Numerical analysis of scale morphology to discriminate between atlantic salmon stocks

Hélène de Pontual [(1)] and Patrick Prouzet [(2)]

[(1)] *IFREMER, Centre de Brest, DRV/AP, B.P. 337, 29273 Brest cedex (France).*
[(2)] *IFREMER, Station Hydrobiologie INRA, B.P. 3, Saint-Pée-sur-Nivelle, 64310 Ascain (France).*

October 19, 1987; accepted January 21, 1988.

Pontual H. de, P. Prouzet, *Aquat. Living Resour.*, 1988, **1**, 17-27.

Abstract

The use of an image analysis microsystem with a specially developed software enabled scale image processing, outline extraction and computation of features which are not available in the conventional approach of scale analysis. Those parameters (shape factors, moment invariants and elliptic Fourier coefficients) were introduced into a discriminant analysis process to test their usefulness in stock identification. The method was applied to European stocks originating from Norway (Etne river) and France (Elorn river). The low misclassification rate obtained indicates that this approach is particularly promising.

Keywords : Atlantic salmon, scale, numerical analysis, stock identification.

Analyse numérique de la morphologie des écailles pour la discrimination de stocks de saumon atlantique.

Résumé

L'utilisation d'un microsystème d'analyse d'image et d'un logiciel développé pour cette application permet l'acquisition et le traitement d'images d'écailles, l'extraction de leur contour et le calcul d'attributs non accessibles par les méthodes classiques d'analyse scalimétrique. Le pouvoir discriminant de ces attributs (facteurs de forme, moments invariants, coefficients elliptiques de Fourier) est ensuite évalué. La méthode est appliquée à deux stocks originaires de rivières européennes, l'Etne (Norvège) et l'Elorn (France). Le faible taux d'erreur de classification obtenu permet de conclure à l'intérêt de cette approche.

Mots-clés : Saumon atlantique, écaille, analyse numérique, identification de stocks.

INTRODUCTION

Atlantic salmon (*Salmo salar* L.) are known to be homing fish. They spend 1 to 5 years in the sea before returning to the native river for spawning. This geographical isolation results in the subdivision of the species into a great number of discrete breeding sub-units or stocks as defined by Ricker (1972).

A major problem in salmon management is that, as shown by results from tagging and recovery experiments, different stocks intermingle on feeding grounds (Went, 1973; Jensen, 1980; Ruggles and Ritter, 1980; Swain, 1980). Fisheries exploiting stocks characterized by different reproduction potentials may seriously deplete or even exterminate the less productive components if a suitable fishing rate is applied to the more productive ones (Ricker, 1958). Identification of the contributing units and estimation of their relative proportions is thus a prerequisite to assess the effects of high sea fisheries on homewater stocks. Though a wide variety of techniques has been examined (Isshen *et al.*, 1981), scale characterization based on counting or measurement of growth rings is most frequently used (Bilton and Messinger, 1975; Lear and Misra, 1978; Reddin and Misra, 1985). Nevertheless, it presents some disadvantages and limitations. It may be sensitive to the scale reader's interpretation (Bilton *et al.*, 1983) and though it was shown to be rather effective in continental classification (Lear and Sandeman, 1980; Reddin and Burfitt, 1983; Reddin and Short, 1986), the discrete nature and narrow range of data used may be insufficient for differentiating between specific stocks (Shearer, 1983; Sych, 1983). Jarvis *et al.* (1978) recently proposed a new approach based on scale information. To discriminate between stocks of walleyes (*Stizostedion vitreum vitreum* Mitchill), they quantified the planar shape of scales by unrolling the digitized scale outlines and describing the resultant functions by Fourier series. Different investigations have since been conducted using this technique on Lake white fish (*Coregonus clupeaformis* L.), (Casselman *et al.*, 1981) and walleyes (Riley and Carline, 1982). A preliminary test on Atlantic salmon stocks (Pontual *et al.*, 1983) suggested that the technique should be improved to obtain more reliable results. Actually, the Jarvis' approach is limited. It is semi-automatic since hand digitization of the outline is required. The Fourier analysis algorithm used may encounter some difficulties limiting its use to simple forms. Finally, shape features such as shape factors and moment invariants commonly used to quantify shape occurring in biology were left out. Thus, the shape analysis system used in the present work was designed to overcome these limitations.

MATERIAL AND METHOD

Data collection

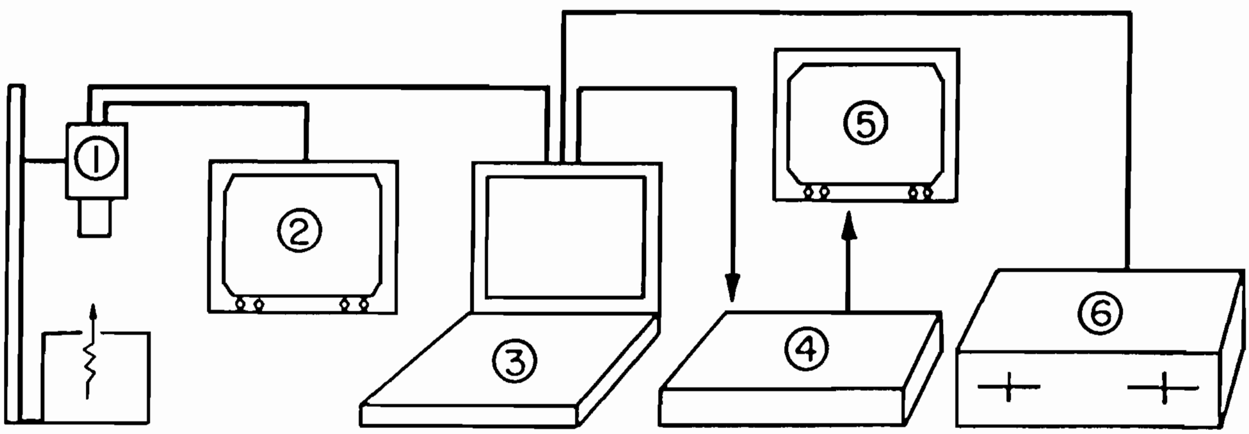
Since they are homing fish, individuals returning to their native river can be used for stock identification purposes. The samples thus consisted of scales from two sea-winter salmon collected in homewaters in 1982 from Etne River, Norway (32 specimens) and Elorn River, France (30 specimens). In order to control the within fish variability in scale shape, 3 scales per fish were selected (eroded and regenerated scales were systematically discarded). Samples were taken form the standard area, namely: "on the left hand side of the first 3-6 rows above the lateral line and on a line extending from the anterior edge of the anal fin to the posterior edge of the dorsal fin" (Anon., 1984). Scales were mounted between two slides after cleaning with sodium peroxyde.

