Precision of acoustic fish abundance estimates: separating the number of schools from the biomass in the schools

Emile Marchal [1] and Pierre Petitgas [2]

[1] *ORSTOM, Institut Océanographique, 195, rue Saint-Jacques, 75005 Paris, France.*

[2] *ORSTOM, Laboratoire d'Informatique Appliquée, 72, route d'Aulnay, 93143 Bondy Cedex, France.*

Marchal E., P. Petitgas. *Aquat. Living Resour.,* 1993, **6**, 211-219.

Abstract

The gulf of Cariaco, Venezuela, is a rather isolated piece of coastal ocean where spatio-temporal variability has been measured. In 1986, three repeated surveys have been carried out in this gulf. The important differences in biomass estimates are attributed to sampling error only. The biological nature of this variability is studied. It is due to a high concentration of the biomass in a few schools and correlatively to the hit or the miss of these large schools, during the sampling. Structural information for predicting them is sought. Also, the number of schools per elementary sampling distance unit (ESDU) and the biomass in each school were measured by echointegration. It is found that there is no clear relation between these two variables. Thus their spatial structures are studied separately, using geostatistics. The number of schools per ESDU is structured in space. Its histogram is not very dissymmetrical. The error variance of its mean estimate is reasonable. The biomass per school is not structured in space. Its histogram is highly skewed. The error variance of its mean estimate is high. The major error variance term on the abundance is due to the imprecision on the estimate of the mean biomass per school. The major quantitative problem is the indeterminancy at the sampling stage of the maximum biomass per school and correlatively the major problem is the characterization of the skewness of the biomass per school histogram. The possibility of using the number of schools as an index of biomass is discussed.

Keywords : Stock assessment, echo surveys, reliability.

Résumé

Le golfe de Cariaco, situé au Venezuela, se présente comme une petite mer assez fermée où l'on peut estimer qu'à court terme les échanges avec l'océan sont négligeables. Pour mesurer la précision des évaluations acoustiques dans cette région, trois campagnes de prospection y ont été réalisées en août 1986. Les différences importantes trouvées dans les estimations de biomasse ont été attribuées aux erreurs d'échantillonnage. L'analyse des données élémentaires montre qu'en fait cette variabilité est à la fois de nature biologique et statistique : en effet la biomasse apparaît comme très concentrée en quelques gros bancs que les hasards de l'échantillonnage vont faire rencontrer ou non. On a alors étudié leur répartition spatiale. Aucune relation entre le nombre de bancs par unité de distance élémentaire (ESDU) et l'énergie rétrodiffusée de chaque banc n'a pu être mise en évidence. Aussi leurs structures spatiales ont-elles été étudiées séparément. Il apparaît une certaine structuration spatiale du nombre de bancs par unité de distance. L'histogramme de ces valeurs n'est pas très dissymétrique, et la variance d'estimation de la moyenne est raisonnable. Par contre la biomasse (énergie) par banc n'est pas structurée spatialement. L'histogramme de ses valeurs est très dissymétrique et la variance d'estimation de la moyenne est élevée. Il est donc clair que le terme principal entrant dans la variance d'estimation de l'abondance est constitué par la biomasse moyenne par banc. Le problème majeur en terme quantitatif provient de l'incertitude sur les très grosses valeurs de biomasse par banc et du caractère extrêmement asymétrique de l'histogramme de distribution des valeurs de biomasse par banc. L'utilisation du nombre de bancs comme indice d'abondance est discutée.

Mots-clés : Evaluation des stocks, campagnes acoustiques, précision.

INTRODUCTION

The problem focused on here concerns the variability of the abundance estimate of pelagic fish stocks derived by acoustic surveys. Quantitative evidence is given here both experimentally and formally that the imprecision on the abundance estimate is caused by the hit or the miss of schools of large biomass during the sampling.

Three surveys were repeated in the same area. The three biomass estimates were very different and the reasons for this are investigated. These differences are attributed to spatial variability. This paper is an attempt to characterize biologically the spatial variability of pelagic biomass by separating fish density into two components, the number of schools per sea surface unit and the biomass in the schools.

Biologically, the school biomass is related to schooling processes of concentration and the density in number of the schools is related to the way the habitat is occupied. The separation is important as the two variables may have different biological meanings. Their implications in the abundance estimation process are characterized and quantified.

In acoustic surveys the data are obtained by accumulating measurements from individual schools and dispersed fish. So the process generating the data may be decomposed into its different components. Such a decomposition is not new: MacLennan and MacKenzie (1988) interpretated their data with a model of school occurrence in the purpose of testing an alternative biomass estimate. Our objective is different. We focus on the decomposition of the variance of the estimate. This structural analysis enables us to state statistically where the major source of imprecision lies and how it could be treated.

