

MOVIES-B: an acoustic detection description software. Application to shoal species' classification

Alain Weill, Carla Scalabrin and Noël Diner

IFREMER – Centre de Brest, DITI/GO, B.P. 70, 29280 Plouzané, France.

Received December 14, 1992; accepted March 10, 1993.

Weill A., C. Scalabrin, N. Diner. *Aquat. Living Resour.*, 1993, 6, 255-267.

Abstract

Software, MOVIES-B, developed by IFREMER (Institut français de recherche pour l'exploitation de la mer), enables the recognition of echogram features as acoustic detection structures according to spatial-energetic contiguity criteria. These structures (shoals, plankton layers) are described by energetic, morphological, spatial and temporal descriptors. Specifying acoustic detections as individual shoals will allow shoal by shoal echo-integration and may lead to better biomass estimation and shoal species description. The conceptual and operational framework of MOVIES-B is presented in the first part of this paper. The second part presents preliminary work performed in order to assess the feasibility of classification and possible species identification of shoals using MOVIES-B software. This research has studied the data set obtained from four fisheries acoustic surveys, in the Bay of Biscay. The data set is a matrix of 29 descriptors from 13,122 shoals. A limited knowledge of species composition was obtained from trawling. Eight species were captured: *Sardina pilchardus*, *Clupea harengus*, *Sprattus sprattus*, *Engraulis encrasicolus*, *Micromesistius poutassou*, *Trachurus* sp., *Scomber* sp. and *Capros aper*. Cluster analysis and linear discriminant analysis were performed on trawled identified shoals data (808 shoals), after a reduction of the space descriptors by a stepwise technique. A partition into 6 clusters was carried out and clusters are described by species composition and continuous variables values. After grouping the six species shoals in two groups, 84% of shoals from the testing data set could be correctly classified using the linear discriminant function.

Keywords: Echo-integration, fish shoal, fisheries acoustics, statistical analysis, clustering analysis.

MOVIES-B : Logiciel pour la description de détections acoustiques. Application à la classification des bancs de poissons.

Résumé

Le logiciel MOVIES-B, développé par l'IFREMER (Institut français de recherche pour l'exploitation de la mer), est capable de reconnaître et d'individualiser des éléments d'un échogramme comme des structures de détection acoustique en fonction des critères de contiguïté spatiale et énergétique. Ces structures (bancs de poissons, couches planctoniques) sont décrites par des descripteurs de nature énergétique, morphologique et spatio-temporelle. La reconnaissance et l'individualisation des détections acoustiques des bancs de poissons a pour objet d'améliorer les techniques d'écho-intégration permettant l'écho-intégration banc par banc. Le cadre conceptuel et opérationnel de MOVIES-B est présenté dans la première partie de cet article. La deuxième partie, présente les résultats d'une recherche préliminaire afin de vérifier la potentialité de l'utilisation de MOVIES-B dans la classification et l'identification des bancs de poissons par espèce. Cette recherche a étudié les données acquises pendant quatre campagnes halieutiques dans le golfe de Gascogne, à l'aide de l'analyse statistique multidimensionnelle. La matrice ainsi constituée est composée de 13 122 bancs de poissons et de 29 descripteurs. Une connaissance limitée sur l'identification de l'espèce des poissons qui forment les bancs a été obtenue par chalutage. Huit espèces furent capturées : *Sardina pilchardus*, *Clupea harengus*, *Sprattus sprattus*, *Engraulis encrasicolus*, *Micromesistius poutassou*,

Trachurus sp. *Scomber* sp. et *Capros aper*. Après une réduction de l'espace des descripteurs par une méthode pas à pas, la matrice composée par les bancs potentiellement chalutés et identifiés (808 bancs) a été soumise à un algorithme de classification et une analyse linéaire discriminante. Une partition en six classes fut demandée et les classes sont décrites par leur composition spécifique et par les variables continues. Après le groupement de six espèces en deux catégories, 84 % des bancs de l'ensemble de test furent bien classés par la fonction linéaire discriminante.

Mots-clés : Echo-intégration, banc de poissons, acoustique halieutique, analyse statistique, classification hiérarchique.

INTRODUCTION

A good assessment of the size of fish stocks is a prerequisite to efficient fisheries management. Among assessment techniques, hydroacoustic assessment by echo-integration is suitable for pelagic species, specially those having a short life cycle such as anchovy (three years), or for under-exploited stocks. Echo-integration techniques rely on the proportionality between back-scattered energy and fish density. If the fish are limited to a narrow depth range, the measurement is described as area back-scattering, but if they are extended in depth the term volume back-scattering (reverberation) is used (Mitson, 1983).

The effectiveness of acoustic surveys is hampered by several factors: variable fish behaviour resulting in target strength uncertainty, non availability of the fish to the acoustic beam, vessel avoidance and uncertainty in the identification of species from echo traces. Some of these problems affect only the absolute estimates of stock abundance. Estimates of relative abundance or stock indices are sufficient for the management of most resources, provided annual surveys are conducted. Behavioural aspects and target strength are often assumed to be constant and ignored in the calculation of stock indices. Error in the identification of species may be an important source of bias in such indices. Large errors may be introduced by the inability to objectively discriminate acoustic targets into species. When the assessment is carried out in a multispecies environment this kind of error is enhanced.

At present, the acoustically determined biomass is split into specific estimates using identification methods based on visual analysis of echograms and species composition of trawls sampled near the acoustic transect. Problems with trawling are caused first by differences in species' catchability and second, because spatial or temporal sampling methods are not directly comparable with acoustic sampling (Thorne, 1987).

Visual interpretation of echograms, with empirical species discrimination, is a subjective and time consuming approach. Several studies have been performed in order to develop methods to automatically discriminate and recognize species or groups of special acoustic features. Some of these employed wide band echosounders to characterize the spectral signature of different targets (Zakharia and Sessarego, 1982; Simmonds and Armstrong, 1987; Lebourges,

1990). Another group of studies is based upon signal attribute extraction techniques from vertical narrow band echosounders (Deuser *et al.*, 1979; Vray *et al.*, 1987; Rose and Leggett, 1988) or pattern recognition techniques from echogram features (Azzali, 1982; Souid, 1988; Nero and Magnuson, 1989; Kieser and Langford (1991); Reid and Simmonds, 1991; Richards *et al.*, 1991). Up to now, none of these has been used on a large scale.

The research described here is part of a project which attempts the automatic recognition, analysis and classification of shoals from acoustic data. An improved assessment of the pelagic resources of the Bay of Biscay would then be possible. Given this information, fisheries might also become more selective. The first step in the program was to develop software which would be able to distinguish shoal acoustic response from other echoes; the surface should also allow storage and visualization of echoes and compute descriptors for each structure recognized as a shoal. Echogram features were constructed from high-resolution data (0.1 m depth by approximately 2 m distance sampling intervals). The second step was to build a shoal data base and perform multivariate statistical analyses. Acoustic findings were supported by independent catch data.

The first section describes the INES-MOVIES echo processing system. The underlying acoustic model (Johannesson and Mitson, 1983; MacLennan and Simmonds, 1991) is briefly presented to describe the standard integration technique by depth layers. This model is then used to describe a shoal by shoal integration technique. The concepts of acoustic shoal and standard shoal descriptors computed by the system are presented. The second section presents an example of shoal species classification using MOVIES-B data collected by fisheries acoustic surveys in the Bay of Biscay.