Image analysis micro system description

Figure 1 shows a diagram of the image analysis microsystem used for this study. It consists of the following hardware:

— a macrophotographic stand with transmitted illumination;

— a Charge Coupled Device (CCD) camera providing either a TV signal or a digital signal whose spatial resolution is 208 horizontal pixels (picture elements) × 144 vertical pixels and the brightness resolution is 64 gray levels;



(1) CC CAMERA
(2)-(5) TV MONITORS
(3) MICRO COMPUTER
(4) RGB 256 MATRAX
(6) STORAGE UNIT

Figure 1. — Image analysis microsystem.

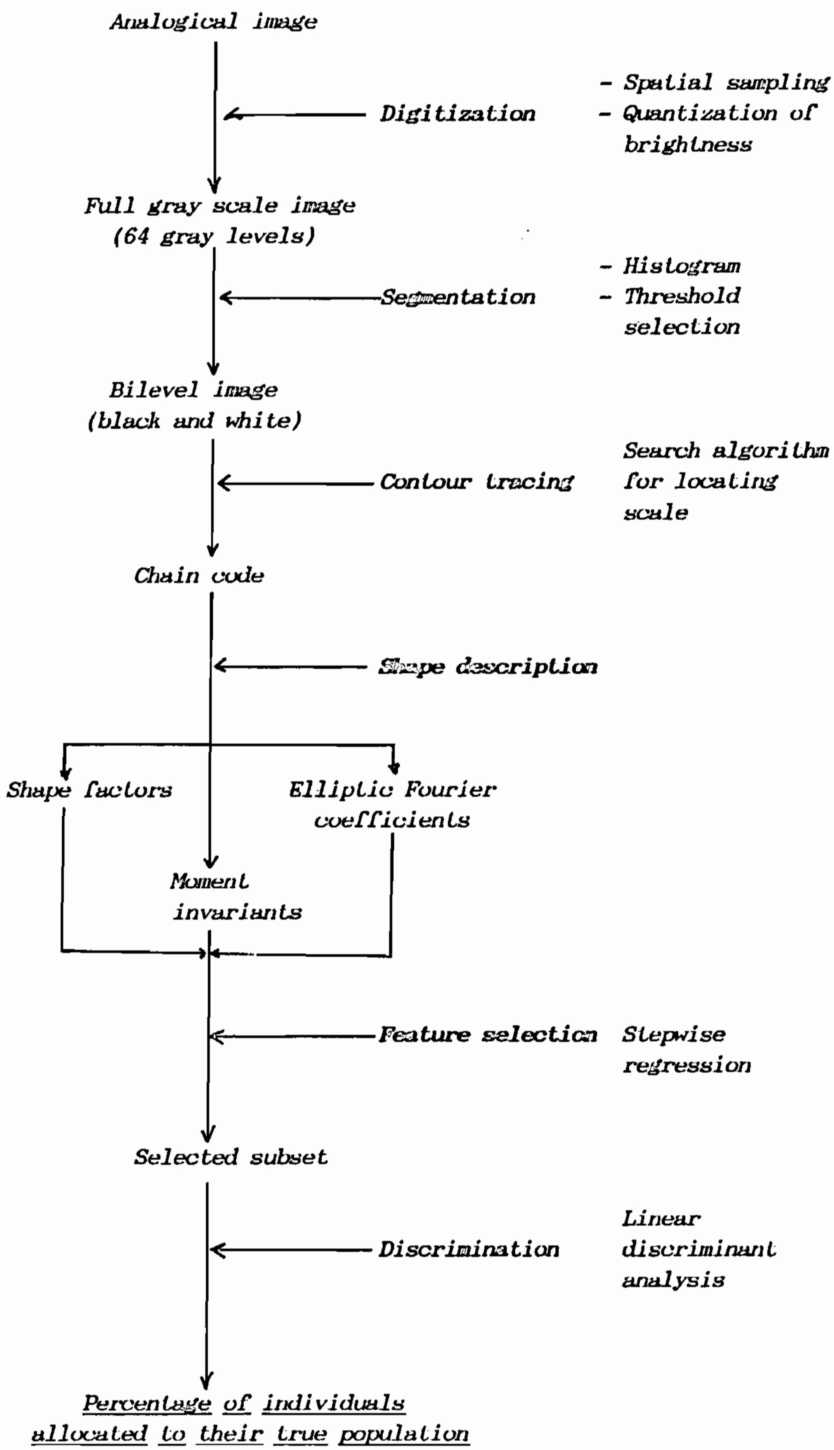


Figure 2. — Scheme for analysis of scale morphology to discriminate between salmon stocks.

— a micro computer (576 kb memory) with two disk-drives (1.2 Mb memory each);

— a digital-to-analog converter to visualize digitized images;

— two TV monitors for input and output images.

Software (using Basic language) was specially developed to enable image processing and feature extraction. The extracted features were then transmitted to a main computer for statistical pattern recognition with the SPAD software package (Lebart *et al.*, 1985).

Figure 2 shows a diagram of the procedures used to process digitized images of scales so as to quantify shapes and test the potential usefulness of shape descriptors to discriminate between stocks. This sequence consists of the following stages:
- image processing;
- shape quantification or feature extraction;
- statistical analysis.

Image processing

The first step (digitization) is the conversion of a continuous picture into a discrete form that may be manipulated by computer. This process consists of spatial sampling (a digital image is here represented with a discrete grid of 208 horizontal by 144 vertical elements) and quantization corresponding to the mapping of brightness into integers called gray levels (64 in this study). Illumination characteristics are chosen so that a digital image of scale typically consists of two phases: a dark silhouette on a light background.

The next step (segmentation) is the selection of a threshold value allowing to subdivide the image into two significant regions, scale and background. In this procedure the brightness value of each pixel is compared to a threshold level and the pixel is assigned to one of the two regions depending on wether the threshold value is exceeded or not. This results in a bilevel picture in which the scale apears like a black silhouette on a white background. In most cases, automatic threshold selection is a nontrivial problem. Indeed, when an image consists of two discernible regions, a threshold level can be chosen from the histogram of the distribution of gray levels that typically presents a bimodal pattern. Concerning salmon scale pictures, the posterior part of the scale may have intermediate gray levels creating an additionnal peak in the histogram. This leads to some difficulties for automatic thresholding which were solved by adaptating the technique proposed by Rosenfeld and de la Torre (1983) to trimodal image histogram (Pontual, 1986).

The contour of the scale can easily be extracted from the bilevel picture obtained after the threshold selection procedure. A convenient algorithm of contour tracing is given in Pavlidis (1982). Outlines are represented by chain codes (Freeman, 1961; 1974) in which two successive points are joined by a vector to which is assigned one symbol corresponding to one out of eight possible directions. Final coded contours are finally stored in memory and later used for shape feature extraction.