MATERIAL AND METHOD OF SAMPLING

The Gulf of Cariaco, Venezuela, has served as an experimental area for measuring spatio-temporal variability. Gerlotto and Petitgas (1991) have given field evidence in this gulf of temporal variability limitations when surveying the area. This implies a necessary compromise between the temporal and spatial scales of variability when designing a survey. The surveys we are concerned with are performed along regularly spaced transects crossing the gulf's width, the inter-transects being coastal. The inter-transect distance is chosen so that the acoustic survey of the entire gulf takes less than 12 hours. Three surveys of the gulf were repeated along exactly the same track (*fig.* 1). The second and third surveys occurred 8 and 9 days after the first one. The biomass consists mainly of pelagic fishes (Clupeidae, Engraulidae, Carangidae) which form schools during the day. The surveys were performed during daytime.
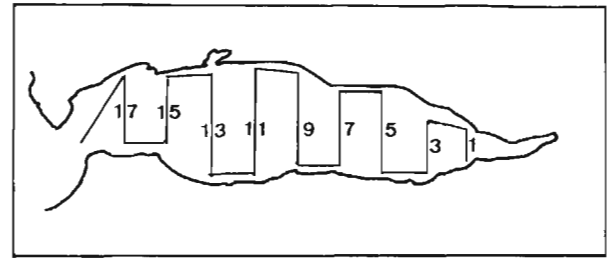

Figure 1. – The repeated ship's track of the acoustic surveys in the Gulf of Cariaco (Venezuela).

Differences between the three surveys can be attributed to temporal variability and/or to sampling error. Differences in the spatial distribution thus cannot be easily interpreted, but differences in the gulf's biomass estimates can be. The Gulf of Cariaco is a rather closed piece of ocean which serves as a nursery area for the young pelagic fishes. Thus it it reasonable to postulate that in and out movements of fish are insignificant during the 9-day period during which the three surveys were repeated. The gulf can be seen as a big ocean pond where the conditions are good for studying the influence of sampling error due to high spatial dispersion, on the abundance estimate.

The echo sounder was an EKS 120 kHz from SIMRAD. An analogic echo integrator QM MKII was used. The echoes were integrated along the ship's track at regular time intervals. The elementary sampling distance unit (ESDU) is on average one nautical mile. The schools were isolated during the survey and the integration was also performed for each school, thus giving a measure of the biomass in each school. The number of schools per ESDU were recorded but not their precise location. The conversion of the integrated echoes (backscattering units) into actual fish biomass was not performed. The calibration and the equipment adjustments were the same for the three surveys. Therefore a direct comparison of abundance measures is possible between the three surveys.
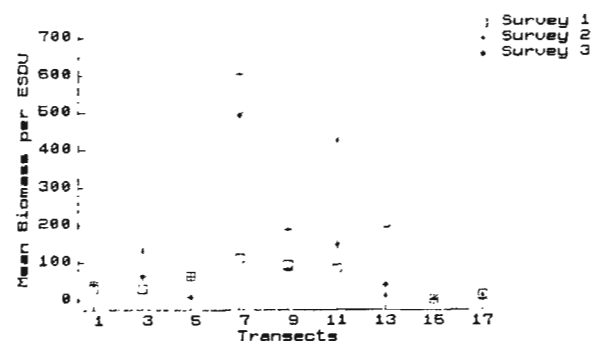

Figure 2. – Mean biomass per ESDU (relative units) by transect for the three surveys. Only parallel transects (odd figures) are put on the graph. No records for transects 15, 17 of the second survey (see text).

In addition, an horizontal sonar was set at 90° of the course of the ship in order to count the number of schools along the track.

For technical reasons the second survey had to be shortened and the extreme west side of the gulf was not surveyed. This represents a lack of information on 10 ESDUs out of 70. This missing part was assumed to contain low biomass according to the results of the two other surveys (fig. 2). It was considered to have no consequence for the following data analysis.

PROBLEM ANALYSIS:
VARIABILITY OF THE ABUNDANCE ESTIMATE

Biomass per ESDU

A between-survey comparison of the basic statistics of the integrated echoes per ESDU illustrates the difficulty of estimating precisely the abundance (see table 1).

As there are repeated surveys, the sampling error may directly be computed on the discrepancy between survey mean estimates. However, the precision of the method is poor because there are only 3 surveys and because the differences between them are high.

Experimental variograms were computed without removing the high value; they showed pure nugget effects and are not shown. No spatial structure was identified. An average sampling error was thus computed using the statistics of the average intra-survey variability. Average mean and variance for the surveys were computed as follows:

$$m = \frac{(n_1 m_1 + n_2 m_2 + n_3 m_3)}{(n_1 + n_2 + n_3)} \qquad (1)$$

$$s^2 = \frac{((n_1 - 1) s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1) s_3^2)}{(n_1 + n_2 + n_3 - 3)} \qquad (2)$$

The average variance is 81 858, the average mean is 115. and an average number of ESDUs is 73. The average sampling error of the mean estimate is:

$$(s/\sqrt{n})/m = 29\%$$

The variability of the mean estimate is not correlated with the time interval between the surveys. This is in agreement with the hypothesis of constant stock size during the period under study.

During the second survey the maximum ESDU value is twice the maximum for the other surveys: mean and variance are considerably increased. We may thus ask ourselves: How high is the highest possible value? Are we still far from it? Are we sure each set is a realistic representation of reality? Can we compute the real sampling error from the data?

We now attempt to understand what causes these high values as they generate such big differences in the survey estimates.

Schooling biomass per ESDU

On the echogram, the deviation of the integrated echoes (energy) is measured for each dense trace. On the histogram computed on all the deviations for the three surveys, a population of small values individualizes under the threshold of 6 backscattering units. We thus define a school as a dense trace on the echogram that generates an integrated echo greater than 6. In doing so, we postulate the existence of a lowest school size and that this size did not vary during the period of the three surveys.

For the three surveys together, the biomass in schools represented 90% of the total biomass. In each ESDU the biomass is measured by the integrated acoustic energy of the echoes; the biomass in schools is measured in the same way but by taking into account the schools only. The biomass per ESDU relates well to the schooling biomass per ESDU. A linear regression explains 99% of the variance of the biomass per ESDU and the slope of the regression is very near to one.

The biomass per ESDU also relates well to the maximum biomass per school in the ESDU. The slope of the linear regression is 1.29 and this regression explains 92% of the variance of the biomass per ESDU. Thus, the biomass in the ESDUs is determined

Table 1. – Basic statistics of the integrated echoes per Elemental Sampling Distance Unit (ESDU) for each survey: (n) number of ESDUs, (m) simple mean of the data, (s^2) estimated variance, (max) maximum integrated value, (s/m) coefficient of variation of the variable.

Survey	Date	n	m	s^2	max	s/m
Survey 1	24/08	68	76	23 311	1 045	2.01
Survey 2	01/09	59	187	187 943	2 258	2.32
Survey 3	02/09	78	81	52 894	1 400	2.84

Table 2. – Basic statistics of the biomass per school (relative units): (n) number of schools sailed over and recorded by the echo-sounder, (m) simple mean of the data, (s^2) estimated variance, (max) maximum value (biomass per school), (s/m) coefficient of variation of the variable.

Survey	Date	n	m	s^2	max	s/m
Survey 1	24/08	76	60	6 885	405	1.38
Survey 2	01/09	87	118	70 147	1 560	2.24
Survey 3	02/09	82	73	40 222	1 360	2.75

greatly by the school containing the biggest biomass. The abundance estimate is dominated by the hit or the miss of the few schools containing a high biomass.

The basic statistics computed on the biomass per school are summarized on *table* 2. In comparison, the average number of schools per ESDU (1.12, 1.47, 1.05) is much less variable than the mean biomass per school. The variability of the biomass per school follows well the one of the biomass per ESDU. We may group together surveys 2 and 3 because high values were sampled, the maxima being of the same order of magnitude. However, survey 3 contains a greater proportion of small and medium values than survey 2 and its coefficient of variation s/m is greater. Thus the mean estimate for survey 3 is closer to the mean estimate for survey 1. Survey 1 indicates alone a low estimate with a low imprecision, survey 3 a low estimate with a high imprecision and survey 2 a high estimate with a high imprecision.

We need to quantify the influence of the schools containing a very high biomass on the precision. The total data set of the three surveys is used as if it gave a precise enough image of an average reality. This analysis gives us a lower bound for the influence because it implies that the variance was finite and well enough known from the samples. The calculation of a high bound for the influence would certainly require simulations and extrapolations of the histogram tail under the hypothesis of infinite variance.

