ICES C.M. 1995/J:22
Baltic Fish Committee

International Council for the
Exploration of the Sea

ON THE DISCRIMINATION OF HERRING STOCKS IN DIVISION IIIA

by
Joachim Gröger
and
Tomas Gröhsler

Federal Research Centre for Fishery
Institute for Baltic Fishery
An der Jägerbäk 2
D-18069 Rostock

ABSTRACT

In 1994 two pelagic hydroacustic herring surveys were carried out by means of the two research vessels R/V Walter Herwig III and R/V Solea. The North Sea data were won during the Walter Herwig III survey in July 1994, the Baltic data were taken during the Solea survey in October 1994. A minor aspect of these cruises was to take account on the discrimination of herring from mixed populations in Division IIIa. In this context the North Sea data under consideration can be considered as representative for the characteristics of North Sea herring whereby the Baltic data represent the characteristics of Baltic herring. The basic idea is to use these as "pure" i.e. locally well separated learning samples which serve as proper information sources in order to model decision rules which are able to detect single individuals of a mixed herring population from Division IIIa either as North Sea or Baltic individuals. In principal, this can be done in different ways and by means of various variables with certain discrimination power. Two different stochastic methods were used here. The first approach is an inverted generalized linear model (GLM) and the second a discrimination rule. To keep it as simple as possible in terms of herring preparation the discrimination variables considered here are vertebral counts of herring. It is well known that also other variables (blood/genetic investigations etc.) have or could have a certain (may be higher) discrimination power but a trade-off between effort and accuracy usually has and had to be made here as well. Both sex/maturity indices and age/length compositions were correspondingly taken but did not play such a role for the discrimination.

INTRODUCTION

Catches of herring in Division IIIa (Kattegat and Skagerrak) are considered mainly to be a mixture of two spawning stocks:

- the Baltic/IIIa spring spawners (Rügen herring) and
- the North Sea autumn spawners.

The component of a local spring spawning herring in Division IIIa is of minor importance (Anon. 1991a).

The North Sea autumn spawners enter Skagerrak and Kattegat as larvae (Anon. 1977, Bartsch et al 1989, Johannesen and Moksness 1991) and migrate back to the North Sea with an age of 2-3 years (Anon. 1991a and Johansen 1927).

The Western Baltic herring enter Division IIIa through the Sound and Belt Sea after spawning on their feeding migration as 2 years of age (Aro 1989, Biester 1979 and Weber 1975) and spread out into the Western part of Skagerrak and the Eastern North Sea. Towards the end of the summer the herrings aggregate in the Eastern Skagerrak and Kattegat before they migrate to the main wintering areas in the southern part of Kattegat, the Sound and the Western Baltic (Anon. 1991a). Due to the mixing of the North Sea autumn spawners and the Western Baltic spring spawners in Division IIIa and IV, the assessment of these two stocks (Division IV, Division IIIa and Sub-Divisions 22-24) requires a method of stock seperation mainly for age groups 0-2. As a routine application in assessement this method in addition should require to be cheap and easy to handle so it can be used on a large scale. Many methods have been studied so far to differentiate between fish stocks:

- Analysis of mean vertebrae number (Anon. 1990, 1991a, 1991b, 1992a and 1993, Heincke 1898, Mann et al. 1983, Popiel 1956).
- Analysis of mean vertebrae number combined with a modal length analysis (Anon. 1988, 1989, 1994 and 1995, Hagström 1984, Rosenberg and Palmen 1981)
- Analysis of morphometric, meristic characters and maturity (Bohl 1962, Heincke 1898, McQuinn 1989, Ojaveer 1980, Petursson and Rosenberg 1982, Pope and Hall 1970, Rosenberg and Palmen 1981 and Schumacher 1967).
- Analysis of size and shape of otoliths (Anon. 1993, Bird et al. 1986, Campana and Casselman 1993, Kompowski 1969, Postuma 1974, Rauck 1964, Schulz 1967, Sosinski 1969).
- Analysis of otolith microstructure (Andersen et al 1969, Fossum and Moksness 1988, Gjoseter and Oiestad 1981, Moksness and Fossum 1991, Munk et al. 1991, Rosenberg and Lough 1977, Rosenberg and Palmen 1981,).
- Analysis of fatty acids (Grahl-Nielsen and Ulvund, 1990).
- Analysis of mitochondrial DNA (Dahle and Eriksen, 1990).
- Analysis of enzymes (Heath and Walker 1985, Jorstad and Pedersen 1986, Odense and Annand 1980 and Zenkin and Lysenko, 1977).
- Analysis of ^{137}Cs isotope (Rasmussen and Lassen 1994, Reinert et al. 1992)
- Analysis of parasite infections (Kühlmorgen-Hille 1983, Lubieniecki 1972, MacKenzie 1988, Stryzewska and Popiel 1974, Tshervontsev et al. 1994).

- Analysis of mark-recapture experiments (Ackefors 1978, Bakken and Ulltang 1972, Biester 1979, Haraldsvik 1967 and Weber 1975).

The mean vertebrae number analysis alone (Anon. 1990, 1991a, 1992a and 1993) or in combination with the modal length analysis (Anon. 1988, 1989, 1994 and 1995) are at moment the only methods which are practically used for management purposes to seperate the stocks in Eastern North Sea and Divsion IIIa. But still in some years these methods have failed to provide confirmation of the stocks concerned (Anon. 1992b).

In 1994 two pelagic hydroacustic herring surveys were carried out by means of the two research vessels R/V Walter Herwig III and R/V Solea. The North Sea data were won during a Walter Herwig III survey in July 1994, the Baltic data were taken during a Solea survey in October 1994. A minor aspect of these cruises was to take account on the discrimination of herring from mixed populations in Division IIIa. In this context the North Sea data under consideration can be viewed as representative for the characteristics of North Sea herring whereby the Baltic data under investigation represent the characteristics of Baltic herring. The basic idea is to use these locally seperated data as proper learning samples in order to model statistical decision rules which can indicate whether a laterly caught single individual from a mixed herring population of Division IIIa stems from a North Sea or a Baltic herring population.

