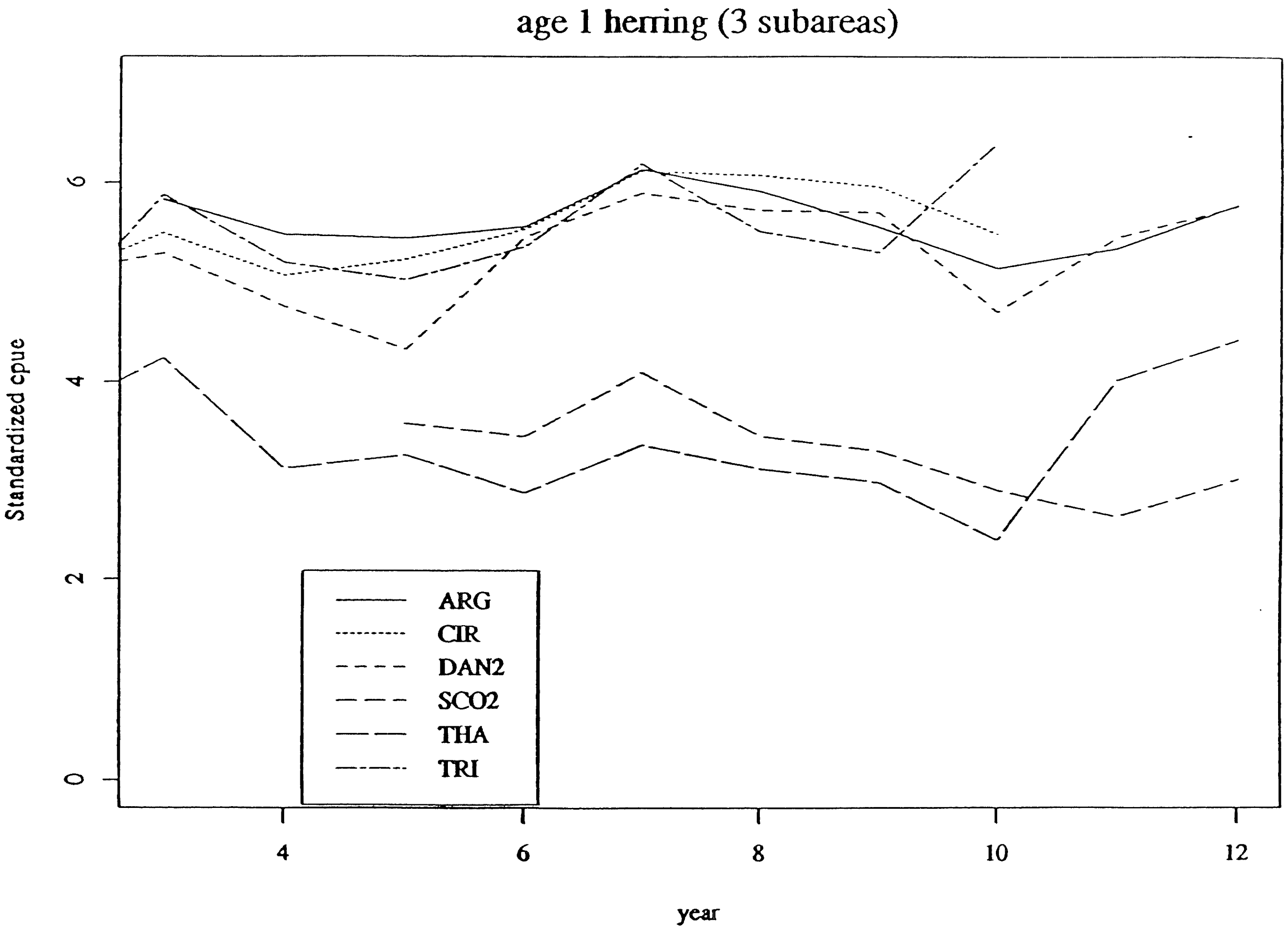


Figure 2.2.1.7. Predicted catch rates of age 1 herring for six research vessels in the North Sea.

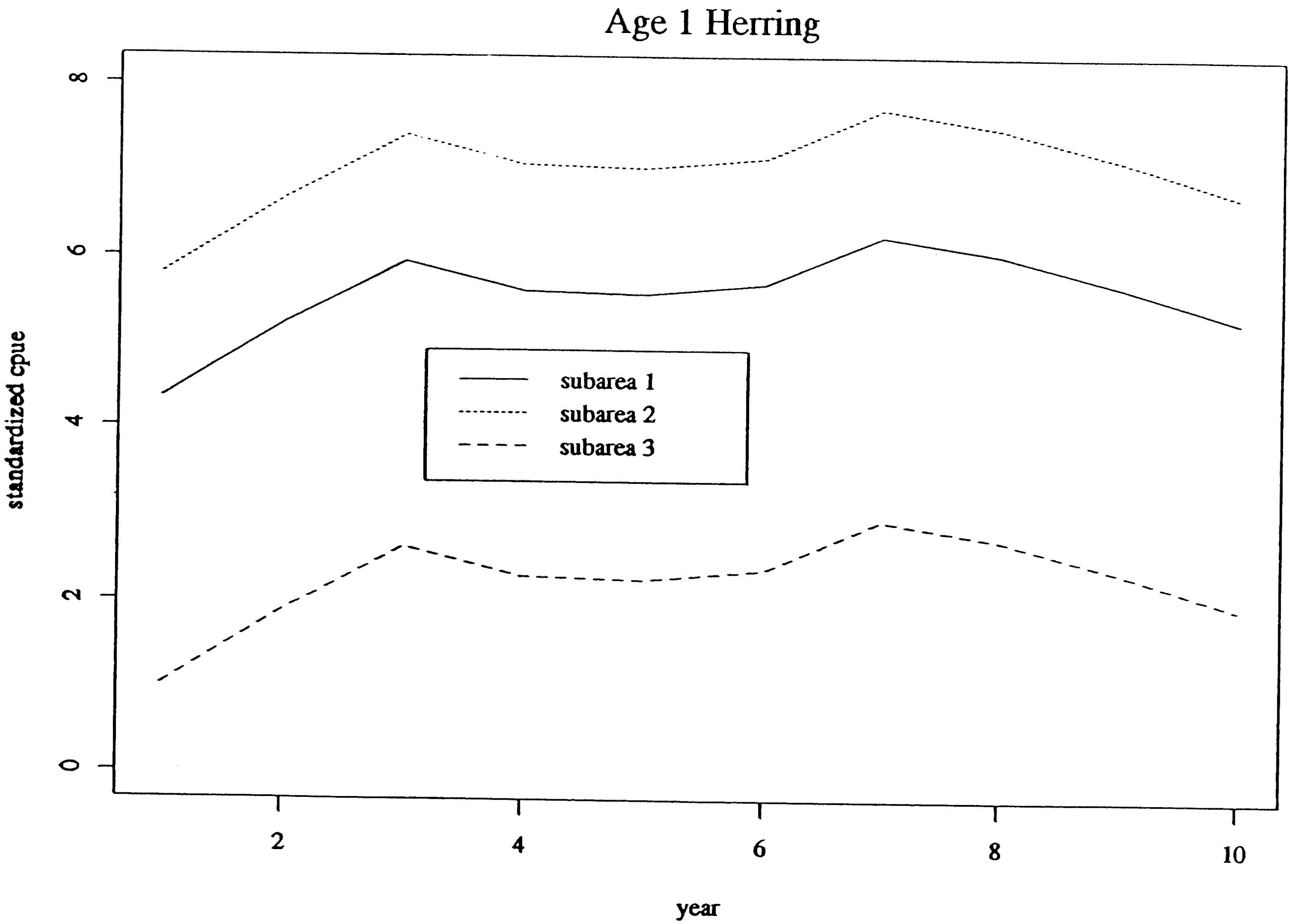


Figure 2.2.1.8. Predicted catch rate from vessel CIROLANA for age 1 herring in the North Sea, by sub-division.

Figure 2.2.1.9. Latitude/longitude locations of IYFS trawl hauls, by research vessel.

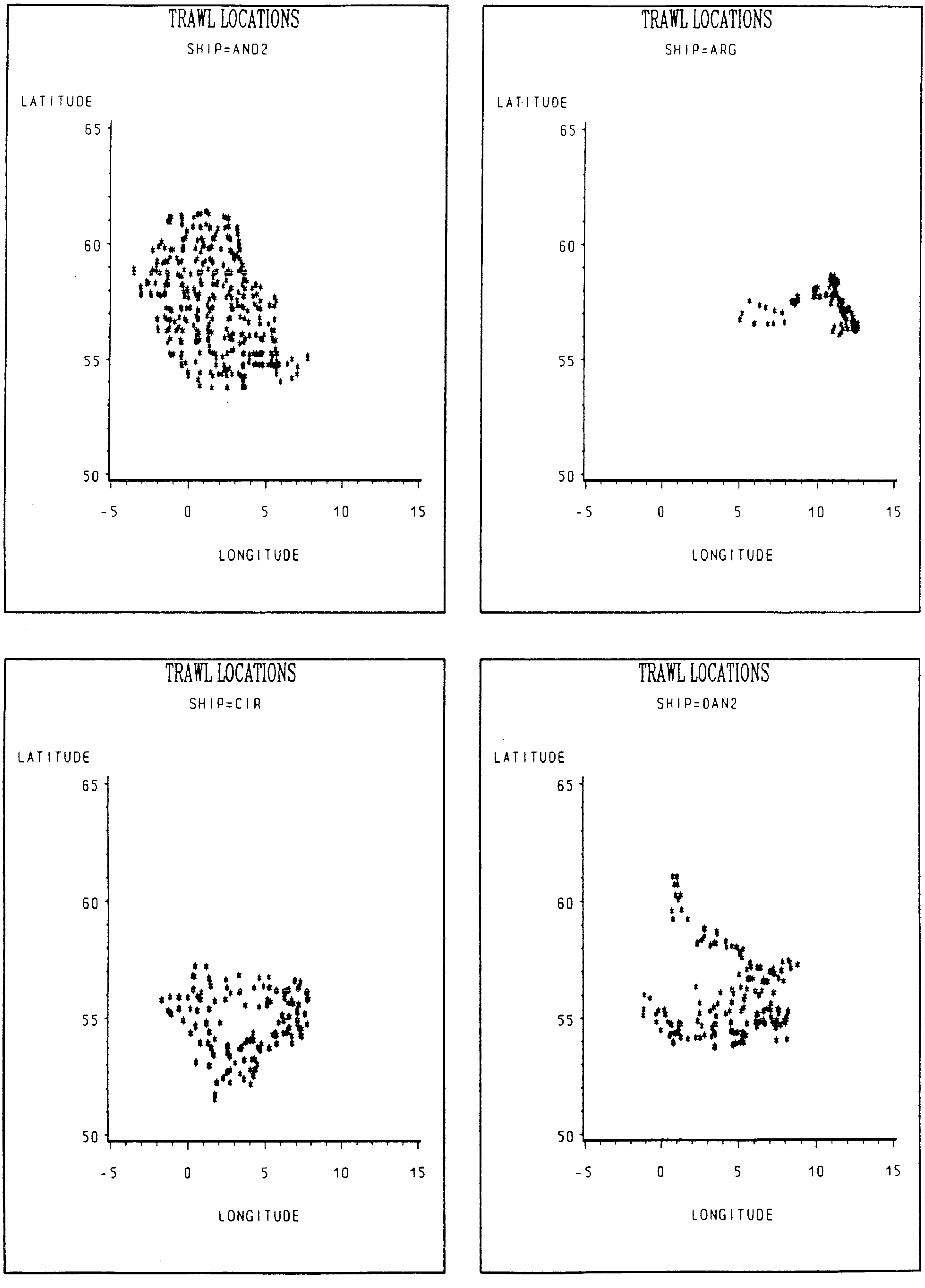


Figure 2.2.1.9. (continued).

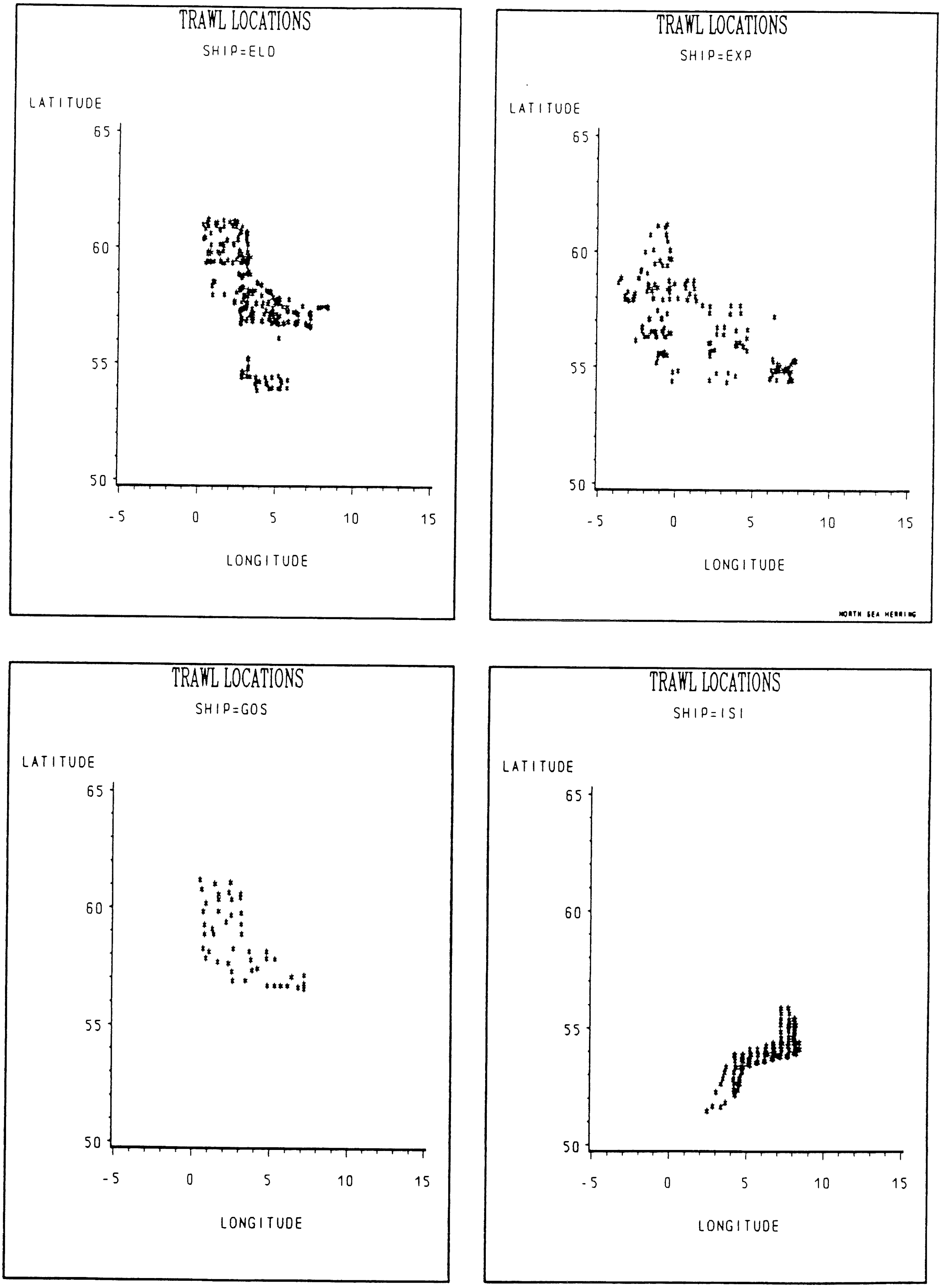


Figure 2.2.1.9. (continued).

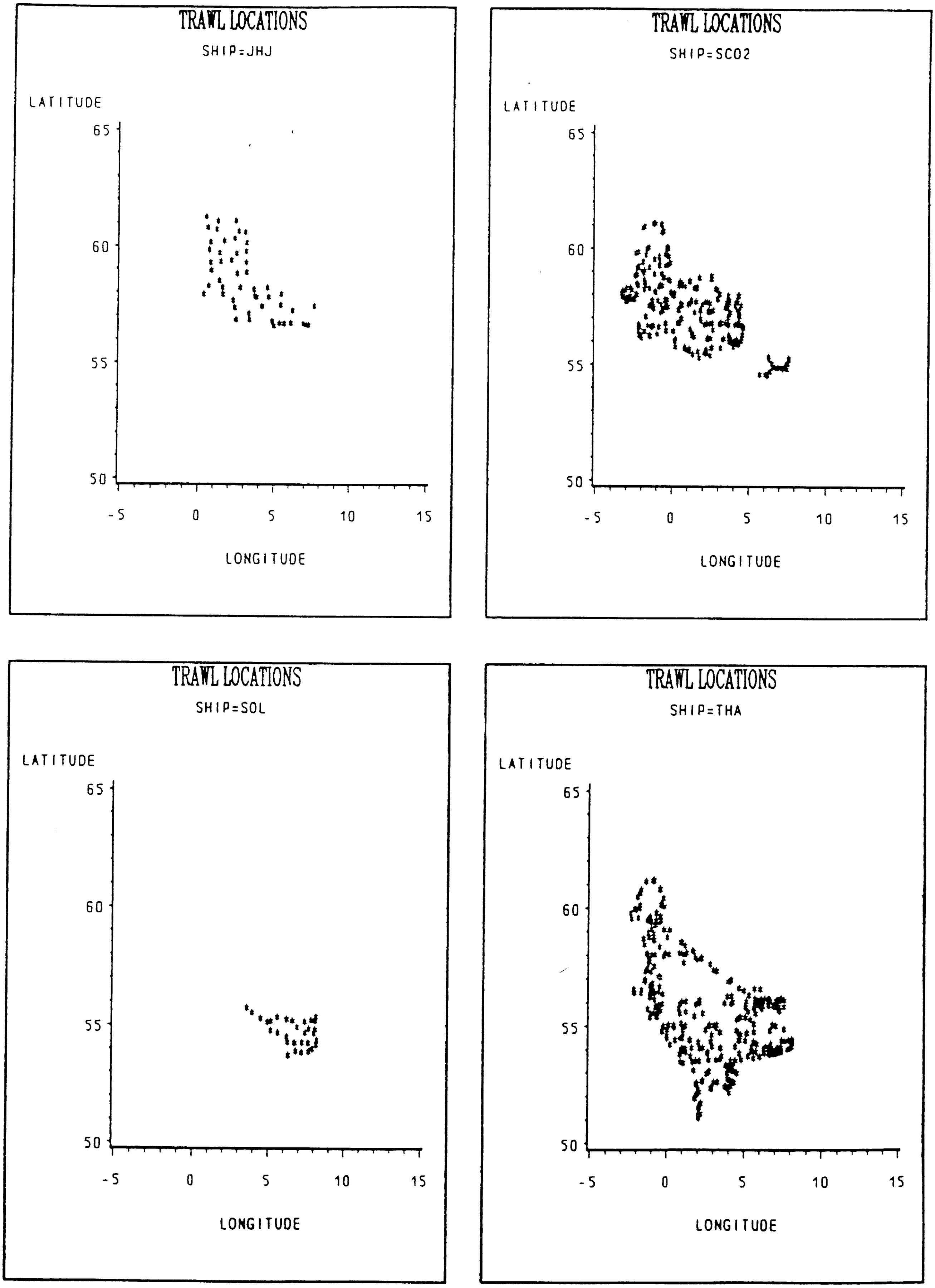
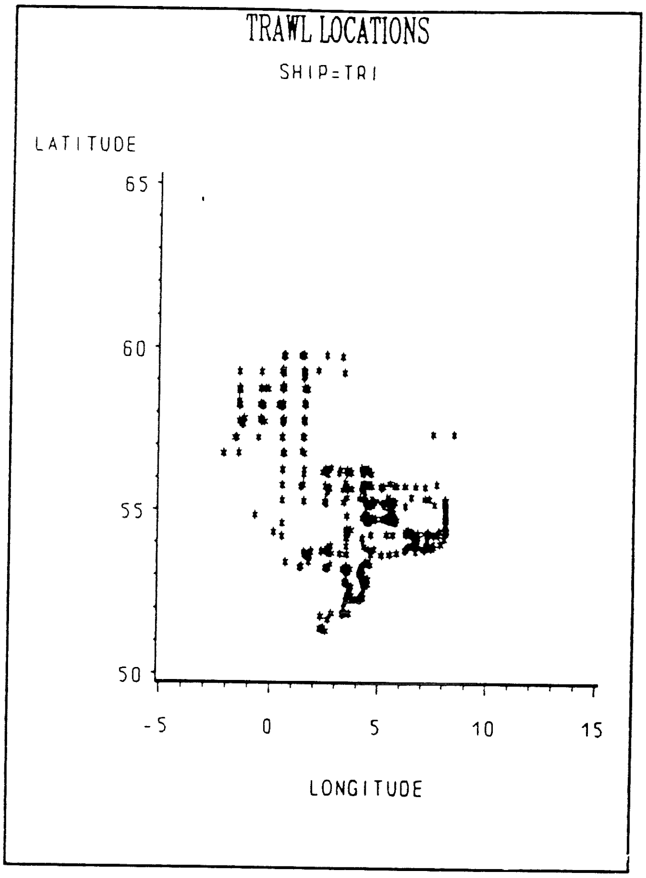
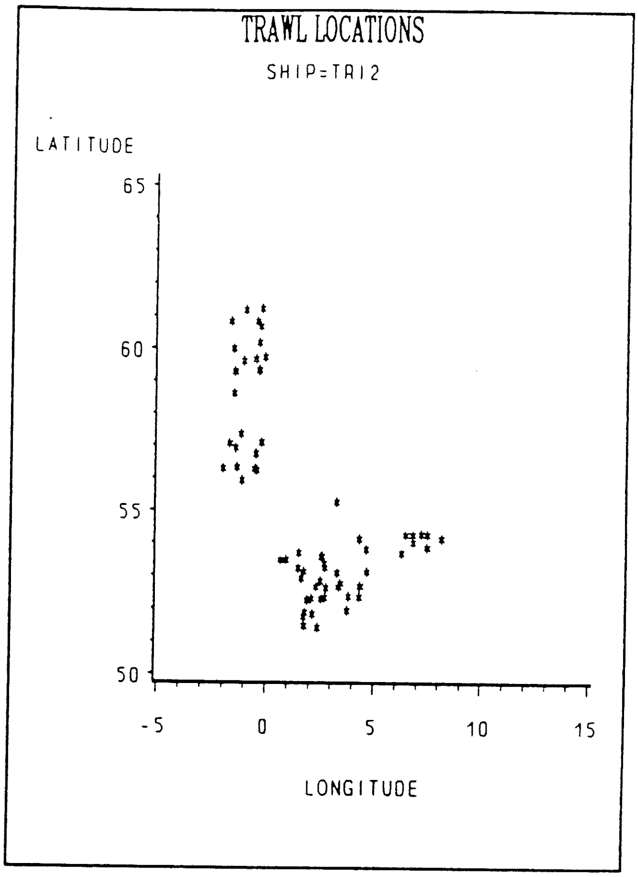
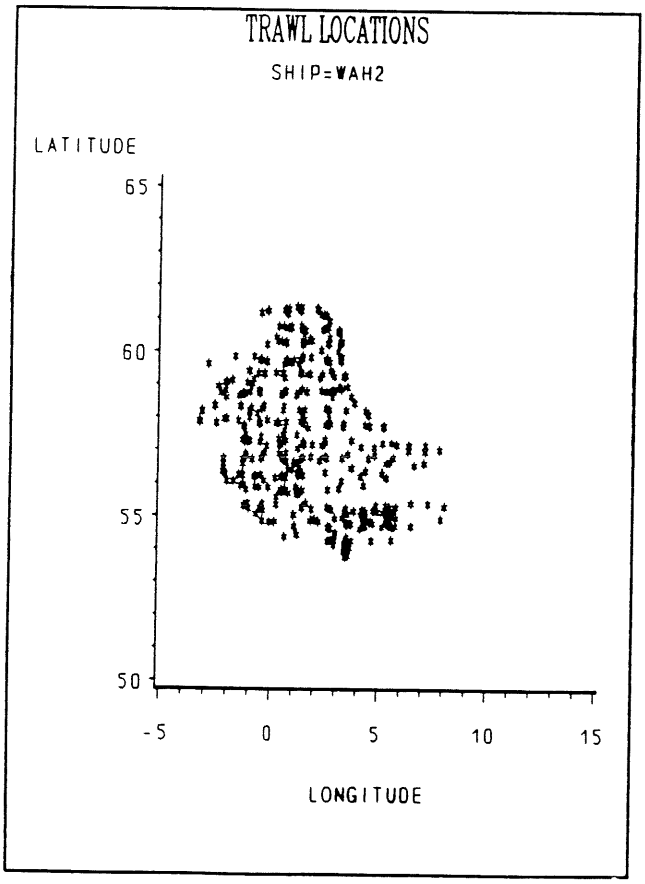


Figure 2.2.1.9. (continued).







