APPLICATION OF GEOSTATISTICS TO GROUNDFISH SURVEY DATA

H.G.PEREIRA, A.O.SOARES
Centro de Valorização de Recursos Minerais
Technical University of Lisbon
Instituto Superior Técnico, Av. Rovisco Pais
1096 LISBOA CODEX PORTUGAL

ABSTRACT. In research bottom trawl surveys conducted in the Portuguese continental waters, a variety of regionalized variables in space and time are sampled. For each sampling station, the variables support is defined by the volume swept by the gear in each haul.

The major objective of those surveys is to provide annual estimations of the abundance of commercial fish species, in order to assess the current status of stocks in large sectors (average area of 500 square nautic miles). