Bundesforschungsanstalt für Fischerei, Hamburg
Bibliothek

Digitalization sponsored by Thünen-Institut

International Council for the Exploration of The Sea

C.M.1991/M:12 Ref.F+J
Anadromous and Catadromous Fish Committee

PAPER

# COMPARISON OF THE VISUAL AND DISCRIMINANT FUNCTION METHODS IN DISTINGUISHING WILD AND HATCHERY-REARED SALMON IN THE NORTHERN PART OF THE BALTIC SEA

by

P. Hiilivirta, E. Ikonen and J. Lappalainen
Finnish Game and Fisheries Research Institute, Fisheries Division,
P.O. Box 202, SF-00151 Helsinki 15, Finland

## ABSTRACT

In this paper an improved discriminant function for distinguishing wild and hatchery-reared Baltic salmon is presented. The new function is based on seven variables instead of the three used in the previous one. Using this function, it was possible to classify correctly 94% of fish of known origin. Use of the visual method resulted in a 99-percent correct classification of the same sample. When the two methods were compared using a scale sample of unknown origin, it was found that the results agreed at a level of 89.49%. The suitability of the methods for stock identification and the discrepancies between them are briefly discussed.

## Introduction

Different characters and methods are used to identify salmon stocks. Discriminant analysis using such characters as the river age, the numbers of circuli in the second annual river zone and the first annual sea zone and the anterior radii of these zones have been used to discriminate between salmon originating from North America and Europe (Lear and Sandeman 1980). Reddin and Burfitt (1983) used similar characters but separated the counts of circuli in the winter and summer growth portions of the first sea year. European salmon stocks have also been discriminated using scale shape measurements (de Pontual and Prouzet 1986 a; de Pontual and Prouzet 1986 b). In Norway, stock identification based on, for instance, scale characters, deformation of fins and pigment analysis is important in assessing the amount of escapees from fish farms (Anon. 1991 a).

In the Baltic Sea area the main problem has been to distinguish wild and hatchery-reared salmon. Antere and Ikonen (1983) described how this can be done visually on the basis of scale structure. Sych and Tuszynska (1983) introduced the use of discriminant analysis in distinguishing stocks of Baltic salmon. After that there have been many attempts to improve the method (Borzecka 1988; Borzecka et al. 1990). Borzecka et al. (1990) used the following characters: the number of circuli in the freshwater annual zones and the width of these zones, the number of circuli in the first annual zone in the sea phase and the width of this zone, and the freshwater age of the fish.

The aim of this study is to test and improve a discriminant function previously derived for the distinguishing of wild and hatchery-reared salmon in the northern part of the Baltic Sea (see Anon. 1991 b) and to compare it with the visual method. This is done in three different phases: in the first phase new scale characters are introduced to the discriminant function to improve its precision, in the second phase the discriminant function method and the visual method are tested with scales of known origin. Finally, both methods are additionally tested using a routine catch sample from the Bothnian Bay.

## Materials and methods

### Scales and scale reading

Most of the scales used in this study were obtained from commercial fishermen. Because the scale sampling was carried out by many different persons, the exact location of the scales on the fish is not known.

Impressions of the scales were made on polycarbonate strips using a mechanical press. For the discriminant analysis, scales were examined with a computer aided scale reading apparatus (henceforth called CSR, for details see Anon. 1991 b). The measurements and circuli counts were made from the focus along the longest axis of the scale (this was not always possible due to flaws in the impressions). Visual classification was carried out using the criteria presented by Antere & Ikonen (1983).

### Calibration data

The material on which the discriminant function is based consists of 304 hatchery-reared and 346 wild fish.

The hatchery-reared fish originated from the stocks of the Rivers Oulujoki (n=100) and Iijoki (n=115). All of these fish were tagged, i.e. their origin was known with certainty.

The origin of 181 wild fishes was known with reasonable certainty: either the fish had been tagged as smolts, or the scales had been collected during the smolt run. All these fish belonged to the River Simojoki stock.

To broaden the basis of the discriminant function, a randomly chosen sample of hatchery-reared (n=89) and wild (n=165) fish from Sub-divisions 29-32 was included. The origin of these fish was determined by the visual method independently by two Finnish scale readers.

The first discriminant analysis (SAS 1988) was performed using the three following characters (see also Anon. 1991 b):
fw1 = the width of the first freshwater annual zone (in mm);
fwc1 = the number of circuli in the first annual zone;
mean = the mean number of circuli in the freshwater zone (=the total number of circuli in the freshwater zone divided by freshwater age).