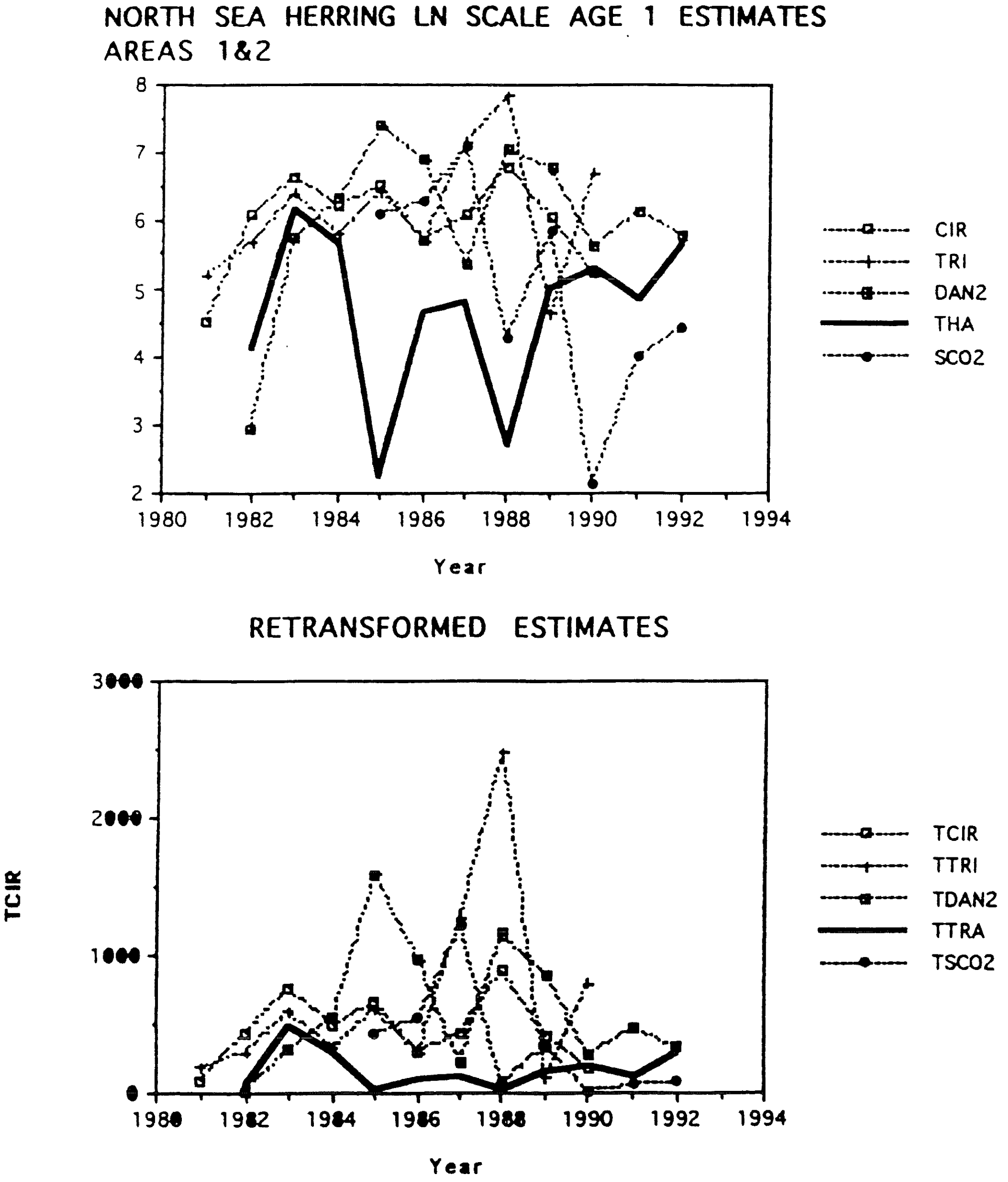


Figure 2.2.1.10. Year effects by research vessel for age 1 herring catches in the North Sea. Top figure ln scale, bottom is re-transformed.

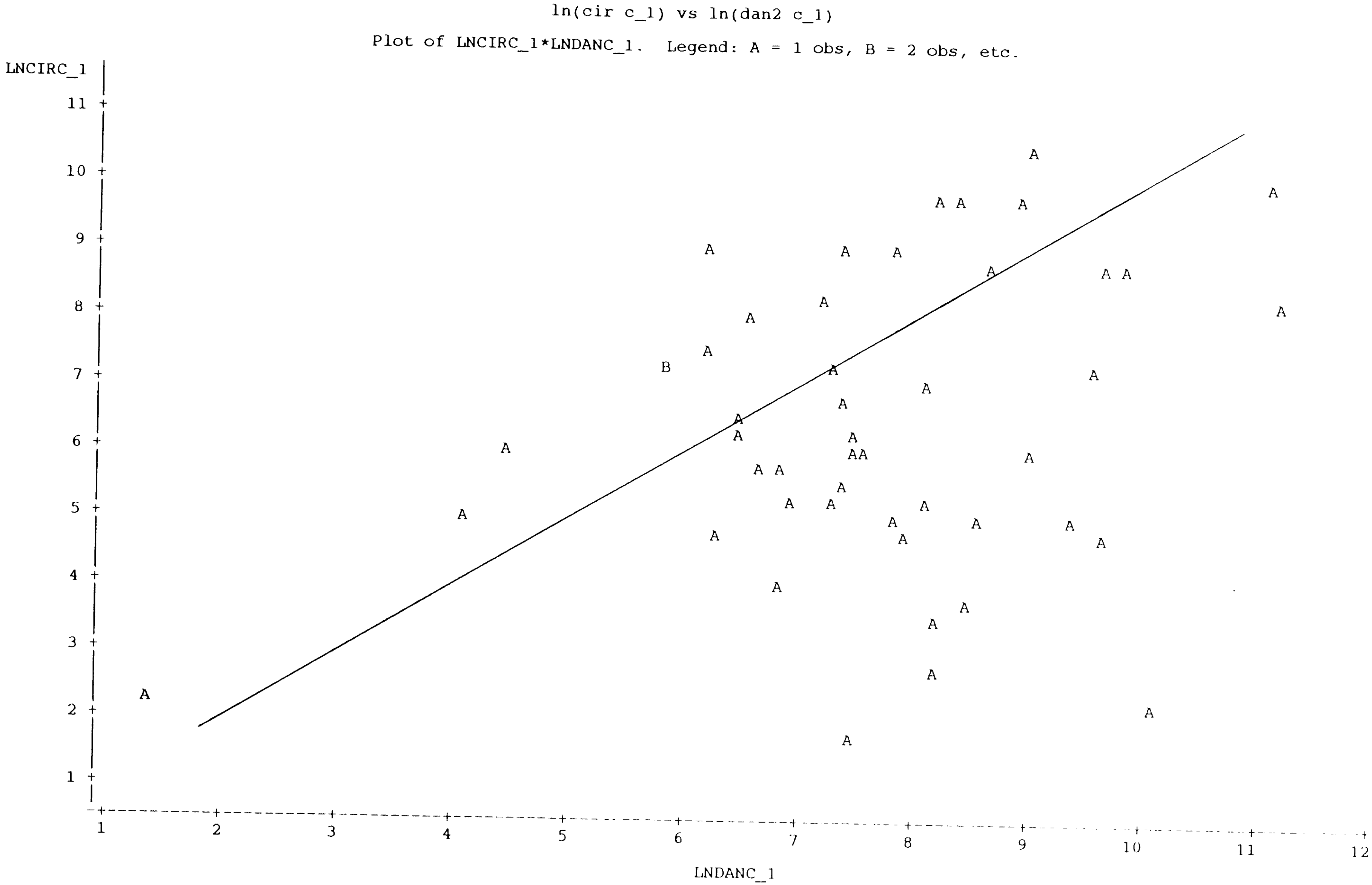


Figure 2.2.2.1a. Ln (CIROLANA age 1) vs. ln (DANA age 1). IYFS herring data. Slope of line = 1.0.

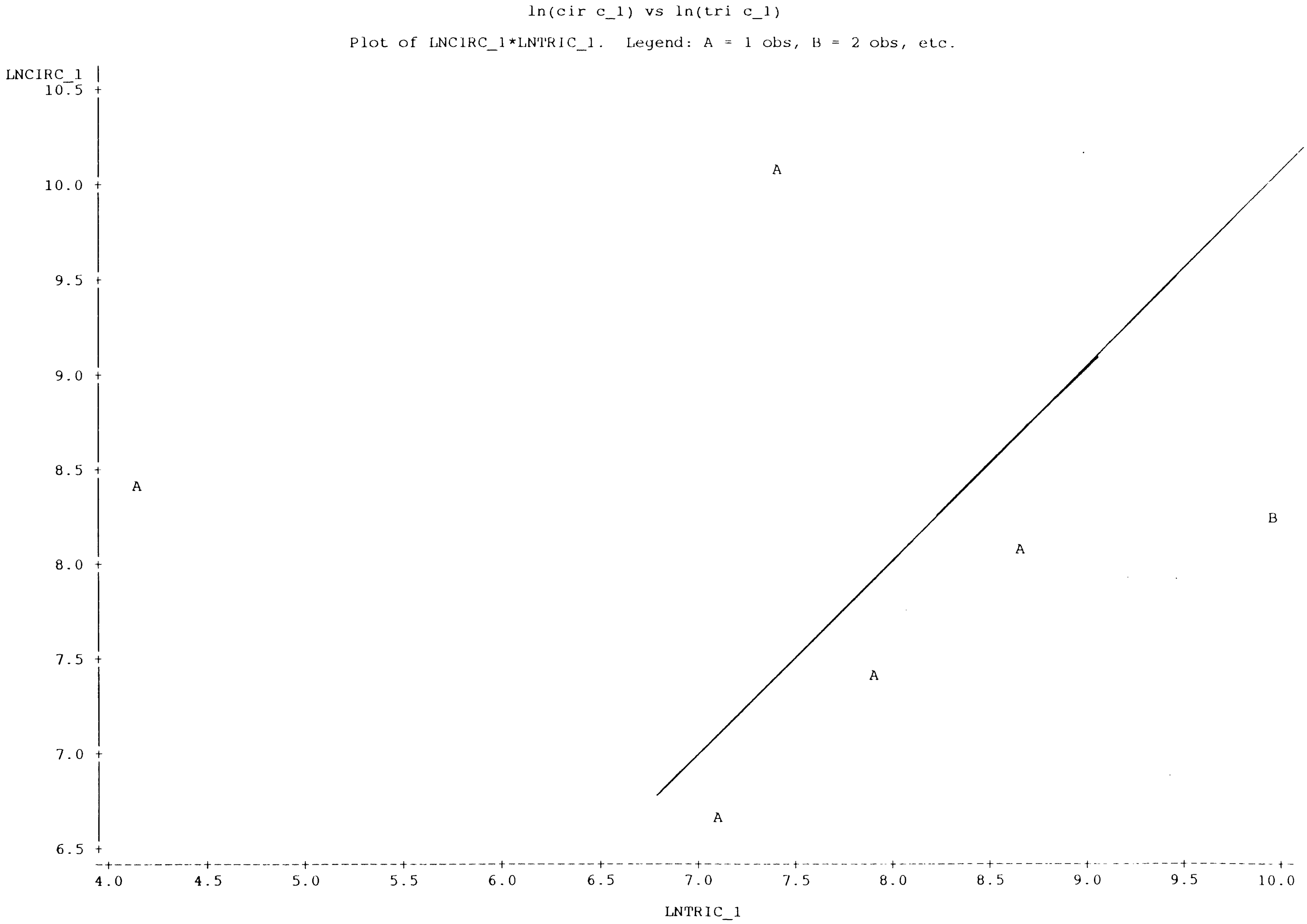


Figure 2.2.2.1b. Ln (CIROLANA age 1) vs. ln (TRIDENS age 1). IYFS herring data. Slope of line = 1.0.

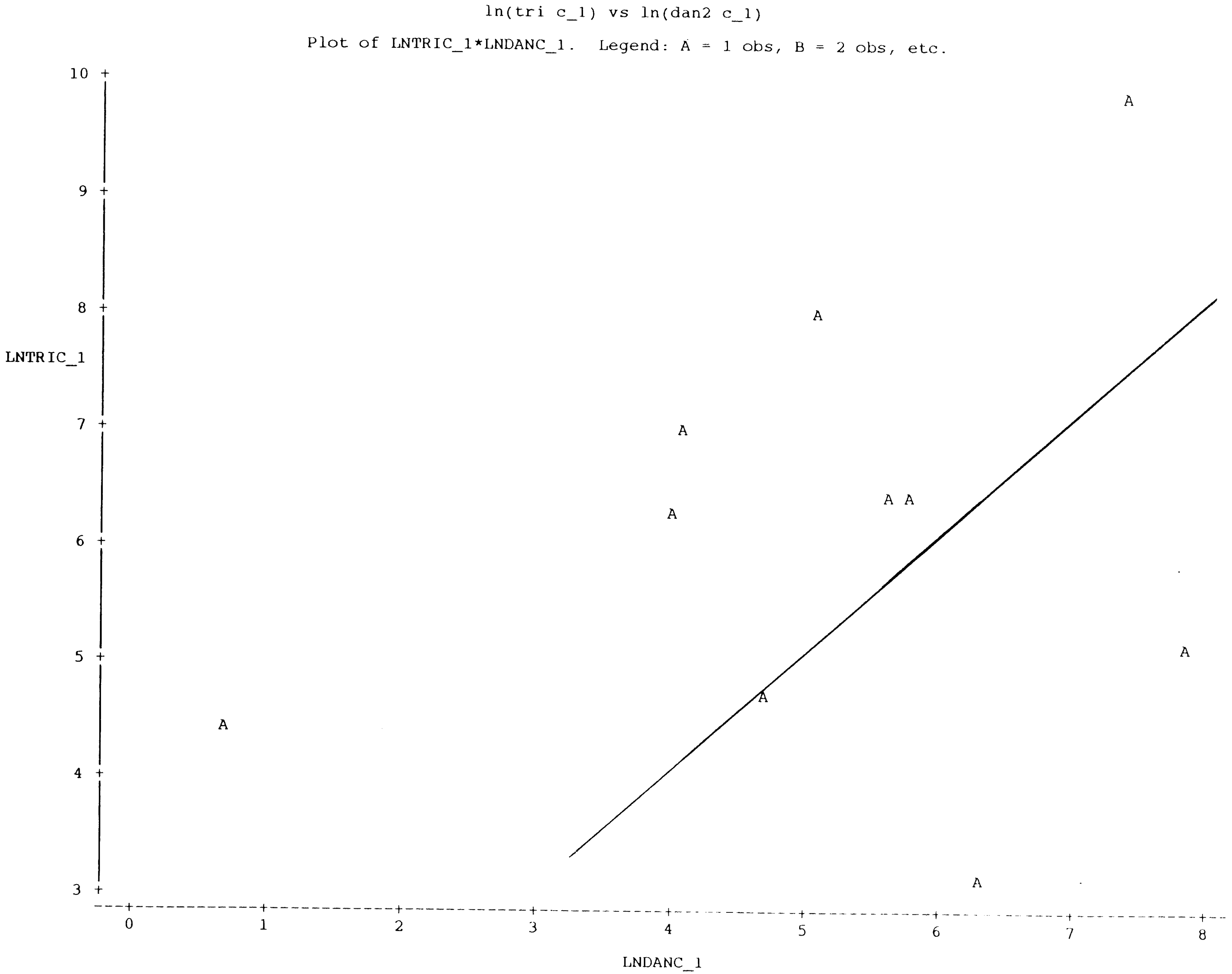


Figure 2.2.2.1c. Ln (TRIDENS age 1) vs. ln (DANA age 1). IYFS herring data.
Slope of line = 1.0.

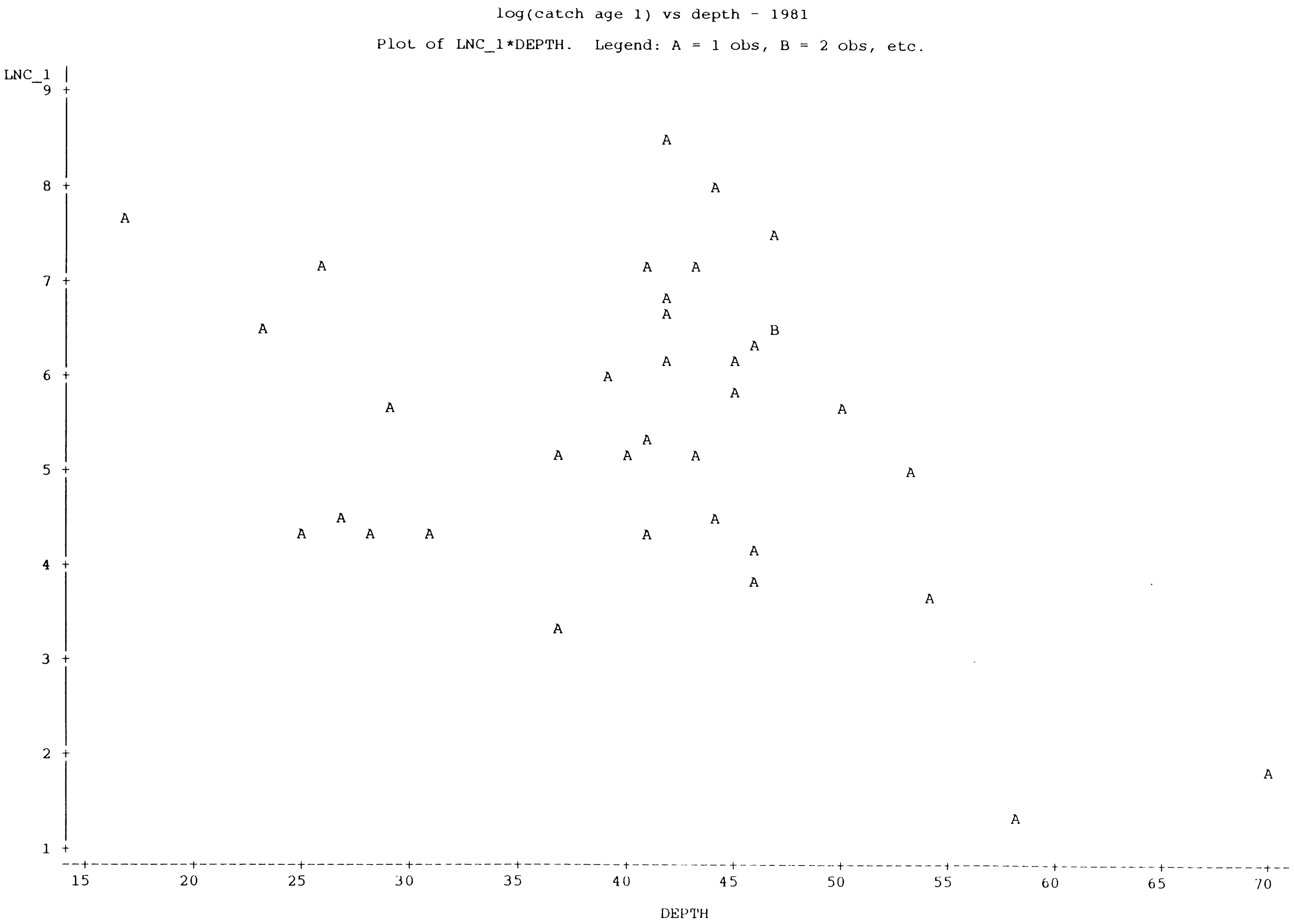


Figure 2.2.2.2. Ln (Catch age 1) vs. depth (meters). IYFS herring data.

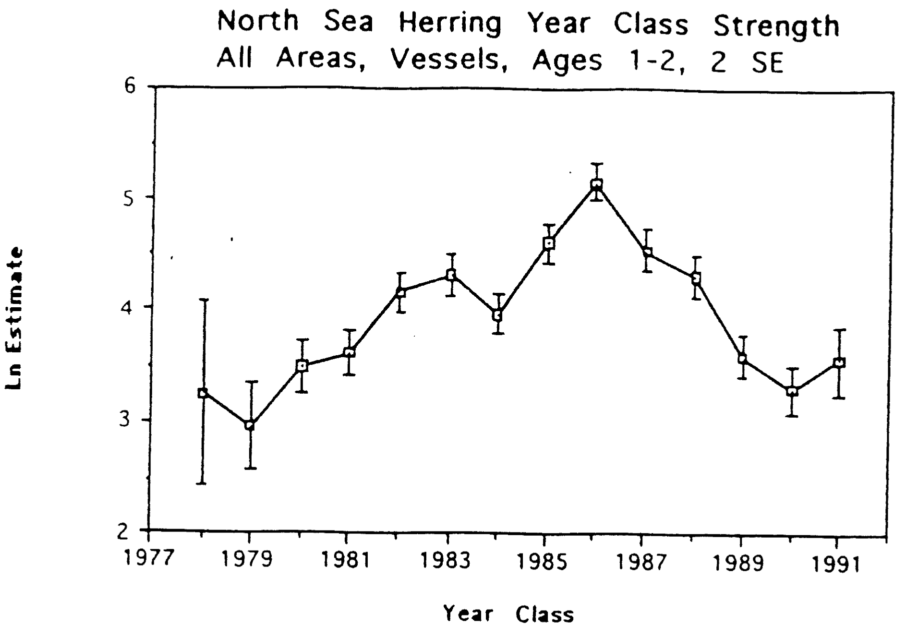


Figure 2.2.4.1. Trends in estimated year class strength for North Sea herring, based on research vessel indices. Indices ±2 SE are plotted.

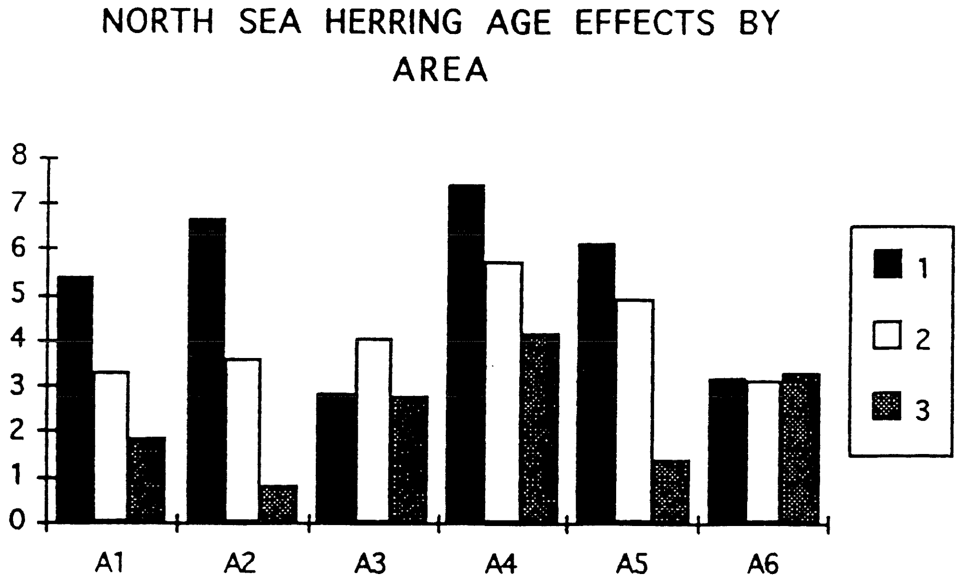


Figure 2.2.4.2. Relative year class strength (age effects) for six herring survey areas in the North Sea, by age group (1, 2, 3).