INES-MOVIES system

The echo-sounder data processing INES-MOVIES system was developed by IFREMER in 1989 (Diner *et al.*, 1989). It was designed to perform echo-integration from any kind of vertical sounder which has analog and trigger outputs (*fig. 1*). The system is IBM-PC compatible and consists of an electronic interface I.N.E.S. (Interface de

Numérisation des Échos de Sondeur – echo-sounder digital interface) and MO.V.I.E.S. software (MOdule pour la Visualisation, l'Intégration des Echos et leur Stockage – visualization, echo-integration and storage module). In addition to sounder signals, the system is able to receive and process information from other on-board equipment including: heading, speed and position.

INES performs root mean square (RMS) echo level detection and digitizes the analog echo level voltage into 12 bits samples. The A/D sampling frequency is 7.5 kHz, corresponding to a 0.10 meter vertical resolution. INES can also provide a TVG (Time Variable Gain) function. The system provides high-resolution color echogram visualization, echo-integration and storage of raw data. The storage function allows digitized data to be replayed from selected transects or surveys. The software is also able to help acoustic calibration and to control the TVG function of the vertical sounder. An automatic bottom recognition algorithm makes possible the distinction between ocean bottom, background noise and aquatic biomass data, a distinction needed for accurate processing of fish echoes.

MOVIES software was designed to perform standard echo-integration by depth layers, the new version, MOVIES-B, also allows shoal by shoal echo-integration. A shoal, or a fish aggregation, is defined as a set of acoustic samples which satisfy a contiguity law and exceeds predetermined energy and size thresholds.

This definition also holds for a plankton aggregation with different thresholds.

Shoal by shoal echo-integration may be useful to measure the contribution of each shoal to the final biomass assessment and is a new tool to describe spatio-temporal distribution of aquatic organisms. It is also possible to compute quantitative descriptors from each shoal and perform a classification process which could be helpful in species recognition. Shoals are generally assumed to be composed of only one species, hence species recognition of shoals would allow species or group echo-integration.

Standard echo-integration by depth layers

MOVIES software can accomplish standard echo-integration by quantifying back-scattered acoustic energy from specified depth and distance intervals along a transect. When depth layer echo-integration is carried out, the water column is divided into layers, 10 surface referenced and 4 bottom referenced. The total layer processed runs from the upper limit of the first surface referenced layer to the lower limit of the deepest bottom referenced layer. It is also possible to define surface and bottom offset which are not quantified for the echo-integration process.

The system computes, for each layer and for the total layer, four descriptors: S_a , S_b , Q_d and R_v .

- S_a number of signal samples above the echo-integration threshold in the layer.

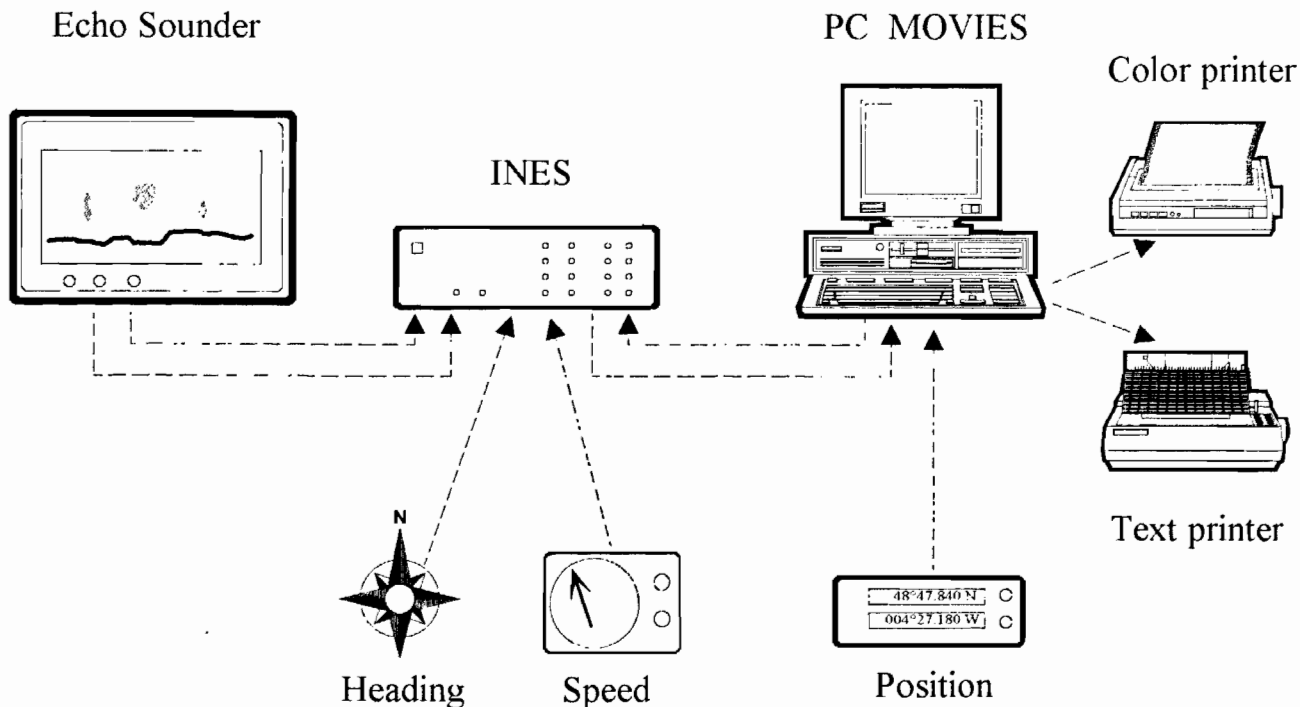


Figure 1. – INES-MOVIES System diagram.

• S_i total number of signal samples in the layer (regardless of threshold value).

• Qd deviation, is related to the back-scattered energy. It is the sum of squared echo amplitude over specified depth intervals and time intervals (distances). It should be vessel speed and echo-sounder pulse rate independent, so the amplitude values measured are weighted by the distance covered between successive pings:

$$Qd = K \sum_{j=1}^N d_j \sum_{i=1}^{n_j} V_{ij}^2 \quad (1)$$

where:

V_{ij} =amplitude of sample i on ping j , measured in volts. $V_{ij}=0$ if the sample value is below echo-integration threshold.

d_j =distance covered between pings j and $j-1$, in meters.

n_j =number of samples integrated for ping j .

N =number of pings over integrated distance.

K =constant which allows the Qd_{ref} convention value calculation. Qd_{ref} has a value of 1000 for 1 volt amplitude signal over a layer of 1 meter height and 1 nautical mile long:

$$K = \frac{1000}{1852 \cdot \eta} \quad (2)$$

The vertical resolution is given by η , representing the number of vertical samples by meter. The distance d_j covered between successive pings may be computed by the relation:

$$d_j = S_j \cdot T_j \quad (3)$$

where S_j is the vessel speed in meters per second and T_j is the ping period in seconds. The Qd is finally given by the relation:

$$Qd = \frac{1000}{1852 \cdot \eta} \sum_{j=1}^N S_j \cdot T_j \sum_{i=1}^{n_j} V_{ij}^2 \quad (4)$$

Rv is a measure of the volume back-scattered energy, it includes the echo-sounder calibration constant. Rv represents the layer volume reverberation index and is expressed in decibels (dB):

$$Rv = 10 \log \frac{Qd}{Qd_{ref} \cdot H \cdot D} + \text{Sounder Constant} \quad (5)$$

where:

D =echo-integration distance in nautical miles.