DATA ANALYSIS: DECOMPOSITION OF THE VARIABILITY

Here we limit the analysis to estimating the biomass standing in schools:

$$\hat{Q} = \hat{N}_s \overline{B}_s$$

where $\hat{N}_s$ is the total number of schools observed and where $\overline{B}_s$ is the mean biomass per school. If the number of schools is uncorrelated with the biomass in the schools, the relative variances are written:

$$\frac{\sigma_Q^2}{Q^2} = \frac{\sigma_{N_s}^2}{N_s^2} + \frac{\sigma_{B_s}^2}{\overline{B}_s^2}$$

Can the number of schools be estimated independently from the biomass in the schools?

Relation between number of schools and biomass in the schools

The biomass in each school is plotted against the number of schools per ESDU on figure 3. There is no clear relation. Rich schools occur in ESDUs where there are many schools as well as in ESDUs were there are only few schools, although the probability of occurrence seems to differ. Experimental frequencies were computed for examining this.
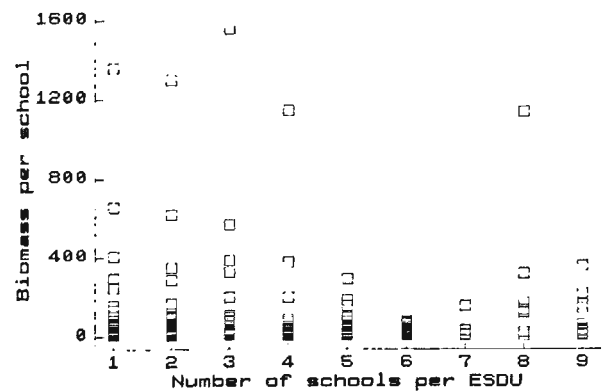

Figure 3. – Plot of the biomass per school against the number of schools in the ESDUs.

The frequency of a rich school is computed conditionally on the number of schools in the ESDU. What is the probability of having at least one rich school in the ESDU for a given number of schools? The ESDUs were ranked in ascending order of school numbers. In each class the number of ESDUs containing at least a rich school (biomass greater than 100) were counted. Results are shown in the following table as the number of occurrences of at least one rich school per ESDU relatively to the number of ESDUs per class.

There are only a few ESDUs containing 4 schools or more. Therefore the experimental values are variable. Nevertheless, the probability of encountering a rich

Schools/ESDU	1	2	3	4	5	6	7	8	9
Nb. of times	7/54	19/23	5/11	2/6	1/5	0/0	1/1	1/1	1/2

Survey	Date	n	m	s^2	max	s/m
Survey 1	24/08	68	1.12	3.40	9	1.65
Survey 2	01/09	59	1.47	3.12	8	1.20
Survey 3	02/09	78	1.05	2.80	7	1.59

school seems to increase when there are more and more schools. Now consider the two variables as independent. There will be rich schools occurring alone, but in areas where there will be many schools, the probability of presence of a rich school will be greater. As an independence property enables the loose relationship observed to be reproduced, we shall consider that the number of schools per ESDU and the biomass per school are uncorrelated.

Number of schools per ESDU

Statistics

The basic statistics computed for each survey are summarized on *table 3*.

The histogram is computed on all the data for the three surveys (*fig.* 4). It is not too asymmetric and could be reasonably modelled by a lognormal statistical distribution as indicated by the linear aspect of the probability plot (*fig.* 5).
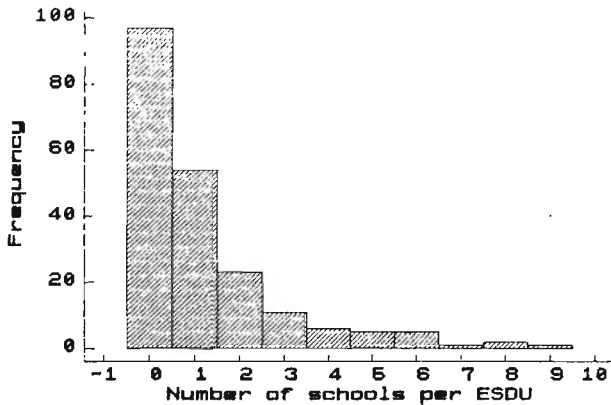

Figure 5. – Normal probability plot of the number of schools per ESDU

Figure 4. – Histogram of the number of schools per elementary sampling distance unit (ESDU).