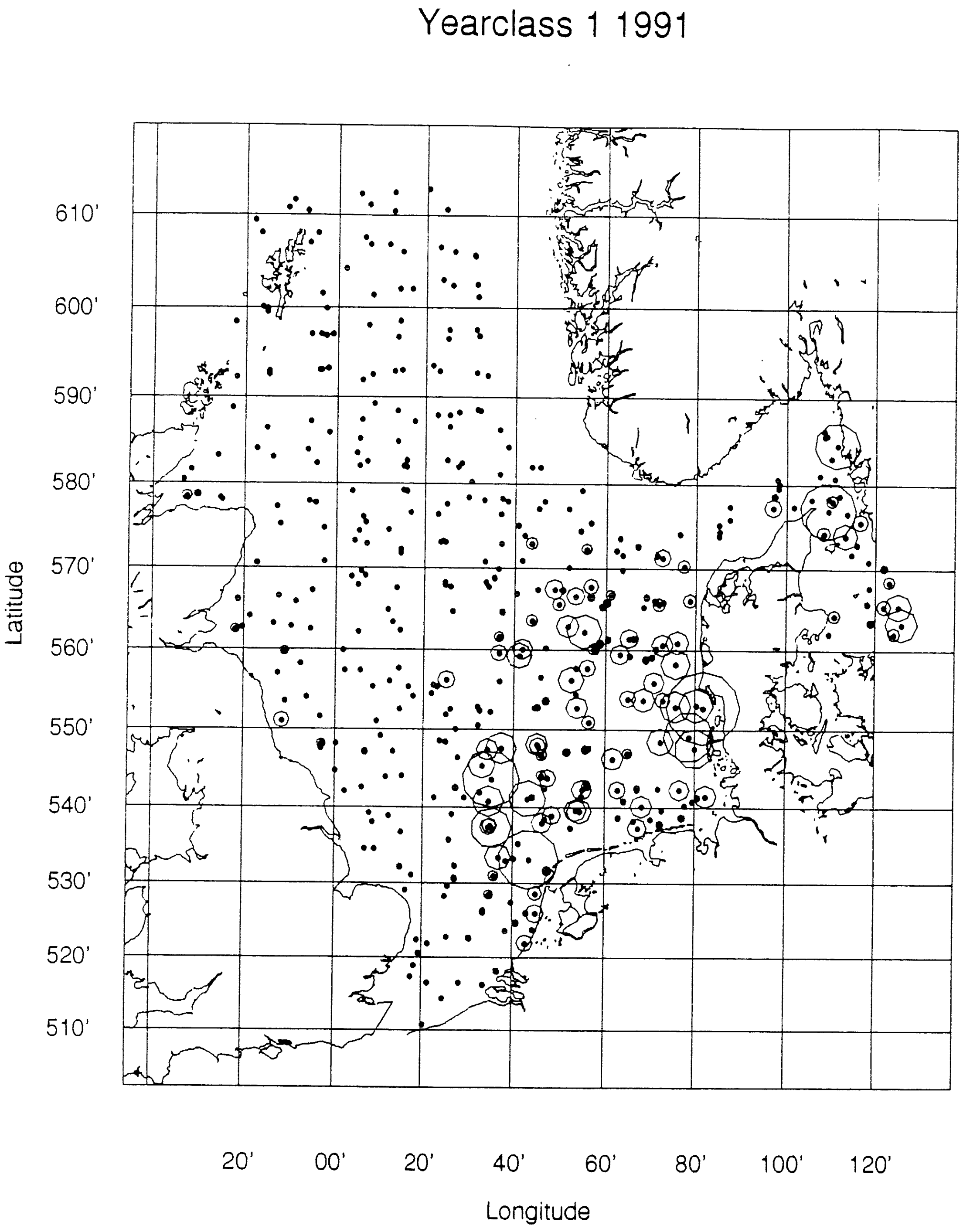


Figure 2.3.1.1a. Relative catch in numbers of age 1 herring in the IYFS, 1991.

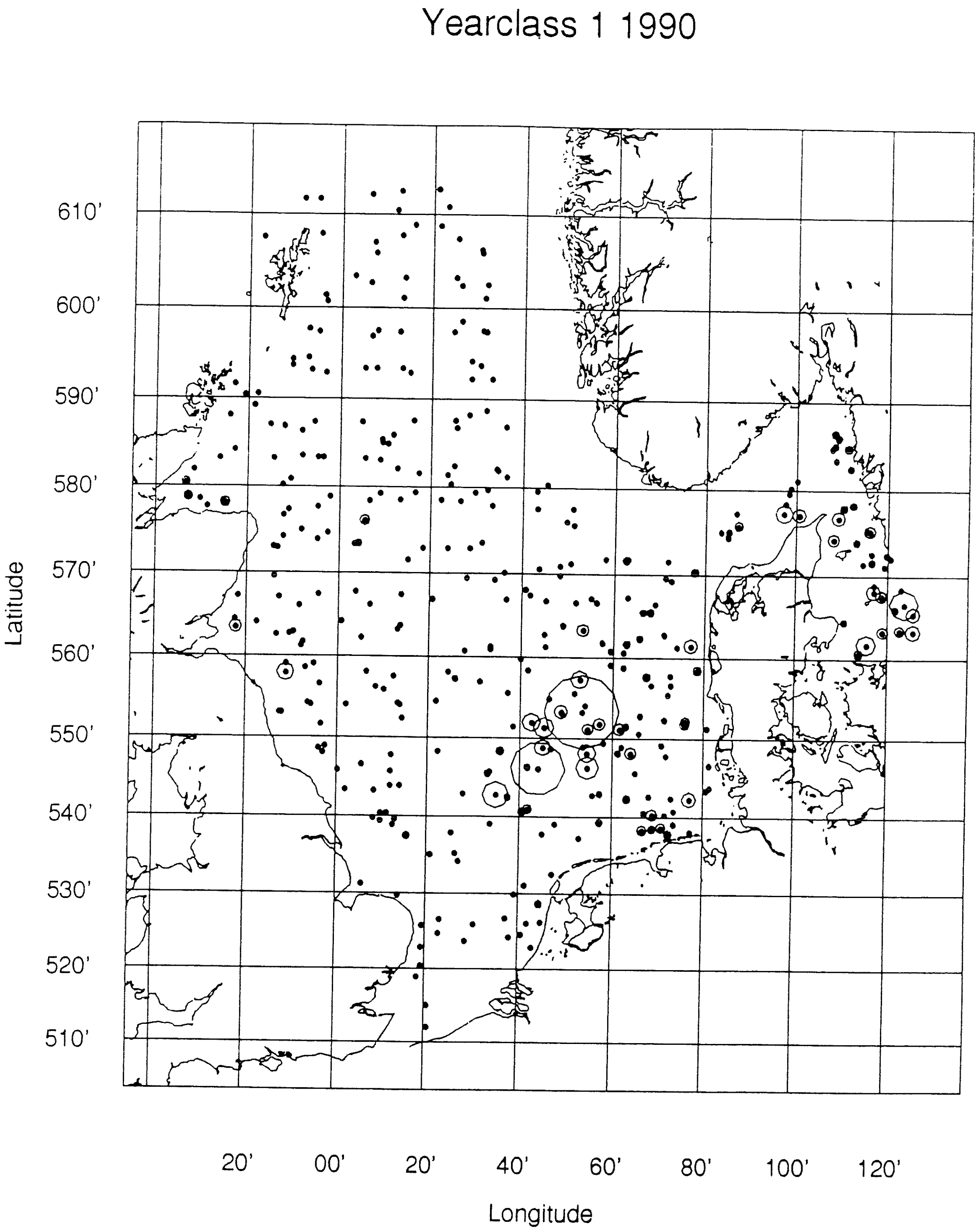


Figure 2.3.1.1b. Relative catch in numbers of age 1 herring in the IYFS, 1990.

Yearclass 1 1989

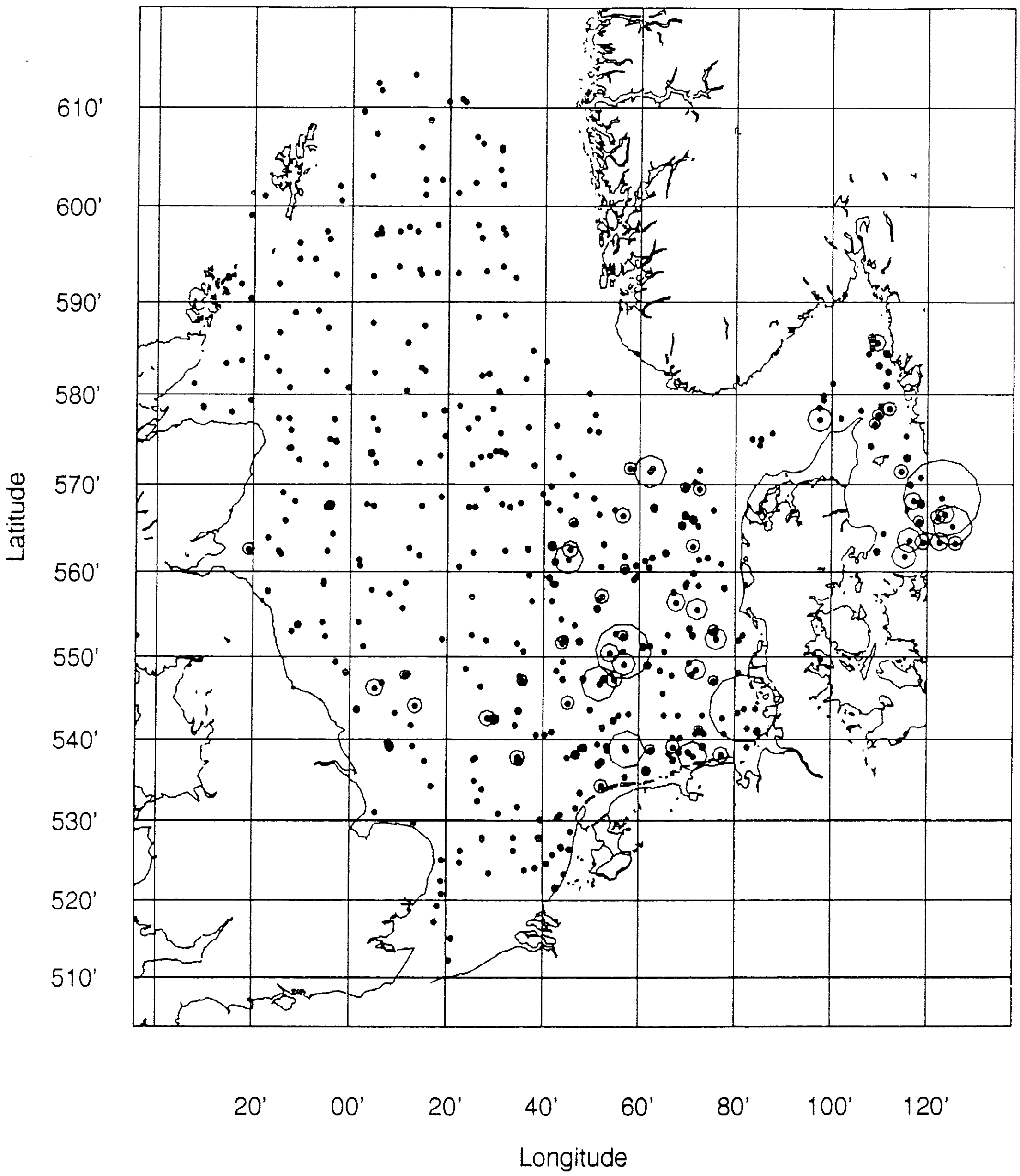
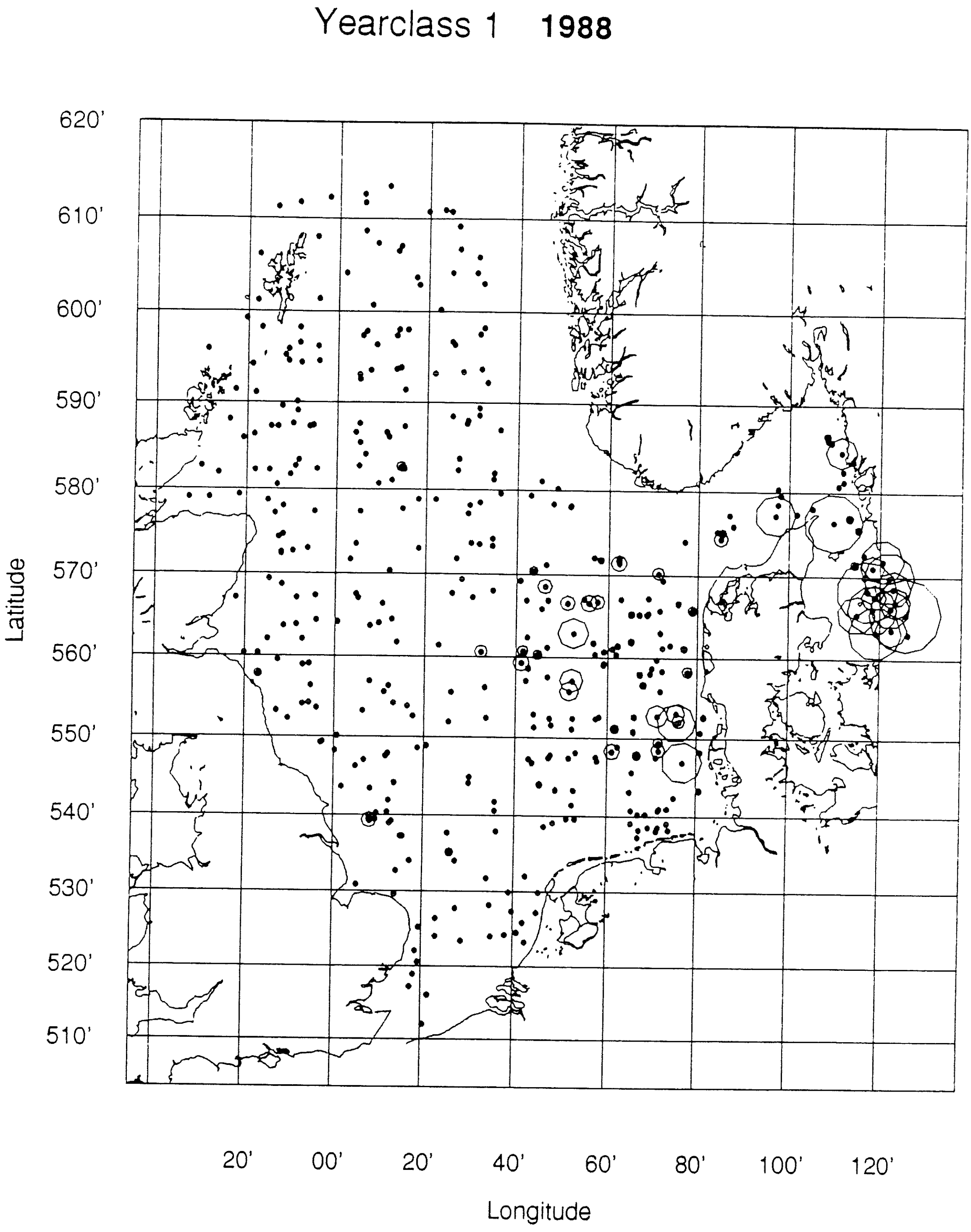


Figure 2.3.1.1c. Relative catch in numbers of age 1 herring in the IYFS, 1989.



Yearclass 1 **1988**

Figure 2.3.1.1d. Relative catch in numbers of age 1 herring in the IYFS, 1988.

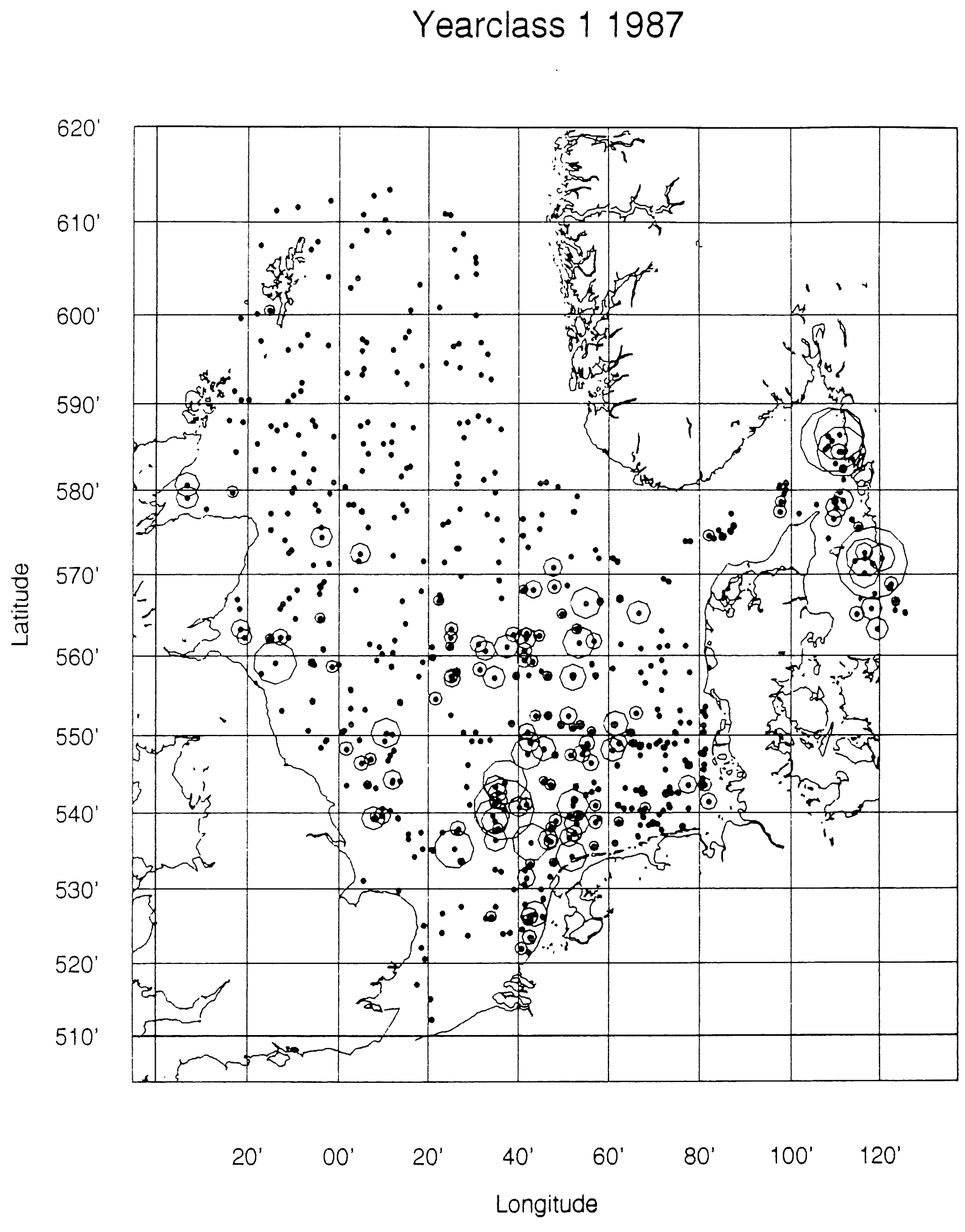


Figure 2.3.1.1e. Relative catch in numbers of age 1 herring in the IYFS, 1987.

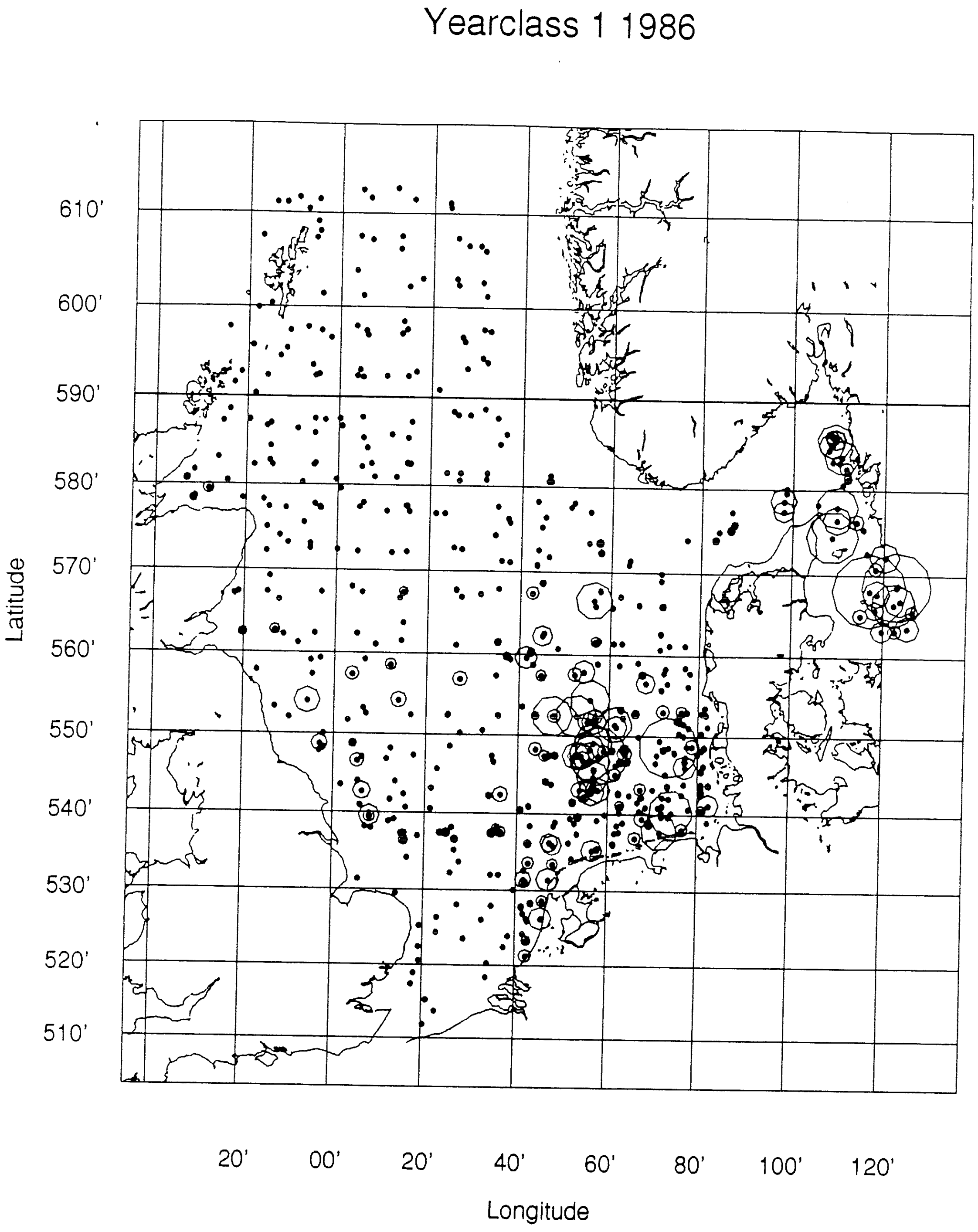

Yearclass 1 1986

Figure 2.3.1.1f. Relative catch in numbers of age 1 herring in the IYFS, 1986.