Of the hatchery-reared salmon, 78.62% were classified correctly, of the wild salmon 92.77% (Table 1).

Table 1. The results of the first discriminant analysis using the three variables fw1, fw2 and mean

| Actual group | | Groups into which the fish were classified: Hatchery-reared | Wild | Total number |
|---|---|---|---|---|
| Hatchery-reared | n | 239 | 65 | 304 |
| | % | 78.62 | 21.38 | 100.00 |
| Wild | n | 25 | 321 | 346 |
| | % | 7.23 | 92.77 | 100.00 |
| Total | | 264 | 386 | 650 |
| Percent | | 40.62 | 59.38 | 100.00 |

The error count estimate, i.e. the probability of misclassification, is 14.30%.

To improve the precision of the analysis, a stepwise discriminant analysis was performed (SAS 1988). The following variables were chosen for the analysis:

fw1 = the width of the first freshwater annual zone in mm;

fw2 = the width of the second freshwater annual zone;

fwc1 = the number of circuli in the first freshwater annual zone;

fwc2 = the number of circuli in the second freshwater annual zone;

fwcmax = the maximum number of circuli per year in the freshwater phase;

fwctot = the total number of circuli in the freshwater phase;

fwmax = the width of the freshwater phase in mm;

mean = fwctot/freshwater age;

fwcmin = the minimum number of circuli per year in the freshwater phase;

fwrel = fw1/fw2.

Since no differences were found in this material in the characters of the first sea year, it was not included in the stepwise analysis.

In the stepwise analysis the variables are entered into the model according to their discriminating power. The variable fw2 has the greatest discriminating power. Variables fwc1, fwc2 and fwcmin were excluded, probably because they are correlated with those remaining in the function (Table 2).

Table 2. Results of the stepwise discriminant analysis

| Step | Variable Entered | Variable Removed | Number In | Partial $R^2$ | F Statistic | Prob > F |
|---|---|---|---|---|---|---|
| 1 | fw2 | | 1 | 0.4142 | 457.401 | 0.0001 |
| 2 | fwl | | 2 | 0.1103 | 80.047 | 0.0001 |
| 3 | fwcmin | | 3 | 0.0698 | 48.392 | 0.0001 |
| 4 | fwcmax | | 4 | 0.0415 | 27.900 | 0.0001 |
| 5 | fwctot | | 5 | 0.0528 | 35.822 | 0.0001 |
| 6 | fwmax | | 6 | 0.1016 | 72.626 | 0.0001 |
| 7 | mean | | 7 | 0.0076 | 4.933 | 0.0267 |
| 8 | | fwcmin | 6 | 0.0009 | 0.594 | 0.4413 |
| 9 | fwrel | | 7 | 0.0051 | 3.268 | 0.0711 |

Use of the new variables fwl, fw2, fwcmax, fwctot, fwmax, mean and fwrel resulted in a better overall discrimination between wild and hatchery-reared salmon, although the proportion of incorrectly classified wild fish increased to some extent. Of the hatchery-reared salmon 87.46% were classified correctly, of the wild salmon 88.15% (Table 3). One hatchery-reared fish apparently stocked at the age of one was excluded from the analysis. The error count estimate is 12.20%, which is 2.1 percentage units less than that obtained using the variables fwl, fw2 and mean. For the wild fish the value of the error count estimate is 11.85%, for the hatchery-reared fish 12.54%.

Table 3. The results of the discriminant analysis using variables fw1, fw2, fwcmax, fwctot, fwmax, mean and fwrel

| Actual group | | Groups into which the fish were classified: Hatchery-reared | Wild | Total number |
|---|---|---|---|---|
| Hatchery-reared | n | 265 | 38 | 303 |
| | % | 87.46 | 12.54 | 100.00 |
| Wild | n | 41 | 305 | 346 |
| | % | 11.85 | 88.15 | 100.00 |
| Total | | 306 | 343 | 649 |
| Percent | | 47.15 | 52.85 | 100.00 |

Thus the linear discriminant functions are:

| | Hatchery-reared | Wild |
|---|---|---|
| CONSTANT | -43.08653 | -33.80480 |
| fwcmax | -0.38607 | -0.78714 |
| fw2 | -100.90667 | -92.36229 |
| fw1 | 107.47671 | 99.09735 |
| fwmax | 31.58036 | 18.35080 |
| fwctot | -0.36464 | 0.22334 |
| mean | 2.31358 | 1.75019 |
| fwrel | 29.78816 | 28.78643 |

**Test I: Comparison of the visual method and discriminant analysis using scales of known origin**

In order to compare the visual method and the discriminant function method, a blind test was made where scales of fish of known origin were analysed by the same scale reader using both methods.