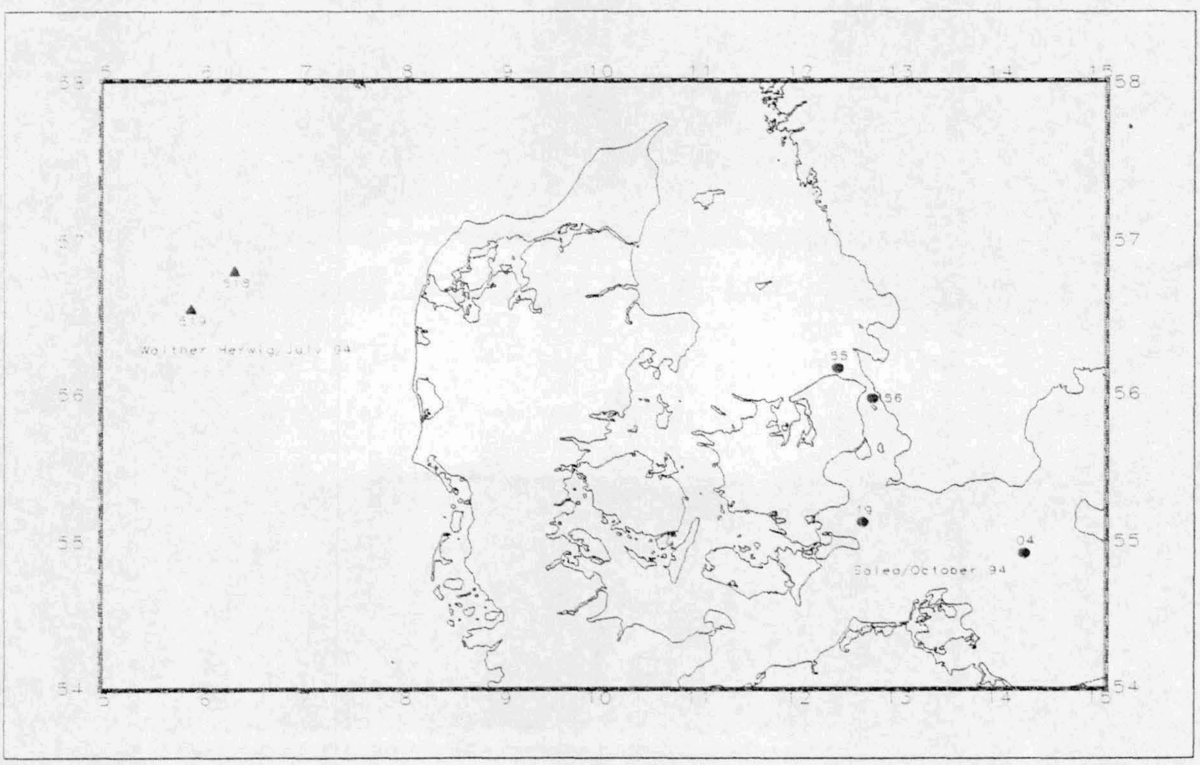

Fig. 1 Fishing Stations

MATERIAL AND METHODS

Data collection method

Herring samples were obtained by pelagic trawling. They were collected at 2 stations at 18th July in the North Sea and at 4 stations between 2nd and 17th October in the Baltic (Fig. 1). Sampling was conducted during time between 15:09 and 3:49 h at depths 24 to 53 m. Trawling periods lasted from 2 to 31 minutes at a trawling speed of 3.6 to 4.2 knots.

Tab. 1 Information on sampled stations

Area	Ship	Station-number	Catch position (Start) Latitude	Longitude	Date	Time	Water-depth (m)	Trawling Time (minutes)	Trawling speed (knots)
North Sea	W. Herwig	518	56°48.49'N	06°18.84'E	180794	149	53	30	4,10
North Sea	W. Herwig	519	56°33.70'N	05°52.73'E	180794	1505	53	2	4,10
Baltic Sea	Solea	4	54°55.07'N	14°11.72'E	21094	1709	42	31	4,20
Baltic Sea	Solea	19	55°07.81'N	12°35.42'E	71094	349	24	30	4,00
Baltic Sea	Solea	55	56°10.11'N	12°20.08'E	171094	1708	24	31	4,10
Baltic Sea	Solea	56	55°58.15'N	12°41.14'E	171094	2005	31	31	3,60

Samples on board of R/V 'Walther Herwig' were taken by a 1600# pelagic trawl. R/V 'Solea' used a pelagic mid-water trawl 'Blacksprutte' with a circumference of 854 meshes of 200 mm bar length in the opening and 10 mm bar length in the codend. Table 1 gives information concerning the trawl stations with detaills of catch-position, time of day, water depth, trawling period and trawling speed. Directly after sampling the total length of the herring was measured to the lower half centimetre. In July it was intended to collect 10 individuals per 0.5 cm group. In October the size group sampling was designed for assessment purposes. Table 2 shows the length frequency distribution of sampled herring per station. 234 and 445 herring were collected in July and October respectively (Table 2). Also total weight (g) of all collected herring per 0.5 cm length group was determined. The herring sampled by size group and station was then deep-frozen on board at -25 °C for analysis later on. After thawing the herring in the laboratory the maturity was estimated, the otoliths were taken for later age reading and the total vertebrae number counted. The VIII degree maturity scale of Heincke (1898) was used. For vertebrae counting all the flesh was cut off to make the vertebrae visible. The number was determinated in counting the space between the vertebrae. As the last vertebra was not quite visible one number was added after counting the last visible space.

Tab. 2 Length frequency of herring per station

Length (mm)	North Sea		Baltic Sea			
	518	519	4	19	55	56
90			2			
95			2			
100			1		2	
105			1		2	
110			2		2	
115			2		2	
120			2		1	
125			2		2	
130			2		2	
135			2		2	
140			2		2	
145			2		2	
150	6		4		2	
155	7					
160	10		1		1	
165	10		3			
170	10		5		1	
175	10	5	14		1	
180	10	9	20		5	
185	10	10	19		5	
190	10	9	15		5	
195	10	8	7		5	
200	9	10	14		5	
205	10	10	10		5	
210	10	10	9		5	
215	8	10	4		8	
220	1	8	6		9	
225	1	8	3		5	

Tab. 2 Length frequency of herring per station

Length (mm)	North Sea		Baltic Sea				
	518	519	4	19	55	56	
230	1	1	4		5		
235		1	1		5		
240			2		5		
245			3		6		
250		1	3	2	7	10	
255			1	4	3	7	
260		1	2	3	2	10	
265			1	4		10	
270				5	2	10	
275			2	3	1	10	
280			2	3		10	
285				1		10	
290			2	3		10	
295				4		8	
300				1		5	
305				1		5	
310						5	
315					1	3	
320						1	**Total**
Sum	133	101	179	34	118	114	**679**

Statistical analysis

To be able to distinguish more objectively, i.e. statistically between the two different herring populations two sligthly different simple mathematical approaches were used. The first is based on a generalized linear model, the second on a discriminant analysis. The results of both approaches can be taken to use them by their own or to compare them with the results of the other method in order to confirm each other.