H =layer height in meters= $S_i/N \cdot \eta$

Qd is defined by equation (4) and Rv may be computed as:

$$Rv = 10 \log \left[\frac{N}{S_i \cdot 1852 \cdot D} \sum_{j=1}^N S_j \cdot T_j \sum_{i=1}^{n_j} V_{ij}^2 \right] + \text{Sounder Constant} \quad (6)$$

Several additional parameters must be included to correctly calculate the volume reverberation index (Johannesson and Mitson, 1983). This acoustic index is appropriate to measure fish density if a $20 \log R$ TVG correction is applied. The transducer directivity must be taken into account by introducing the equivalent beam angle $10 \log \psi$ in the echo-sounder calibration constant. This constant is automatically computed by the MOVIES calibration process according to relation:

$$\text{Sounder Constant} = 20 \log Rm + 2 \alpha Rm - (SL + VR) - 10 \log \psi - 10 \log \left(\frac{1}{2} \tau c \right) \quad (7)$$

where

Rm =TVG range in meters.

α =working frequency absorption coefficient in decibels per meter(dB/m).

$SL+VR$ =echo-sounder on axis acoustic calibration factor, in decibels. This is the sum of source level and voltage response.

ψ =beam factor.

τ =pulse duration, in seconds.

c =speed of sound in water, in meters per second.

Shoal by shoal echo-integration

Accurate stock assessment is essential to improve fisheries management (Andersen and Sutinen, 1984). When the echo-integration technique is used, an improved identification of acoustic detections from fish aggregations and species groups would improve the accuracy of fish biomass assessment. Partitioning the water column into separate depth layers is already an improvement compared to a global echo-integration technique (over the whole water column). But it is an artificial partition and fish or plankton aggregations seldom go along these layers in a natural way. In southern European or in tropical waters, detections are representations of a multispecies environment and the standard echo-integration technique by depth layers usually does not allow discrimination of the acoustically measured biomass by species. This is the main reason for trying to distinguish and measure, in real time, the input of each detection to the final

result and, in the future, recognize the specific nature of each detection.

In order to achieve this objective, a new version of the MOVIES software was developed on 1991 and named MOVIES-B. MOVIES-B performs all of the previous tasks and includes the shoal echo-integration. For each shoal detection a set of quantitative parameters is computed which may be used as input for multivariate statistical analyses and classification.

Acoustic shoal concept

An acoustic shoal detection is defined as a set of samples which forms an echogram feature. The sample set for each detection is composed only by samples which have amplitude values above the echo-integration threshold. The samples also must satisfy a contiguity law, for both vertical and horizontal axis.

To define an echogram feature as a shoal detection, the algorithm must find contiguous samples along the same ping (vertical contiguity) and contiguous samples from one ping to the previous one (horizontal contiguity).

Vertical segments composed by samples with amplitude value above the echo-integration threshold are defined for each ping. Each segment presents vertical contiguous samples, but discontinuities shorter than the number of samples corresponding to the half of the pulse duration are accepted (*i. e.* a segment cannot have any hole of more than $\frac{1}{2}(\tau c \eta)$ samples). To verify horizontal contiguity a search is then made over the values of the processed segment to find at least one sample which has the same depth as other segment sample values from the previous ping. If this search is successful, the two segments are considered as belonging to the same shoal detection. If the detection sample number is not increased after the new ping, the shoal detection is considered as finished and the set of descriptor values is computed. The resolution for horizontal contiguity is the distance covered between successive pings and for vertical contiguity over the same ping it is the distance corresponding to half of the pulse duration.

Echo-integration

In order to compare shoal by shoal echo-integration and standard echo-integration measurements, it is important to use the same acoustic concepts. A unique layer, surface and bottom referenced, is then defined for the shoal echo-integration technique. This layer is equivalent to the total layer measured by the depth layer echo-integration technique. Values of Qd , Rv , S_a and S_l are computed and edited for this unique layer, by echo-integration distance unit. These unique layer values are the same as the total layer values from the standard echo-integration by depth layers. Besides these values, measures of Qd_s , Rv_s , S_{as} and S_{ts} are also computed for each shoal detection.

S_{as} and S_{ts} are, respectively, the number of samples above echo-integration threshold and the total number

of samples inside the contour from each shoal detection.

At the end of each echo-integration distance unit, the sum of shoal Qd_s values increased by Qd values from echoes non classified as shoals, must be the same as the unique layer Qd total value. But it may happen that a shoal detection is not completely finished at the end of the echo-integration distance (see shoal number 4 on *fig. 4*). In this case, the total Qd_s value of this shoal is transferred to the next echo-integration distance. Differences may thus be found between Qd values computed from depth layers and shoal by shoal echo-integration at the end of each distance unit.

Another problem arises when shoals are formed by very long layer aggregations, such as plankton layers or horse-mackerel layers in autumn. Their length may substantially exceed the echo-integration distance unit and a termination convention is needed in order to define the shoal and to compute the descriptor values. The termination convention used is that the shoal length must not exceed two distance units. A shoal already processed at the beginning of the echo-integration distance unit will be automatically truncated at the end of it, if the shoal is not finished. A new shoal will be computed with the remaining samples. Shoal energy values are computed and allocated to echo-integration distance units as previously. Shoals will also be truncated if their boundaries exceed the limits of the echo-integrated layer and the excluded values will not be included in the computation. In both situations, truncated shoals are pointed out as incomplete.

As in the depth layer echo-integration, Qd is related to the back-scattered energy, but shoal Qd_s value is computed only with samples used to define the shoal detection:

$$Qd_s = \frac{1000}{1852 \cdot \eta} \sum_{j=1}^{N_s} S_j \cdot T_j \sum_{i=1}^{n_{j_s}} V_{ij}^2 \quad (8)$$

where:

N_s = number of pings used to define the shoal.

n_{j_s} = number of shoal samples integrated for each ping.

Rv_s is the volume reverberation index of the shoal:

$$Rv_s = 10 \log \frac{Qd_s}{K_s \cdot A} + \text{Sonder Constant} \quad (9)$$

where

A = shoal cross sectional area (m^2).

K_s = unity normalization constant:

$$K_s = \frac{Qd_{ref}}{1852}$$

Table 1. — Symbol definition list.