For the test 143 scale impressions were randomly chosen by a person not participating in the scale reading. The impressions were numbered using a table of random numbers. Of these, 43 had to be discarded owing to regenerated scales, poor quality of the impressions or similar reasons, so that the final test material consisted of 100 impressions of which, by pure chance, 50 were of wild and 50 of hatchery origin.

The scales of hatchery-reared fish were collected in connection with tag returns. Of these, 46 belonged to the River Iijoki stock (stocking age three years) and four to the River Oulujoki stock (stocking age two years). The fish were tagged in the years 1978-1981 and in 1985.

Most of the scales of wild fish (n=45) used in this test originated from the Merikoski rapids in the River Oulujoki running into the Bothnian Bay. The scales had been collected by T.H. Järvi in 1947. Considering the situation in post-war Finland, it is quite reasonable to assume that these fish were of wild origin. An additional sample of five fish collected during routine sampling was also included in the test. The origin of these fish had previously been visually determined by an experienced scale reader. Three of the fish were caught on 25.-26.7.1985 in the Bothnian Bay (Ii, Laitakari, ICES statistical rectangle 59H5). The two remaining fish were caught in the 1980's, but the place and date of capture are unknown.

The origin of the fish was determined visually using a Bell & Howell SR-VIII microfilm reading device (magnification ca. 37.5 x). After this, the scale reader examined the scale impressions with the CSR recording the scale data.

**Test II: Comparison of the visual method and discriminant analysis using a routine catch sample**

The improved discriminant function was also compared with the visual method using part of a routine catch sample from the Bothnian Bay near the mouth of the River Simojoki (Isle of Montaja, statistical rectangle 60H5). The material consisted of scale samples from 257 fishes collected in 1990 during the following periods: 28.5.-6.6., 23.6.-30.6. and 5.7.-28.7.. These periods were chosen to ensure the inclusion of wild fish in the material. Because hatchery-reared fish have a tendency to run later in the season and because a relatively large part of the material consists of fish running early, the material does not give a true picture of the proportions of wild and hatchery-reared salmon. Adipose fin clipped fish were excluded from the analysis for two reasons. Firstly, the freshwater phases of scales of fish stocked when one-year-old tend to be almost identical to those of scales of wild fish and, secondly, the origin of adipose fin clipped fish is known anyhow. The material was analysed like in test I, except that the visual examination of the scales was conducted from the video monitor while recording the scale data with the CSR.

## Results

### Test I

Using the visual method, 99% of the fish of known origin were classified correctly. The only incorrectly classified fish was a fish (age 4.1+) previously classified wild by the visual method which was now classified hatchery-reared (Table 4).

Table 4. Results of the visual classification (fish of known origin)

| Actual group | | Groups into which the fish were classified: Hatchery-reared | Wild | Total number |
|---|---|---|---|---|
| Hatchery-reared | n | 50 | 0 | 50 |
| | % | 100.00 | 0.00 | 100.00 |
| Wild | n | 1 | 49 | 50 |
| | % | 2.00 | 98.00 | 100.00 |
| Total | | 51 | 49 | 100 |
| Percent | | 51.00 | 49.00 | 100.00 |

The discriminant analysis classified correctly 96% of the fish of known origin (Table 5). One hatchery-reared fish was incorrectly classified wild, and three wild fish were incorrectly classified hatchery-reared. It should be noted that the fish classified incorrectly by the two methods were different. On the basis of this test it seems that the discriminant function may slightly underestimate the proportion of wild fish.