Feature extraction

From the coded contours of scales, three types of shape descriptors are computed that are commonly used to describe and compare shape quantitatively, particularly those occuring in biology. All of them are independent of the size (property of similitude), translation and/or the rotation of the scale in the field of the image and translation of the starting point of the trace of the contour. Indeed, these properties are required when comparisons have to be done.

Shape factors

The perimeter, area, maximum length and maximum width of an object can be easily extracted from a coded contour (Freeman, 1974). These geometric parameters are used to compute dimentionless ratios or shape factors such as the following ones used in this study:

F1 = Perimeter/square root of the area;
F2 = Perimeter/length;
F3 = Perimeter/width;
F4 = Square root of the area/length;
F5 = Square root of the area/width;
F6 = Width/length;
F7 = Area/area of the minimum circumscribed rectangle.

These shape factors commonly used for pattern recognition (*see* for example Jeffries *et al.*, 1984) measure gross shape properties such as elongation, compactness and so on and may be sufficient to differentiate between quite different forms. However some of them are not robust since they may yield similar numerical values for quite different form (Young *et al.*, 1974). Moreover, some of their properties are susceptible to be altered when measured on digitized object (Rosenfeld and Kak, 1976).

Moment invariants

Let us define a bilevel image of scale by a function $f(x, y)$ such as:

$$f(x, y) = 1 \quad \text{if} \quad P \in A$$
$$f(x, y) = 0 \quad \text{if} \quad P \notin A$$

where x, y are the spatial coordinates of a pixel P and A corresponds to the digitized scale. The function $f(x, y)$ which describes this scale can be uniquely determined by the infinit set of two dimensional centered moments M_{pq} of order $(p+q)$ and conversely the set M_{pq} is uniquely determined by $f(x, y)$. The moments M_{pq} given by equation 1 can therefore be used as shape features (Hu, 1962).

$$M_{pq} = \int_x \int_y (x-\bar{x})^p (y-\bar{y})^q f(x, y)\, dx\, dy \qquad (1)$$

where $\bar{x}$ and $\bar{y}$ are the coordinates of the center of gravity of the scale.

As a result of the definition of the function f, the moment M_{00} obviously measures the area of the digitized scale, since in the discrete form $M_{00} = \sum_x \sum_y f(x, y)\Delta x \Delta y$, with $\Delta x = \Delta y = 1$. Moreover from equation 1, the dimension of moment M_{pq} appears to be L^{p+q+2}, L being a length. L can be characterized by $M_{00}^{1/2}$ and therefore, the corresponding nondimensional moments N_{pq} are computed by dividing M_{pq} by M_{00}^{γ} with $\gamma = 1+(p+q)/2$. Hu (1962) showed that 6 moment invariants for spatial orientation can be directly computed from the moments N_{pq} of order 2 and 3. They are the following:

$$\begin{aligned}
\mu_1 &= N_{20}+N_{02}\\
\mu_2 &= (N_{20}-N_{02})^2+4N_{11}^2\\
\mu_3 &= (N_{30}-3N_{12})^2+(3N_{21}-N_{03})^2\\
\mu_4 &= (N_{30}+N_{12})^2+(N_{21}+N_{03})^2\\
\mu_5 &= (N_{03}-3N_{12})(N_{30}+N_{12})\\
&\quad \times[(N_{30}+N_{12})^2-3(N_{21}+N_{03})^2]\\
&\quad +(3N_{21}-N_{03})(N_{21}+N_{03})\\
&\quad \times[3(N_{30}+N_{12})^2-(N_{21}+N_{03})^2]\\
\mu_6 &= (N_{20}-N_{02})[(N_{30}+N_{12})^2-(N_{21}+N_{03})^2]\\
&\quad +4N_{11}(N_{30}+N_{12})(N_{21}+N_{03}).
\end{aligned} \tag{2}$$

These moment invariants μ_i therefore satisfy the properties of independence of orientation and size and have been used in a number of studies dealing with biological shapes (Butler, 1964; Berman *et al.*, 1984; Jeffries *et al.*, 1984).

These functions may be considered as similar to the moments of a statistical distribution. According to Hu (1962) the first two functions may be interpreted as "spread" and "slenderness", but, in general, their physical meaning is not easy to evaluate. The μ_i usually have a very large dynamic range including both negative and positive signs. Consequently, Hsia (1981) suggested using the logarithm of their absolute values. Therefore, the six invariants M_i used are given by:

$$M_i = \mathrm{Log}\,|\mu_i|, \qquad i=1,2,\ldots,6.$$

Elliptic Fourier coefficients

A planar shape can be described to whatever degree of precision is required using the decomposition of its outline by way of Fourier series or harmonic analysis. In this approach, the empiric contour is partitioned in a series of components called harmonics whose coefficients may be used as shape descriptors. The gross shape is determined by harmonics of low frequency and the addition of successively higher order harmonics increases the accuracy of shape description. Different methods of Fourier analysis of a closed contour have been proposed depending mainly on the functions used to describe the contour being processed: polar coordinates (Younker and Erlich, 1977; Jarvis *et al.*, 1978), cumulative change of a vector tangent to the outline (Zahn and Roskie, 1972), Dual Axis Fourier Shape Analysis (DAFSA) where equally spaced points of the contour are represented by complex numbers (Moellering and Rayner, 1983). Another method has been proposed by Kuhl and Giardina (1982) which uses a direct procedure to obtain Fourier coefficients from coded contours of simple or concave forms and permits complete regeneration. Elegant procedures of normalization relative to orientation and size can be done, based on the intrinsic shape properties. This technique has been compared to others by Rohlf and Archie (1984) to describe the shape of mosquitoe wings and appeared to be the most appropriate and powerful. Thus, it was adopted here with some slight modifications. A short description of the arithmetic process is proposed below. Full details are given in Kuhl and Giardina (1982) and Pontual (1986).