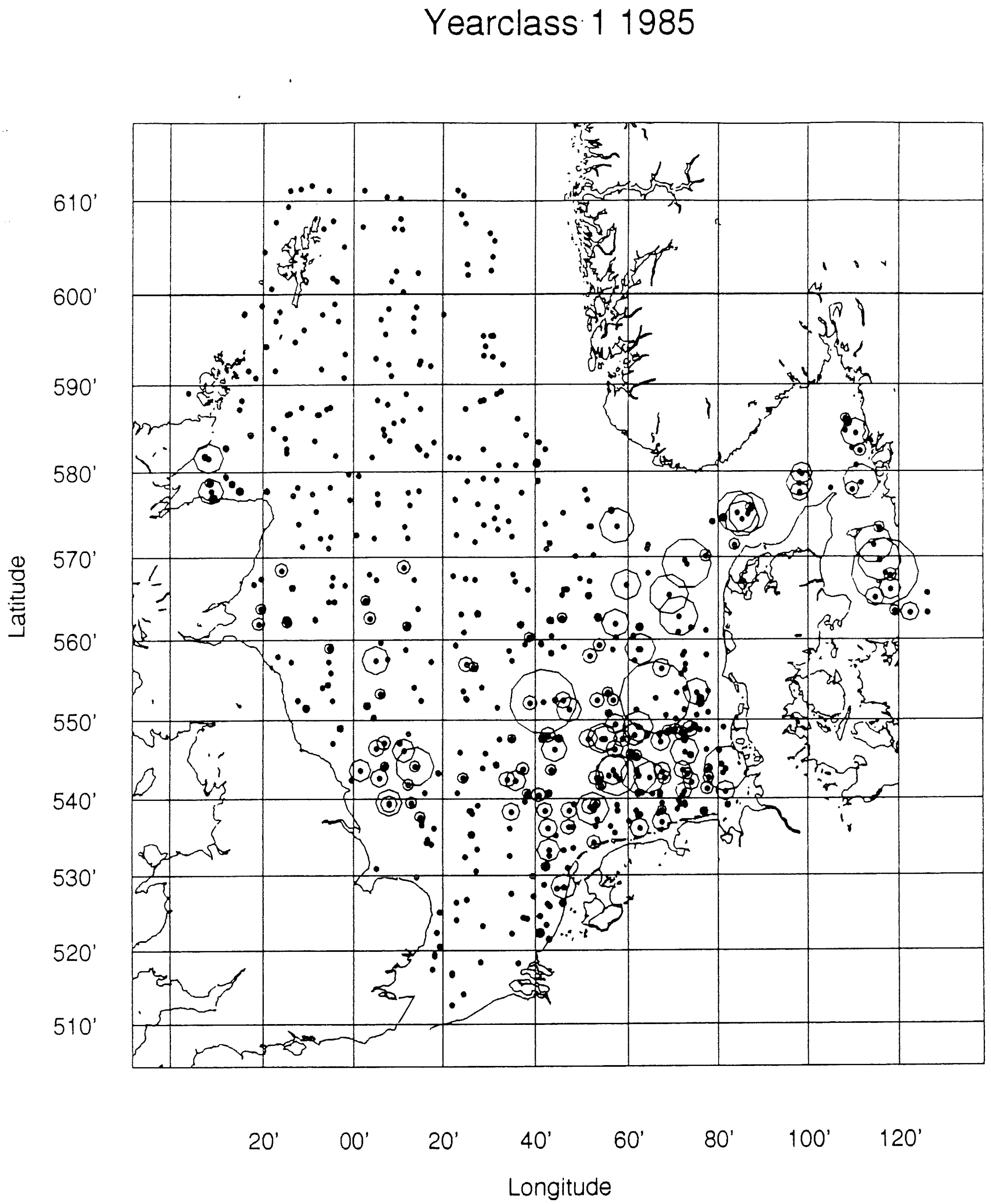


Figure 2.3.1.1g. Relative catch in numbers of age 1 herring in the IYFS, 1985.

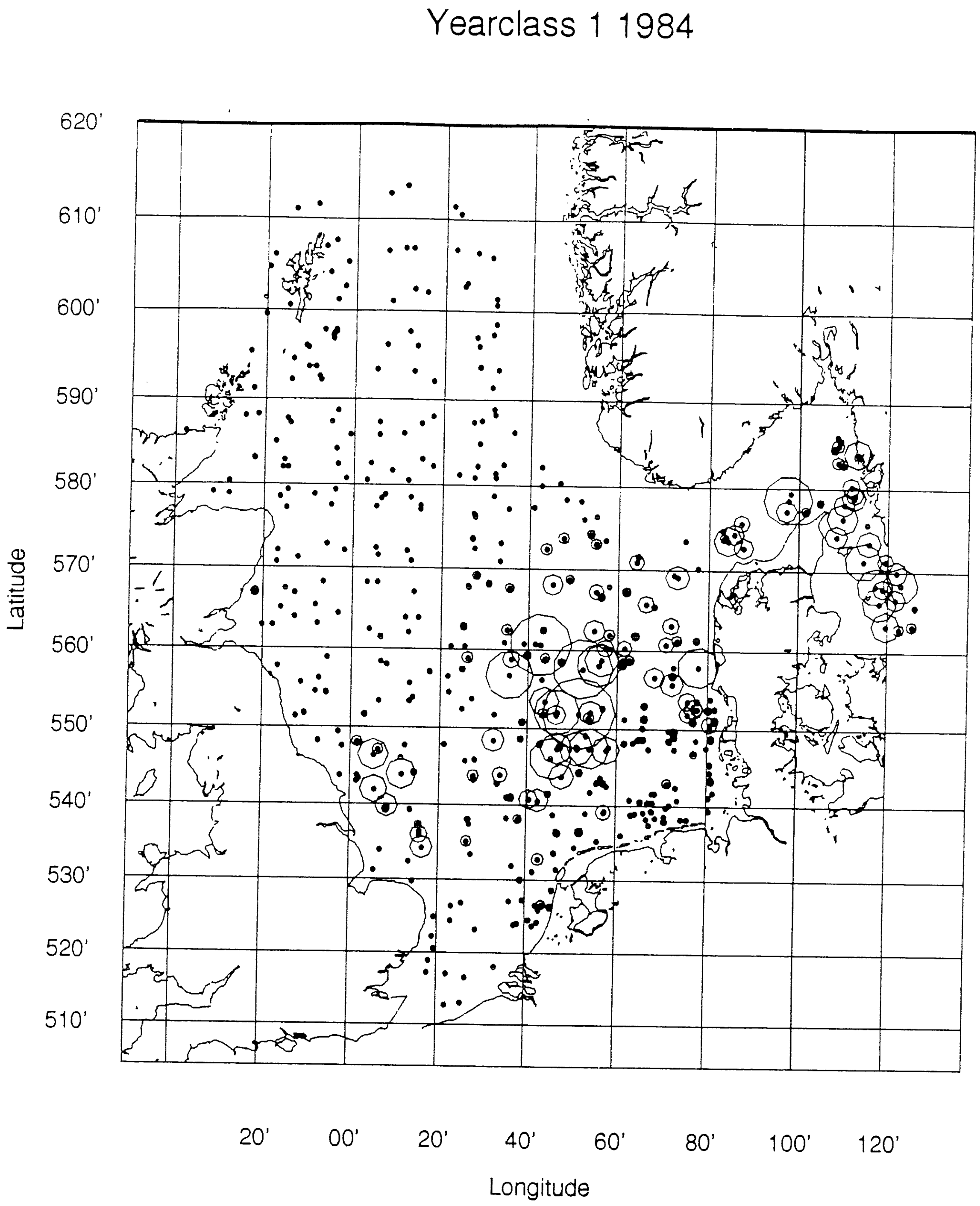


Figure 2.3.1.1h. Relative catch in numbers of age 1 herring in the IYFS, 1984.

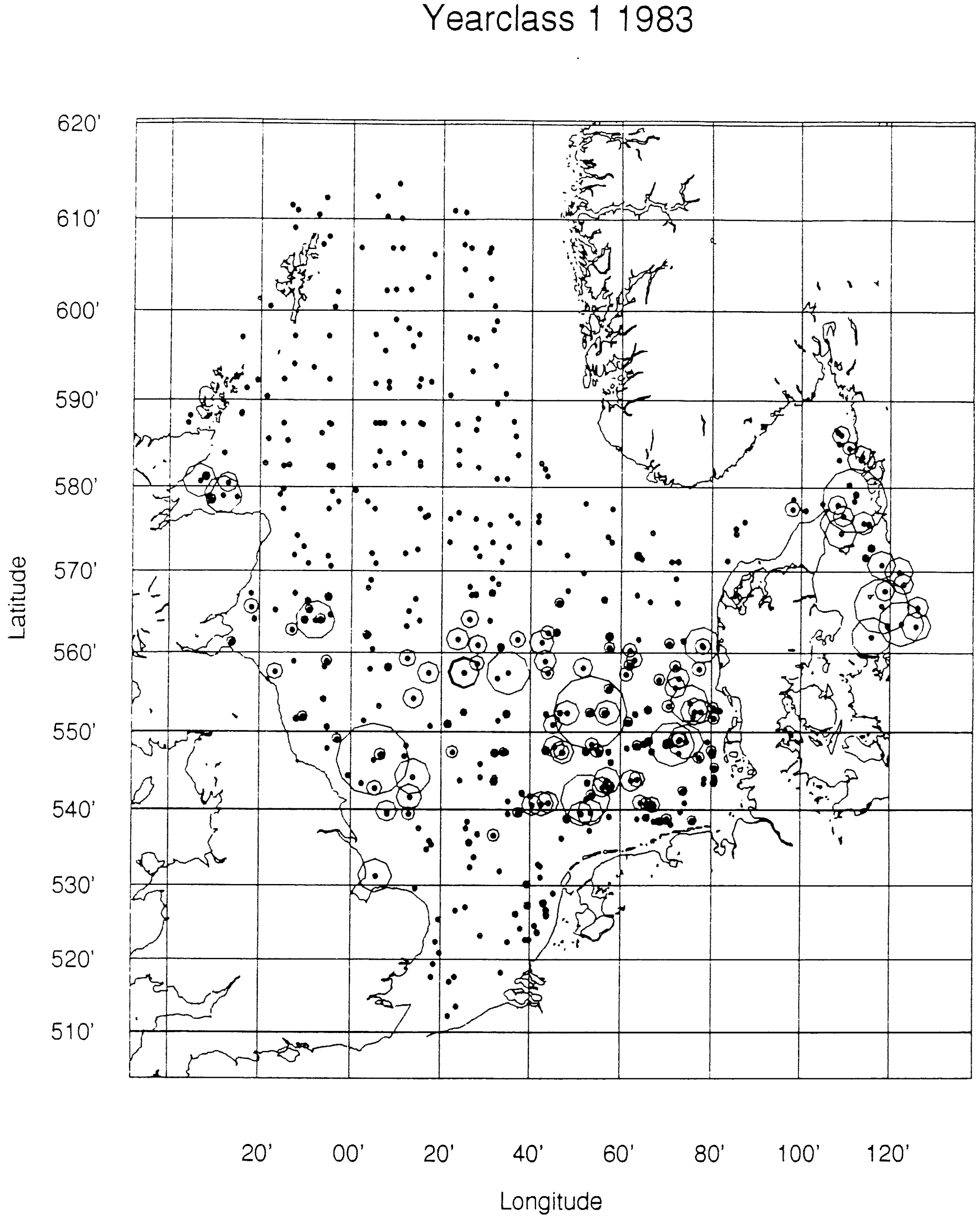


Figure 2.3.1.1i. Relative catch in numbers of age 1 herring in the IYFS, 1983.

Yearclass 1 1982

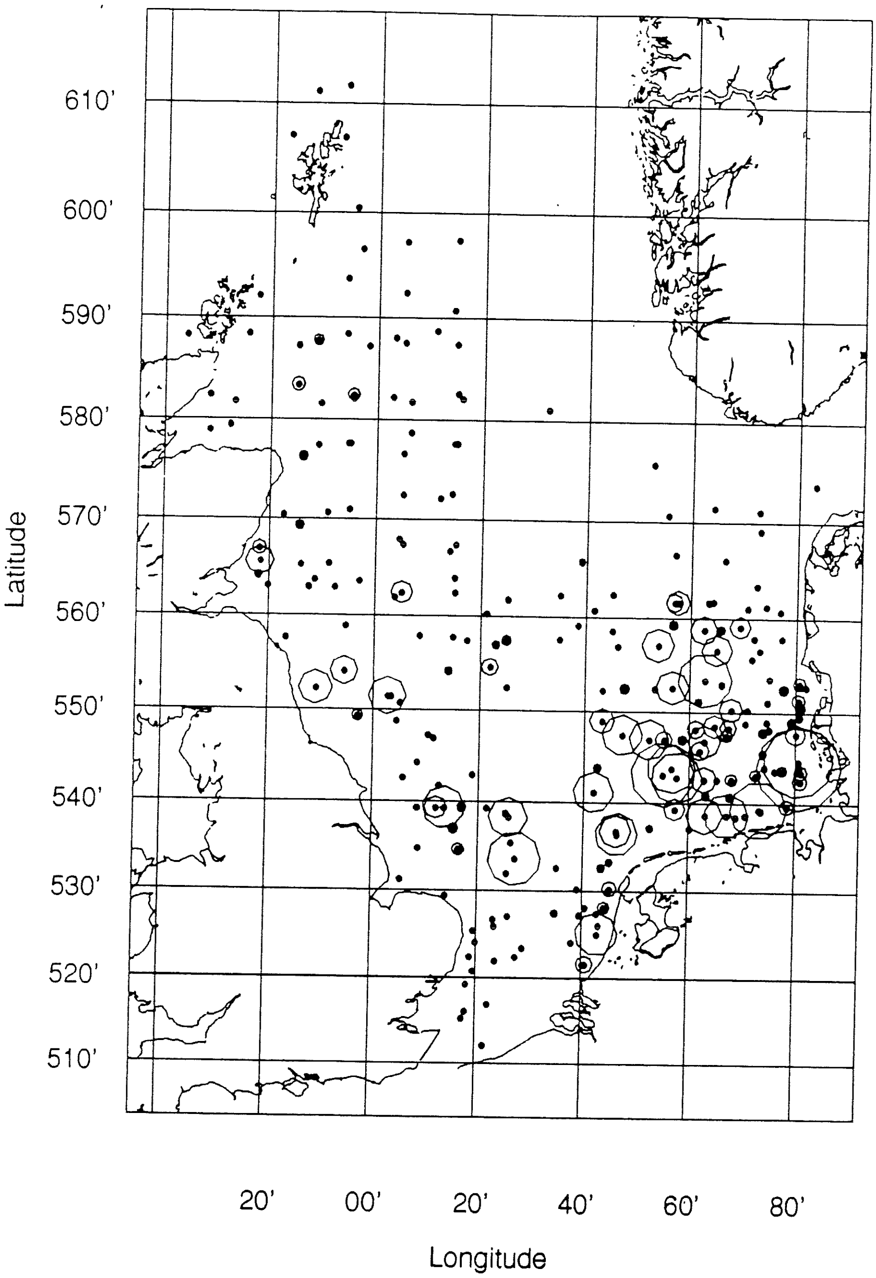


Figure 2.3.1.1j. Relative catch in numbers of age 1 herring in the IYFS, 1982.

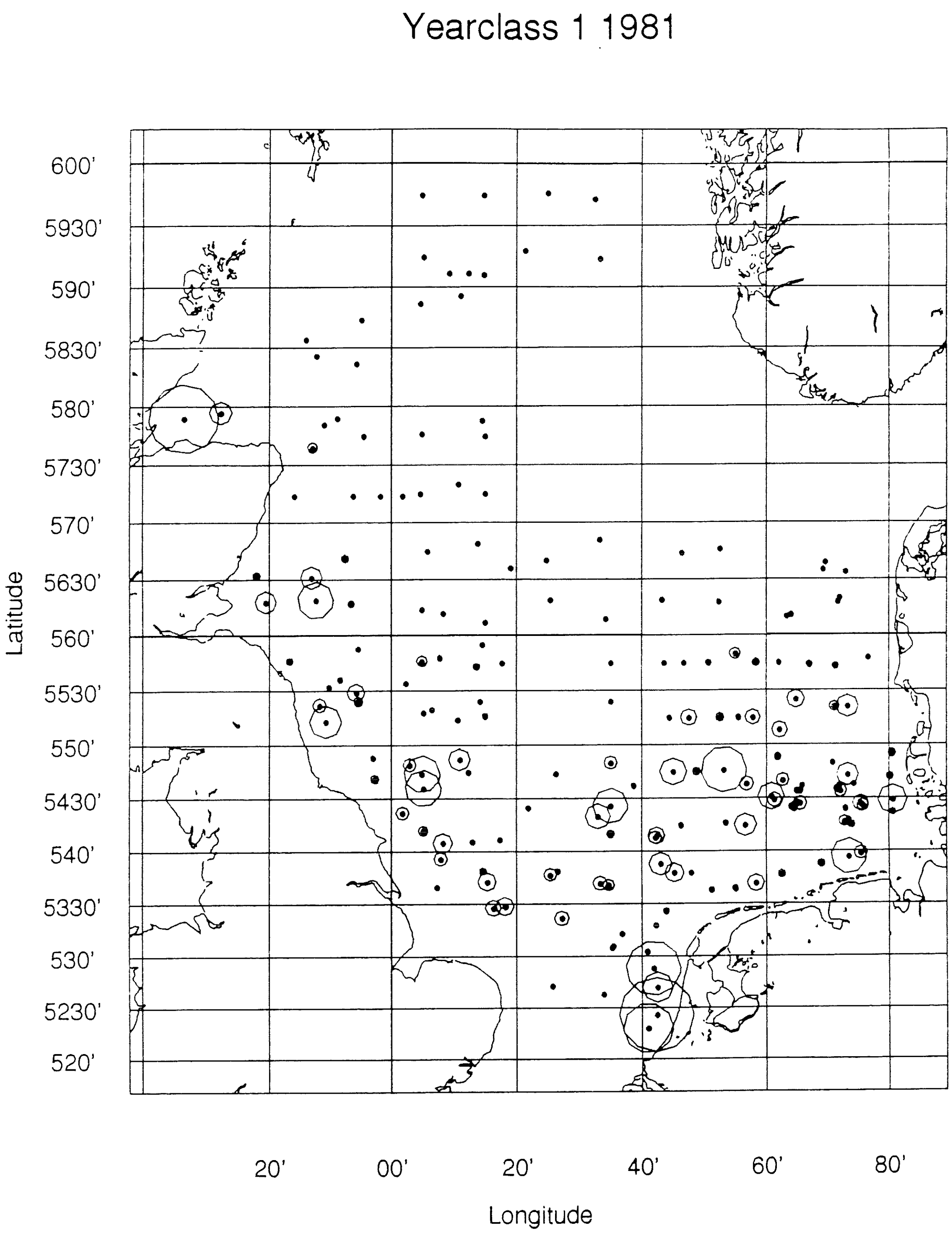


Figure 2.3.1.1k. Relative catch in numbers of age 1 herring in the IYFS, 1981.

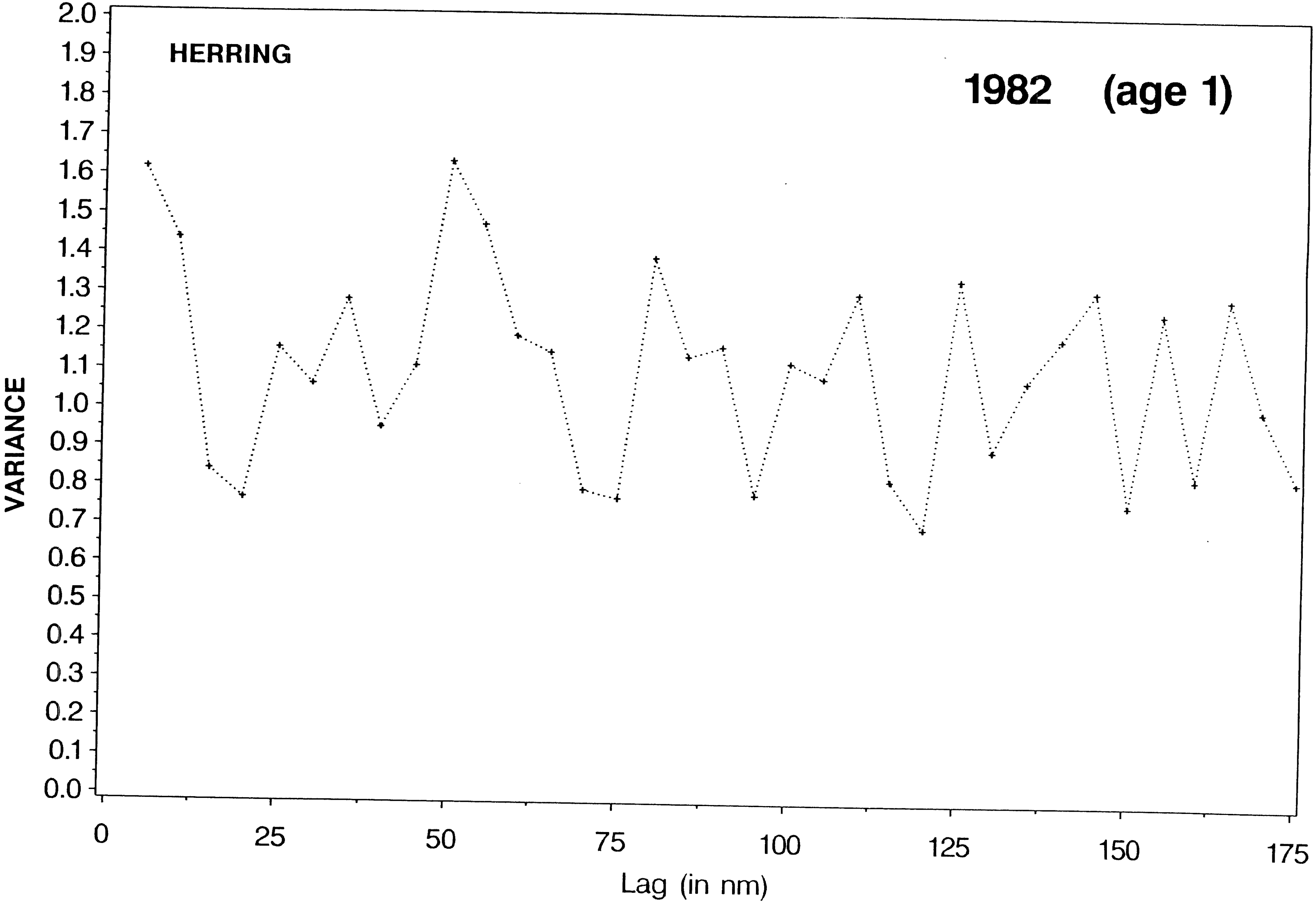


Figure 2.3.1.2. Normalized variogram for age 1 herring from the IYFS data.