Table 5. Results of the discriminant analysis (fish of known origin)

| Actual group | | Groups into which the fish were classified: Hatchery-reared | Wild | Total number |
|---|---|---|---|---|
| Hatchery-reared | n | 49 | 1 | 50 |
| | % | 98.00 | 2.00 | 100.00 |
| Wild | n | 3 | 47 | 50 |
| | % | 6.00 | 94.00 | 100.00 |
| Total | | 52 | 48 | 100 |
| Percent | | 52.00 | 48.00 | 100.00 |

**Test II**

Visual classification and discriminant analysis gave different origins for 27 fishes of the routine catch sample (Table 6): 22 fishes classified hatchery-reared by the visual method were classified wild by the CSR, and 5 fishes classified wild by the visual method were classified hatchery-reared by the CSR. The difference between the two methods is 10.51% (27 differently classified fish out of 257) when the whole material is concerned. If, on the basis of the previous test, it is considered that the visual method gives the "right" result, the differences are respectively 10.48% and 10.64% when hatchery-reared and wild fish are concerned. In this case the discriminant function method tends to overestimate the proportion of wild fish in comparison with the visual method.

Table 6. Comparison of the CSR and visual method (catch sample)

| Classification by the visual method | | Classification by CSR: Hatchery-reared | Classification by CSR: Wild | Total number |
|---|---|---|---|---|
| Hatchery-reared | n | 188 | 22 | 210 |
| | % | 89.52 | 10.48 | 100.00 |
| Wild | n | 5 | 42 | 47 |
| | % | 10.64 | 89.36 | 100.00 |
| Total | | 193 | 64 | 257 |
| Percent | | 75.10 | 24.90 | 100.00 |

## Discussion

It has been shown that it is possible to visually distinguish wild and hatchery-reared salmon on the basis of the structure of the freshwater zone of the scales (Antere & Ikonen 1983). The use of this method requires no special equipment, but it does include a subjective element even if a skilfull scale reader is employed. It is also impossible to formulate it mathematically. On the other hand, discriminant analysis gives the user functions which can be used to discriminate between wild and hatchery-reared salmon even by a person with limited experience in scale reading.

Using the scale data on which the discriminant function derived in this study is based, the error count estimate is 12.2%, which is of the same order of magnitude as the misclassification rates obtained in other studies (Lear & Sandeman 1980, Reddin & Burfitt 1983, Borzecka et al. 1990, Anon. 1991 b). However, the misclassification rates of wild and hatchery-reared fish were almost equal, being 11.85% and 12.54%, respectively (previously

7.23% and 21.38%). This is far more important than the reduction of the error count estimate by 2.1 percentage units, especially because the ratio of wild salmon to hatchery-reared salmon in the Baltic Sea is approximately 1:10 (Anon. 1991 c). An example will illustrate the situation. A sample containing 100 wild and 1000 hatchery-reared salmon is analysed using the new discriminant function. If the fish are misclassified according to the above mentioned misclassification rates of 11.85% and 12.54%, 213 fishes would be classified wild and 887 hatchery-reared. If, on the other hand, the misclassification rates were 7;23% and 21.38%, 307 fishes would be classified wild and 793 hatchery-reared. Thus, the relative abundances are estimated much more accurately with the improved discriminant function although the error count estimates of the the functions do not markedly differ. Of course, every effort should be made to keep the misclassification rate as low as possible.

The results obtained using scales of known origin indicate that the differences between the visual and discriminant function method are very small, the error rates being 1% and 4%, respectively. Compared with the misclassification rate (12.2%) of the basic material these figures seem rather low. The reasons for the better discrimination of the material used in test I are not clear, but one may be the greater homogeneity of the material, which contained almost only fish of northern stocks. In this connection it must be noted that the sea age of the fish from the Merikoski Rapids in test I was mostly three or four years. In contrast to this, the sea age of the hatchery-reared fish in test I was mostly one or two years. In spite of the fact that the appearances of the Merikoski scales were very typically "wild", the difference in the sea ages may have given an advantage to the scale reader. However, this advantage has apparently been of little importance, since the discriminant analysis misclassified only three (= 6.7%) of the 45 Merikoski fishes.

The results of test II more are difficult to interpret because the origin of the fish is unknown. It seems, however, that no great discrepancies between the two methods exist, since the overall difference between them was no more than 10.51%. One reason for the greater discrepancy in this case may be the fact that the stocking age of hatchery-reared fish is sometimes very difficult to determine, due to the irregular structure of the freshwater zone of the scale. It is virtually certain that the proportion of hatchery-reared 3-year-old fish was overestimated because only about 1-2% of the salmon stockings in Finland consist of 3-year-old smolts (Anon. 1986, 1988, 1989, 1991 d and 1991 e). In the catch sample, 8.1 % of the fish visually classified hatchery-reared were considered to be three years old when stocked. It is conceivable that some of these fish have been classified erroneously wild by the discriminant analysis, because of the lower counts of circuli and widths of annual zones in the freshwater phase, resulting from the overestimation of the stocking age. This is supported by the fact that the only hatchery-reared fish misclassified in test I was a smolt of the River Iijoki stock which was stocked at the age of three.