Variography

An isotropic variogram was computed for each survey then the mean was calculated for each lag (*fig.* 6). The variogram value at the first lag gives an estimate of the variance between adjacent ESDUs (2.5): it is important and represents 81% of the data variance. Even if the variability at short distance explains the major part of the spatial variability, the variogram shows a detectable range after which it
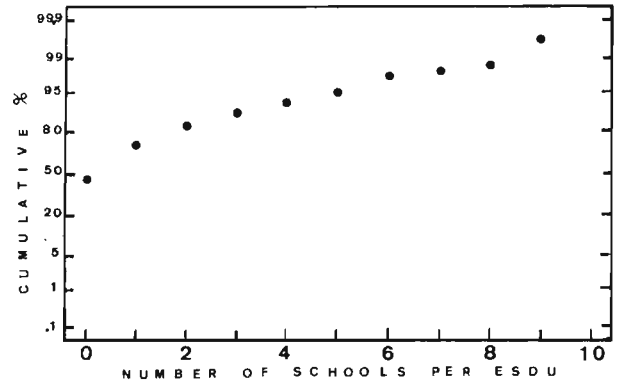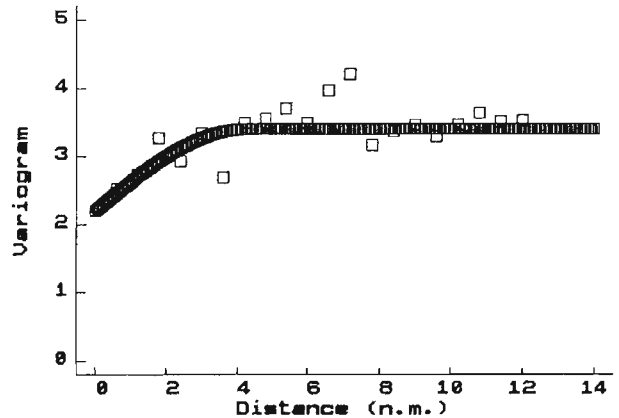

Figure 6. – Variogram of the number of schools per ESDU. The model is: nugget (2.2) + spherical (sill = 1.2, range = 4 nautical miles).

stabilizes around a sill a little greater than the variance. This is to be expected when the range is not small in comparison to the dimensions of the area surveyed. The experimental variogram was modelled by a nugget effect of 2.2 and a spherical structure of sill 1.2 and range about 4 nautical miles.

Biomass per school

Statistics

The average statistics calculated for a survey are: $n = 82$ (number of schools), $m = 85$, $s^2 = 40\ 525$, max = 1 560, s/m = 2.37.

The histogram of the biomass per school computed on all the data for the three surveys gives an average estimate of an underlying histogram (*fig.* 7). It is highly asymmetrical with a tail higher than for a lognormal distribution as can be easily seen by plotting the cumulative frequencies on a probit scale. The sampling error on the histogram tail is very big as survey 1 did not find it at all (no large schools were encountered). The maximum biomass per school is not well known, thus the mean is not and the variance is even less known. Should a model with finite or infinite variance be chosen?
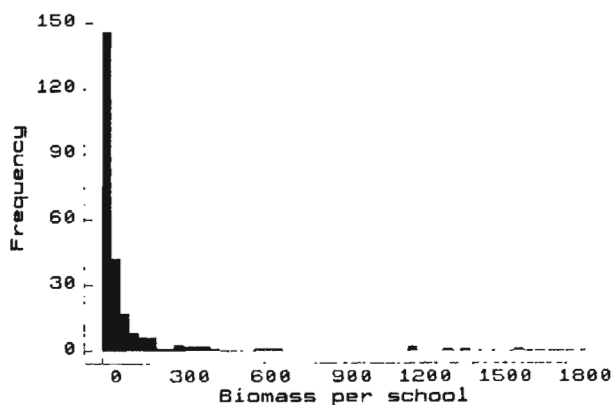

Figure 7. – Histogram of the biomass per school.

Variography

The location of each single school was not recorded. Thus a variogram cannot be computed for the school biomass. The intra ESDU variability can nevertheless be computed and compared to the overall spatial variability. Although the result will depend on the ESDU length, it will nevertheless give an order of magnitude of the value of the variogram for a distance equal to the ESDU. As such distance is small, it will give an idea of the nugget effect. In each ESDU we compute the mean and the sum of squares of the differences to the mean. The sum of squares of all ESDUs are added. On the total set we compute the sum of squares of the differences of each value to the global mean. Results for each survey are:

	Sum of squares of the differences to the mean SSD (ESDU)/SSD (total)
Survey 1	29.6%
Survey 2	53.3%
Survey 3	41.3%

A high percentage of the variability stands inside the ESDUs. Large schools stand near smaller ones.

Inside an ESDU, a large school may occur alone or with other schools. These may be small schools only, large and small, or large only. The probability for each of these three cases was estimated on the total data

set. A large school is defined as sustaining a biomass greater than 100.