Variable and data selection:

In order to lead to unbiased results the variables used for detecting any discrimination are only allowed to show pure as well as significant differences in the biology of the two different herring populations and are not wanted to reflect differences due to different sampling frame conditions in space and time. From that point of view we have not included age or length indices as variable(s) into our approach despite the fact that any inclusion of age/length composition data of the two different herring populations may lead to higher discriminatory results. The two main reasons are: at first age/length compositions were found to be completely different in july (North Sea, R/V Walther Herwig III) and october (Baltic, R/V Solea). In the North Sea we found a range of 1 to 4 age/length groups with a varying number of individuals per age/length class, in the Baltic 1 to 9 age/length groups[1]. This might be the normal case but also could be a random artefact induced by different frame conditions of the two cruises. A second reason is that no uniform interpretation of age rings exist between the different departments concerned with age reading of North Sea and Baltic samples. Sex and maturity indices were also excluded from any further analysis since no significant discriminatory effect of these two variables could be detected as pre-investigations showed. Therefore, the measured variables left here as useful for inclusion into all statistical analyses are vertebra counts and area index (code for North Sea and Baltic, description see below).

Regression approach:

In order to receive a proper discrimination between North Sea and Baltic herring in Division IIIa it may be helpful to formulate the following questions: Have the area conditions (= independent or exogenous variable) in North Sea and Baltic any effect on the vertebra numbers (= dependent or endogenous variable) of the corresponding herring populations on average? If yes, how can a single herring or a larger number of individual herring (laterly sampled) be identified as North Sea or Baltic herring on the basis of this model? Mathematically this can be written as

$$\textit{vertebra counts} = \textit{ordinate} + \textit{slope} \times \textit{area} + \textit{residuals}$$

or **(1)**

$$y = a + b\,x + u$$

which is equivalent to a simple linear regression approach with coefficients a (ordinate of regression line) and b (slope of regression line). Since matrix notation helps to make unidimensional problems easier to generalize this approach can be rewritten in matrix algebra as

1 From the North Sea sample age class 4 was dropped since it consisted of only one individual with an extreme high number of vertebra biasing the overall result. From the Baltic sample the age groups 8 and 9 were excluded since the corresponding coefficients of variation indicated a much higher variability of vertebra counts compared to all other age groups as pre-investigations showed.

$$y = [1 \ x] \begin{bmatrix} a \\ b \end{bmatrix} + u \qquad (2)$$

$$= X \quad \beta + u$$

where the design matrix X now consists of a variable x (area code) and a vector of ones. The regression coefficients a and b are now contained in the column vector ß. If one would consider the first four actual Baltic herring vertebra counts and the last four from the North Sea y variable (vertebra counts) and design matrix X would then be

$$\begin{bmatrix} 56 \\ 55 \\ 55 \\ 54 \\ \vdots \\ 57 \\ 57 \\ 57 \\ 56 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \beta + u \qquad (3)$$

The first column in X consisting of ones is necessary in order to calculate the value of the ordinate a. The second column in X represents the area variable. It is a binary or dichotomous indicator variable (also called dummy variable) where "1" means "Baltic" and "0" "not Baltic" (i.e. "North Sea")[2]. Dummy coding is used here in order to permit a simple model formulation as well as easier model transformations. An obvious reason is that the model can be easily inverted as follows after a and b having calculated

$$\hat{area} = \frac{\hat{a} - vertebra\ counts}{\hat{b}} \qquad (4)$$

The hats on a and b mean that these are estimations. Once the coefficients are estimated, they are known and can be handled as constants within mathematical operations. The minus sign results from the coding of the model combined with the fact that the average number of vertebrae of Baltic herring is expected to be smaller than

2 Both columns can also be interpreted as selection variables selecting all (first column) or only a part (second column) of the values from the y variable.

for North Sea herring. This results in a negative slope of the regression line.

The inverted model enables us to decide whether a newly sampled single herring caught in the Kattegat or Skagerak (for which the actual area of origin is unknown) belongs to a North Sea or a Baltic herring population, only on the basis of counted vertebra numbers. Simply spoken, the regression line works as a pointer which indicates the most probable area.

The estimated ordinate $\hat{a}$ in equation (4) is nothing else than the average of the vertebra counts when the area code is "0"[3]. I.e., it is the mean value of vertebra counts of the North Sea herring learning sample. This follows from the fact that the regression line does not only go directly through the overall means of the variables area code and vertebra counts but also through the partial means of the North Sea and the Baltic sample[4]. The estimated slope $\hat{b}$ can be interpreted as the difference between the average of the vertebra counts of the North Sea sample and that of the Baltic[5]. The better the discrimination the larger will be $\hat{b}$. Three different main solutions are possible:

- a value of the vertebra counts which equals $\hat{a}$ exactly would result in a value "0" of the formula leading to the conclusion that this individual belongs to a North Sea population

- a difference of ($\hat{a}$ - vertebra number) which equals $\hat{b}$ exactly would result in a value "1" of the formula which leads to the conclusion of belonging to a Baltic population[6].

---

3 To see this consider the North Sea case when area code is 0, i.e.

$$\begin{aligned} \textit{vertebra counts} &= \hat{a} - \hat{b} \times \textit{area} \\ &= \hat{a} - 0 \\ &= \hat{a}.. \end{aligned}$$

4 This is due to the fact that using the individual values for vertebra counts and area code of the original samples is equivalent to using the separately computed means of the North Sea and the Baltic sample in order to calculate a regression line through 2 points.The procedure can be summarized as follows:

- Calculate the 2 averages for vertebra counts and the 2 averages for area code of the North Sea and the Baltic learning sample separately whereby the mean for the variable area code of the North Sea sample will be 0 and that of the Baltic sample 1.
- Calculate a regression line with only 2 points through the 2 means of the variable vertebra counts.

5 Since the results depend strongly on local means the stability of the results is ensured if the vertebra numbers of North Sea / Baltic herring keep constant on average.

6 Note that these two cases are integer solutions which go conform with the original (0,1)-coding of the variable area.

- usually, we will not find the above two extreme situations in reality. The following case will be the most probable: for a single North Sea herring we will normally find a vertebra number more or less far from the average of the learning sample (the same applies to the Baltic herring). Since that a resulting value of the formula near "1" (i.e. larger than 0.5) indicates a Baltic herring whereby a value near "0" (i.e. smaller than 0.5) indicates an individual belonging to the North Sea population[7].

In order to ensure the quality of the estimated parameters and to verify the fitted model, respectively, significance tests and calculations of other measures (for instance r^2, confidence intervals etc) have been carried out.