A closed coded contour can be represented using two series $x(t)$, $y(t)$ corresponding to the projections of the contour on the x-axis and y-axis respectively while it is being traced as a function of the arc length (t) measured from an arbitrary starting point. Fourier series approximation of the x-projection is given by:

$$X(t) = A_0 + \sum_{n=1}^{N}\left(a_n \cos\frac{2n\pi t}{T} + b_n \sin\frac{2n\pi t}{T}\right) = A_0 + \sum_{n=1}^{N} X_n$$

with:

$$\begin{aligned}
a_n &= \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\times\left(\cos\frac{2n\pi t_p}{T}-\cos\frac{2n\pi t_{p-1}}{T}\right)\\
b_n &= \frac{T}{2n^2\pi^2}\sum_{p=1}^{k}\frac{\Delta x_p}{\Delta t_p}\times\left(\sin\frac{2n\pi t_p}{T}-\sin\frac{2n\pi t_{p-1}}{T}\right)
\end{aligned} \tag{3}$$

where N is the number of harmonics used to approximate $x(t)$, k the number of links in the chain code, t_p the arc length of the first p links, $T=t_k$ the basic period of the chain (equivalent to the perimeter) and Δx_p the change in the x-coordinate as the link p is traversed. The approximation of the y-projection profile and the corresponding coefficients c_n and d_n are found in the same way using the incremental change in the y-direction. The pair (A_0, C_0) where C_0 corresponds to A_0 for the y-projection denotes the location of the geometric center of the object (*see* Kuhl and Giardina, 1982 for equations giving A_0 and C_0);

Giardina and Kuhl (1977) showed that the points whose coordinates are (X_n, Y_n) describe an ellipse as t varies. As a consequence, a closed contour can be expressed as a summation of N ellipses, N being the number of harmonics needed to approximate the empiric contour to the desirable degree of accuracy. On the basis of this property, Kuhl and Giardina (1982) proposed normalization procedures yielding invariant coefficients useful to compare shapes. It is obvious that coefficients a_n, b_n, c_n and d_n vary according to the starting point of the trace of the contour, the spatial orientation and the size of the object being processed. For the n-th ellipse the corrected coefficients are given by:

$$\begin{bmatrix} a_n^* & b_n^* \\ c_n^* & d_n^* \end{bmatrix} = \begin{bmatrix} \cos\Phi_1 & \sin\Phi_1 \\ -\sin\Phi_1 & \cos\Phi_1 \end{bmatrix} \times \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \times \begin{bmatrix} \cos n\theta_1 & -\sin n\theta_1 \\ \sin n\theta_1 & \cos n\theta_1 \end{bmatrix} \quad (4)$$

where

$$\theta_1 = 0.5 \arctan \frac{2(a_1 b_1 + c_1 d_1)}{a_1^2 + c_1^2 - b_1^2 - d_1^2}, \qquad \Phi_1 = \arctan \frac{c_1 \cos\theta_1 + d_1 \sin\theta_1}{a_1 \cos\theta_1 + b_1 \sin\theta_1}. \quad (5)$$

Instead of using those corrected coefficients as features for pattern recognition, we computed the geometric parameters of the corresponding ellipses. Each set a_n, b_n, c_n, d_n characterizes the n-th ellipse which can also be described by the amplitudes of its semi-major and semi-minor axes, respectively A_n and B_n, the orientation Φ_n of its major axis with respect to the major axis A_1 of the first ellipse and a phasor θ_n. The parameters Φ_n and θ_n are computed from equation (5) where Φ_1, θ_1, a_1, b_1, c_1 and d_1 are replaced by Φ_n, θ_n, a_n^*, b_n^*, c_n^* and d_n^*. Parameters A_n and B_n are given by:

$$\begin{aligned} A_n^2 &= (a_n^* \cos\Phi_n + b_n^* \sin\Phi_n)^2 + (c_n^* \cos\Phi_n + d_n^* \sin\Phi_n)^2 \\ B_n^2 &= (a_n^* \cos\alpha_n + b_n^* \sin\alpha_n)^2 + (c_n^* \cos\alpha_n + d_n^* \sin\alpha_n)^2 \end{aligned} \quad (6)$$

with $\alpha_n = \Phi_n + \pi/2$.

Invariance for similitude is obtained by dividing A_n and B_n by the amplitude A_1 of the first semi-major axis. Finally $(4N-3)$ normalized parameters are avalaible to describe a contour approximated with N harmonics. Twenty harmonics were extracted here since, as shown further, this number is sufficient to describe very accurately the shape of salmon scales. The elliptic Fourier coefficients utilizable as shape features were the following:

$$A_2/A_1, \ldots, A_N/A_1, B_1/A_1, \ldots, B_N/A_1, \Phi_2, \ldots, \Phi_N, \theta_2, \ldots, \theta_N.$$

Hence, the parameters A_n/A_1 and B_n/A_1 will be called A_n and B_n respectively.

This approach quite similar to that proposed by Tai *et al.* (1982) is attractive since it allows to select some independent parameters having a physical significance.

Statistical treatment

After extraction of the features, the data matrix consisted of 186 scale shapes each described by 7 shape factors, 6 moment invariants and twenty harmonics (77 coefficients). To take into account the within fish variability in scale shape (Pontual, 1986), the statistical analysis used the mean vector of the three vectors (each corresponding to one scale) of features available per fish and therefore the data matrix was reduced to 62 "mean scales" described by 90 variables.

Our aim was to determine whether numerical analysis of scale morphology might be helpful and informative to distinguish between salmon stocks. Such an evaluation is generally made using discriminant analysis. Based on the assumptions of multivariate normal distribution and homoscedasticity, the linear approach is the most commonly used and was adopted here. The first step of the analysis consists in selecting the subset of variables which best fits the observed data. This procedure avoids taking into account redundant information and using a number of variables which exceeds the maximum number (depending on the sample size) allowed to get a significant separability. A two group linear discriminant analysis is in fact a particular case of multiregression (Lebart *et al.*, 1982) and the selection can thus be done using a stepwise multiregression procedure. The best subset of variables is then used to compute the discriminant function whose objective is here to be predictive since it may used to allocate individuals whose origin is not known. The size of the avalaible samples was too small to split data into two sets, one as a base (called training set) to construct the discriminant function, the other as a test sample to evaluate its powerfulness. We thus used the bootstrap classification method developed by Efron (1982) which allows to estimate and thus to remove the bias inherent in calculating estimate of the misclassification rate on the data set used to determine the classification function. Notice that this approach, particularly useful when no test samples are avalaible, is proposed in the two group linear discriminant analysis program of SPAD software package (Lebart *et al.*, 1985). In the performed analyses, 200 bootstrap samples were taken from each original data set (a bootstrap sample consisted of a sample of n elements randomly chosen, but with replacement in the original sample of size n). Four analyses were conducted so as to evaluate the powerfulness of each type of shape descriptors. Three more analyses were used to determine whether shape descriptors taken all together do or do not increase the performance of discrimination. The features introduced in each selection procedure and the corresponding selected subsets are listed in *table* 1. The stepwise multi-regression software used here did not allow to introduce more than about 30 features (such an operation is indeed rather time consuming). This created difficulties in optimizing the

Table 1. – Accuracy of different subsets of features to discrinate between Elorn River and Etne River stocks; percentage of individuals correctly classified with (standard deviation).