Results are given as a fraction of the number of occurrences on the number of observations:

alone	with other schools		
	small	small and large	large
7/26	6/26	11/26	2/26

On average, the immediate vicinity of a large school is made of other large or small schools with a probability of 0.65. Thus even inside the ESDUs the presence of a large school does not influence the biomass in the neighbouring schools. The area of influence of a big school shall thus be considered negligible here.

The inter-ESDU spatial structure for different quantiles of the histogram are now studied. Each quantile of the histogram is coded by an indicator associated to a threshold. The variogram of an indicator gives information on the structure of the geometrical sets defined in space by the corresponding threshold. Thus the geometrical properties of a spatial distribution can be studied by computing indicator variograms (Rivoirard, 1990; Petitgas, 1993). As we work on the histogram for the three surveys, the same cutoffs are used for each survey. We work on the average process that concentrates the biomass in the schools. For each ESDU, the indicator will equal one if the ESDU contains at least one school showing a biomass greater or equal to the threshold. If all the schools in the ESDU have a biomass strictly lower than the threshold then the indicator equals zero for this ESDU. The thresholds looked at are: 6, 100, 200, 500, 1000.

The isotropic variograms of the previous indicators are given for each survey on figures 8, 9, 10. The indicator variograms associated with thresholds lower than 200 are a little structured (surveys 1 and 3), the range being about 2 nautical miles (2 ESDUs). However, for thresholds greater than 200, no structure is detectable any more (surveys 2 and 3). For survey 2, all variograms look like pure nugget effects. The spatial structures are small scale for the low values and are not very repeatable from one survey to the other. Thus we shall consider that the biomass per school is not structured in space.

MODEL OF SCHOOL: OCCURRENCE AND RESULTS

The biomass per ESDU can be interpreted as the product of two independent variables: the number of schools and the biomass in the schools. The number of schools per ESDU has a lognormal type of statistical distribution and shows a spatial structure of a spherical type. The biomass per school is extremely
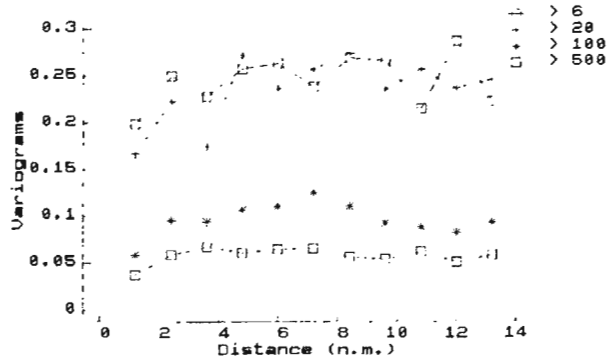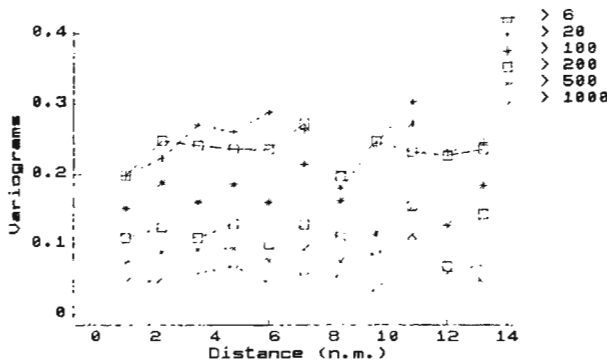
Figure 8. – Indicator variograms - survey 1.

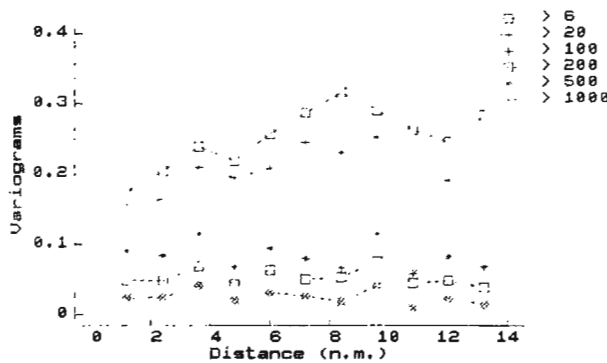
Figure 9. – Indicator variograms - survey 2.

Figure 10. – Indicator variograms - survey 3.

heterogeneous, it shows a statistical distribution with a very long tail and has no spatial structure. This model is used for computing the variance of the abundance estimate.

Relative precision of the mean number of schools per ESDU

As there is a spatial structure, the variance is calculated using the variogram and the geostatistical formulae.