Discriminant analysis:
In this case, the objective of applying the discriminant analysis is to find a decision rule which enables us to allocate a newly sampled herring from Skagerak/Kattegat on the basis of its vertebra number either into the North Sea or into the Baltic herring group. The main point is to calculate discriminant functions on the basis of two pure learning samples of vertebra counts (Baltic, North Sea). In terms of vertebra counts as criterion these learning samples must be as far away from each other as possible. Or equivalently, instead of calculating discriminant functions one can calculate two related distance functions. The latter measure the number of vertebra of the newly sampled herring (x_{new}) as difference from the mean vertebra count of either the Baltic ($\overline{x}_{Baltic}$) or the North Sea herring population ($\overline{x}_{N.S.}$). These two distance functions are:

$$d_{Baltic}(x_{new}) = -\frac{1}{2}\,(x_{new}-\overline{x}_{Baltic})'\,\Sigma_{Baltic}^{-1}\,(x_{new}-\overline{x}_{Baltic}) - \frac{1}{2}\ln|\Sigma_{Baltic}| + \ln p(Baltic)$$

$$d_{N.S.}(x_{new}) = -\frac{1}{2}\,(x_{new}-\overline{x}_{N.S.})'\,\Sigma_{N.S.}^{-1}\,(x_{new}-\overline{x}_{N.S.}) - \frac{1}{2}\ln|\Sigma_{N.S.}| + \ln p(N.S.) \qquad (5)$$

The basic idea is to allocate the single herring into that group which receives the highest probability of being allocated by the decision rule i.e. for which the difference between mean and vertebra number of a single herring is smallest or for which the distance function is largest. The corresponding decision rule is :

---

7 Note that this is a non-integer solution which does not go conform with the binary definition of the variable area. Anyhow, by inserting the vertebra counts of a single herring of a mixed population from Division IIIa in eq. (4) the outcome can be alternatively interpreted as that probability with which this herring belongs to a Baltic herring population (in such a case some operation on eq. (4) has to restrict it on the interval [0,1] in order to fulfill the properties of a probability function). Or by inserting an average number of vertebra counts of a Division IIIa sample of herring in eq. (4) the outcome can be interpreted as that fraction of herring that belongs to a Baltic herring population.

allocate a newly sampled single herring into the Baltic herring group if

$$d_{Baltic}(x_{new}) > d_{N.S.}(x_{new}) \qquad (6)$$

and into the North Sea herring group vice versa.

The two distance functions above are based on the Bayes rule (Hartung et al. 1987). They are also quadratic since inhomogenous learning samples from North Sea and Baltic will be expected. Quadratic forms are always more complicated than linear, also their interpretation. It might be better to reduce the numerical effort and to simplify the interpretation by using linear forms. In order to check whether we can do so or not a likelihood-ratio-test on the uniformity of the two [1x1] within covariance matrices $\sum_{Baltic}$ and $\sum_{N.S.}$ will be applied. Null and alternative hypotheses of such a test are

$$H_0\text{: } \Sigma_{Baltic} = \Sigma_{N.S.} \text{ versus } H_1\text{: } \neg H_0 \qquad (7)$$

The homogeneity of the data will be assumed under the null hypothesis. The corresponding test statistic is

$$-2 \ln \lambda = N \ln|\Sigma| - N_{Baltic} \ln|\Sigma_{Baltic}| - N_{N.S.} \ln|\Sigma_{N.S.}| \qquad (8)$$

where $\sum$ is the pooled [1x1] covariance matrix of both data groups (Baltic and North Sea together). This test statistic is approximately χ^2 distributed. If the null hypothesis of homogeneity will not be rejected we are allowed to use the easier linear distance functions (see Lütkepohl 1992).

There is also an a-priori probability term in the Bayes formula by which the bias due to different sizes of the Baltic and the North Sea learning samples can be considered and reduced.

The quality of the decision rules will be checked by calculating classification rates on the basis of posterior probabilities. The data used for this evaluation will be selected either by jackknifing (which leaves out exactly one single herring from the calculation of the decision rules) or by bootstrapping (which excludes randomly a larger subset of herring data from the calculation of the decision rules)[8]. From these results a classification matrix and error counts will be derived.

---

8 A single herring from the excluded data set with vertebra number x_{excl} will be allocated, for instance, into the Baltic herring group only if the posterior probability for this is larger than 0.5. This is formally given by eq. (9). This result will be compared with the known true membership. If the comparison fails it counts as misclassification. See also eq. (6). Something similar appears for the North Sea herring membership.

$$P(Baltic \mid x_{excl}) = \frac{e^{(d_{Baltic}(x_{excl}))}}{e^{(d_{Baltic}(x_{excl}))} + e^{(d_{N.S.}(x_{excl}))}} > .5 \qquad (9)$$

RESULTS

Results of the linear regression

While the first part of table 4 shows the overall means and corresponding cv's[9] of herring vertebra counts for North Sea (areacode=0) and Baltic (areacode=1) the second part displays the average vertebra counts of herring and their cv's splitted up by age. The total of 679 items (see table 2) was reduced downto 631 items by the following procedure: only age groups with more than three individuals and a cv smaller or equal than 1.5% have been included,

- for the North Sea age groups 1 to 3 with 234 individuals (4 age groups have been sampled in total)
- for the Baltic agegroups 1 to 7 with 397 individuals (10 agegroups have been sampled in total).

Tab. 4 Average vertebra counts by age and area (North Sea: areacode = 0, Baltic: areacode = 1)

Analysis Variable : VERTEBRA COUNTS

AREACODE	N Obs	Mean	CV
0	228	56.50	1.18
1	392	55.79	1.41

AREACODE	AGE	N Obs	Sum	Mean	CV
0	1	121	6838.0	56.51	1.25
	2	101	5703.0	56.47	1.11
	3	6	341.00	56.83	0.72
1	1	45	2507.0	55.71	1.19
	2	85	4743.0	55.80	1.46
	3	87	4865.0	55.92	1.50
	4	79	4400.0	55.70	1.39
	5	46	2561.0	55.67	1.42
	6	33	1845.0	55.91	1.29
	7	17	951.00	55.94	1.18

The total of 631 items was further decreased by a random procedure which selected 11 items as bootstrapping sample (5 items from the North Sea, 6 items from the Baltic). While the resulting 620 items served as data basis for the entire statistical analysis, the bootstrapping sample was choosen to verify the outcome of the discriminant analysis (details see below).