Some error may also have been due to the fact that it was not always possible to measure the annual zones and count the circuli along the longest axis of the scale. Also this may have resulted in underestimating the widths of the freshwater annual zones and in overestimation of the proportion of wild fish. The possible use of nonpreferred scales may also have had some influence on the results, but its significance is very difficult to assess.

## Conclusions

On the basis of the present study it seems that the visual method is quite suitable for the discrimination of wild and hatchery-reared salmon in the northern parts of the Baltic Sea. However, in some instances, it may have drawbacks that can be avoided by the use of discriminant analysis. The present results indicate that it is possible to derive a discriminant function which distinguishes wild and hatchery-reared salmon in the above mentioned area almost as accurately as the visual method. If more data were added to the material on which the discriminant function is based, it is reasonable to believe that the accuracy of the discriminant function method may even exceed that of the visual method, at least in situations where scale readers of limited experience have to be employed.

**Acknowledgements**
We thank Mr. Pekka Vakkari and Mr. Mikael Hildén for comments on the manuscript. Mr. Pekka Vakkari gave us also valuable assistance in statistical methods.

## References

Anon. 1986. Statistics on fishing and fish farming in Finland in 1983. Suomen Kalatalous 52: 47-135.

Anon. 1988. Statistics on fishing and fish farming in Finland in 1984. Suomen Kalatalous 54: 49-165.

Anon. 1989. Statistics on fishing and fish farming in Finland in 1985. Suomen Kalatalous 55: 34-137.

Anon. 1991 a. Report of the Workshop on Identification of Fish Farm Escapees and Wild Salmon. ICES, C.M. 1991/M:2.

Anon. 1991 b. Baltic Salmon Scale Reading. Report of the Baltic Salmon Scale Reading Workshop. ICES, C.M.1991/M:7 Ref.J.

Anon. 1991 c. Report of the Baltic Salmon and Trout Assessment Working Group. ICES, C.M.1991/Assess:13.

Anon. 1991 d. Statistics on fishing and fish farming in Finland in 1986. Suomen Kalatalous 58: 47-136.

Anon. 1991 e. Statistics on fishing and fish farming in Finland in 1987. (Manuscript).

Antere, I. and Ikonen, E. 1983. A method of distinguishing wild salmon from those originating from fish farms on the basis of scale structure. ICES, C.M. 1983/M:26.

Borzecka, I. 1988. Preliminary information on the test of discrimination between the wild and hatchery-reared salmon by discriminant function. ICES, C.M. 1988/Assess:19. Appendix 1.

Borzecka, I., Ikonen, E. and Torvi, I. 1990. Attempts of using the discriminant function based on scale structure of Baltic salmon to distinquish between the wild and hatchery-reared smolts. ICES, C.M. 1990/M:21.

de Pontual, H. and Prouzet, P. 1986 a. Atlantic salmon stocks discrimination by scale shape analysis. ICES, C.M. 1986/M:11.

de Pontual, H. and Prouzet, P. 1986 b. Scale shape measurements to discriminate between Atlantic salmon stocks. ICES, C.M. 1986/M:12.

Lear, W. and Sandeman, E. 1980. Use of scale characteristics and discriminant functions for identifying continental origin of Atlantic salmon. In ICES/ ICNAF Joint Investigations on North Atlantic salmon. Rapp. P.-v Reun. Cons. int. Explor. Mer, 176: 68-75.

Reddin, D. and Burfitt, R. 1983. An Update: The use of scale characters and multivariate analysis to discriminate between Atlantic salmon (Salmo salar L.) of North American and European Origin Caught at West Greenland. ICES, C.M. 1983/M:11.

SAS 1988. SAS/STAT User's Guide, Release 6.03 Edition. Cary, NC: SAS Institute Inc. 1028 pp.

Sych, R. and Tuszynska, L. 1983. Attempts of using the scale characteristics for separation of some Baltic Salmon and Sea Trout stocks. ICES, C.M.1983/M:29, Appendix 6.