Trial n°	Original features	Selected subset	Training set			Bootstrap estimations		
			Elorn R.	Etne R.	Total	Elorn R.	Etne R.	Total
1	Shape factors	F1 F3 F4 F6	100.0	100.0	100.0	99.1 (3.2)	98.4 (1.4)	98.1 (1.7)
2	Moments invariants	M1 M3	76.7	68.7	72.6	75.0 (7.8)	68.5 (7.2)	71.6 (6.1)
3	$A2 \rightarrow A10$, $B1 \rightarrow B10$	A6 A9 B3 B10	96.7	93.7	95.2	96.1 (3.8)	92.7 (3.7)	94.3 (3.0)
4	$A2 \rightarrow A8$, $B1 \rightarrow B8$ $\Phi2 \rightarrow \Phi8$, $\theta2 \rightarrow \theta8$	A2 A3 A7 B1 $\Phi4$	90.0	90.6	90.3	86.8 (4.8)	89.2 (4.0)	88.1 (3.5)
5	$F1 \rightarrow F7$, $M1 \rightarrow M6$ $A2 \rightarrow A8$, $B1 \rightarrow B8$	F1 F3 F5 A6 A7 A8	100.0	100.0	100.0	99.5 (1.5)	99.4 (0.5)	99.5 (0.7)
6	$F1 \rightarrow F7$, $M1 \rightarrow M6$ $A2 \rightarrow A8$, $B1 \rightarrow B8$	F1 F3 F5 M3 A6 A7 A8	100.0	100.0	100.0	99.1 (1.6)	99.8 (0.0)	99.5 (0.7)
7	$F1 \rightarrow F7$, $M1 \rightarrow M6$ $A2 \rightarrow A8$, $B1 \rightarrow B8$	F1 F3 F5 M3 A6 A7 A8 B3	100.0	100.0	100.0	99.2 (0.6)	99.4 (0.5)	99.3 (0.4)

selection procedure and explains why eight or ten harmonics including or not the parameters Φ_n and θ_n were used when Fourier descriptors were involved.

RESULTS

One advantage of shape description by means of Fourier series is that of feasible regeneration which allows to evaluate the quality of the approximation. *Figure* 3 shows the results of such an operation. Five to eight harmonics are sufficient to reconstruct the empiric contour with a good precision. Addition of higher order components permits to include more local information such as protuberances. Twenty harmonics generate a smooth version of the original contour and addition of higher frequencies might be considered as addition of noise partly attributable to the digitization procedure. *Figure* 4 shows a representation of the relative contributions of the successive harmonics which rapidly decrease with increasing frequencies. This is not surprising since it means that the processed signal presents simple characteristics. This is indeed the case with salmon scales whose shapes are not particularly complex.

Results of the different discriminant analyses are summarized in *table* 1. Performance may be evaluated and compared using the so-called re-substitution method (training set columns) which has been previously reported to be biased giving in most cases underestimations of misclassification rates. Therefore, it is more suitable to examine the corresponding bootstrap estimates. Their associated standard deviations indicate their degree of precision which increases with decreasing values. They may be used to calculate confidence intervals. Thus, the most powerful subsets are those which yield a high value of the estimated percentage of individuals correctly classified, associated with low standard deviations. The results of trials 1 to 4 suggest that the most discriminating features are shape factors (trial 1) and amplitudes of the first ten harmonics (trial 3). Moment invariants and Fourier parameters θ_n and Φ_n do not appear as very informative for stock identification. Nevertheless, the best results are obtained by selecting data among the three available types of features, i. e. shape factors, moment invariants and amplitudes of the first eight harmonics (trial 5 to 7 where estimated rates of correct classification are higher than 99%). This fact is clearly shown by comparing the individual group membership probabilities for trials 1 (shape factor only) and 6 (three types of features) (*table* 2). One can see that some scales (Elorn river 10 and 27 for example), are correctly classified in both analysis but with a very less significant probability in trial 1 where features were selected only among shape factors. Consequently, even if the discriminant scores do not appear significantly different (they are both very high), the selected subset 6 has to be considered as the most powerful. Conversely, there are no significant differences between trials 5, 6 and 7 and trial 5 would have been chosen to perform predictive analysis since the discrimant function that minimizes the misclassification rate with the fewest number of features is considered as the most successful one.

DISCUSSION

A previous paper using Fourier coefficients that were not invariant for similitude, *i. e.* size-dependent, showed variations due to age and origin (Pontual *et al.*, 1983). Pure shape descriptors have been found to be as much or even more informative than the corresponding size-dependent features (Pontual, 1986). They were thus used assuming that removing the size effect which is susceptible to vary with the change in environmental and climatic conditions, should endow morphological features with a higher degree of temporal stability. This is one of the advantages of the technique proposed since the conventional