A main aspect to be cleared is: is it better to use data aggregated by age or nonaggregated data? From table 4 can be inferred that within one area, either North Sea or Baltic, the means by age seem to be homogenously distributed (even the cv's does not differ dramatically). Checking this by a simple ANOVA[10] separatly carried out both for the North Sea and the Baltic gives: since in both areas the means does not differ significantly between the age groups the following analyses will be performed by inaggregated data (see table 5).

---

9 cv = coefficent of variation in %

10 simple ANOVA = unbalanced oneway analysis of variance with factor age group

Tab. 5 Unbalanced ANOVA to see whether means between age groups differ significantly

Dependent Variable: VERTEBRA

----------------------------------- AREACODE=0 -----------------------------------

Number of observations in by group = 228

Class	Levels	Values
AGE	3	1 2 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.80654884	0.40327442	0.91	0.4058
Error	225	100.19345116	0.44530423		
Corrected Total	227	101.00000000			

----------------------------------- AREACODE=1 -----------------------------------

Number of observations in by group = 392

Class	Levels	Values
AGE	7	1 2 3 4 5 6 7

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	3.85212091	0.64202015	1.04	0.3969
Error	385	236.99481787	0.61557096		
Corrected Total	391	240.84693878			

$$\textit{vertebra counts} = 56.50 - 0.71 \times \textit{area} \quad \text{or} \quad \hat{area} = \frac{56.50 - \textit{vertebra counts}}{0.71} \qquad (10)$$

A first inspection of the average vertebra counts in table 4 shows that these are in general slightly higher for North Sea than for Baltic herring in all age groups. The performed regression analysis confirms this obsrevation statistically. Table 6 compresses the results of the linear model fit. It can be seen that the estimation of the ordinate $\hat{a}$ = 56.50 is identical to the mean of the vertebra counts of North Sea herring. With slope $\hat{b}$ = 0.71 the regression model is given in equation (10). An ad hoc way to verify the model is to simply use the inverted version of equation (10). Inserting the average vertebra number 56.50 for North Sea herring leads to the correct solution "area = 0", vice versa inserting the average vertebra number 55.79 for Baltic herring into leads to solution "area = 1" which is also the correct result.

Tab. 6 Generalized Linear Regression Modell

Number of observations in data set = 620

Dependent Variable: VERTEBRA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1947978.1531	973989.0765	99999.99	0.0
Error	618	341.8469	0.5532		
Uncorrected Total	620	1948320.0000			

	R-Square	C.V.	Root MSE	VERTEBRA Mean
	0.999825	1.326886	0.7437408	56.05161290

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ORDINATE	1	1947905.6516	1947905.6516	99999.99	0.0
AREACODE	1	72.5014	72.5014	131.07	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ORDINATE	1	727833.00000	727833.00000	99999.99	0.0
AREACODE	1	72.50145	72.50145	131.07	0.0001

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
ORDINATE	56.50000000	1147.08	0.0	0.04925544
AREACODE	-0.70918367	-11.45	0.0001	0.06194511

Due to the 2-point structure of the model the correlation coefficient is near 1 ($r^2 = 0.999$) which indicates a proper fit. This will be confirmed by various tests on the estimated regression coefficients which display a high significance on the 5% level: the partial t-values and associated p-values indicate this for the particular regression coefficients, the F-value and the corresponding p-value indicates this in more general terms for the entire model (see table 6).

Tab. 7 Normal restriction

```
              Plot of RESIDUAL*AREACODE.  Symbol used is 'u'.
                 (NOTE: 609 obs hidden.)
        RESIDUAL |
               5 +
                 |  u
                 |
                 |  u                                         u
                 |  u                                         u
               0 +  u                                         u
                 |  u                                         u
                 |                                            u
                 |                                            u
                 |
              -5 +
                 ---+-----------------------------------------+--
                    0                                         1
                                     AREACODE

                           Histogram                          #
                 3.75+*                                       1
                     .
                     .
                     .**                                      6
                     .**                                      6
                     .***********                            55
                     .*********************                 108
                     .****************************************  196
                     .*********************                 103
                     .*************************             122
                     .**                                     10
                     .***                                    12
                     .
                -2.75+*                                       1
                      ----+----+----+----+----+----+----+----+
                      * may represent up to 5 counts

                              Normal Probability Plot
          3.75+                                               *
              |
              |
              |                                             ***
              |                                           **+++
              |                                    **********
              |                              *******+
              |                        *********
              |                    *****++
              |        *************
              |      ***++++
              |******
              |
         -2.75+*
              +----+----+----+----+----+----+----+----+----+----+
                   -2        -1         0        +1        +2
```

N_OBS	MEAN	MEDIAN	MODE	MIN	MAX	STD	SKEWNESS	KURTOSIS	SHA/WIL
620	0	0.20918	0.20918	-2.79082	3.5	0.74314	0.16875	0.92686	0.92909

From table 7 it can be inferred that the normal restriction of the model does not seem violated: mean, median and mode of the residuals lie near each other, skewness and (modified) kurtosis of their empirical distribution is not far from 0, the Shapiro-Wilks test statistic (Shapiro et al. 1968) results in a value near 1. Also histogram and probability plot of the residuals do not indicate any violation of the normal constraint. The plot of the estimated residuals û does not show any obvious artefacts or implicit systematics.

Results of the discriminant analysis

In order to calculate the two distance functions for the North Sea and the Baltic learning samples we only need the two means, the two variances (as [1x1] within covariance matrices) and the two sample sizes (as a-priori probabilities) of the herring vertebra counts. Table 4 contains the means, table 8 summarizes the remaining measures. Inserting these values into equation (5) gives equation (11). The latter is the decision rule by which a newly sampled single herring (or more than one individual) from Division IIIa can be allocated to the North Sea or Baltic. Table 8 also shows that the [1x1] covariance matrices are heterogenous. This is the reason why the more general quadratic form has been choosen here.