Symbol	Definition
A	Shoal cross-sectional area (m ²).
c	Speed of sound in water (m/s).
d	Distance covered between successive pings (m).
D	Echo-integration distance in nautical miles.
H	Shoal or layer height (m).
K	Constant which allows the Qd_{ref} convention value calculation.
K_s	Shoal unity normalization constant.
n	Shoal or layer sample number integrated for each ping.
N	Number of pings over integrated distance or shoal length.
S_a	Number of shoal or layer signal samples above echo-integration threshold.
S_t	Total number of shoal or layer signal samples.
Qd	Shoal or layer deviation, value positively related to the back-scattered energy.
Qd_{ref}	Qd convention value of 1 000 for 1 Volt amplitude signal over a layer of 1 metre height and 1 nautical mile long.
Rm	TVG range (m).
Rv	Shoal or layer volume reverberation index.
S	Vessel speed (m/s).
$SL+VR$	Source level and voltage response, factor obtained by the echo-sounder acoustic calibration, in decibels.
<i>Sounder Constant</i>	Echo-sounder calibration constant.
T	Ping period in seconds.
TVG	Time variation gain.
V	Integrated sample amplitude, measured in Volt.
α	Working frequency absorption coefficient (dB/m).
η	Number of signal samples by vertical metre.
τ	Pulse duration, in seconds.
ψ	Beam factor.

Substituting Qd_s and K_s , it is possible to calculate shoal Rv_s , expressed in decibels, by the relation:

$$Rv_s = 10 \log \left[\frac{1}{A \cdot \eta} \sum_{j=1}^{N_s} S_j \cdot T_j \sum_{i=1}^{n_{j_s}} V_{ij}^2 \right] + \text{Sounder Constant} \quad (10)$$

Shoal detection descriptors

For each shoal detection a set of descriptors is computed, stored and printed, in real time. Descriptors are stored in two different files. The first one has a MOVIES-B visual and editing format, the second one is an exporting ASCII free format file. Descriptors may be classified by the nature of their information: general, acoustic, time and space position, morphological, bathymetric and energetic (table 2).

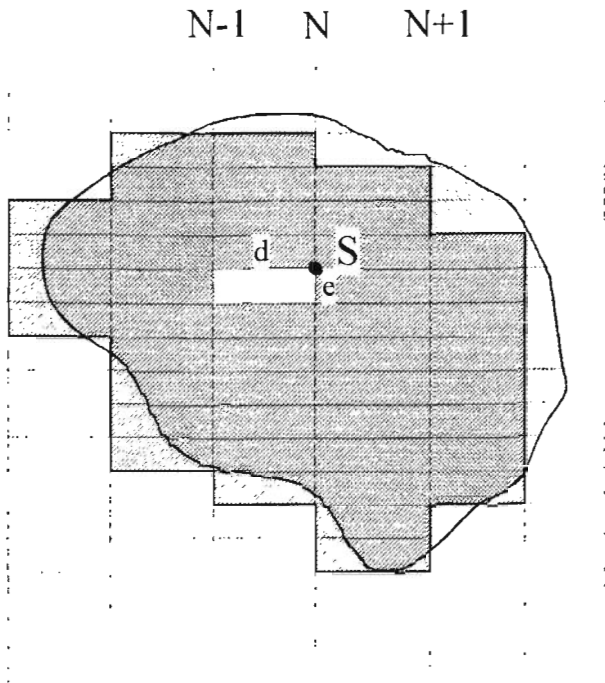
Echogram features may be seriously altered by the vessel movement (heave). This problem could be solved by a software correction for the vessel movement using a heave sensor. But, in the absence of such a device, it seems preferable to take the sea floor as reference.

A very simple model is used for shoal cross section calculations. A rectangle is associated to each sample S with amplitude value above the echo-integration threshold. The rectangle surface S_e is computed as the product of the horizontal distance covered since the precedent ping, d_j , and the vertical distance covered since the precedent sample, e , where $e = \frac{1}{\eta}$, (fig. 2):

$$S_e = d_j \cdot e \quad (11)$$

Table 2. — Summary of descriptors computed for each shoal by INES-MOVIES System.

Abbreviation	Shoal Descriptor	Unit
General		
	File identifier and shoal feature number	
	Survey	
<i>Loch</i>	Relative distance covered (~ESDU)	0.1 mile
S	Vessel speed	knot
Acoustic		
F	Echo-sounder frequency	kHz
N	Number of pings	
S_a	Sample number above echo-integration threshold	
S_t	Total sample number	
Time and space position		
<i>Year</i>	Year	
<i>Day</i>	Day of the year	
<i>Hour</i>	Time	decimal
<i>Quad</i>	Geographic quadrant	
<i>Lati</i>	Latitude	decimal
<i>Long</i>	Longitude	decimal
Morphological		
H	Height	meter
L	Length	m
P	Perimeter	m
A	Cross-sectional area	m ²
$DFrct$	Fractal dimension	
$Elon$	Elongation	
Bathymetric		
$Depth$	Bottom depth	m
$Dmin$	Shoal depth	m
$Amin$	Shoal minimal altitude	m
$Arel$	Shoal altitude index	
Energetic		
Qd	Deviation	
Rv	Volume reverberation index	decibel (dB)
E	Back-scattered energy	mV ²
A_{max}	Amplitude sample maximal value	mV
A	Amplitude mean value	mV
A_s	Amplitude standard deviation	mV
A_{cv}	Amplitude variation coefficient	




 Digitized shoal
Actual shoal

Figure 2. – Shoal digitalization.

These rectangles are used to compute shoal descriptor values. Shoal perimeter is computed as the sum of external rectangle faces located at the shoal boundary. Shoal area is the sum of S_e . If there is a hole inside the shoal (a discontinuity longer than $\frac{1}{2}(\tau c \eta)$ samples) or adjacent pings with incomplete vertical coverage, the perimeter is increased by the hole perimeter and the shoal area is decreased by the hole area.

Some special shoal points are used as references to compute morphological and bathymetric values (fig. 3 and 4). The fractal dimension (Mandelbrot, 1983) is a descriptor used to characterize the unevenness of the shoal detection outline. It is a relationship between the surface of a square, with perimeter equivalent to the shoal perimeter, and the shoal surface (Nero and Magnuson, 1989), see table 2 for parameter definitions.

$$DFrct = 2 \cdot \frac{\ln \frac{P}{4}}{\ln A} \quad (12)$$

A fractal dimension value of 1 represents a very smooth or square outline shape shoal and a value of 2

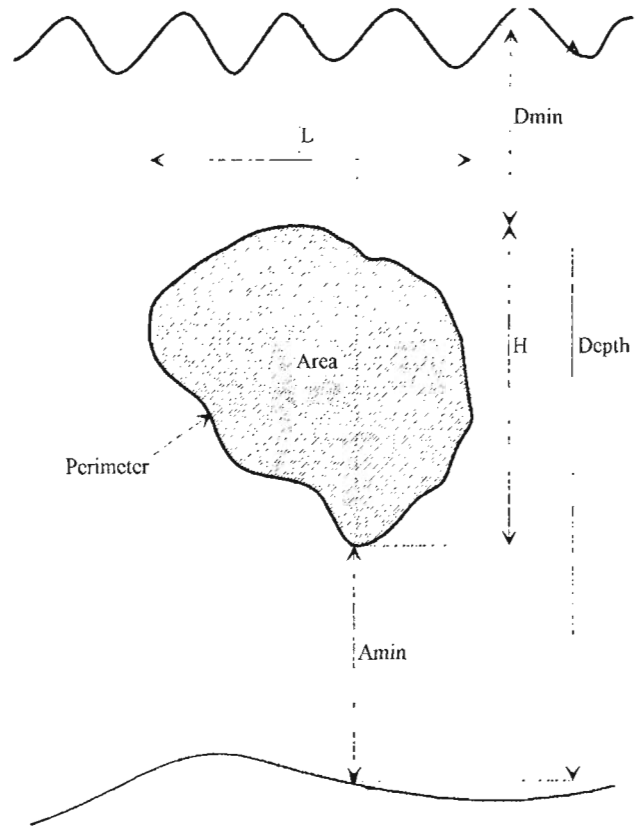


Figure 3. – Reference marks of shoal descriptor computations.