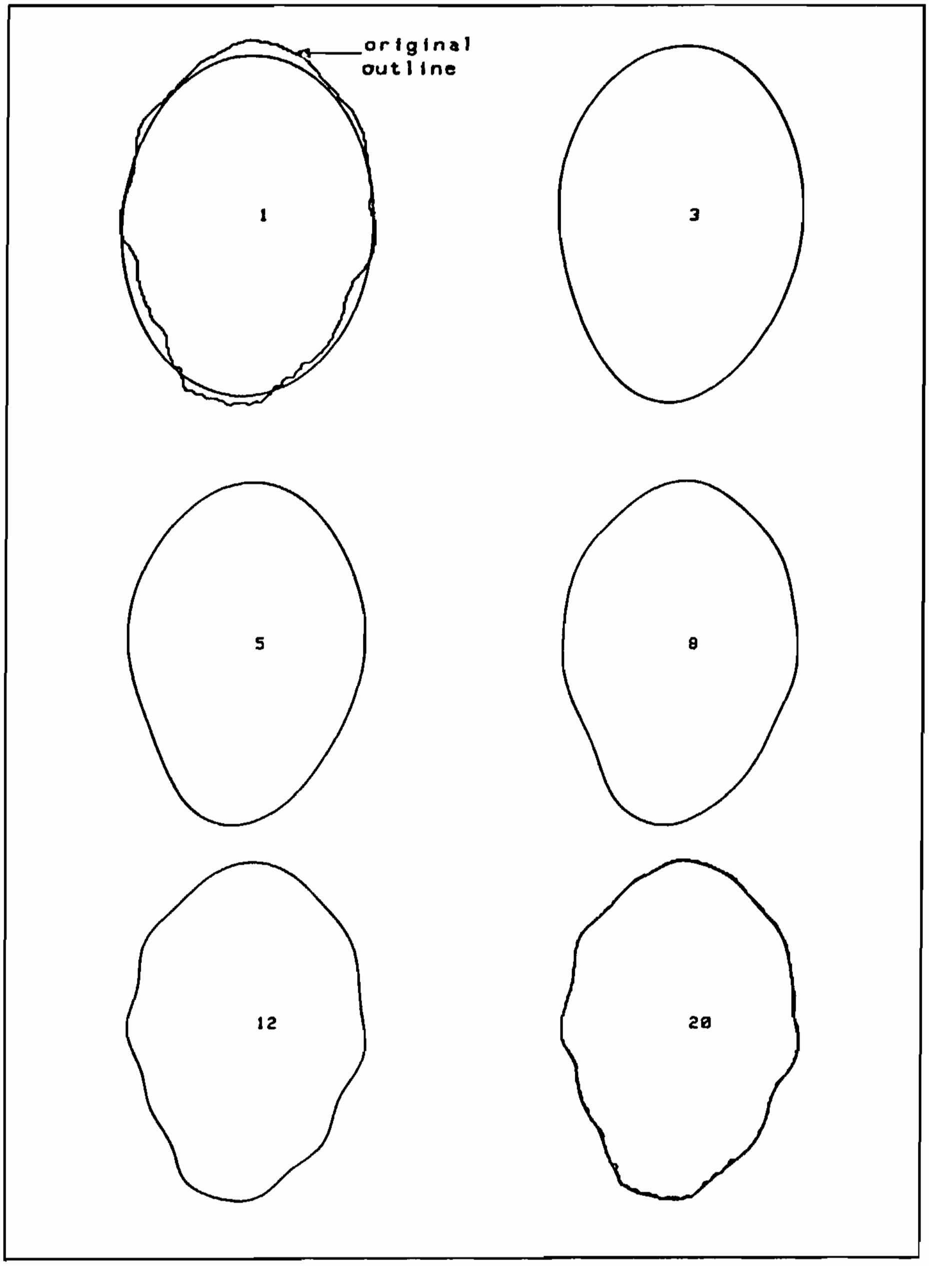


Figure 3. — Reconstruction of a scale outline based on 1, 3, 5, 8, 12 and 20 harmonics.

scale characters whose expression depends on environmental conditions (Reddin, 1981) have been shown to be sensitive to their fluctuations (Reddin and Short, 1986).

The continuous nature of shape information is another factor of powerfulness. Actually, only a small number of conventional scale character are usable. Counts of circuli are used instead of measurements because of their too large associated standard deviations yielding overlap between classes (Reddin, 1982). River zone information is useless as far as the discrimination is susceptible to involve hatchery reared smolts (Lear and Sandeman, 1980). That is why the technique described by Reddin and Burfitt (1983) only uses two features which consist of counts of circuli in winter and summer zones in the first sea year area. Though rather efficient in continental classification, this method is obviously too discrete to give valid and reliable results when the purpose is to identify a number of specify stocks in a mixed stock fishery (Shearer, 1983; Sych, 1983).

Jarvis' approach for scale shape analysis is based on Fourier series decomposition of hand digitized contours. Our results suggest that the computation

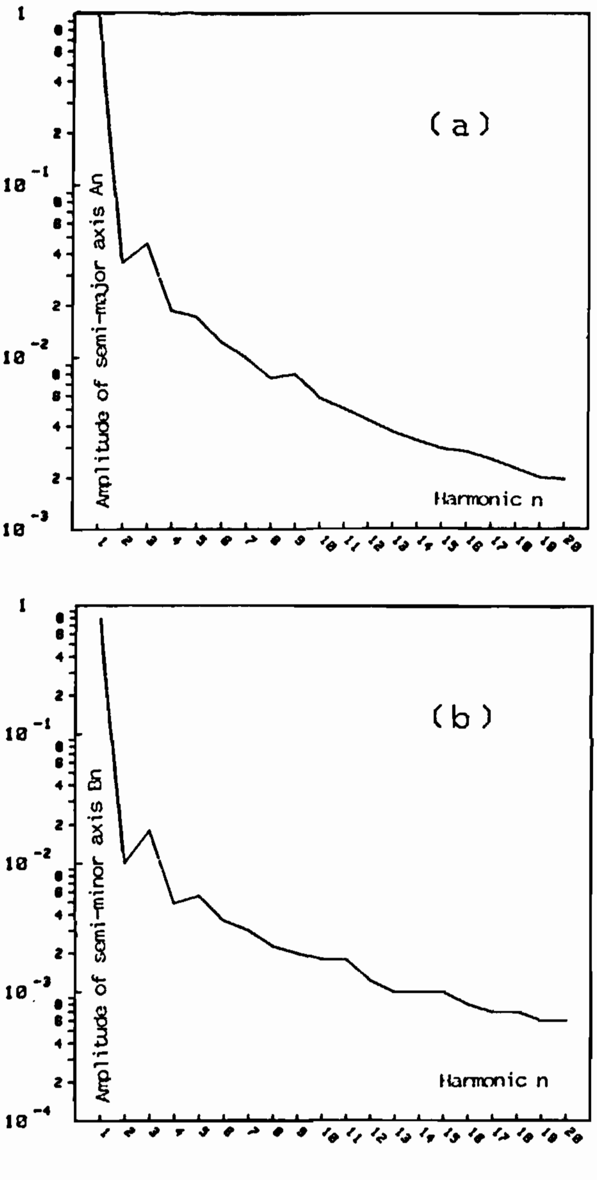


Figure 4. — Average amplitude of Fourier features (*a*) A_n and (*b*) B_n for the first 20 harmonics computed on 186 scales.

Table 2. — Group membership probabilities: training set of trials n^{os} 1 and 6.

Etne R. scale n°	Trial n° 1	Trial n° 6	Elorn. R. scale n°	Trial n° 1	Trial n° 6
1	1.000	1.000	1	0.994	1.000
2	1.000	1.000	2	0.954	1.000
3	0.882	1.000	3	1.000	1.000
4	0.948	0.959	4	1.000	1.000
5	0.956	0.985	5	0.979	1.000
6	0.999	1.000	6	0.755	0.911
7	0.997	1.000	7	0.974	1.000
8	0.998	1.000	8	0.904	1.000
9	0.991	1.000	9	0.962	1.000
10	0.989	1.000	10	0.562	0.968
11	0.951	0.993	11	1.000	0.996
12	0.995	1.000	12	1.000	1.000
13	0.999	1.000	13	1.000	1.000
14	0.945	1.000	14	1.000	1.000
15	1.000	1.000	15	1.000	1.000
16	1.000	1.000	16	1.000	1.000
17	0.901	1.000	17	0.995	1.000
18	0.960	0.993	18	0.998	1.000
19	0.996	1.000	19	0.998	1.000
20	1.000	1.000	20	0.978	1.000
21	1.000	1.000	21	1.000	1.000
22	0.999	1.000	22	0.966	1.000
23	0.946	1.000	23	0.996	0.997
24	1.000	1.000	24	1.000	1.000
25	0.999	1.000	25	0.995	1.000
26	0.881	0.999	26	0.998	1.000
27	1.000	1.000	27	0.507	0.999
28	1.000	1.000	28	0.996	1.000
29	0.836	1.000	29	1.000	1.000
30	0.897	1.000	30	1.000	1.000
31	0.988	1.000	–	–	–
32	0.999	1.000	–	–	–