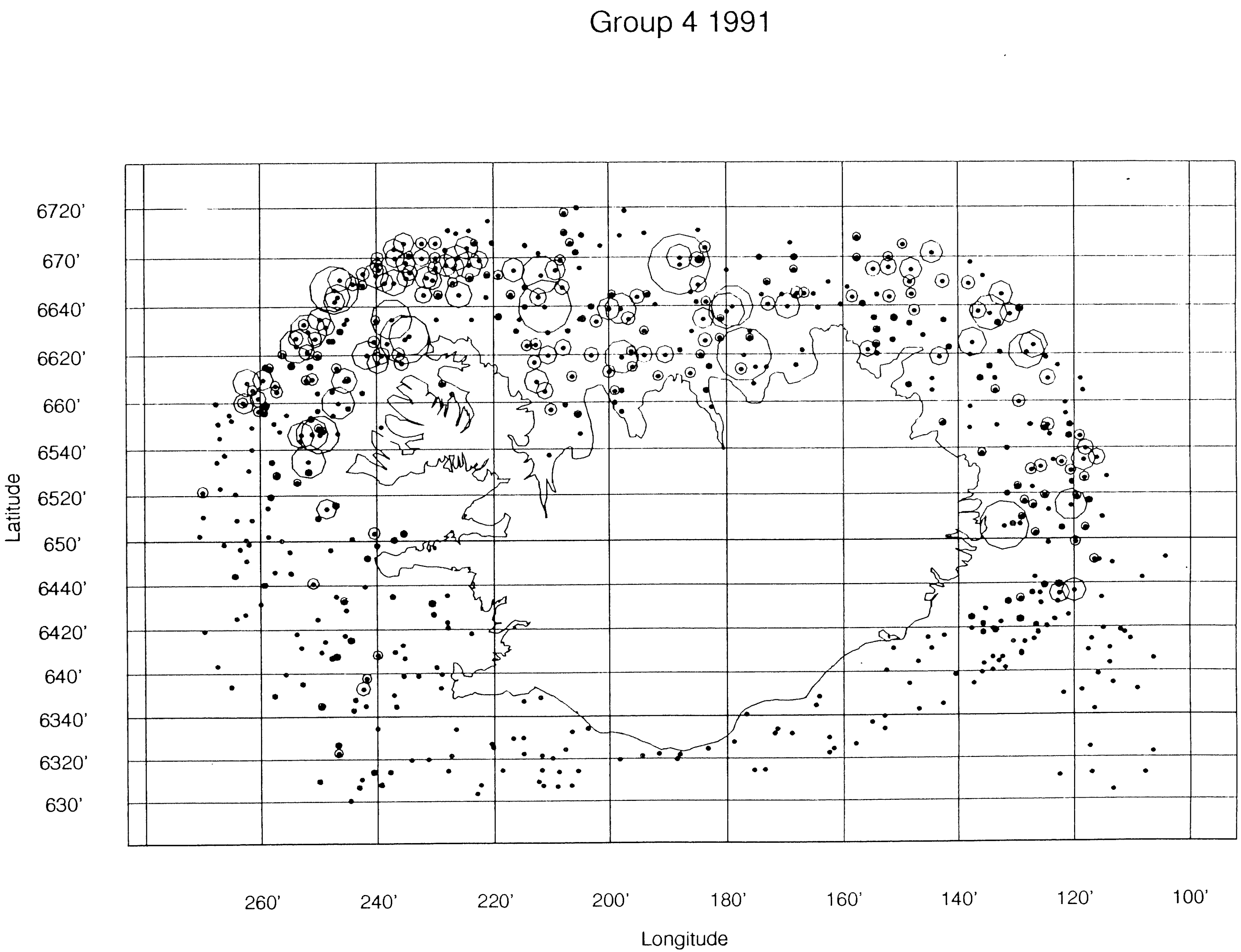


Figure 2.3.1.3a. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, in `91

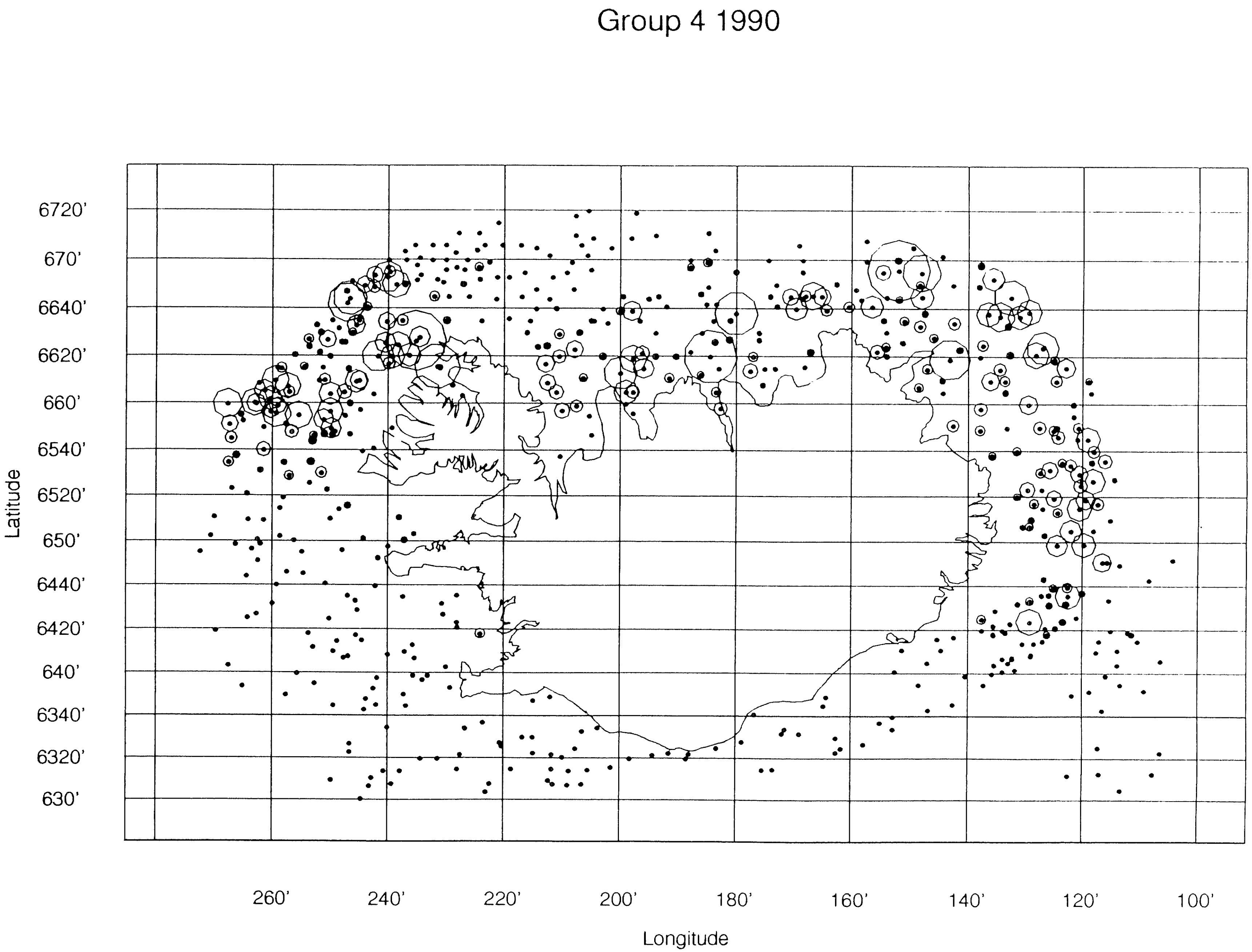


Figure 2.3.1.3b. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey in 1990

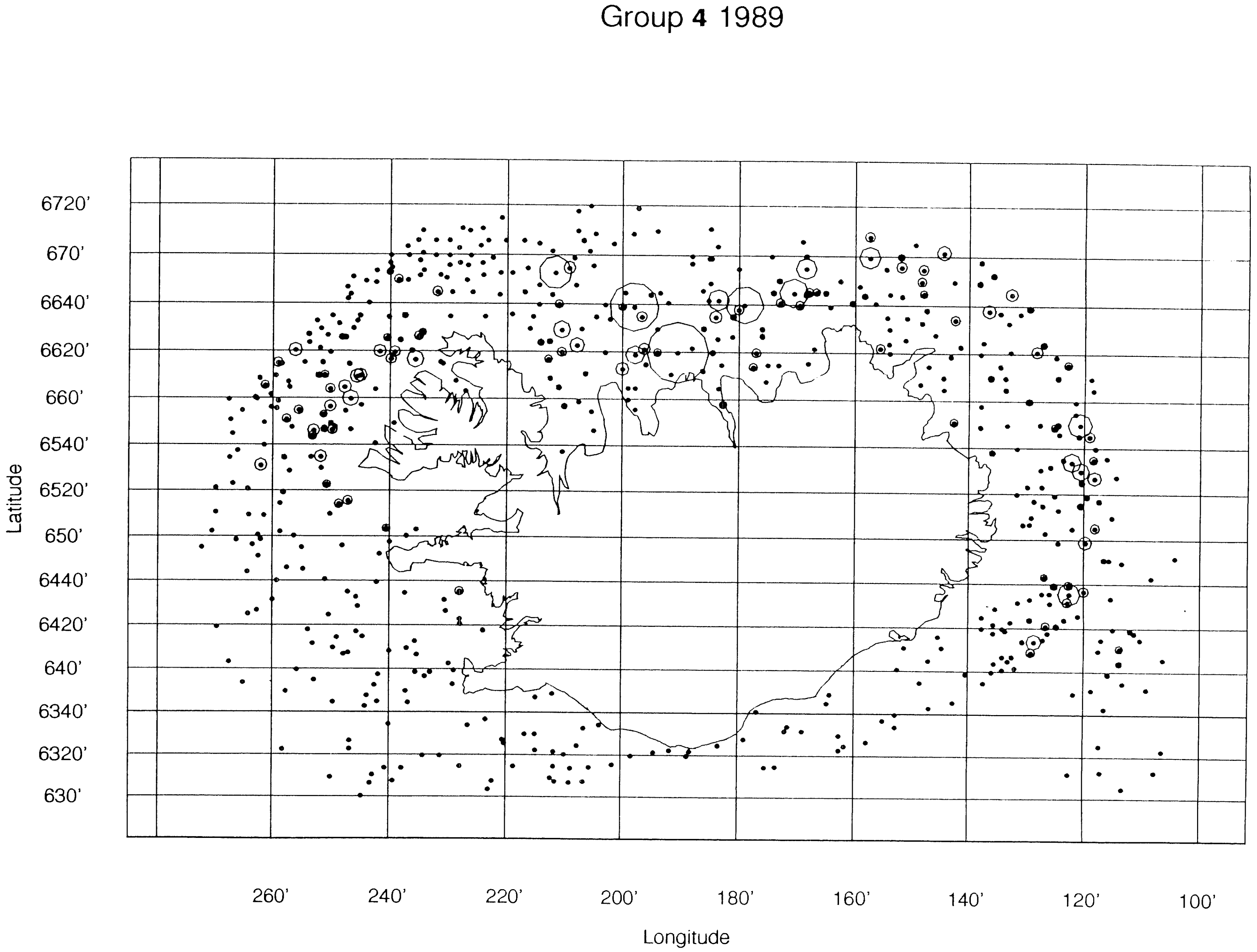
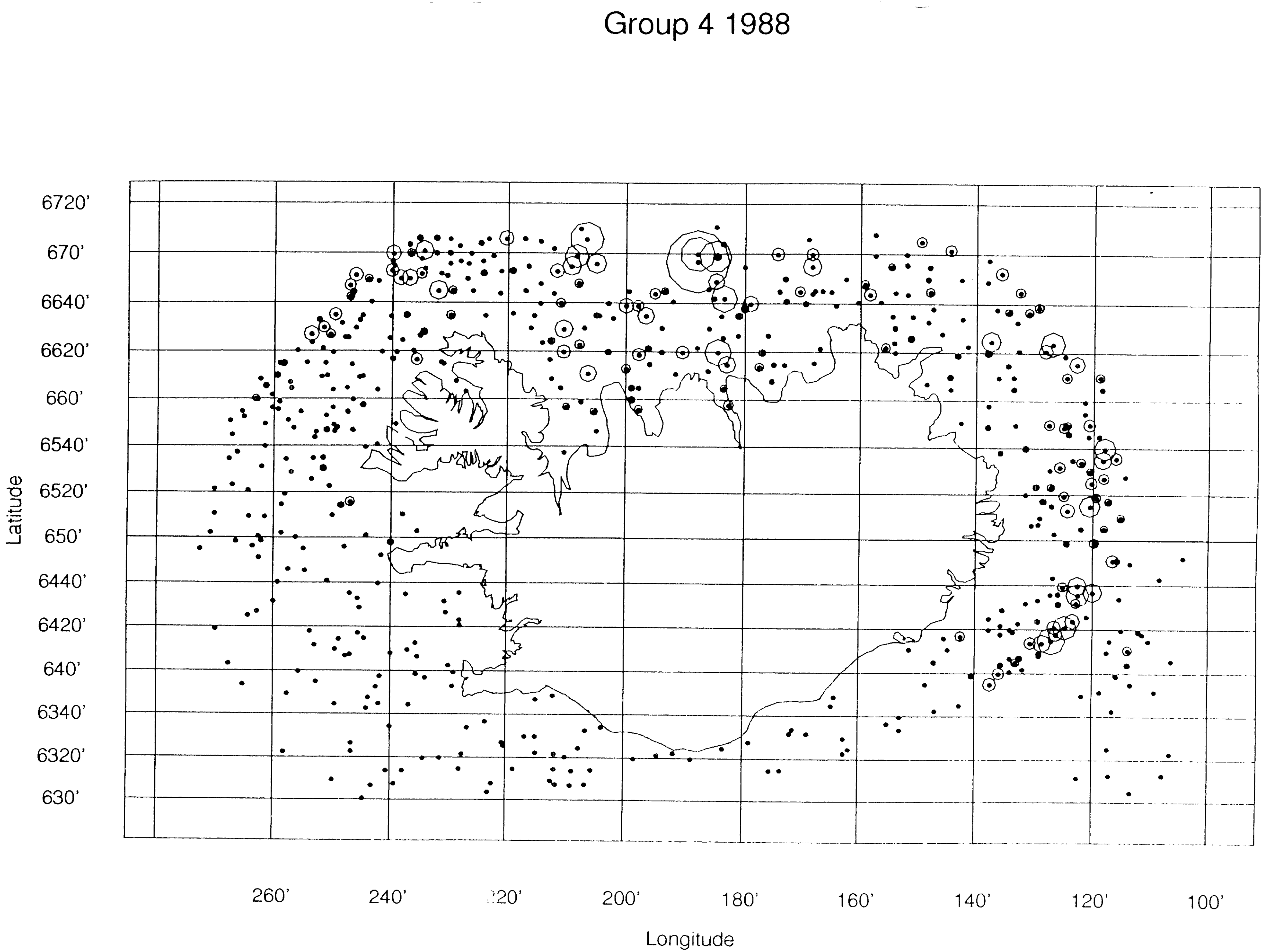


Figure 2.3.1.3c. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, in 1989.



Group 4 1988

Figure 2.3.1.3d. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, in 1988.

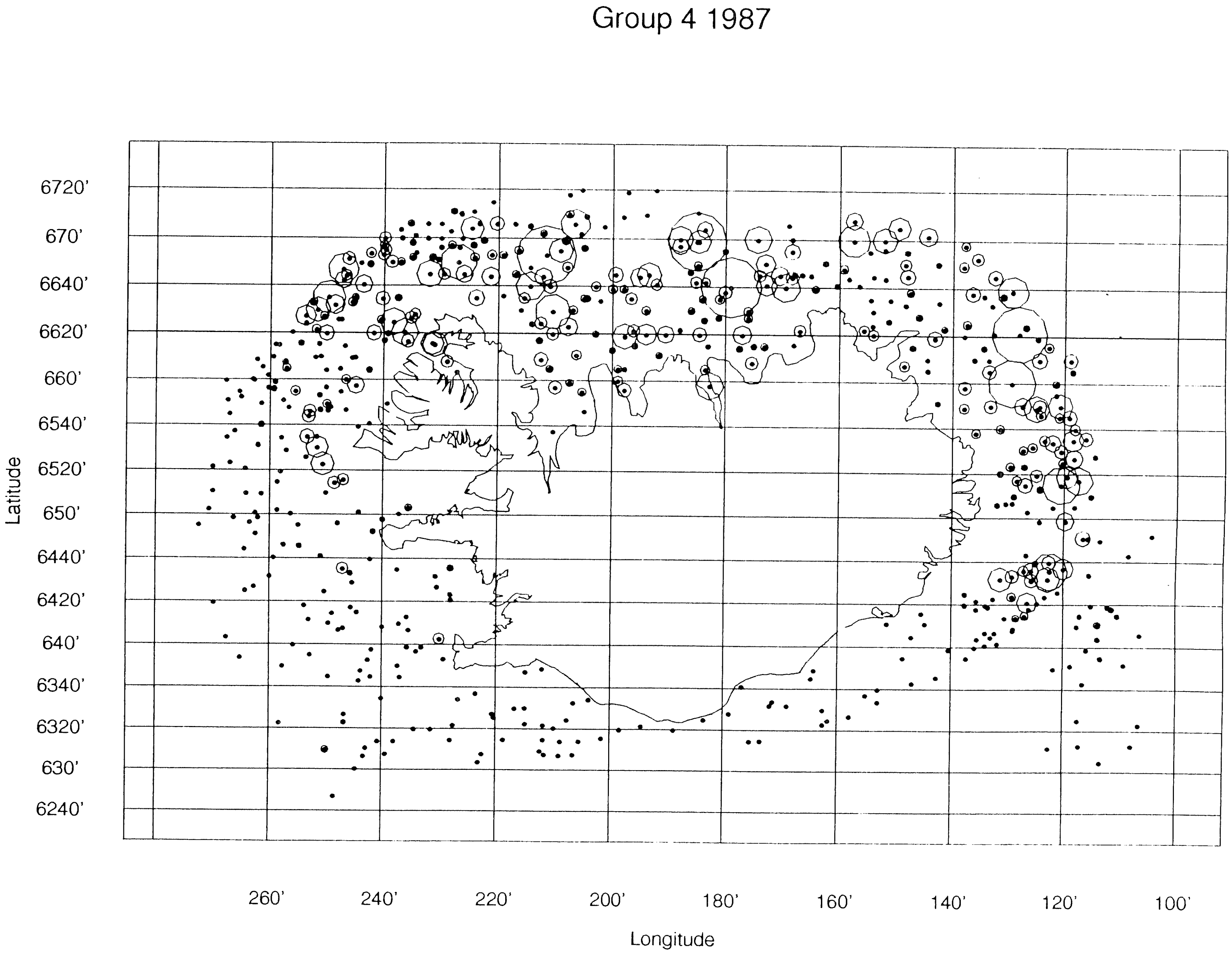


Figure 2.3.1.3e. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, i 19

group 4 1986

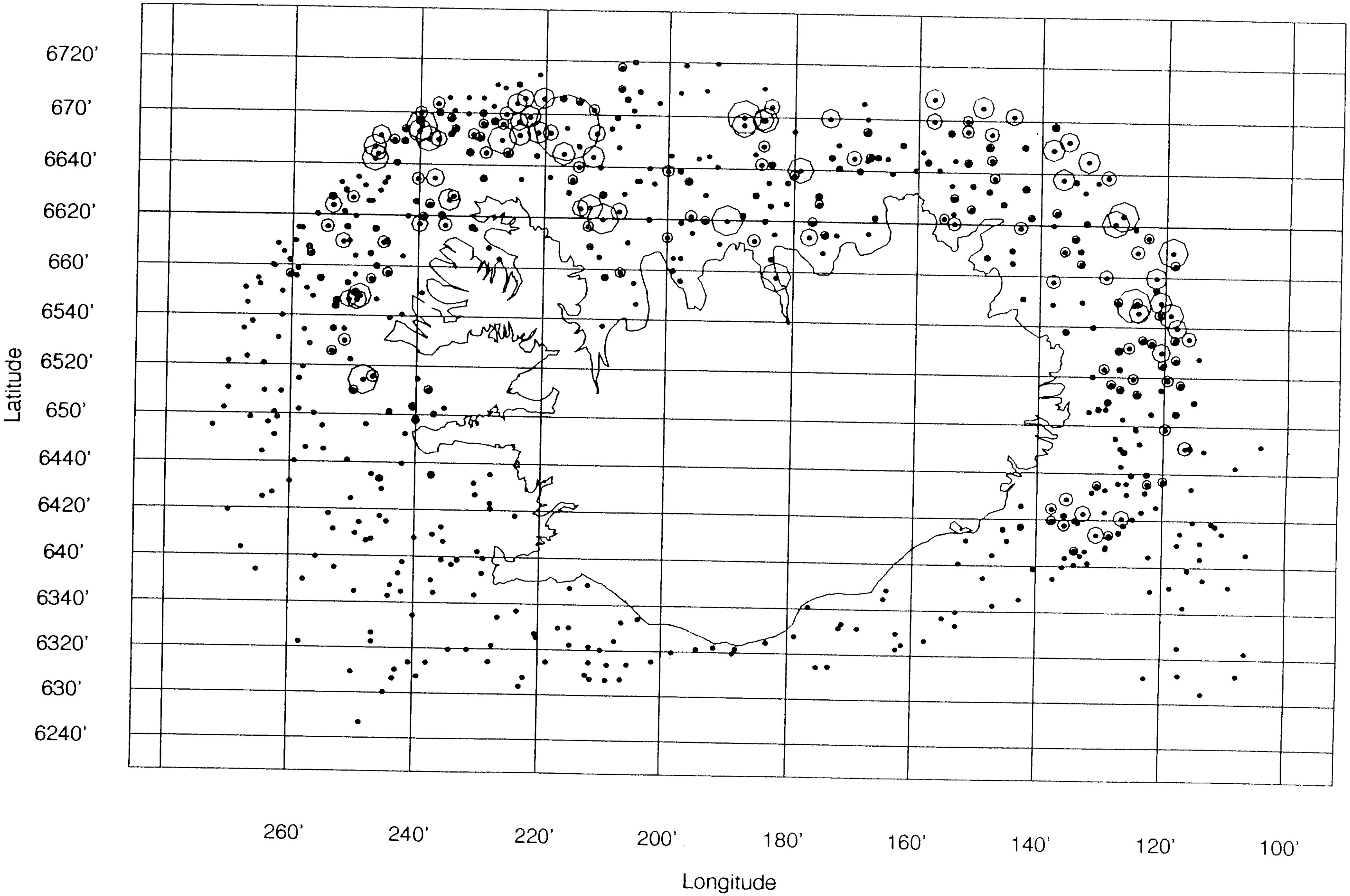


Figure 2.3.1.3f. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, in 1986.

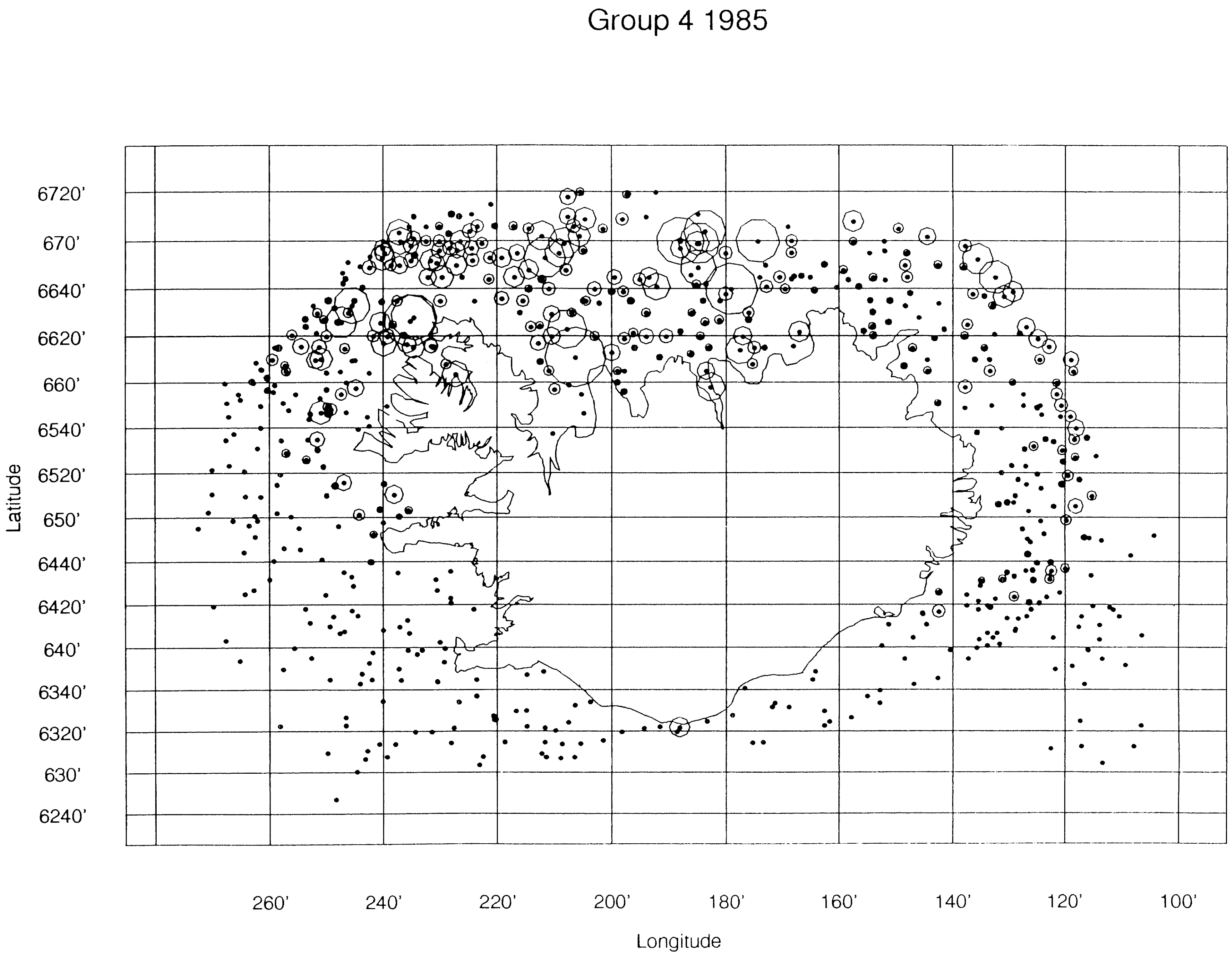


Figure 2.3.1.3g. Relative catch in numbers of age 4 cod in the Icelandic groundfish survey, in 1985.