The entire gulf is an area of fish presence. The data on the coastal inter-transects were included in the estimation of the mean and the mean is estimated by

the simple average of the data. An estimate of the associated precision can be derived using the variogram. In order to include the inter-transects in the computation of the estimation variance in a simple way, the ship's sampling track was arranged as a straight line. The estimation configuration becomes the following: the mean in the rectangle is estimated by the mean of the continuously sampled median line. The length of the rectangle equals the length of the ship's track. The width is the inter-transect distance. The formal equations for this configuration have been solved by Matheron (1971). A detailed explanation can be found in Journel and Huijbregts (1978). These authors give tables were the variance of estimation can be read for different vairograms and rectangle dimensions. Here we computed these tables for a spherical variogram using the formulas given in Guibal (1973).

Using the average variogram and the following rectangle dimensions, we derive an estimate of precision that one may expect on average using a full sampling grid in the gulf.

The number of samples along the sailing track is:

$n = 73$ ESDUs (nearly 73 Nm).

The inter-transect distance is:

$a = 4$ ESDUs (nearly 4 Nm).

The nugget constant is:

$Co = 2.2$

The variogram has two structural components, the estimation variance is thus the sum of two variances, each coming from a structural component:

the variance coming from the spherical structure is:

$$v_1 = 0.008$$

the variance coming from the nugget component is:

$$v_0 = Co/n = 0.03$$

So the relative error on the mean estimate is:

$$\frac{\sigma_{\overline{N}_s}}{\overline{N}_s} = 16.1\,\%.$$

The nugget contributes 79% to the estimation variance.

The average relative precision of a survey was also calculated as if there was no spatial structure. The average mean and variance are computed (see equations 1 and 2). This gives a relative precision of:

$$\left(s/\sqrt{n}\right)/m = 17.2\,\%.$$

Due to the high nugget, the two values are close.

Comparison with the precision of sonar school counts

Schools were also detected by sonar. A comparison of the precision between the two methods can be made.

Using the sonar, the schools were counted and summed by transect instead of ESDU. Thus it was

not possible to compute the same statistics as for the other variables. However, it remains interesting to look at the total number of schools .recorded during each survey (see *table* 4).

We can see that the relative precision of the total number of schools detected by sonar is comparable with the one of the number of schools per ESDU recorded by echosounder, even a little better.

Relative precision of the mean biomass per school

The biomass per school is not structured in space. Therefore the expected estimation variance of the mean estimate is calculated using the average mean and variance for a survey:

$$\frac{\sigma_{\overline{B}_s}}{\overline{B}_s} = \frac{s/\sqrt{n}}{m} = 26.2\,\%.$$

Relative precision for the abundance estimate

The abundance estimate is the product of two uncorrelated variables:

$$\hat{Q} = \hat{N}_s\,\overline{B}_s = A\overline{N}_s\,\overline{B}_s$$

where A is the gulf surface, $\overline{N}_s$ is the mean number of schools and $\overline{B}_s$ is the mean biomass per school.

As the entire gulf can be considered as fish habitat, there is no sampling error made on the surface A which is a scalar. Secondly, considering there are no border effects on the number of schools, the relative variance of the number of schools equals the relative variance of the mean number of schools:

$$\frac{\sigma_{N_s}^2}{N_s^2} = \frac{\sigma_{\overline{N}_s}^2}{\overline{N}_s^2}$$

Thirdly, as the number of schools per surface unit and the biomass per school are considered uncorrelated we now have:

$$\frac{\sigma_Q}{Q} = \sqrt{\frac{\sigma_{\overline{N}_s}^2}{\overline{N}_s^2} + \frac{\sigma_{\overline{B}_s}^2}{\overline{B}_s^2}} = 30.7\,\%$$

This is in good agreement with the relative precision calculated on the biomass per ESDU directly (29%).

The relative variance of the school count ($\sigma_{N_s}^2/\overline{N}_s^2$) represents 27.4% of the relative variance of the abundance estimate (σ_Q^2/Q^2). The precision on the school counts could be increased, for example by sampling a greater seawater volume using lateral sonar equipment. However, even if all schools were counted without error, the relative precision on the abundance would equal the one on the biomass per school and the estimate would stay imprecise (26%). The problem is to estimate precisely the variance of the biomass per school, *i. e.* the shape of the histogram of the biomass per school.

DISCUSSION AND CONCLUSION

On the present data set, the sampling was performed during daytime and the biomass was in schools. The number of schools per ESDU and the biomass per school were found to be two different processes with two different behaviours both statistically and spatially.

The number of schools per ESDU is spatially structured and its histogram is not very skew. The information in this variable may relate biologically to occupation of the habitat. The biomass per school had a different behaviour: it was a much more dispersed variable and had no structural information in it. The process which concentrate biomass in the schools was here non-spatial. This may relate biologically to processes other than habitat occupation. The model is thus the following. The schools are located in space in a aggregated and structured manner. In each school the biomass follows a probability distribution function that is very skew.