an uneven or complex outline shape shoal. Elongation is determined from the length to height relationship:

$$Elong = \frac{L}{H} \quad (13)$$

Altitude index is used to measure the shoal position in the water column and may be more useful than absolute altitude values. It is expressed in percentage by the relation:

$$Arel = 100 \frac{\left(Amin + \frac{H}{2} \right)}{Depth} \quad (14)$$

Shoal back-scattered energy is measured as the sum of squared sample amplitude values above echo-integration threshold over each ping and for all shoal pings. Distances between pings may not be constant, so each sample amplitude value is weighted by d . Amplitude values are corrected by the sounder constant in order to get independent acoustic equipment measures. Shoal back-scattered energy is not expressed in decibels but in linear units, square millivolts, by the relation:

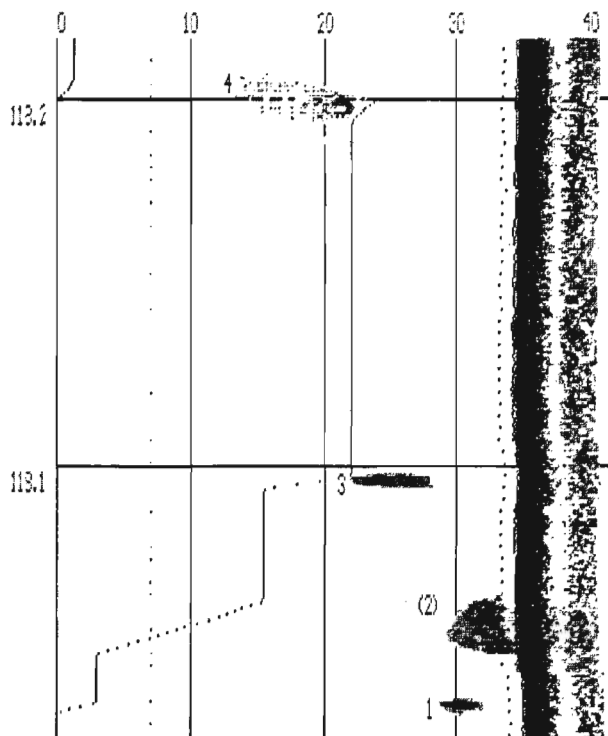


Figure 4. – Example of echogram processed by MOVIES-B with recognized shoals. Some descriptor values from shoal number 4 are: Qd_s : 39.6, N_s : 21, S_{as} : 1314, $Depth$: 34.5 m, $Dmin$: 13.2 m, H : 11.0 m, L : 24.5 m, $Amin$: 10.3 m, P : 78.2 m, A : 163.85 m².

$$Energy = \frac{1}{\eta} \sum_{j=1}^N S_j \cdot T_j$$

$$\sum_{i=1}^{n_j} V_{ij}^2 \cdot 10^6 \cdot 10^{Sounder\ Constant/10} \quad (15)$$

Multiple scattering of echoes inside the shoal is partially corrected to compute morphological and bathymetric descriptor values. It is known that echogram features from high density shoals may present a downward tail. This tail is composed by multiple scattering of signals inside the shoal before being received by the transducer. Delayed echoes seem to come from below the shoal and may artificially increase perimeter, cross sectional area and height values or modify the true value of shoal bathymetric position.

The selected correction removes, from the lower part of the shoal segment analyzed, all samples with amplitude value lower than 10% of segment amplitude maximal value. The number of samples removed is displayed. The correction is carried out before computing morphological and bathymetric descriptor values, but it is not used in computing energy values. Indeed, all back-scattered energy should be recovered

to compute values correlated with the shoal fish density and multiple scattering is a signal of higher density.

No account is taken of depth dependent distortions arising from the sounder beam width nor of pulse length elongation distortions, as corrections may be made by *a posteriori* processing methods such as described in (Scalabrin and Massé, 1993).

SHOAL CLASSIFICATION

Data collection at sea was performed using the INES-MOVIES System, connected to a 38 kHz narrow-band echo-sounder with TVG amplification of 20 log R. The pulse duration was 1 ms. The A/D sampling frequency of the echo envelope was 7.5 kHz which enables a 0.10-m vertical resolution. The detection data from four fishery acoustic surveys performed in the Bay of Biscay in 1989 and 1990 have been processed by an earlier version of MOVIES-B and a data matrix was built. This matrix contains 13,122 shoals detected during trawling or, when surveying, after or before the catches. Only day-time data, when fishes are aggregated in shoals, have been processed. Multivariate statistical analyses were performed using the SPAD.N package (Lebart et al., 1988) on a Cyber 992-31 computer. Methods are described elsewhere (Morrisson, 1976; Lebart et al., 1984).

Eight species were present in the catches: sardine *Sardina pilchardus*, sprat *Sprattus sprattus*, herring *Clupea harengus*, anchovy *Engraulis encrasicolus*, blue-whiting *Micromesistius poutassou*, mackerel *Scomber* sp., horse-mackerel *Trachurus* sp. and *Capros aper*.

The data matrix (13,122 shoals × 29 descriptors) was subjected to a principal components analysis which was applied as a method designed to summarize the information contained in the data matrix. Among descriptors, 17 were active continuous variables and the others were supplementary variables (geographic position, date, hour, species' name, survey identifier). Continuous variables were standardized in order to remove the effects of different scaling and no assumptions were made about their distributions.

The early MOVIES-B version used to generate shoal descriptor values did not compute volume reverberation index, energetic descriptors nor fractal dimension as described in the first part of this paper. Circularity, unevenness 1 and unevenness 2 were descriptors used to measure the degree of outline smoothness of shoal detections. Circularity is the relationship between the radius of a circle with the same perimeter as the shoal and the radius of a circle with the same area as the shoal and is computed by:

$$Circularity = \frac{P^2}{\pi \cdot A} \quad (16)$$

Unevenness 1 measures the difference between the shoal perimeter and the rectangle perimeter computed from the shoal height and length normalized by the number of pings used to define the shoal:

$$Unevenness\ 1 = \frac{P - 2(L + H)}{N} \quad (17)$$

Unevenness 2 is the relationship between the shoal perimeter and the rectangle perimeter computed from the shoal height and length:

$$Unevenness\ 2 = \frac{P}{2(L + H)} \quad (18)$$

Shoal energetic characteristics are measured by total energy and two energy indexes. The first is a ratio between shoal energy and shoal cross sectional area, the second is a ratio between shoal energy and shoal length:

$$Energy\ index\ 1 = \sqrt{\frac{1.852 E}{A}} \quad (19)$$

$$Energy\ index\ 2 = \frac{E}{L} \quad (20)$$

Computations are described in (Scalabrin, 1991) and detailed results of the principal components analysis in (Scalabrin *et al.*, 1992).