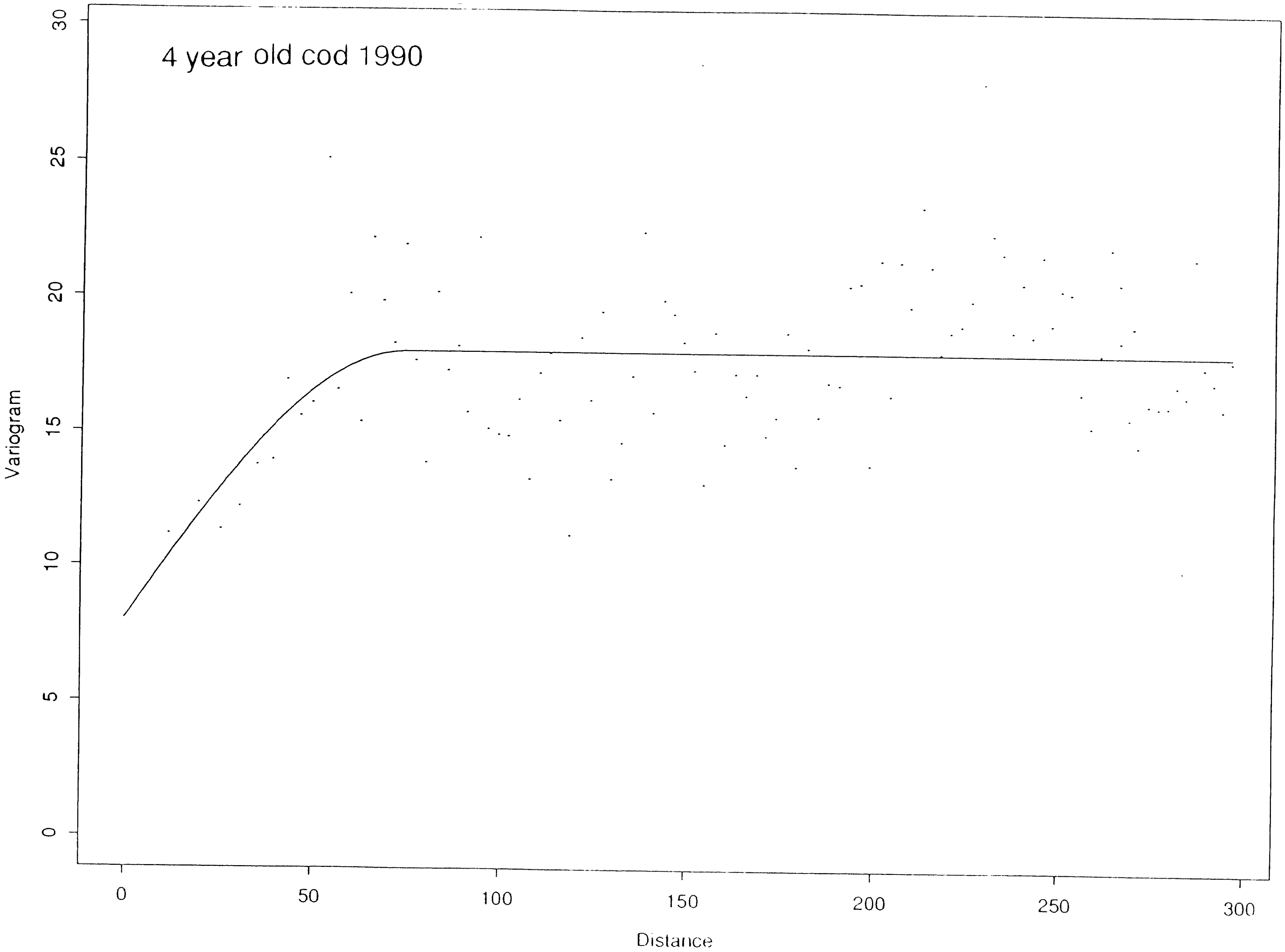
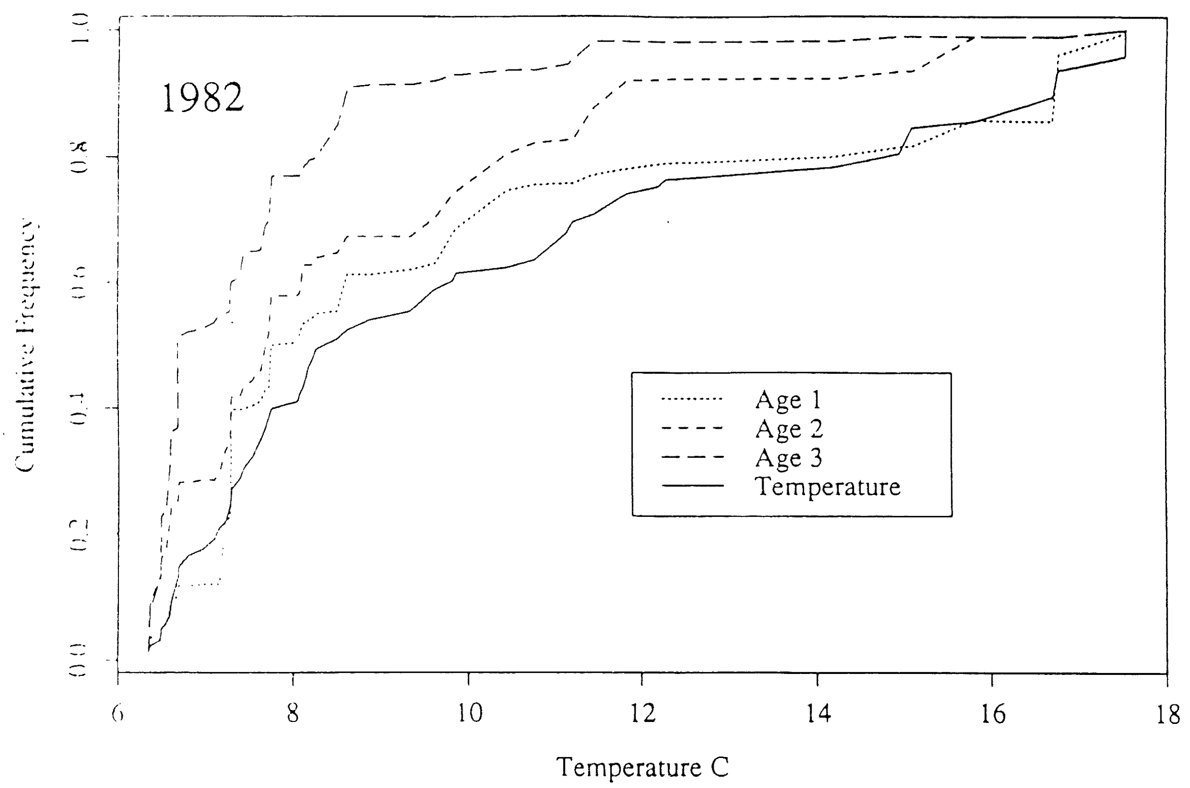


Figure 2.3.1.4. Variogram of catch in numbers of age 4 cod in the Icelandic groundfish survey in 1990.



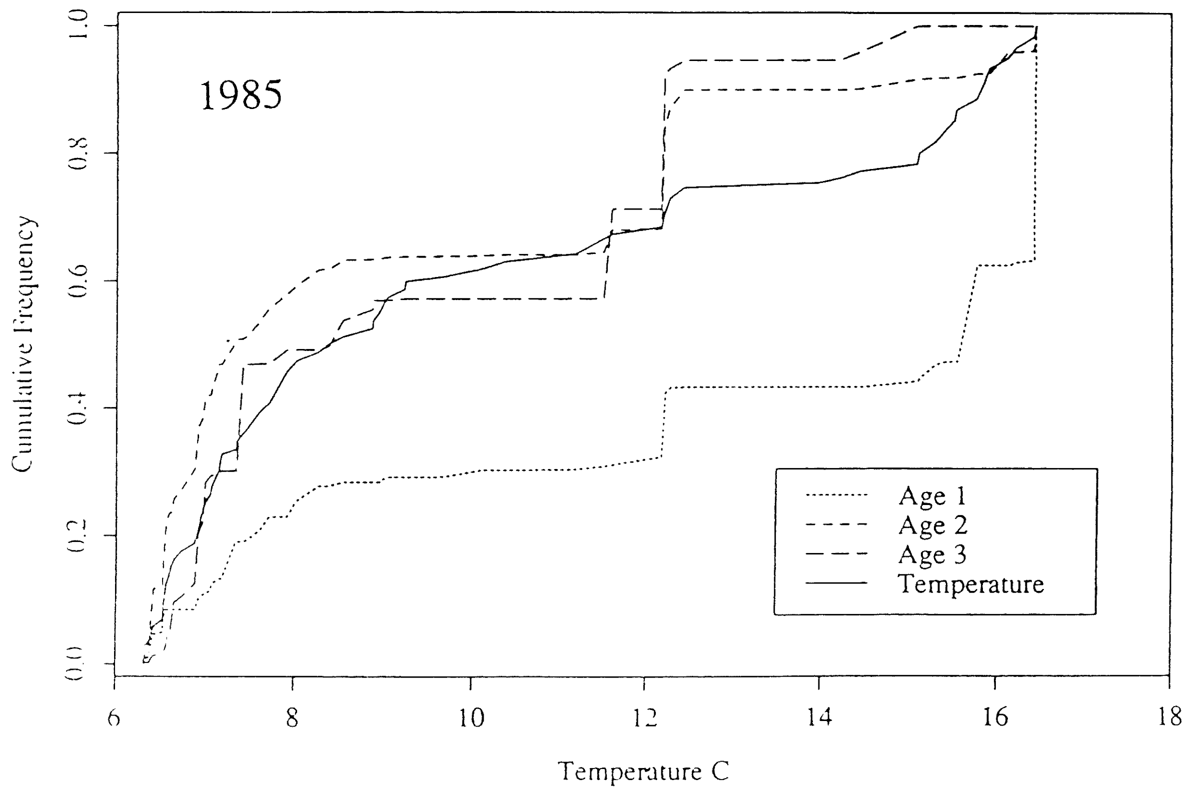
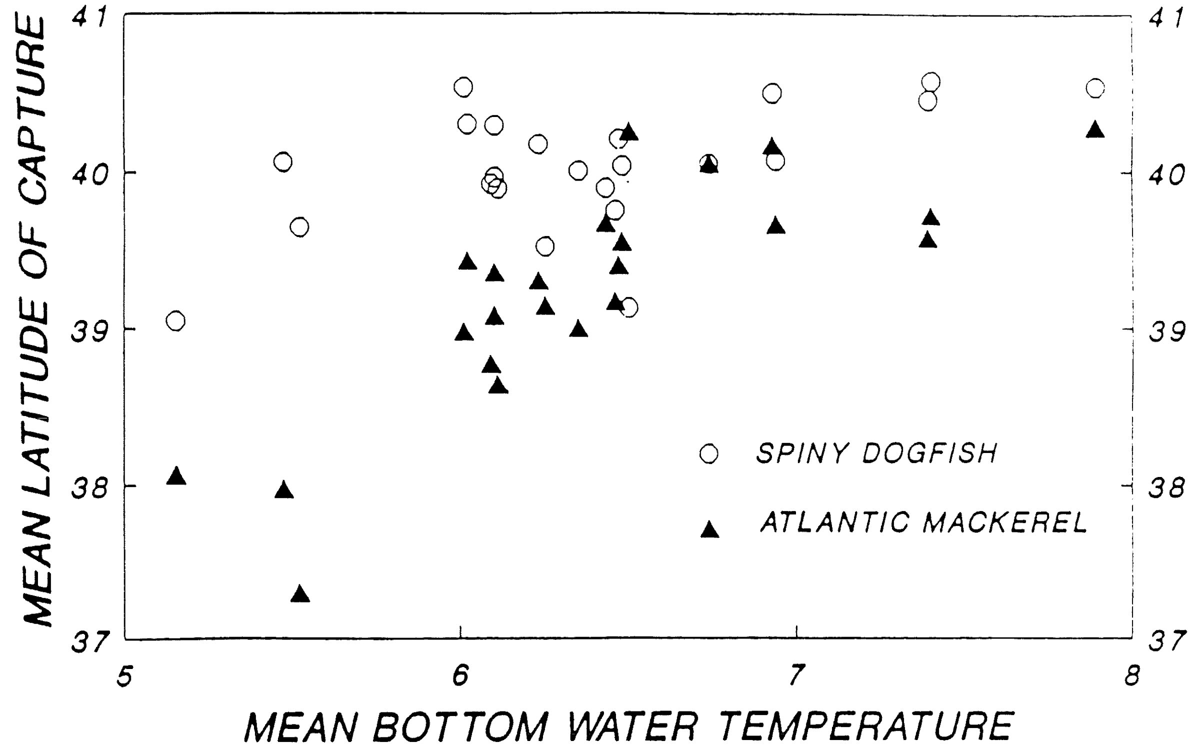


Figure 2.4.4.1. Empirical frequency distributions of bottom water temperatures taken in the English groundfish surveys of the North Sea, and the cumulative weighted frequency distribution of catches of cod age groups 1-3, in 1982 and 1985.



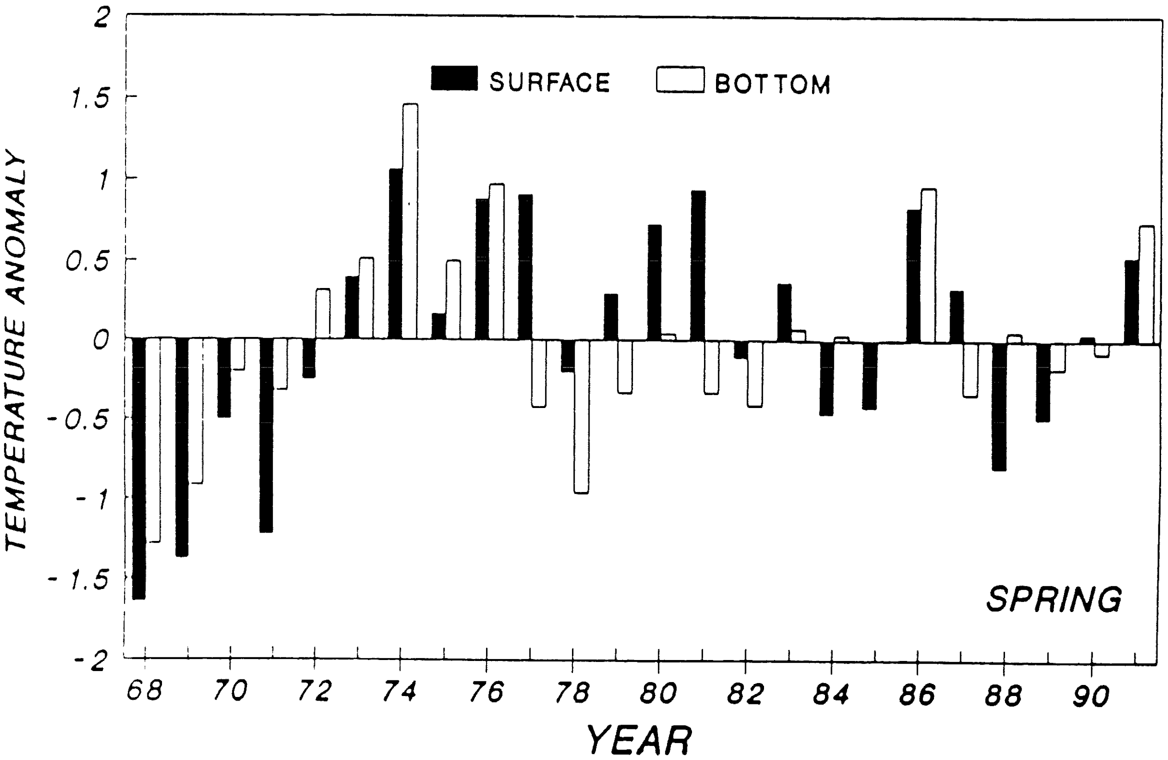


Figure 2.4.4.2. Latitudinal centroids of the distributions of spiny dogfish and Atlantic mackerel, in relation to geographically-weighted mean bottom water temperatures from USA bottom trawl surveys in Spring. Annual Spring temperature anomalies for the period 1968-1991 are given below.

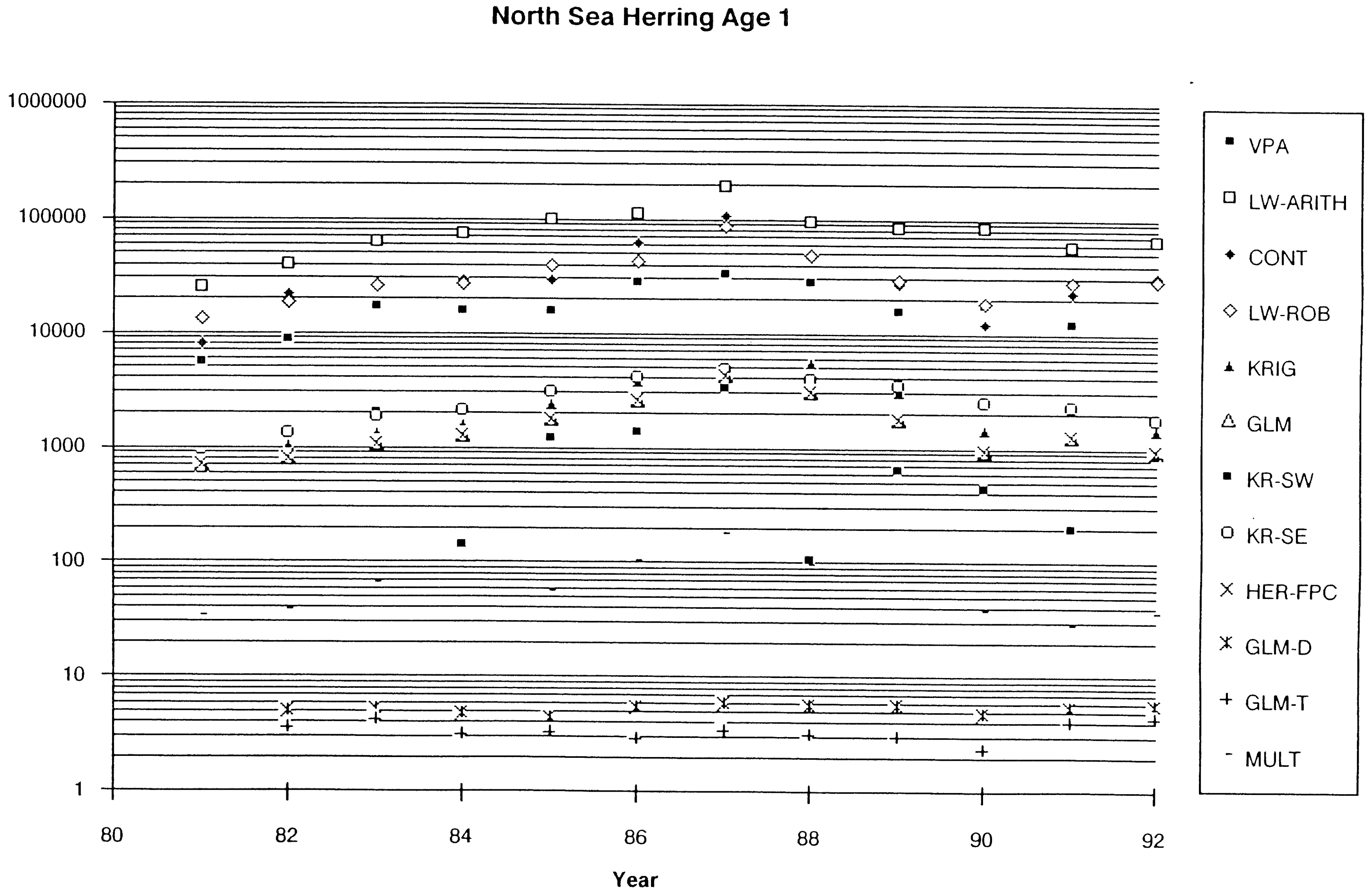


Figure 2.5.1.1. Abundance indices and VPA stock sizes for age 1 herring in the North Sea. Key to the indices is given in Table 2.5.1.1.

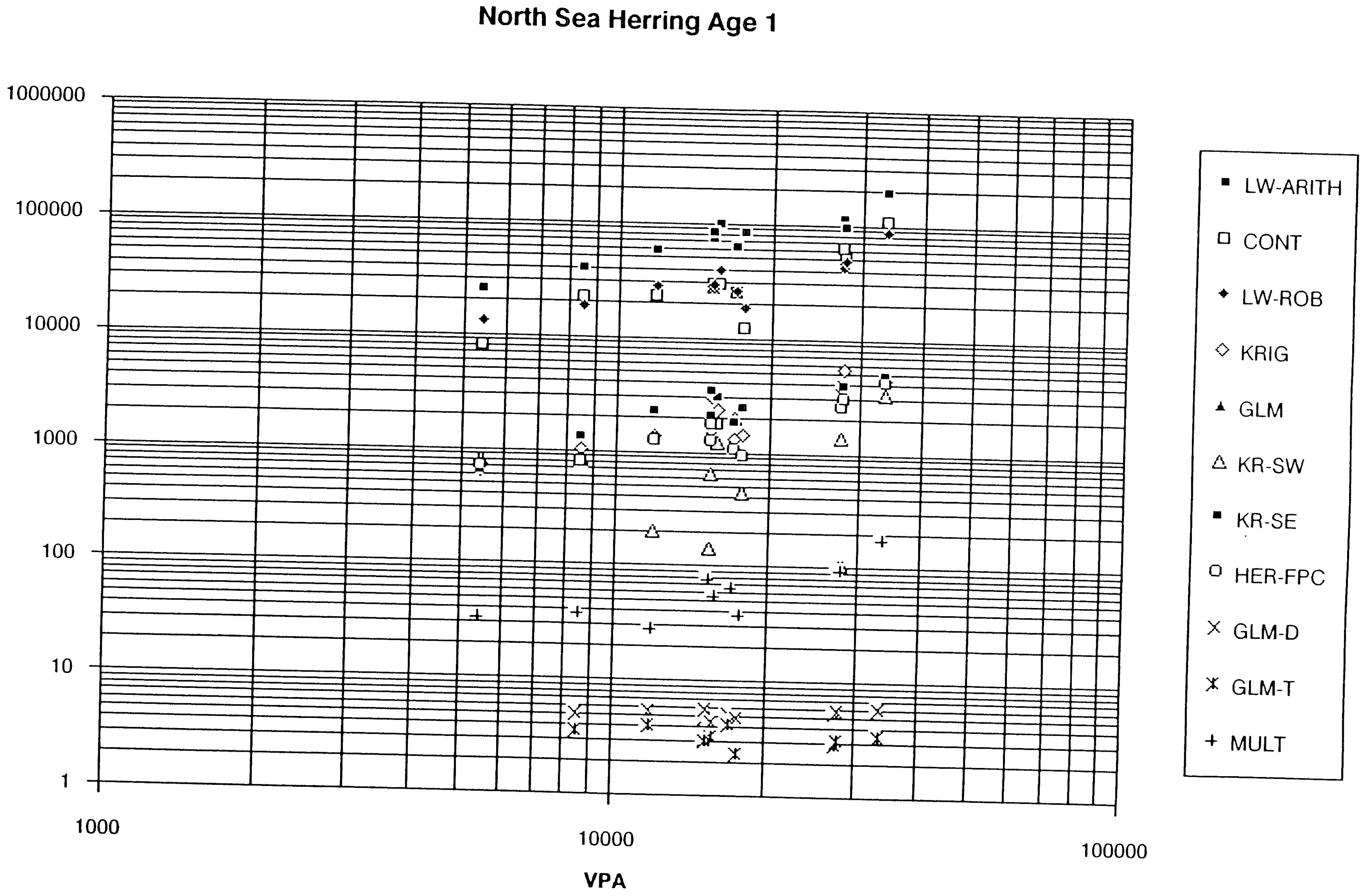


Figure 2.5.1.2. Regression of VPA stock size in numbers vs. abundance indices for age 1 herring in the North Sea. Key to the indices is given in Table 2.5.1.1.