On a simulation schools would form aggregates in space. The biomass in each school would be a random outcome of a very skew statistical distribution. Thus in areas where there are more schools the probability of having a high biomass would be greater than elsewhere. However, there would also be rich schools in areas where there are only few schools and the location of these will be difficult to predict. In statistical terms, biomass may therefore

Table 4. – Statistics of the number of schools recorded by the lateral sonar: (*m*) average of the three surveys, (*s*) standard deviation of the values, (*n*) number of surveys, (CV) coefficient of variation of the mean. The "reduced number" corresponds to the common area, *i. e.* excluding transects 14-18 not covered during the second survey

Survey	Date	Total number	Reduced number
Survey 1	24/08	655	624
Survey 2	01/09	502	502
Survey 3	02/09	561	554
Mean	*m*	577	560
CV	$(s/\sqrt{n-1})/m$	13.4%	10.9%

be considered "hidden". Is this statistical property biologically interesting for a fish population?

In their decomposition of the biomass per ESDU, MacLennan and MacKenzie (1988) used the same variables as us. They also considered them uncorrelated. They also found that the biomass per school had a very skew statistical distribution, but they did not look at the spatial structure of their variables. In particular they did not consider that the number of schools per ESDU could be structured. It is our thinking that the spatial structure of the schools is an important piece of information for understanding the biology of the fish as well as for designing a survey.

The statistical characteristics of our tropical data are in good agreement with the ones observed by MacLennan and MacKenzie (1988) on North Sea data. We are therefore led to consider that these characteristics are of general interest for pelagic schooling fish.

The present study gives evidence that the major difficulty in estimating the abundance of pelagic fish stocks is in modelling the histogram of the biomass per school. The fluctuations of the histogram due to sampling are higher on the tail than on other portions. The imprecision on the abundance estimate is due to the experimental fluctuations on the tail. In contrast, the number of schools is much less erratic in space and much less dispersed on the histogram. The number of schools can be estimated with a good precision. This result leads us to direct further research towards investigation of the relation between abundance, number of schools and biomass per school variability. If the biomass per school can be considered as a non-spatial variable which has a specific histogram, then the number of schools may be a good index of abundance.

The problem is to characterize the skewness of the histogram of the biomass per school and to relate this to biological knowledge. What is it that the variability of the biomass per school characterizes? The skewness is related to the variability of the packing density in the schools. In situ experimental measurements (Gerlotto *et al.*, 1988; Misund, 1991) have shown that it is extremely variable. However, in statistical terms for a population, can the skewness of the histogram be a characteristic of the fish population? Is this a species specific characteristic or an ecological descriptor? If a specific fractioning of the biomass into different school sizes is as strategy of the population in response to a specific environmental situation, then skewness may be a good descriptor of what the population is doing. As it is difficult to study this experimentally because of the fluctuations on the tail of the histogram due to sampling, simulations may help choose an appropriate class of distribution models. Skew type histograms could be generated by making reasonable statistical assumptions on the schooling itself.

REFERENCES

Gerlotto F., P. Fréon, 1988. Influence of the structure and behaviour of fish school on acoustic assessment, ICES/CIEM, CM88/B:53, 28 p.

Gerlotto F., P. Petitgas, 1991. Some elements on time variablity in acoustic surveys through the example of a single transect repeated during 24 hours. ICES/CIEM, CM91/B:15, 9 p.

Guibal D., 1973. Les fonctions auxiliaires à deux dimensions pour le schéma sphérique. Note N-347, Centre de Géostatistique, École des Mines de Paris, ENSMP, Fontainebleau, France, 18 p.

Journel A., C. Huijbregts, 1978. Mining geostatistics. London, Academic Press, 368 p.

MacLennan D., I. MacKenzie, 1988. Precision of acoustic fish stock estimates. *Can. J. Fish. Aquat. Sci.*, **45**, 605/616.

Matheron G., 1971. The theory of regionalised variables and its applications. Les cahiers du Centre de Morphologie Mathématique, fasc. 5, Ecole des Mines de Paris, 212 p.

Misund O., 1991. Swimming behaviour of schools related to fish capture and abundance estimation, *Ph. D. Thesis*, Univ. Bergen, Norway, 132 p.

Petitgas P., 1993. Use of a disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. *Aquat. Living Resour.*, **6**, 201-209.

Rivoirard J., 1990. Introduction to disjunctive kriging. Course C-118, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, France, 84 p.