of different types of shape descriptors may be more relevant. Actually, when Fourier coefficients are used alone, discrimination between the involved groups may require a great number of features which is not suitable from a statistical point of view. This difficulty might be overcome by introducing a small number of shape descriptors (such as shape factors) susceptible to differentiate between gross shapes whereas Fourier components are used when refinements to the description are required. The number of components to be employed largely depends on the sensitivity necessary to detect differences in the scale shape patterns to differentiate. But, as noticed by Bird *et al.* (1986) the potential powerfulness of shape analysis is all the more obvious as variations are found in the lower harmonic frequencies which indicates differences in overall shape. In this way, we had no a priori knoledge concerning the number of harmonics needed to detect differences between the two involved populations and that is why, although twenty harmonics were extracted as suggested by Jarvis *et al.* (1978), only the first eight ones were used in the discriminant procedure since they were found to be sufficient to yield high level of correct classification.

A limitation of the method is the difficulty to relate the shape quantification to some fundamental properties of the analysed scale. In the present study, the two involved stocks show differences in terms of compactness and circularity. This appears when comparing the mean values of shape factors F1 and F6 respectively which indicate that scales from Etne river are less compact and more elongated than those from Elorn river. A more precise interpretation is difficult because the physical meaning of a given numerical value taken by a given shape descriptor such as Fourier coefficient or invariant moment is in most cases somewhat unclear. Moreover, because a simple change in form does not result in simple change in shape parameter as pointed out by Bookstein *et al.* (1982), this approach cannot be used to devise an a priori model of scale shape development. But, as noticed by Rohlf and Ferson (1983), if the purpose is to find some descriptors whose variations are sufficient to distinguish between given groups, this limitation is not a real problem.

Discriminant analysis yielding estimated misclassification rate lower than 1% can be considered as very successful. This high degree of efficiency and the capability, with an appropriate image processing equipment, of rapidly examining large samples automatically suggest that numerical analysis of scale morphology could provide a useful tool for fishery management. Nevertheless, further research is needed to investigate such a promising area. From a technical point of view, the reported experiment involved only two stocks and performance have to be evaluated when a great number of components are mixed. It would be also quite interesting to see how the technique allocates fish independent of the data base used for the construction of the classification rule and compare the results with those given by the bootstrap estimation method. This would require to process a greater number of scales in each studied stock. Moreover, such tests may require the use of a discriminant procedure different from that used in the present study such as for example the quadratic approach. Because it is not based on the assumption of homoscedasticity, this technique has been proposed for studies using conventional scale analysis (Reddin and Burfitt, 1983) instead of the linear discriminant analysis first used by Lear and Sandeman (1980) which gave a lower efficiency on a yet less complex data base. Finally, optimizing the feature selection procedure is an essential requirement since this step obviously conditions the performance of the resulting classification. This will be all the more necessary as slight differences will have to be recognized which might require to investigate the potential usefulness of a greater number of Fourier components than that used in the present work. From a biological point of view, much has to be done especially to understand the biological bases of the observed variations in scale shape as well between as within stocks. Such knowledge might be very useful to settle realistic and efficient strategies for investigation of mixed stock fisheries. Such problems extend beyond the scope of this paper and are discussed with the results of a more general study on salmon stock discrimination based on scale shape analysis (Pontual and Prouzet, 1987).

Acknowledgements

We would like to thank Mrs Bouroche for her helpful comments concerning the traduction of this paper.

REFERENCES

Anonymous, 1984. Atlantic salmon scale reading. International Council for Exploration of the Sea. Report of the Atlantic salmon scale reading workshop. Aberdeen, Scotland 23-28 April 1984.

Berman M., J. R., Green, K. Sherman, 1984. Application of image analysis to the marine ecosystem studies. Int. Coun. Explor. Sea. C.M. 1984/L:8.

Bilton H. T., M., Flain, F. Lucas, P. Kearton, R. Gard, 1983. Tests on the accuracy of ageing New Zealand Quinnat salmon (*Oncorhynchus tschawytscha*) from their scales. *Can. Techn. Rep. Fish. Aquat. Sci.* n° 1199.

Bilton H. T., H. B. Messinger, 1975. Identification of Major British Columbia and Alaska runs of age 1.2 and 1.3 sockeye from their scale characters. *Int. North Pacific Fish. Comm. Bull.*, **32**, 109-129.

Bird J. L., D. T. Eppler, D. M. Checkley, 1986. Comparisons of herring otholiths using Fourier shape analysis. *Can. J. Fish. Aquat. Sci.*, **43**, 1228-1234.

Bookstein F. L., R. E. Stauss, J. M. Humphries, B. Chernoff, R. L. Elder, G. R. Smith, 1982. A comment upon the use of Fourier methods in systematics. *Systematic Zool.*, **31**, 85-92.

Butler J. W., M. K. Butler, A. Stroud, 1964. Automatic classification of chromosomes. *In:* Data acquisition and processing in biology and medecine K. Enslein Ed., Pergamon Press.

Casselman J. M., J. J. Collins, E. J. Crossman, P. E. Isshen, G. R. Spangler, 1981. Lake white fish (*Coregonus clupeaformis*) stocks of the Ontario waters of Lake Huron. *Can. J. Fish. Aquat. Sci.*, **38**, 1772-1789.

Efron B., 1982. The Jacknife, the Bootstrap and others resampling plans. *Regional conference series in applied mathematics.* Society for industrial and applied mathematics. Philadelphia Pensylvania, 19103, 92 p.

Freeman H., 1961. On the encoding of arbitrary geometric configurations. *IRE transactions on computers*, Vol. EC-**10**, 421-432.

Freeman H., 1974. Computer processing of line drawing images. *Computing surveys*, **6**, 57-97.

Giardina C. R., F. P. Kuhl, 1977. Accuracy of curve approximation by harmonically related vectors with elliptic loci. *Computer graphics and image processing*, **6**, 277-285.

Hsia T. C., 1981. A note on invariant moments in image processing. *IEEE Transactions on systems, man and cybernetics*, Vol. SMC-**11**, 831-834.

Hu M. K., 1962. Visual pattern recognition by moment invariants. *IRE Transactions on information theory*, Vol. IT-**8**, 831-834.