From 13,122 shoals, 808 were identified with the highest level of certainty. These shoals were those detected during trawling when the catch was monospecific (up to 95% of one species), and potentially caught by the trawl (localized with the netzond). No herring or sprat shoal were included in this new matrix since these species never appeared alone in the catches.

Only identified trawled shoals (808 shoals) were selected as input data to cluster and linear discriminant analysis. A stepwise linear regression was carried out in order to find the best set of discriminant variables, using the squared multiple correlation coefficient as the selection criterion. After the space dimension reduction, only eight continuous variables were retained: bottom depth, shoal minimal altitude, altitude index, shoal height, perimeter, elongation, circularity and energy index 1.

The clustering technique used a mixed algorithm applied over shoal factorial coordinates computed by the principal components analysis. A first clustering is performed by a *k*-mean algorithm. Clusters formed by this method are then aggregated in new clusters by a hierarchical classification algorithm. The strategy of clustering is based on the minimum variance criterion (Lebart *et al.*, 1984). The final partition of the population into 6 classes was defined from the analysis of the dendrogram; the species' composition and

the continuous variables values are used to describe classes.

To perform a linear discriminant analysis and calculate the coefficients for the linear function it was necessary to group species in two classes. The method used here is that developed previously by IFREMER in analysing echo-integration data (Massé, 1988). Horse-mackerel, blue-whiting and *Capros aper* are grouped as DEP1 class and sardine, anchovy and mackerel as DEP2 class.

RESULTS

Clustering analysis

After the space dimension reduction, a principal components analysis (PCA) was performed on the new matrix (808 shoals × 8 descriptors). PCA was useful in setting up a structure for the data and in providing a new descriptor formed by shoal coordinates in the descriptors space. The resulting interpretation of a PCA is based upon two characteristics: (a) the correlation matrix is symmetrical and the axes are the orthogonal eigenvectors which are linear independent directions in the observation space; (b) the eigenvalues of the correlation matrix measure the quantity of variance which is explained by each eigenvector or axis (Legendre and Legendre, 1984).

In the analysis of the shoal matrix, 76% of the variance is explained by the first three axes. The first axis corresponding to the eigenvalue of 38%, is characterized by perimeter, circularity and shoal height, which are correlated descriptors. The second axis, with an eigenvalue of 22%, describes the bathymetric position of the shoal. Energy index 1 is the descriptor characterizing the third axis, which has an eigenvalue of 16%.

It is important to note that these groups of variables are independent of each other (Lebart *et al.*, 1984). So the position of the shoal in the water column is not correlated with its size nor with its energetic characteristics.

Figure 5 represents the centroids of the individual species observations which were identified by monospecific trawlings. On the first factorial plane it is possible to observe an opposition between blue-whiting and the other species. This opposition is characterized by the fact that blue-whiting shoals are larger and with higher altitude values than shoals from other species. The second group of species is opposed by the first axis. Sardine and anchovy shoals are smaller than mackerel, horse-mackerel and *Capros aper* shoals.

There are two groups of species on the second factorial plane. Sardine, anchovy and horse-mackerel are opposed to blue-whiting, mackerel and *Capros aper* by the third axis. Sardine shoals have greater values for energy index 1 than blue-whiting shoals, which are not so dense.

Table 3. – Species composition of clusters obtained by the hierarchical classification (shoals from monospecific hauls), (a) distribution of shoal observations among clusters, expressed in percentage of the total number of shoals for each species, (b) species composition of each cluster expressed in percentage of the cluster total number of shoals (N: number of shoals).

Cluster	Sardine	Anchovy	Mackerel	Horse mackerel	Blue whiting	<i>Capros aper</i>	(%)
1	21.7 (a)	61.4	61.9	48.3		96.3	
N=419	1.2 (b)	50.8	12.4	23.2		12.4	100.0
2	13.0	1.2		25.4			
N=58	5.2	6.9		87.9			100.0
3	65.2	37.2	31.0	24.4			
N=219	6.9	58.9	11.9	22.4			100.0
4					69.7		
N=69					100.0		100.0
5		0.3	4.8	2.0	26.3	3.7	
N=37		2.7	10.8	10.8	70.3	5.4	100.0
6			2.4		4.0		
N=6			33.3		66.7		100.0
N=808	100	100.0	100.0	100.0	100.0	100.0	

Thirty stable clusters were found by the clustering algorithm, but a great number with few shoals. A partition into 6 clusters was performed after analysis of the dendrogram and the hierarchical tree indices histogram. The species' composition of clusters is presented in table 3 and statistics in table 4.

Table 4. – Clusters statistical description - *t* V (test Value) is a two-tailed test analogous to "Student's *t*-test", associating mean comparisons. The hypothesis of significant differences among descriptors or axes is possible if the value for the test is outside the range $-t_{0.995}$ to $t_{0.995}$ (or -2.58 to 2.58), with a significance level at 0.005 and ∞ degrees of freedom. The test value sign indicates the relative position of the class mean value: above (+) or below (-) the general mean value. \cos^2 is the relative contribution value.

Cluster	Descriptor			Axis	
	Mean value			\cos^2	<i>t</i> -V
	cluster	general	<i>t</i> V		
Cluster 1					
axis 2				0.625	-17
axis 3				0.196	-11
AMIN	5.5	13.9	-18		
AREL	7.7	14.2	-14		
Cluster 2					
axis 3				0.457	14
axis 2				0.353	-11
ELON	24.2	8.3	16		
IND1	5.5	2.3	14		
Cluster 3					
axis 2				0.587	18
axis 1				0.272	-9
AREL	32.7	14.2	24		
AMIN	27.4	13.9	17		
Cluster 4					
axis 4				0.315	15
axis 2				0.222	9
DEPTH	382.6	120.3	22		
AMIN	31.4	13.9	11		
Cluster 5					
axis 1				0.932	15
HEIGHT	23.8	7.0	17		
CIRC	111.9	20.8	12		
PERI	1625.2	252.0	12		
Cluster 6					
axis 1				0.851	18
PERI	6636.5	252.0	23		
CIRC	404.6	20.8	21		

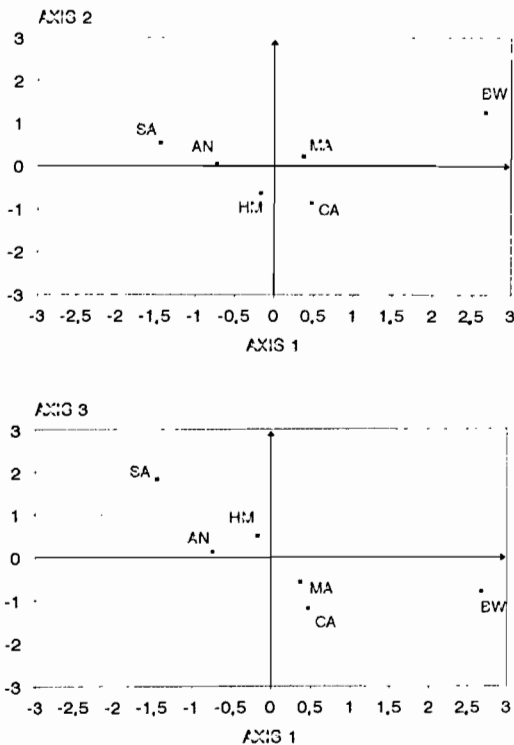


Figure 5. – Barycentres of fish species on the factorials planes. *a*: first factorial plane (axis 1 \Rightarrow size and shape descriptors; axis 2 \Rightarrow shoal bathymetric position). *b*: second factorial plane (axis 3 \Rightarrow shoal energy index). AN: anchovy; BW: blue-whiting; CA: *Capros aper*; HM: horse-mackerel; MA: mackerel; SA: sardine.