Tab. 8 Homogeneity of the covariance matrices

DISCRIMINANT ANALYSIS

620 Observations	619 DF Total
1 Variables	618 DF Within Classes
Classes	1 DF Between Classes

Class Level Information

AREACODE	Frequency	Weight	Proportion	Prior Probability
0	228	228.0000	0.367742	0.367742
1	392	392.0000	0.632258	0.632258

WITHIN-CLASS COVARIANCE MATRICES

AREACODE = 0 DF = 227

Variable	VERTEBRA
VERTEBRA	0.4449339207

AREACODE = 1 DF = 391

Variable	VERTEBRA
VERTEBRA	0.6159768255

WITHIN COVARIANCE MATRIX INFORMATION

AREACODE	Covariance Matrix Rank	Natural Log of Determinant of the Covariance Matrix
0	1	-0.80983
1	1	-0.48455
Pooled	1	-0.59213

TEST OF HOMOGENEITY OF WITHIN COVARIANCE MATRICES

Test Chi-Square Value = 7.342177 with 1 DF Prob > Chi-Sq = 0.0067

Since the chi-square value is significant at the 0.1000 level, the within covariance matrices will be used in the discriminant function.

$$d_{Baltic}(x_{new}) = -\frac{1}{2}(x_{new} - 55.79)' \, 0.62^{-1} \, (x_{new} - 55.79) - \frac{1}{2} \ln|0.62| + \ln(0.63)$$
$$= -(x_{new} - 55.79)^2 \times 0.807 - 0.223$$
$$d_{N.S.}(x_{new}) = -\frac{1}{2}(x_{new} - 56.50)' \, 0.44^{-1} \, (x_{new} - 56.50) - \frac{1}{2} \ln|0.44| + \ln(0.37)$$
$$= -(x_{new} - 56.50)^2 \times 1.136 - 0.584 \qquad \textbf{(11)}$$

The second step is to check and evaluate the quality of the two distance functions. This can be done in several ways. The first was to randomly select a bootstrap test sample for calculating bootstrap error rates. This was done on the basis of unambiguous index numbers generated by a uniform random number generator. These index numbers were used to point to associated items (vertebra counts) of the learning samples (North Sea and Baltic) which then were excluded from the calculation of the two distance functions. The selected vertebra counts were laterly inserted as x_{new} in the two distance functions of equation (11). Since the origin of the selected vertebra counts is well known the computed allocation is compared with the actual membership. Table 9 contains the selected items, table 10 the individual posterior probabilities, the partial and overall error rates as well as the corresponding classification matrix for the bootstrapping experiment. Under consideration of the a-priori probabilities (different sample sizes) the overall error rate is 14.71% which means that about 85% of all selected 11 items are correctly classified. This is a very high rate. A closer inspection shows that the herring data from the North Sea are mainly responsible for the larger amount of uncertainty: a partial error rate of 40% indicates a "non-pure" North Sea herring learning sample. Compared with that all herring data from the Baltic could be correctly allocated leading to a 0% partial error rate[11].

Tab. 9 Randomly selected testdata

Random Selection of Testdata

OBS	INDEX	AREACODE
1	210	0
2	229	0
3	230	0
4	306	0
5	328	0
6	366	1
7	408	1
8	441	1
9	477	1
10	513	1
11	627	1

[11] In order to verify the inverted linear model empirically a similar procedure was applied by inserting the same randomly selected values into eq. (10). While the classification rate of 83.33% for the Baltic was nearly the same as in the discriminant analysis, the rate of 80% correct classifications for the North Sea was much better than that of the discriminant analysis. Since the regression model does not contain any weighting in terms of a-priori probabilities (sample sizes) the overall error rate is slightly worse (0.367).

Tab. 10 Bootstrapping error rates and corresponding classification matrix

Posterior Probability of Membership in AREACODE:

Obs	From AREACODE	Classified into AREACODE	0	1
1	0	0	0.6287	0.3713
2	0	1 *	0.3487	0.6513
3	0	1 *	0.3487	0.6513
4	0	0	0.6287	0.3713
5	0	0	0.6287	0.3713
6	1	1	0.0832	0.9168
7	1	1	0.0832	0.9168
8	1	1	0.3487	0.6513
9	1	1	0.3487	0.6513
10	1	1	0.0832	0.9168
11	1	1	0.3487	0.6513

* Misclassified observation

Number of Observations and Percent Classified into AREACODE:

From AREACODE	0	1	Total
0	3	2	5
	60.00	40.00	100.00
1	0	6	6
	0.00	100.00	100.00
Total	3	8	11
Percent	27.27	72.73	100.00
Priors	0.3677	0.6323	

Error Count Estimates for AREACODE:

	0	1	Total
Rate	0.4000	0.0000	0.1471
Priors	0.3677	0.6323	

A second way performed here is the jackknife method where exactly one item of the learning samples is excluded from the calculation of the distance functions. Thereafter this item is inserted as x_{new} into equation (11) and its computed membership is compared with its actual. This is done for each of the items of the two learning samples. The resulting error rates and the corresponding classification matrix are displayed in table 11.

Tab. 11 Jackknife error rates and corresponding classification matrix

Number of Observations and Percent Classified into AREACODE:

From AREACODE	0	1	Total
0	115	113	228
	50.44	49.56	100.00
1	61	331	392
	15.56	84.44	100.00
Total	176	444	620
Percent	28.39	71.61	100.00
Priors	0.3677	0.6323	

Error Count Estimates for AREACCDE:

	0	1	Total
Rate	0.4956	0.1556	0.2806
Priors	0.3677	0.6323	

In general the bootstrapping results have confirmed the jackknife procedure. Despite "jackknife error rates" are usually relatively optimistic in comparison with bootstrap error rates the overall error rate of about 28% in this case is slightly higher than that of the bootstrapping method (about 13%-points). This means that 72% of all 620 herring were correctly classified. A closer look at table 11 gives a more detailed picture. Only 50% of North Sea herring could be correctly allocated but about 84% of Baltic herring. I.e., the Baltic herring learning sample seem to be purer and produced a more appropriate distant function which led to a better classification than that from the North Sea.

DISCUSSION

Statistical Methods:

The results of the ANOVA in table 5 say that it does not matter whether to take the data either aggregated by age or individually. But there are some advantages of processing inaggregated data for which they were used here: in this context inaggregation means

- a larger information basis which leads to more stable results in statistical tests (in this case an aggregation by age would reduce the total sample size from 679 downto 10 data items),
- that all classifying equations in this paper can be used on an individual level and not only on a level where means are only allowed to be inserted,
- that any interpretation could be easier performed with the individual than with the transformed data (in this case it might be easier to focus on an individual than on a group of individuals),
- the fact of working with unbalanced data[12] does not play such a role with larger than with smaller data sets (in this case the aggregation by age would lead to a sample size of 3 instead of 228 for the North Sea learning sample and of 7 instead of 392 for that of the Baltic).

On the other hand aggregation could smooth out internal age group variability and make the data more homogenous.