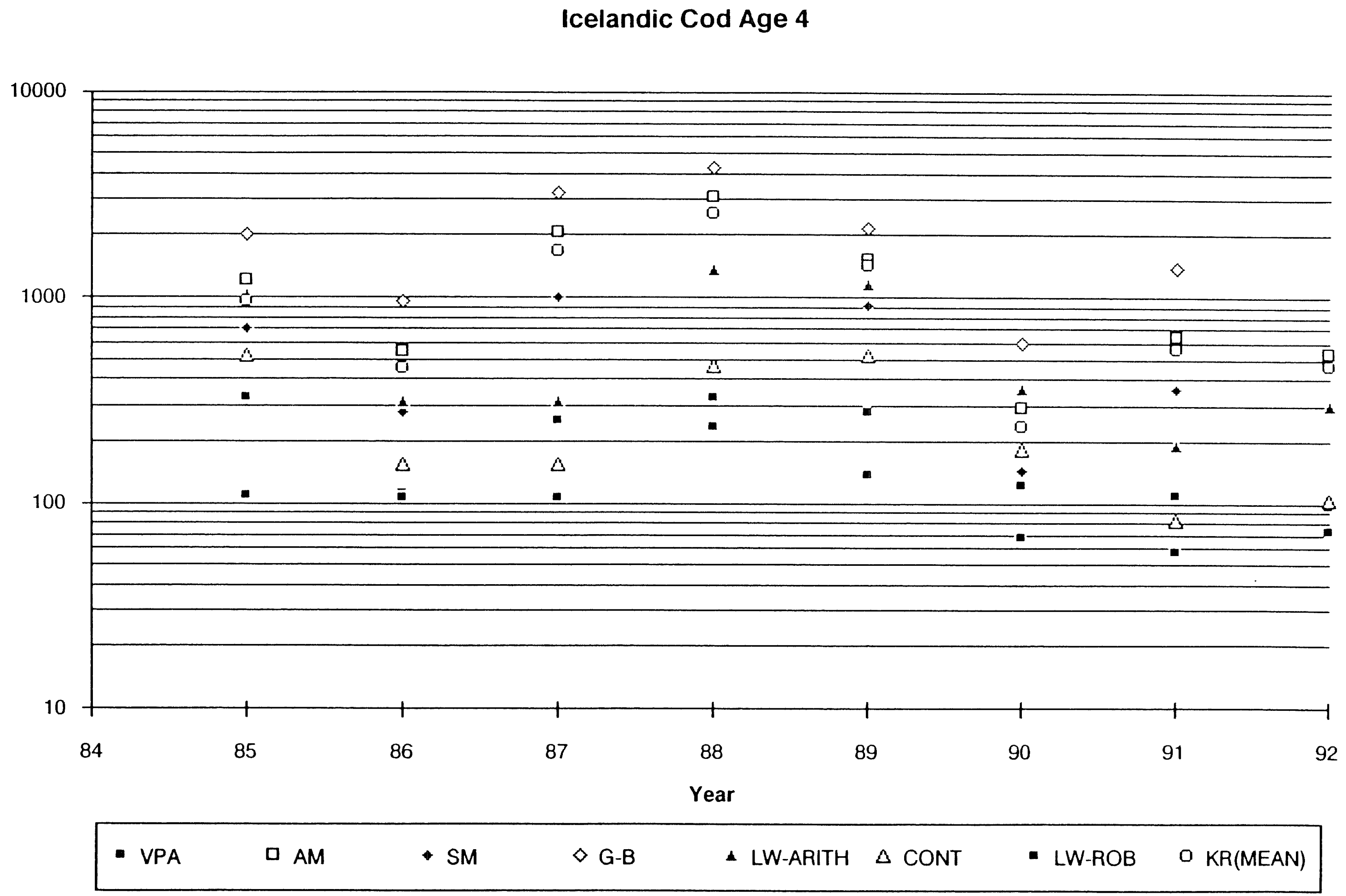


Figure 2.5.2.1. Abundance indices and VPA stock sizes for age 4 Icelandic cod. Key to the indices is given in Table 2.5.2.1.

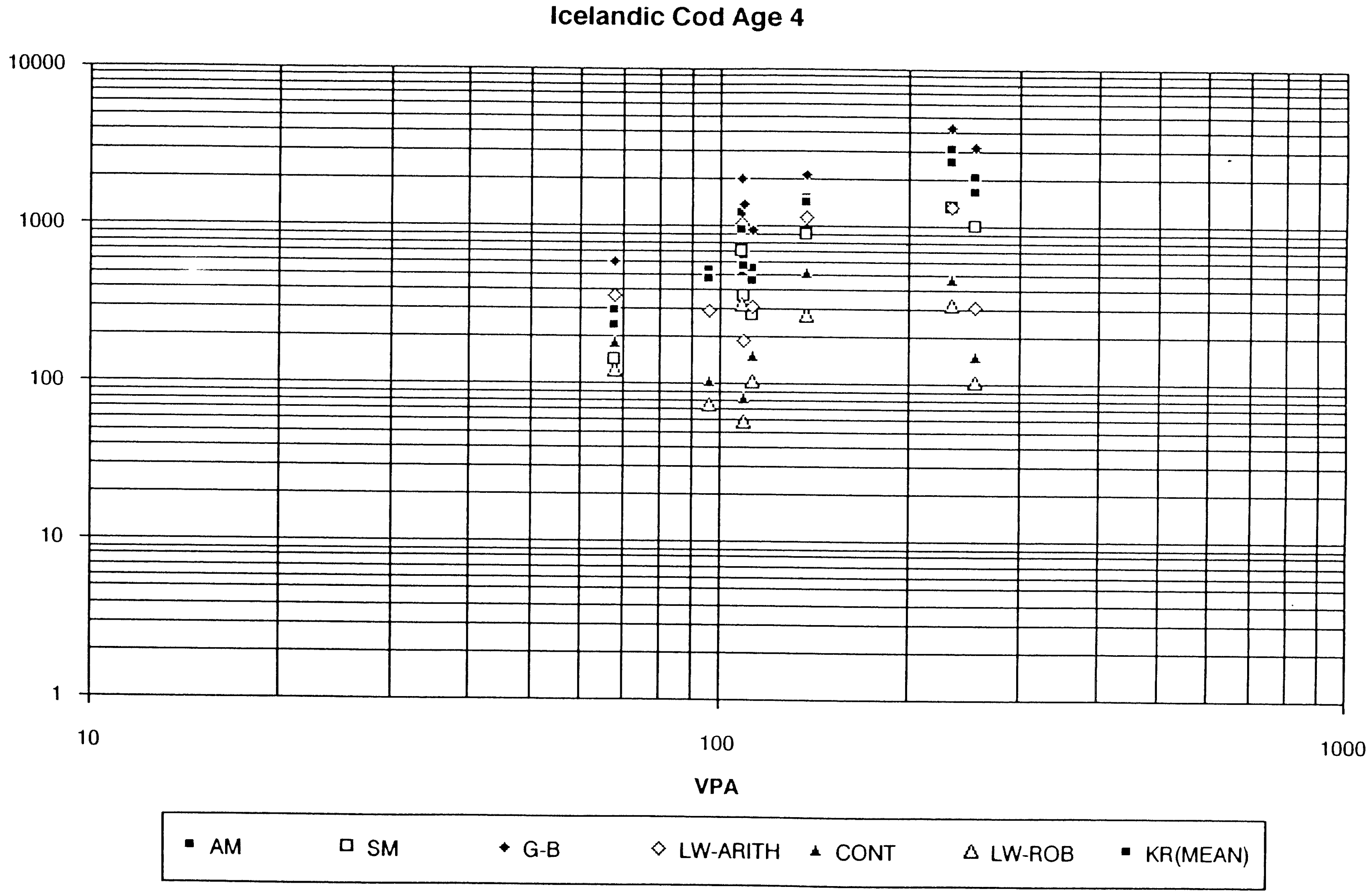


Figure 2.5.2.2. Regression of VPA stock size in numbers vs. abundance indices for age 4 Icelandic cod. Key to the indices is given in Table 2.5.2.1.

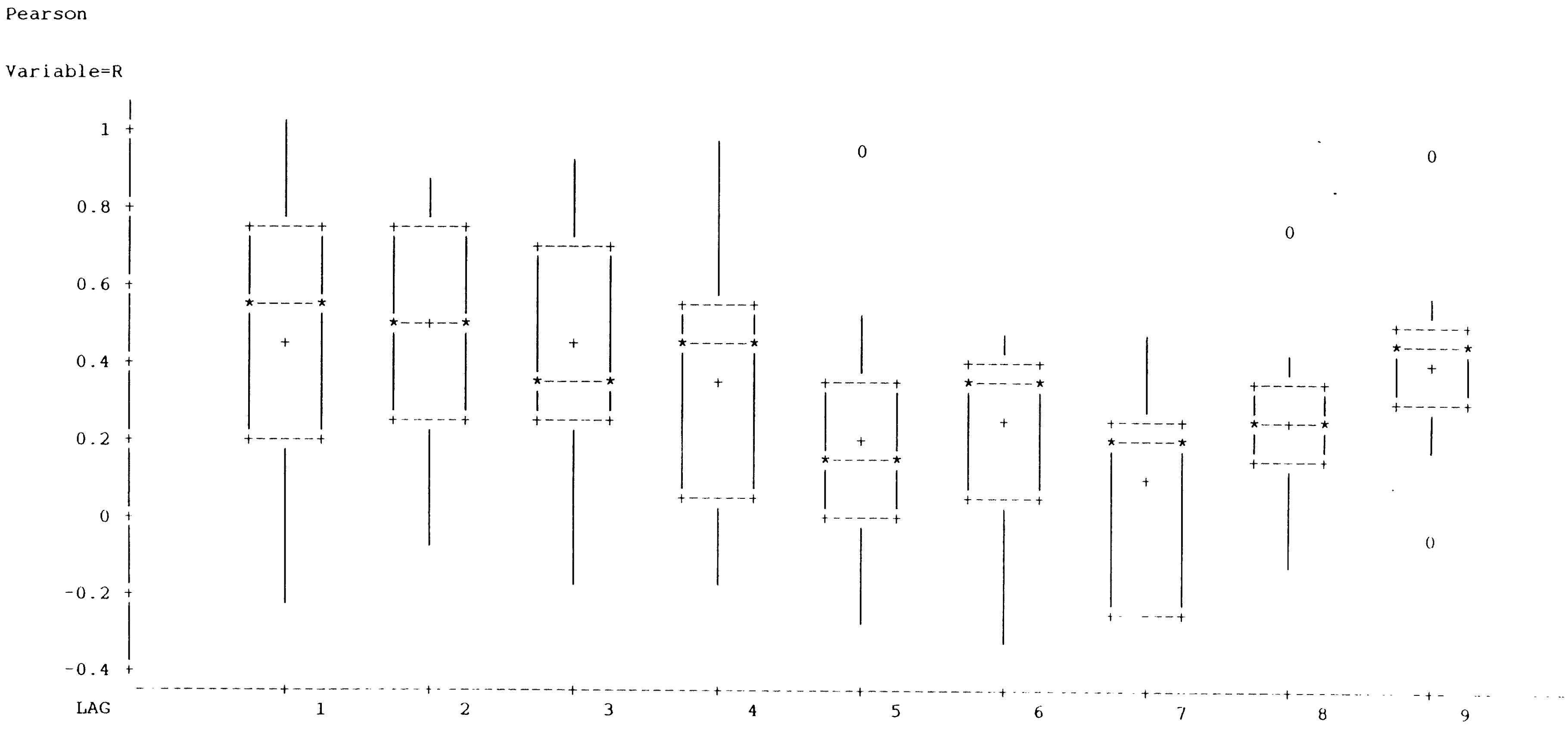


Figure 3.3.4.1. Between-station correlations over years (pearson correlation coefficient) at various time lags for southern Gulf of St. Lawrence cod.

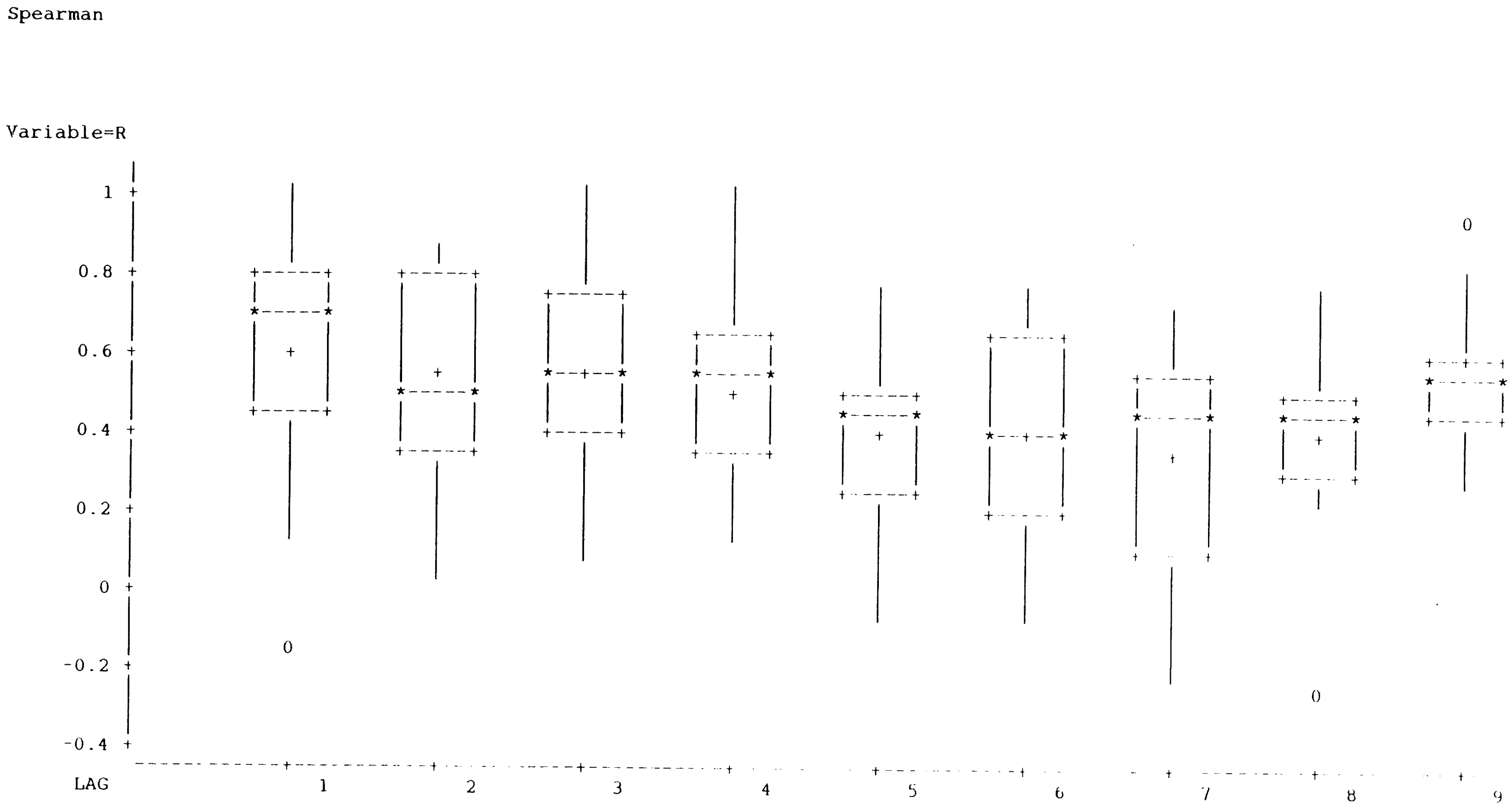


Figure 3.3.4.2. Between-station correlation over years (spearman correlation coefficient) at various time lags for southern Gulf of St. Lawrence cod.

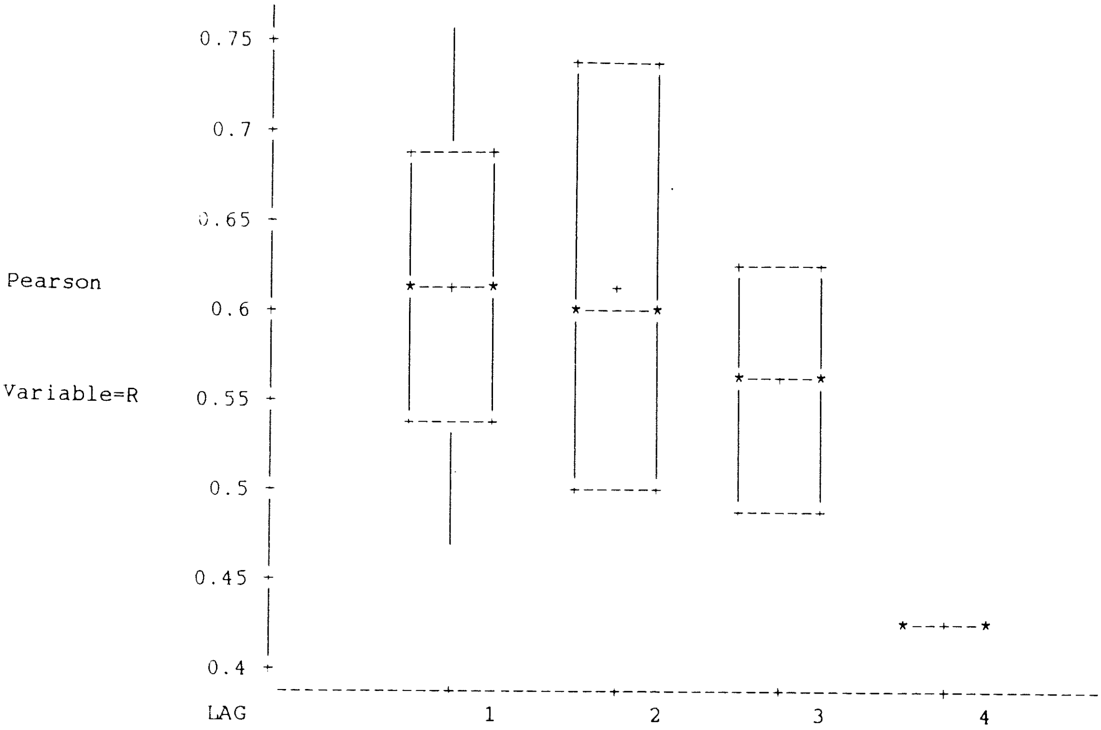


Figure 3.3.4.3. Between-station correlation over years (pearson correlation coefficient) at various time lags for age 1 cod in the English groundfish survey.

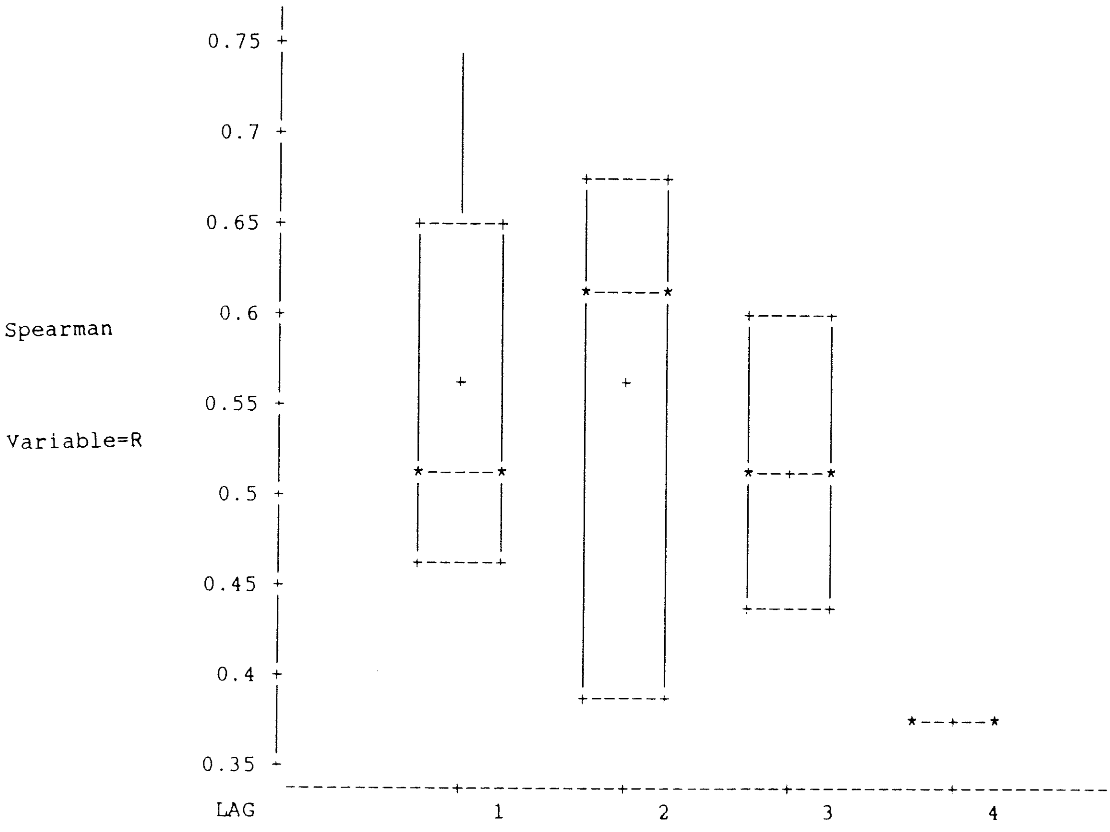


Figure 3.3.4.4. Between-station correlation over years (spearman correlation coefficient) at various time lags for age 1 cod in the English groundfish survey.