The first cluster is composed of shoals from five species, but anchovy and horse-mackerel are the most important. The second axis, with the highest relative contribution value, has a greater influence over the position of this cluster than other axes. This second axis represents shoal bathymetric position and shoals from this cluster are characterized by small altitude values.

The second cluster is basically composed of horse-mackerel shoals and the third axis (energy) has the highest relative contribution value. These high values of energy index for horse-mackerel shoals are not common and may be related to a seasonal behaviour, since all horse-mackerel shoals from the data base were detected by late summer. At this period the species migrates towards offshore (Quéro, 1984) and it is possible that this general movement may alter shoal behaviour giving higher density values and elongated shoals.

The third cluster is a mixed one, composed of anchovy, horse-mackerel, mackerel and with 65% of sardine shoals. It is also characterized by the bathymetric position axis, with shoals in the middle of the water column. The fourth cluster is composed only of blue-whiting shoals, characterized by the highest bottom depth values. Cluster number five is mixed, composed basically by blue-whiting, horse-mackerel and mackerel shoals. It is characterized by the first axis with the highest relative contribution. Shoals have high values for height, circularity and perimeter. The last cluster has only six shoals: 4 from blue-whiting and 2 from mackerel. Here also the first axis has the greatest influence over the cluster position with shoals presenting very high perimeter and circularity values.

Linear discriminant analysis

The linear discriminant function was computed from the same eight descriptors used by the clustering analysis. Species were grouped in two classes: horse-mackerel, blue-whiting and *Capros aper* are grouped as DEP1 class and sardine, anchovy and mackerel as DEP2 class. Shoals from the learning set were chosen by a random process, and remaining shoals (30% of shoals) formed the testing set.

This linear function was able to well classify 84% of shoals from the testing set. But DEP2 shoals were better classified (97%) than DEP1 shoals (66%) (table 5). Histograms of linear discriminant function values (fig. 6 a and b) show that the DEP1 class presents a bimodal distribution, the right mode is composed of blue-whiting shoals, and the left one by horse-mackerel shoals. DEP2 distribution is unimodal with anchovy shoals values closer to zero. This means that poorly classified DEP1 shoals are mostly horse-mackerel shoals, which may be mislabelled as anchovy shoals.

Table 5. – Results from the linear discriminant function when applied to the shoal testing set, expressed in number of shoals. Numbers inside brackets correspond to percentages of the total number of shoals.

Groups	Classification		
	well classified	poorly classified	Total
DEP1	70 (66.0)	36 (34.0)	106 (100.0)
DEP2	134 (98.5)	2 (1.5)	136 (100.0)
Total	204 (84.3)	38 (15.7)	242 (100.0)

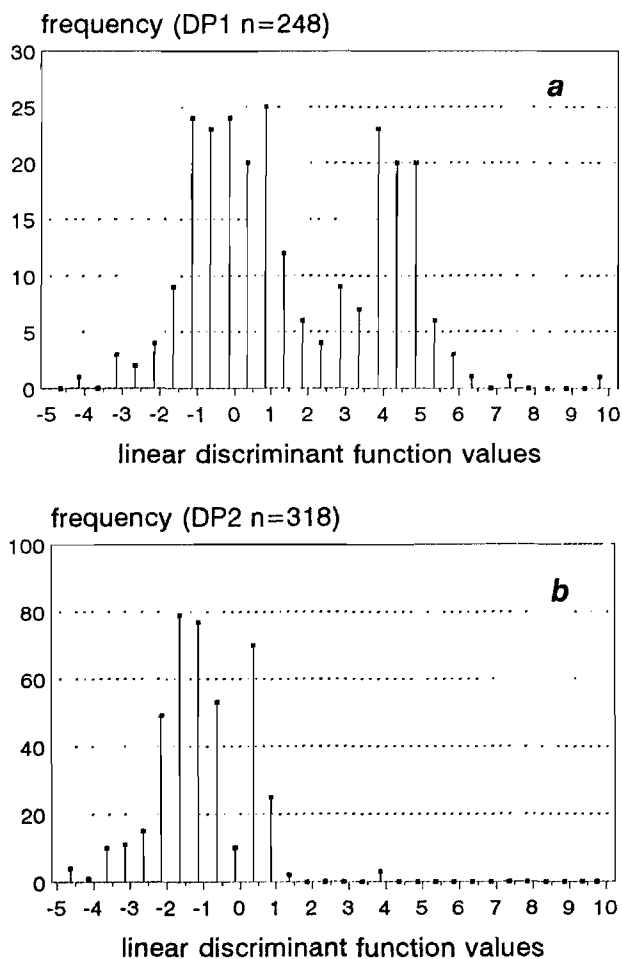


Figure 6. – Linear discriminant function values histograms for DEP1 (a) and DEP2 (b) groups.

DISCUSSION AND CONCLUSION

The system INES-MOVIES and the MOVIES-B software have been presented and described. They present some major features which can be useful in the fisheries research or in the fisheries industry:

- MOVIES-B uses MS-DOS as its exploitation system allowing a high level of portability;

- besides sonar signals, the system is able to receive and process information from other on board equipment including: heading, speed and position. These input data assign real and automatic time-space positions to each detection;

- all computations are accomplished in real time;
- and the most important characteristic of MOVIES-B software is the ability to perform shoal by shoal echo-integration. Results from standard integration by depth layer and shoal by shoal integration are equivalent and provide the same acoustic biomass assessment.

The automatic shoal recognition algorithm is an important difference between the MOVIES-B approach and other methods. These other methods are based on a subjective decision approach, requiring an operator intervention to decide which set of samples may define an interesting acoustic patch (Foote *et al.*, 1991; Reid and Simmonds, 1991).

It is known that operators or fishermen tend to choose only well defined and characteristic acoustic patches when trying to discriminate echoes among fish species. These well defined and characteristic acoustic patches represent a small percentage of acoustic detections of shoals, considering variability in shoaling formation and variability observed in echogram features. In order to improve knowledge about shoal behaviour or to discriminate echoes among fish species, a probabilistic approach might be used, requiring an automatic decision rule to define what is an acoustic shoal detection. MOVIES-B algorithm relates acoustic samples by a contiguity law without operator intervention. The set of samples is then defined as a shoal and descriptor values are computed.

An automatic approach is also adopted by research which quantifies and characterizes patches along acoustic transects in order to find a correlation between echo patches, biological structures and oceanographic features (Nero and Magnuson, 1989; Baussant *et al.*, 1993) or by research which classifies fish assemblages (Richards *et al.*, 1991). The algorithms proposed by these authors operate with automatic decision rules, but averaging acoustic data. Horizontal resolution varies from 250 meters (Richards *et al.*,

1991) to 1.4 km (Nero and Magnuson, 1989). This resolution level is not appropriate to study shoals as independent entities. The MOVIES-B shoal recognition algorithm operates with the highest possible horizontal resolution: the distance covered by the vessel between pings, approximately 2 meter distance sampling intervals. The horizontal resolution depends on the vessel speed and on the ping period.