The coding of the design matrix X in the regression approach can be different from that used here: a vector of ones as first column and a dummy or binary coded variable for the area code as second column. One alternative could be effect or (1,-1) coding for the area code variable. A second alternative could be the introduction of two binary x variables, one for the Baltic[13] and one for the North Sea[14]. In such a case the regression analysis must be performed without calculating an ordinate due to problems of complete multicollinearity. Absolutely seen, in all variations it would give different estimations of the parameters. Also the null and the alternative hypotheses must be

---

12 i.e. data sets of different sizes

13 with "1" for Baltic herring data and "0" for non-Baltic herring data

14 with "1" for North Sea herring data and "0" for non-North Sea herring data

differently formulated in order to test the significance of the parameters. But the relative outcome of the inverse regression model (see equations (10) and (4)) would be the same: in all three cases it would give the same result whether a newly sampled single herring from Division IIIa stems from a North Sea or a Baltic population.

The matrix notation of the various equations implies that the regression approach can be easily generalized by inclusion of other variables with good discrimination power which might influence the specific number of vertebra or which might be influenced by the characteristics of the different areas under investigation (Baltic or North Sea). In the first (multiple) case of more than one independent x variables the methods of verification have to be extended by diverse analyses concerning the interaction between the included x variables which in particular lead to a detailed inspection of the corresponding covariance matrices. In the second (multivariate) case of more than one dependent y variable the whole set of verification procedures related to the multidimensional Y and the corresponding error matrix has to be carried out.

The same applies to the discriminant analysis. From the statistical point of view it is easy to generalize this approach by including more than one variable with discriminatory power (if those exist) as x variables. In such a case the investigations concerning the covariance matrices have to be more detailed.

Data and results:

In general the data contain a strict discrimination between the Baltic and the North Sea herring samples. The discrimination is highly significant as the t and F test results concerning the parameter estimates of the regression approach show. The low overall error rates and high rates of correct classification, respectively, of the two classification procedures confirm this result on a common scale. A further confirmation comes from an external source: with $\hat{a} = 56.50$ and $\hat{b} = 0.7$ the parameter estimates given in Anon. (1994) are very close to our estimations.

All constraints of the analyses methods seem to be more or less fulfilled.

Despite the good overall results, the partial error rate of 50% for the North Sea herring data is poor compared with that of the Baltic (16%). This leads to the conclusion that the North Sea herring learning sample is not a pure one i.e. not free of immigrated Baltic individuals. One consequence can be to take another learning sample of North Sea herring but this time in a subarea of the North Sea which is more uninfluenced by Baltic individuals. A correspondingly taken Baltic herring learning sample must ensure that the frame conditions in both learning samples are kept nearly the same (sampled age groups, used gear and ships, experience and knowledge of the crew etc.). This could lead to a smaller partial error rate for the North Sea and would increase the overall rate of correct classifications.

LITERATURE

Ackefors, H. 1978. Swedish herring tagging experiments, 1968-1970 in the Kattegat. Meddelande fran Havsfiskelaboratoriet, Lysekil, no. 241: 10 pp.

Andersen, K.P.; Burd, A.C.; Popp Madsen, K.; Saville, A.; Schubert, K. and Zijlstra, J.J. (1969). Report of the North Sea Young Herring Working Group. Coop.Res.Rep.Series A, no. 14: 87 pp.

Anon. 1977. Review on the distribution in relation to zones of extended fisheries jurisdiction of the following species: North Sea herring; Celtic Sea herring; Division VIa herring; North Sea sprat; Skagerrak, Kattegat and Norwegian fjord sprat; herring in Division IIIa Skagerrak; blue whiting (*Micromesistius poutassou*). ICES CM 1977/H:3 - APPENDIX.

Anon. 1988. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1988/Assess: 17.

Anon. 1989. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1989/Assess: 15.

Anon. 1990. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1990/Assess: 14.

Anon. 1991a. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1991/Assess: 15.

Anon. 1991b. Report of the Planning Group on Acoustic Surveys in Sub-Area IV and Division IIIa. ICES C.M. 1991/H:40.

Anon. 1992a. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1992/Assess: 11.

Anon. 1992b. Report of the Workshop on Methods of Forecasting Herring Catches in Division IIIa and the North Sea. ICES CM 1992/H:5, Ref.:J.

Anon. 1993. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1993/Assess: 15.

Anon. 1994. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1994/Assess: 13.

Anon. 1995. Report of the Herring Assessment Working Group for the Area South of 62°N. ICES CM 1995/Assess: 13.

Aro, E. 1989. A review of fish migration patterns in the Baltic. Rapp.P.-v.Reun.Cons.int. Explor.Mer, 190: 72-96.

Bakken, E. and Ulltang, O. 1972. Migrations and Stock Size of Herring in the Northern Sea in 1966. Analysis of Norwegian Tagging Experiments. ICES CM 1972/H:29.

Bartsch, J.; Brander, K.; Heath, M.; Munk, P.; Richardson, K. and Svendsen, E. 1989. Modelling the advection of herring larvae in the North Sea. Nature Vol. 340: 632-636.

Biester, E. 1979. Der Frühjahrshering Rügens- seine Rolle in der Fischerei der Ostsee und in den Übergangsgebieten zur Nordsee. Doctoral Thesis. Wilhelm-Pieck Universität Rostock (1979): 236 pp.

Bird, J.L.; Eppler, D.T. and Checkley, D.M., Jr. 1986. Comparisons of herring otoliths using Fourier series shape analysis. Can.J.Fish.Aquat.Sci., Vol 43: 1228-1234.

Bohl, E. 1962. Vergleichende Untersuchungen an Otolithen von Jungheringen der Nordsee im Frühjahr 1960. Ber.Dt.Wiss.Komm.Meeresforsch. XVI, 3: 161-176.

Campana, S.E. and Casselman, J.M. 1993. Stock discrimination using otolith shape analysis. Can.J.Fish.Aquat.Sci. Vol. 50: 1062-1083.

Dahle, G. and Eriksen, A.G. 1990. Spring and autumn spawners of herring (*Clupea harengus* L.) in the North Sea Skagerrak and Kattegat. Populations genetic analysis.Fish.Res.9: 131-141.

Fahmeir, L.; Hamerle, A. 1984. Multivariate Statistische Verfahren. Walter de Gruyter, Berlin.

Fossum, P. and Moksness, E. 1988. Distinguishing between spring and autumn spawned herring from daily increments, a preliminary report. ICES CM 1988/H:15.

Gjosater, H. and Oiestad, V. 1981. Growth patterns in otoliths as an indication of daily growth variations of larval herring (*Clupea harengus*) from an experimental ecosystem. ICES CM 1981/H:31.

Grahl-Nielsen, O. and Ulvund, K.A. 1990. Distinguishing populations of herring by chemometry of fatty acids. Am.Fish.Soc.Symp. 7: 566-571.

Hagström, O. 1984. Separation of young herring index in spring- and autumn spawned components in the Skagerrak-Kattegat area. ICES 1984/H:13.