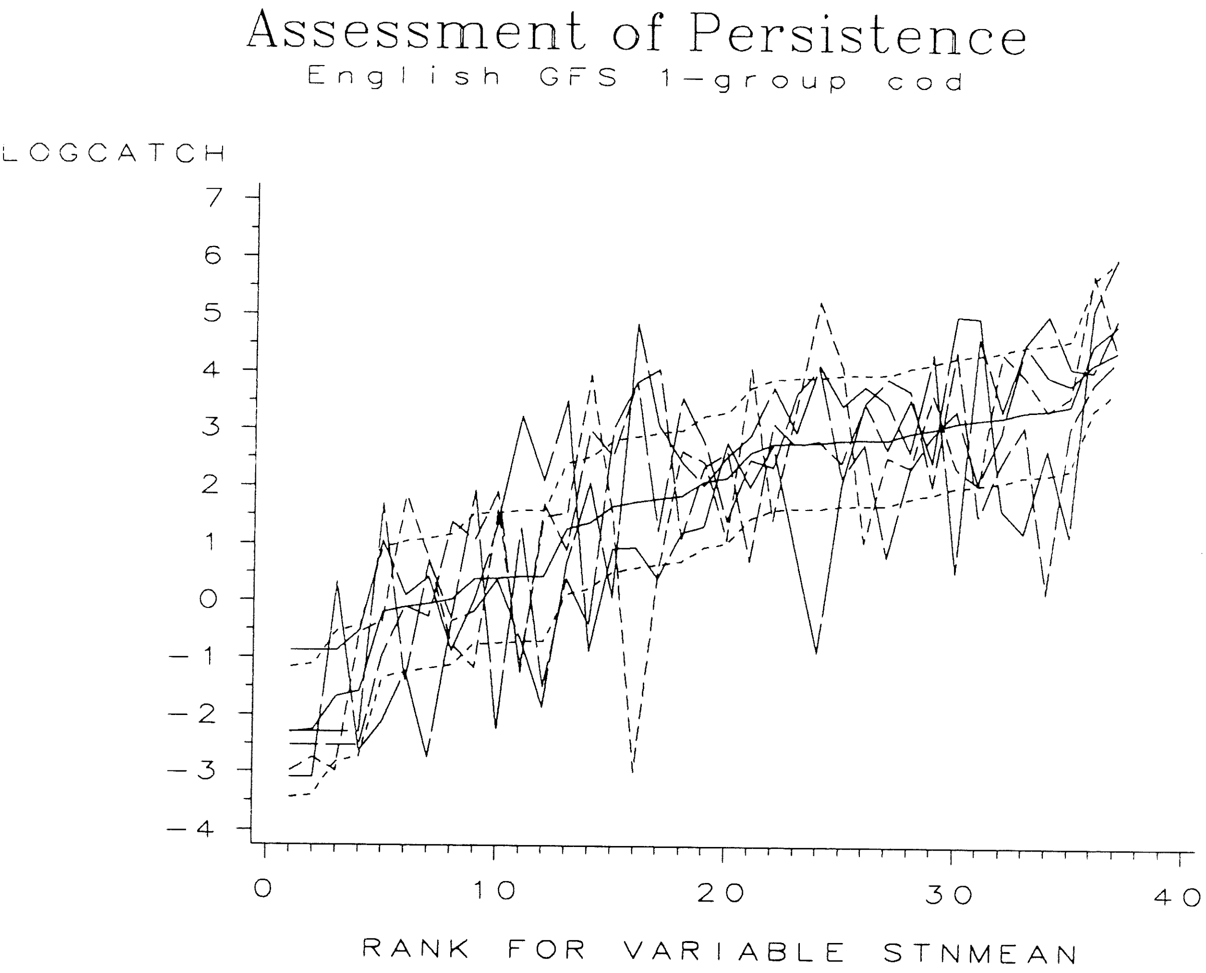


Figure 3.3.5.5. Plots of the persistence of station-by-station catches for age 1 cod in the English groundfish survey.

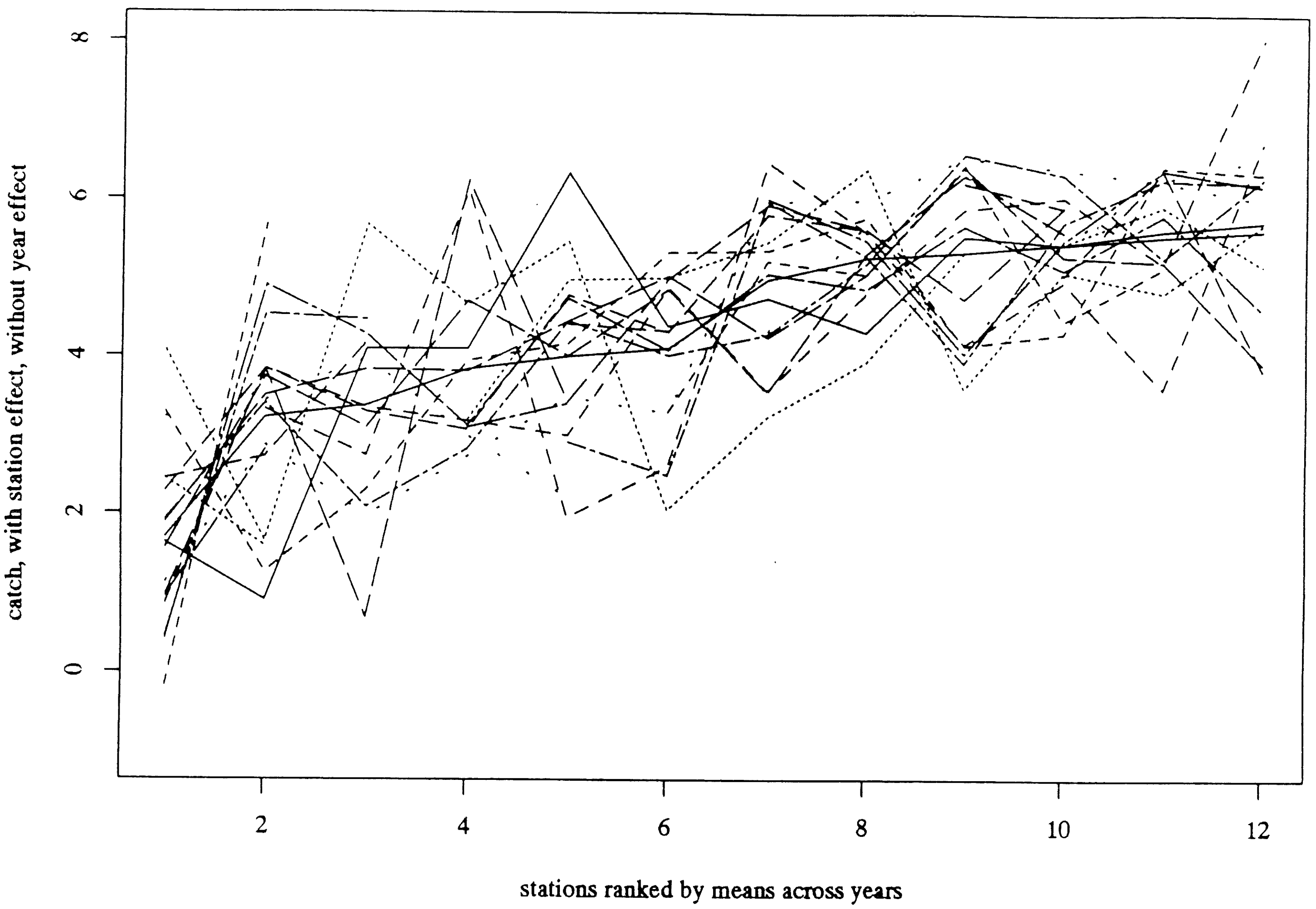


Figure 3.3.5.2. Plots of the persistence of station-by-station catches for cod in the southern Gulf of St. Lawrence, 1971-1988.

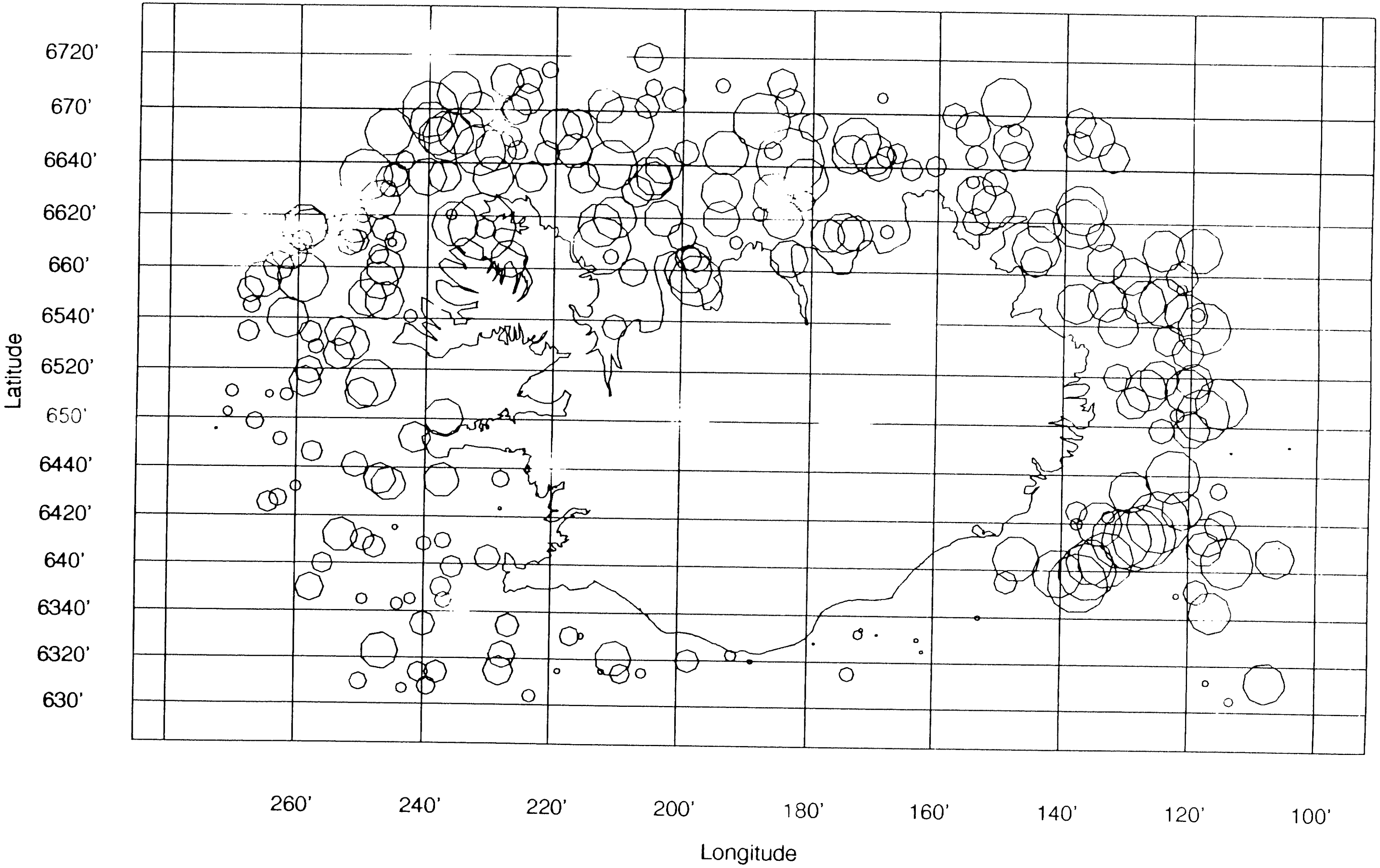


Figure 3.3.6.1. Slope of local population response at each random trawl station for age 4 cod in the Icelandic groundfish survey.

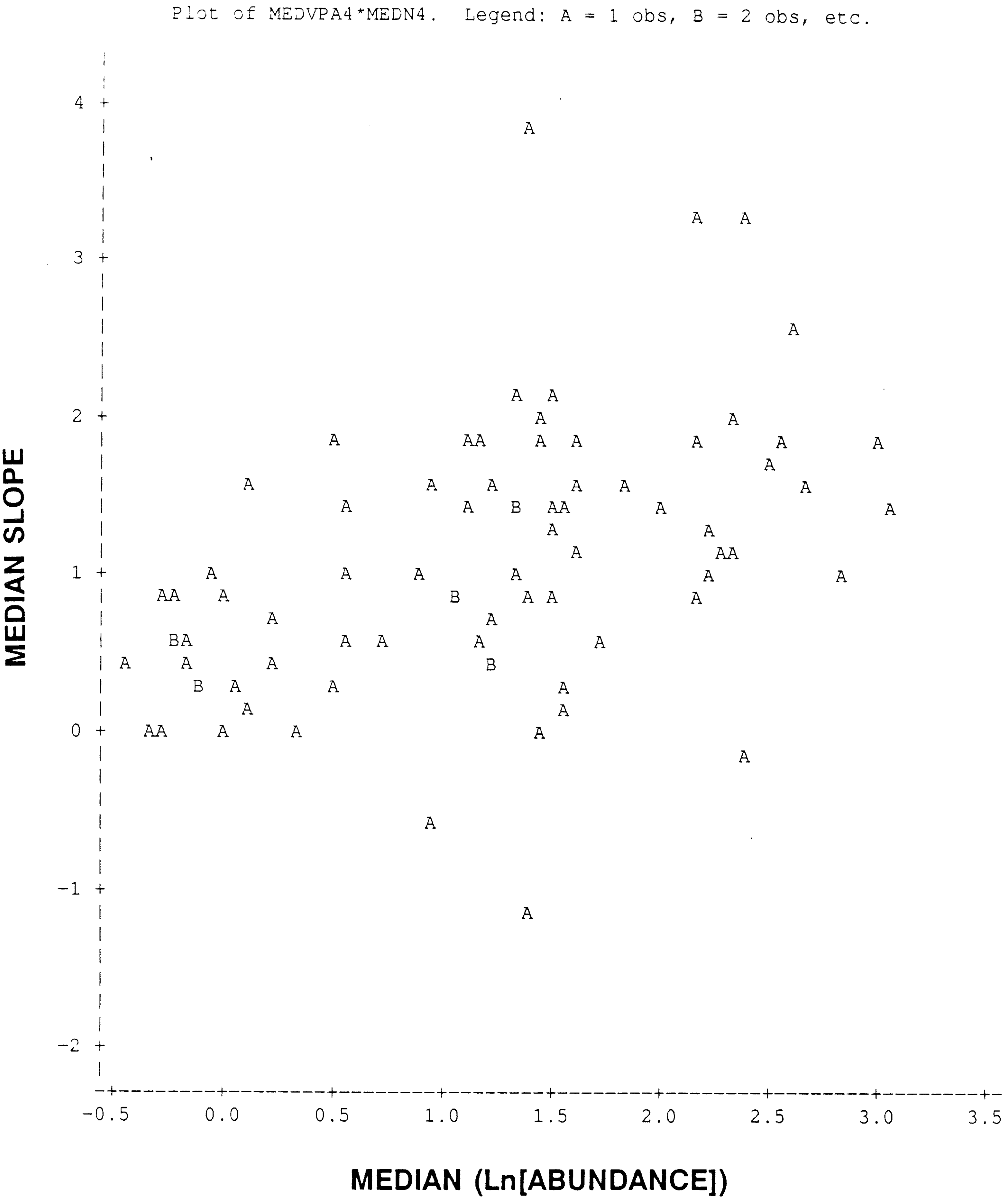


Figure 3.3.6.2. Relationship between the median slope and median Ln abundance by statistical square, for age 4 cod in the Icelandic groundfish survey.

APPENDIX A) List of Working Papers ICES Trawl Survey Workshop

WP-1	Gagnon	Constrained Optimal Stratified Random Sampling of Marine Populations
WP-2	Warren	Sampling with Partial Replacement
WP-3	Pennington & Godø	Measuring the Effect of Changes in Catchability on Variance of Marine Surveyabundance Indices
WP-4	Smith	Bootstrapping Quantities from Stratified Random Surveys
WP-5	Kinani, Fogarty & DeAljeris	The Effects of Post-stratification on the Precision of Stock Abundance Estimation for Atlantic cod, Haddock, Winter flounder and Yellowtail flounder on Georges Bank
WP-6	Sinclair & Chouinard	Application of Multiplicative Model to 4TVn Cod Research Survey Data
WP-7	Sinclair	Application of a Multiplicative Model to 4VsW Cod Research Survey Data
WP-8	Sparholt	Using GLM Analysis on the IYFS Herring Data for the North Sea
WP-9a	Forrester	A Trawl Survey Conversion Coefficient Suitable for Log Normal Data
WP-9b	Byrne & Forrester	Relative Fishing Power of NOAA R/V's Albatross IV and Delaware II
WP-9c	Byrne & Forrester	Relative Fishing Power of Two Types of Trawl Doors
WP-10	Smith & Nicholson	Investigations Into the Hydrographic Preferences of Cod: Evidence from the English Goosefish Survey
WP-11	Murawski	Factors Influencing the Spatial Distribution of Marine Fish Population, as Measured by Research Vessel Surveys
WP-12	Fogarty, Idoine, Almeida & Pennington	Modelling Trends in Abundance Based on Research Vessel Surveys

WP-13	Gavaris, Sinclair & Chouinard	Repeated Sampling in Bottom Trawl Surveys: A Case Study for 4TVn Cod
WP-14	Ona, Pennington, & Vølstad	A Dynamic Stratification Scheme for Trawl Surveys

APPENDIX B) Notation

NOTE: This standard (and largely mnemonic) notation is followed so far as possible, but not slavishly. Other usages and variations may be defined in the text. Array elements are denoted by means of either indices or suffices, whichever is more convenient. The same character may be used as both an index or a variable, if no confusion is likely.

Suffices and Indices

y	indicates	year
f	"	fleet
a	"	age
t	"	last (terminal) year
g	"	oldest (greatest) age group
l	"	length
k	"	year class
$	"	summation over all possible values of index (usually fleets)
#	"	summation over fleets having effort data
@	"	an average (usually over years)
*	"	a reference value

Quantities (all may have as many, and whatever, suffices are appropriate).

C(y,f,a)	Catch in numbers (including discards)
E(y,f)	Fishing effort
F(y,f,a)	Fishing mortality
F.(y,f)	Separable estimate of overall fishing mortality
q	Catchability coefficient (as in $F=qE$)
Y	Yield in weight
W.	Weight of an individual fish in the catch
B	Biomass
P	Population number (also fishing power)
E	Fishing effort
U	Yield or landings per unit of effort
C_w	Catch in weight of fish (including discards)
N	Stock in numbers of fish
F	Instantaneous fishing mortality rate
M	Instantaneous natural mortality rate
Z	Instantaneous total mortality rate
S	Selection coefficient defined as the relative fishing mortality (over age)
R	Recruitment
f	Relative F (e.g.,F/F*)
y	Relative yield (e.g., Y/Y*)
d	Fraction discarded
b	Fraction retained ($b=1-d$)
h	Hang-over factor
G	Instantaneous growth rate (in weight)

L	Landings in numbers (excludes discards)
I	Length
I_∞	Von Bertalanffy asymptotic length
K	Von Bertalanffy "growth rate"
r	Recruit index
MSY	Maximum sustainable yield
F_{msy}	Fishing mortality associated with MSY
E_{msy}	Fishing effort associated with MSY
B_{max}	Pristine stock biomass
m	Shape parameter for various surplus production models

APPENDIX C A Robust Weighted Mean Which Approximates the Median

Consider a robust weighted mean μ, using a hypobulic weighing function

$$w_i = 1/abs\ (x_i - \mu)$$

then

$$\mu = \frac{\Sigma x_i/abs\ (x_i - \mu)}{\Sigma\ 1/abs\ (x_i - \mu)}$$

and

$$\Sigma\ (x_i - \mu)\ /abs\ (x_i - \mu) = \Sigma\ sgn\ (x_i - \mu) = 0$$

If n+ is the number of observations exceeding μ, and n- that less than μ, this indicates that n+ = n-, so μ is the median of the data.

In practice this is not a practical procedure because if μ is at any time very close to any observation, the weighting locks onto this value and never departs (except via a zero-divide error!). To prevent this the hyperbolic weight needs to be bounded. This is most easily done by modifying the weight to be

$$w_i = 1/\ \{a + abs\ (x_i - \mu)\}$$

where a is some estimate of a scale parameter for the distribution. In this case all observations lying within ± a of μ are given roughly equal weight, while those further away are progressively down weighted. Thus μ departs from the median somewhat, and tends towards a trimmed mean.

A suitable measure of the scale parameter is the mean or median absolute deviation - for consistency this may as well be estimated in the same way a μ, ie, writing d_i for $(x_i - \mu)$

$$a = \Sigma\ w_i\ abs\ (d_i)/\Sigma\ w_i$$

where

$$w_i = 1/\{a + abs\ (d_i)\}$$

Both a and μ may be estimated via a very simple iterative reweighting algorithm, listed below.

[Note: in the subroutine rw(i) is the robust weighting, and w(i) is any extra weighting (e.g., distance weighting) required]

Maximum Likelihood Motivation

The use of a median - like estimate of the location parameters and a mean absolute deviation estimate of the scale parameters may be motivated (ie. justified in retrospect) by a maximum likelihood argument.

Consider data distributed according to the very long-tailed non-central exponential distribution: such data is very likely to include extreme observations.

The p.d.f. is

$$P(x) = \frac{1}{2a} \exp\{- \frac{1}{a}\ abs\ (x - \mu)\}$$

and the log-likelihood is

$$L = -n \ln\ (2a) - \Sigma \frac{abs\ (x_i - \mu)}{a}$$

thus

$$\frac{dL}{d\mu} = - \frac{1}{a} \Sigma\ sgn\ (x_i - \mu)$$

which is zeroed when n+ = n- and μ is the median.

Similarly

$$\frac{dL}{da} = -\frac{n}{a} + \frac{1}{a^2} \Sigma \; abs \; (x_i - \mu)$$

$$\therefore \; a = \frac{1}{n} \Sigma \; abs \; (x_i - \mu)$$

i. e., the Mean Absolute Deviation.

```
*    subroutine robmean (n,x,w,xbar,a,nloop)

     calculates robust weighted mean of a set of (possibly weighted) data points, and the
     robust weighted mean absolute deviation (a)

     written by J. Shepherd, June 1992

     implicit none

     integer n,nloop,i,loop
     real x(500),w(500),rw(500,d(500)
     real sx,sw,sd,xbar,a

     do loop=1,nloop
        sx=0.0
        sw=0.0
        do i=1,n
          rw(i)=1.0
          if (loop.gt.1) then
             rw(i)=a/(a+abs(d(i)))
          endif
          sx = sx + w(i)*rw(i)*x(i)
          sw = sw + w(i)*rw(i)
        enddo

        xbar=0.0
        if(sw.gt.0)   xbar=sx/sw

        sd=0.0
        do i=1,n
          d(i)=x(i)-xbar
          sd = sd + w(i)*rw(i)*abs(d(i))
        enddo

        a=0
        if(sw.gt.0)  a=sd/sw

     enddo    !   iterative loop

     return
     end
```

APPENDIX D Guidelines for Exchangeable Data Sets

The following guidelines were submitted as a suggestion. They were not reviewed by the workshop participants.

1) KEEP IT SIMPLE & STRAIGHT FORWARD

2) Provide single flat files in ASCII Text format (definitely not binary or application-specific except by prior agreement among the consenting adults involved). Avoid any special non-printable characters like the plague. Data in multiple files can usually only be read after additional programming by the recipient.

3) Keep the files reasonably small (not more than about 100 K bytes, ideally) since many editors, spreadsheets, etc., are size-limited. Sensible splitting of files (e.g., one per year or whatever) causes little difficulty, since they can usually be concatenated by a single operating system command, while splitting a large file may be a time consuming editing job.

4) Use a regular columnar format with data right-justified in the columns, and a fixed (numeric or text) format in each column. This makes it vastly easier to spot errors just by eye-balling the data.

5) Do not mix numeric and non-numeric formats in the same column. This will cause problems for many commercial and home-grown applications. Avoid special characters for missing values if possible (although these can usually be fixed by a global replacement edit, it saves time if this can be avoided).

6) Use a comma and a space as the field delimiter if possible. Many applications (especially spreadsheets and BASIC programs) have trouble with spaces as delimiters. Comma-separated files can be read by almost everything - the space assists enormously in readability (and therefore error checking).

7) Do not give unnecessarily large numbers of decimal places, or use unnecessarily wide columns (or trailing blanks). These make the files larger (and therefore, more difficult to handle) and the former makes it much more difficult to spot errors.

8) Do not use combinations or concatenated fields (such as dates as ddmmyy - especially not mmddyy!!). These will almost always have to be unscrambled programmatically unless the package has powerful data entry facilities. Keep data items clearly separated.

9) If possible give latitude and longitude as decimal degrees (positive for N and E, negative for S and W). Any other form is likely to require unscrambling (and this calculation is error-prone!).

10) Put any textual comments at the ends of the records, or at the bottom of the file, or in a clearly defined header.

11) Do not use more than one record-type, or records spanning multiple lines, except

by prior agreement.

12) Keep the records short. 80 characters is great (one can eyeball the whole thing on screen), 128 is OK, 256 is the maximum for some editors, etc.