Isshen P. E., H. E. Booke, J. M. Casselman, J. M. Mc Glade, N. R. Payne, F. M. Utter, 1981. Stock identification: materials and methods. *Can. J. Fish. Aquat. Sci.*, **38**, 1838-1855.

Jarvis R. S., H. F. Klodowski, S. P. Sheldon, 1978. New method of quantifying scale shape and an application to stock identification in walleye (*Stizostedion vitreum*). *Transact. Amer. Fish. Soc.*, **107**, 528-534.

Jeffries H. P., M. S. Berman, A. D. Poularikas, C. Katsinis, I. Melas, K. Sherman, L. Bivins, 1984. Automated sizing, counting and identification of zooplancton by pattern recognition. *Mar. Biol.*, **78**, 329-334.

Jensen J. M., 1980. Recaptures from international tagging experiments at West Greenland. *Rapp. P.-v. Réun. Cons. int. Explor. Mer*, **176**, 114-121.

Kuhl F. P., C. R. Giardina, 1982. Elliptic Fourier features of a closed contour. *Computer graphics and image processing*, **18**, 236-258.

Lear W. H., R. K. Misra, 1978. Clinal variation in scale characters of Atlantic salmon (*salmo salar*) based on discriminant function analysis. *J. Fish. Res. Bd Canada*, **35**, 43-47.

Lear W. H., E. J. Sandeman, 1980. Use of scale characters and discriminant functions for identifying the continental origin of Atlantic salmon. *Rapp. P.-v. Réun. Cons. int. Explor. Mer*, **176**, 68-75.

Lebart L., A. Morineau, J. P. Fenelon, 1982. Traitement des données statistiques. Méthodes et programmes. Dunod, Paris.

Lebart L., A. Morineau and collaborators, 1985. SPAD: Système Portable pour l'Analyse de Données. CESIA, Paris, 257 p.

Moellering H., J. N. Rayner, 1984. The Dual Axis Fourier Shape Analysis of closed cartographic forms. *Cartographic J.*, **19**, 53-59.

Pavlidis T., 1982. Algorithms for graphics and image processing. Springer Verlag, 416 p.

Pontual H. de, 1986. Essais de discrimination de stockes de saumon atlantique (*salmo salar*) par reconnaissance de la forme de leurs écailles. Thèse de doctorat, Université P. et M. Curie (Paris-VI), 167 p.

Pontual H. de, P. Prouzet, 1987. Atlantic salmon, *salmo salar* L., stock discrimination by scale shape analysis. *Aquacult. fish. manag.*, **18**, 277-289.

Pontual H. de, P. Prouzet, C. Liacopoulos, 1983. Essai de différenciation de deux stocks de saumon atlantique (*Salmo salar*, L.) par reconnaissance de la forme de leurs écailles. Int. Coun. Explor. Sea, C.M. 1983/M:18.

Reddin D. G., 1981. Stock identification. Int. Coun. Explor. Sea, C.M. 1981/M:2, 23-41.

Reddin D. G., 1982. Some general information on discriminant functions and accuracy for identifying North American and European Atlantic salmon caught at West Greenland. Int. Coun. Explor. Sea, C.M. 1982/M:15.

Redding D. G., R. F. Burfitt, 1983. An update: the use of scale characters and multivariate analysis to discriminate between Atlantic salmon (*Salmo salar*) of North American and European origin caught at West Greenland. Int. Coun. Explor. Sea, C.M. 1983/M:11.

Reddin D. G., R. K. Misra, 1985. Hotelling's T to identify the origin of Atlantic salmon (*Salmo salar*) in a mixed stock fishery. *Can. J. Fish. Aquat. Sci.*, **42**, 250-255.

Reddin D. G., P. B. Short, 1986. Identification of North American and European Atlantic salmon (*Salmo salar* L.) caught at West Greenland in 1985. Int. Coun. Explor. Sea, C.M. 1986/M:9.

Ricker W. E., 1958. Maximum sustained yield from fluctuating environments and mixed stocks. *J. Fish. Res. Bd Canada*, **15**, 991-1066.

Ricker W. E., 1972. Hereditary and environmental factors affecting certain salmonid populations. *In*: The stock concept of Pacific salmon, R. C. Simon, P. A. Larkin Ed., H. R. MacMillan Lectures in Fisheries, University of British Columbia, Vancouver, British Columbia, 19-160.

Riley L. M., R. F. Carline, 1982. Evaluation of scale shape for identification of walleye stock from Western Lake Erié. *Transact. Amer. Fish. Soc.*, **111**, 736-741.

Rohlf F. J., J. W. Archie, 1984. A comparison of Fourier methods for description of wing shape in mosquitoes (*Diptera:* Culicidae). *Systematic zool.* **33**, 302-317.

Rohlf F. J., S. Ferson, 1983. Image analysis. *In*: Numerical taxonomy, J. Felstein Ed., Springer verlag, 583-599.

Rosenfeld A., A. C. Kak, 1976. Digital picture processing. Academic Press, New York, 457 p.

Rosenfeld A., P. de la Torre, 1983. Histogram concavity analysis as an aid in threshold selection. *IEEE Transactions on systems, man and cybernetics*, Vol. **SMC-13**, 231-235.

Ruggles C. P., J. A. Ritter, 1980. Review of North American smolt tagging to assess the Atlantic salmon fishery off West Greenland. *Rapp. P.-v. Réun. Cons. int. Explor. Mer*, **176**, 82-92.

Shearer W. M., 1983. The use of scale characteristics and multivariate analysis to distinguish between stocks of fish. Int. Coun. Explor. Sea, C.M. 1983/M:21.

Swain A., 1980. Tagging of salmon smolts in European rivers with special references to recaptures off West Greenland in 1972 and earlier years. *Rapp. P.-v. Réun. Cons. int. Explor. Mer*, **176**, 93-113.

Sych R., 1983. Attempts of using the scale characteristics for separation of some Baltic salmon and sea trout stocks. Int. Coun. Explor. Sea, C.M. 1983/M:29, 43-79.

Tai H. T., C. C. Li, S. H. Chiang, 1982. Application of Fourier shape descriptors to classification of fine particules. 6th International Conference on Pattern Recognition, Munich, Germany, 19-22.

Went A. E. J., 1973. Movements of *Salmo salar* (L.) to and from Irish waters. *J. Fish. Biol.*, **5**, 659-671.

Young I. T., J. E. Walker, J. E. Bowie, 1974. An analysis technique for biological shape. *Information and Control*, **25**, 357-370.

Younker J. L., R. Ehrlich, 1977. Fourier biometrics: harmonic amplitudes as multivariate shape descriptors. *Systematic Zool.*, **26**, 336-342.

Zahn C. T., R. Z. Roskies, 1972. Fourier descriptors for plane closed curves. *IEEE Transactions on computers*, Vol. **C-21**, 269-281.