Haraldsvik, S. 1967. Norwegian herring tagging experiments in the Northern North Sea. ICES CM 1967/H:26.

Hartung, J.; Elpelt, B.; Klösener, K.-H. 1987. Statistik. R. Oldenbourg, München.

Heath, M.R. and Walker, J. 1985. Drift of larval herring (*Clupea harengus* L.) studied by analysis of gene frequency data. ICES CM 1985/H:30.

Heincke, F. 1898. Naturgeschichte des Herings. Teil I. Die Lokalformen und die Wanderungen des Herings in den europäischen Meeren. Deutscher Seefischerei-Verein. Band II. Heft 2. Proc.Int.Herring Symp., Oct. 1990, Anchorage, Alaska.

Johannesen, A. and Moksness, E. 1991. Herring larvae (*Clupea harengus*) in the Skagerrak area from December 1987 to April 1988. Fish.Res., 11 (1991): 155-170.

Johansen, A.C. 1927. On the migration of herring. Extrat du Journal du Conseil International pour l'exploration de la Mer. Vol. II, no. 1.

Jorstad, K.E. and Pedersen, S.A. 1986. Discrimination of herring populations in a Northern Norwegian fjord: Genetic and biological aspects. ICES CM 1986/H:63.

Kompowski, A. 1969. Types of otoliths of Southern Baltic herring. ICES CM 1969/H:12.

Kühlmorgen-Hille, G. 1983. Infestation with larvea of *Anisakis spec.* as a biological tag for herring in Sub-Division 22, Western Batic Sea. ICES CM 1983/J:11.

Lubieniecki, B. 1972. The occurrence of *Anisakis sp.* larvea (Nematoda) in herring from the Southern Baltic. ICES CM 1972/H:21.

Lütkepohl, H. 1992. Multivariate Statistische Verfahren, Wintersemester 1991/92. Unveröffentlichtes Vorlesungsskript. Institut für Statistik und Ökonometrie. Christian-Albrechts-Universität zu Kiel.

MacKenzie, K. 1988. Digenean parasites as indicators of seasonal migrations of North Sea herring. ICES CM 1988/H:63.

Mann, R.C.; Hand, Jr., R.E. and Braslawsky, G.R. 1983. Parametric analysis of histograms measured in flow cytometry. Cytometry, 4: 47-82.

McQuinn, I.H. 1989. Identification of spring- and autumn-spawning herring (*Clupea harengus harengus*) using maturity assigned from a gonadosomatic index model. Can.J.Fish.Aquat.Sci., Vol. 46: 969-980.

Moksness, E. and Fossum, P. 1991. Distinguishing spring- and autumn-spawned herring larvae (*Clupea harengus*) by otolith microstructure. ICES. J.Mar.Sci, 48.

Munk, P.; Heath, M. and Skaarud, B. 1991. Regional and seasonal differences in the growth of larval North Sea herring (*Clupea harengus* L.) estimmated by otolith microstructure analysis. Continental Shelf Research: 641-654

Odense, P.H. and Annand, M.C. 1980. Herring (*Clupea harengus*) isoenzyme studies. ICES CM 1980/H:25.

Ojaveer, E.A. 1980. On morphometric differences in the spring and autumn Baltic herring, *Clupea harengus membras*, in the Gulf of Riga. J.Ichtyol. 20 (3): 38-44.

Petursson, P.M. & Rosenberg, R. 1982. Maturity and growth of the spring spawning herring in the Kattegat-Skagerrak. Ophelia, 21(2):195-204.

Pope, J.A. and Hall, W.S. 1970. A statistical analysis of morphometric characters in the Buchan and Kobbergrund herring. ICES CM 1970/H:32.

Popiel, J. 1956. The differentiation of biological groups of the herring in the Southern Baltic. ICES Meeting on 'Herring races', No. 21.

Postuma, K.H. 1974. The nucleus of the herring otolith as a racial character. J.Cons.Int.Explor.Mer 35:121-129.

Rasmussen, E. and Lassen, H. 1994. Identifying Rügen herring in Skagerrak using ^{137}CS as tag. ICES CM 1994/H:15, Ref. J.

Rauck, G. 1964. Die Struktur der Otolithen des Ostseeherings als Hifsmittel bei der Trennung biologischer Gruppen. Doctoral Thesis, Christian-Albrechts-Universität Kiel: 93 pp.

Reinert, J.; Hansen, B. and Joensen, H.P. 1992. Stock identification of *S. marinus* L. and *S. mentella* Travin in the NorthEast-Atlantic based on measurements of CS-137 content in the fish. ICES CM 1992/G:28.

Rosenberg, A.S. and Lough, G. 1977. A preliminary report on the age and growth of larval herring (*Clupea harengus*) from daily growth increments in otoliths. ICES CM 1977/L:26.

Rosenberg, R. and Palmen, L.E. 1981. Composition of herring stocks in the Skagerrak-Kattegat and the relations of these stocks with those of the North Sea and adjacent waters. Fish.Res. 1, (1981/1982): 83-104.

Schuchard-Ficher, Chr.; Backhaus, K.; Lohrberg, W.; Plinke, W.; Schreiner, W. 1983. Multivariate Analysemethoden. Springer, Berlin.

Schulz, H. 1967. Further photometric observations on herring otoliths. ICES CM 1967/H:4.

Schumacher, A. 1967. Bestandskundliche Untersuchungen am Nordseehering (Clupea harengus L.). Ber.Dt.Wiss.Komm.Meeresforsch. XIX, H.1:13-63. Sosinski, J. 1969. The types of otoliths of the Dunmore herring (South Ireland). ICES CM 1969/H:13.

Shapiro, S.s.; Wilk, M.B.; Chen, M.J. 1968. A comparative study of various tests for normality. Journal of the American Statistical Association. Vol. 63: 1343-1372.

Stryzewska, K. and Popiel, J. 1974. Characteristics of herring infested by anisakis larvae caught in the Gulf of Gdansk. ICES CM 1974/H:15.

Tshervontsev, V; Fetter, M and Vismanis, K. 1994. The eastern Baltic herring invaded by *Anisakis simplex* (Rudolphi, 1809) larvae. ICES CM 1994/J: 11, Ref. H.

Weber, W. 1975. A tagging experiment on spring-spawning herring of the Kiel Bay. Ber.Dt.Wiss.Komm.Meeresforsch. 24 (1975): 184-188.

Zenkin, V.S. and Lysenko, A.F. 1977. Determination of population specificity of the North Sea young herring based on genetic-biochemical and morphological analysis. ICES CM 1977/H:5.