This new echo-integration technique may improve assessment accuracy if a key is worked out to classify shoal acoustic detections by species. Species identification could then be related automatically to each shoal and would become independent of quantitative trawling results. There would be compatibility between the level of acoustic sampling and species identification. This is not the case when identification is based on trawling results and extrapolated over the whole survey area. In order to give an indication of this possibility, two classifying methods were performed using survey data: clustering and linear discriminant analysis.

The results produced by this research show a multidimensional structure for the data analysed, indicating that values from shoal descriptors contain a certain amount of information and are not random. Classification results are of the same order as those presented by other studies (Vray *et al.*, 1987; Rose and Leggett, 1988). Nevertheless, the information present in the data set was not sufficient to allow perfect discrimination among all species, specially between horse-mackerel and anchovy. This was the major reason for improving the MOVIES-B shoal energetic structure description.

The quality of the analysis was affected by two factors. First, the identification of shoal species. To obtain a higher level of certainty in the identification, the size of the data base must be restricted (13,122 \Rightarrow 808 shoals). There were large variations in the numbers of identified shoals of each species and sardine was represented only by 23 shoals. Secondly, the survey sampling, which may not have covered all the variability in shoal formation in a representative manner.

Acknowledgments

This research was partially financed by the Commission of the European Communities within the framework of the EEC Research Programme in the Fisheries Sector, contract number MA.2.542. We would like to thank the anonymous reviewers for their helpful comments and M. Peter Davies for the English improvements.

REFERENCES

- Anderson P., J. G. Sutinen, 1984. Stochastics Bioeconomics: a review of basic methods and results. *Mar. Res. Econ.*, **1**, 117-136.
- Azzali M., 1982. Regarding the possibility of relating echo signal features to classes of marine organisms: tests carried out in the north and middle Adriatic sea. *In: Symposium on Fisheries Acoustics*, 21/24 June, Bergen, Norway.
- Baussant T., F. Ibanez, M. Etienne, 1993. Numeric analysis of planktonic patterns revealed by echograms. *Aquat. Living Resour.*, **6**, 175-184.
- Deuser L. M., D. Middleton, T. Plemons, J. Vaughan, 1979. On the classification of underwater acoustic signal: II. Experimental applications involving fish. *J. Acoust. Soc. Am.*, **65**, 445-455.
- Diner N., A. Weill, J. Y. Coail, J. M. Coudeville, 1989. INES-MOVIES: a new acoustic data acquisition and processing system. ICES CM 1989/B:45, Fish Capture Committee.
- Foot K. G., H. P. Knudsen, R. J. Korneliusen, 1991. Postprocessing system for echo sounder data. *J. Acoust. Soc. Am.*, **90**, 37-47.
- Johannesson K. A., R. B. Mitson, 1983. Fisheries acoustics. A practical manual for aquatic biomass estimation. *FAO Fish. Tech. Pap.*, **240**, 249 p.
- Kieser R., G. Langford, 1991. An application of spatial analysis to fisheries acoustics. *In: Proceedings of GIS'91 Symposium*, Vancouver, Canada, 335-339.
- Lebart L., Morineau, K. M. Warwick, 1984. Multivariate Descriptive Statistical Analysis. John Wiley & Sons, New York, 231 p.
- Lebart L., A. Morineau, T. Lambert, 1988. SPAD.N: Système Portable pour l'Analyse des Données, version 1.2. CISIA, Sèvres, 306 p.
- Lebourges A., 1990. Utilisation de la spectroscopie ultrasonore en vue d'identifier les espèces de poisson. Thèse dr. Univ. Paris-VI, France.
- Legendre L., P. Legendre, 1984. Écologie Numérique. Masson et Presses de l'Université du Québec, Paris, volumes 1 and 2, 260 et 335 p.
- MacLennan D. N., E. J. Simmonds, 1991. Fisheries Acoustics. Chapman & Hall eds., England, 325 p.
- Mandelbrot B., 1983. The fractal geometry of nature. Freeman & Cie, New York, 468 p.
- Massé J., 1988. Utilisation de l'écho-intégration en recherche halieutique. Rapp. DRV-88.030/RH, Nantes, IFREMER.
- Mitson R. B., 1983. Fisheries Sonar. Fishing News Books Ltd, Farnham, England, 287 p.
- Morrison D. F., 1976. Multivariate statistical methods. McGraw-Hill, New York, 415 p.
- Nero R. W., J. J. Magnuson, 1989. Characterization of patches along transects using high-resolution 70 kHz integrated acoustic data. *Can. J. Fish. Aquat. Sci.*, **46**, 2056-2064.
- Quéro J. C., 1984. Les poissons de mer des pêches françaises. Jacques Grancher éd., Paris, 394 p.
- Reid D. G., E. J. Simmonds, 1991. The application of image analysis techniques to the study of fish shoal structure using acoustic survey data. FAST Working Group, Fish Capture Committee, Ancona.
- Richards L. J., R. Kieser, T. J. Mulligan, J. R. Candy, 1991. Classification of fish assemblages based on echo integration surveys. *Can. J. Fish. Aquat. Sci.*, **48**, 1264-1272.
- Rose G. A., W. C. Leggett, 1988. Hydroacoustics signal classification of fish schools by species. *Can. J. Fish. Aquat. Sci.*, **45**, 597-604.
- Scalabrin C., 1991. Recherche d'une méthodologie pour la classification et l'identification automatiques des détections acoustiques des bancs de poissons. Rapp. IFREMER, DITI/NPA 91.23.
- Scalabrin C., A. Weill, N. Diner, 1992. The structure of multidimensional data from acoustic detection of fish schools. *In: European conference on underwater acoustics*, ed. M. Weydert, Elsevier Applied Science, 141-146.
- Scalabrin C., J. Massé, 1993. Acoustic detection of the spatial and temporal distribution of fish shoals in the Bay of Biscay. *Aquat. Living Resour.*, **6**, 269-283.
- Simmonds E. J., F. Armstrong, 1987. A wide band echo-sounder: measurements on cod, saithe, herring and mackerel from 27 to 54 kHz. *In: International Symposium on Fisheries Acoustics*, June 22-26, Seattle, Washington, USA.
- Souid P., 1988. Automatisation de la description et de la classification des détections acoustiques de bancs de poissons pélagiques pour leur identification. Thèse dr. Univ. Aix - Marseille-II, France.
- Thorne R. E., 1987. Hydroacoustics and ground truth. *In: International Symposium on Fisheries Acoustics*, June 22/26, Seattle, Washington, USA.
- Vray D., G. Gimenez, R. Person, 1987. Attempt of classification of echo-sounder signal based on the linear discriminant function of Fisher. *In: International Symposium on Fisheries Acoustics*, June 22/26, Seattle, Washington, USA.
- Zakharia M., J. P. Sessarego, 1982. Sonar target classification using a coherent echo processing. *In: IEEE International Conference on acoustics, speech and signal processing*, p. 331-334, Paris.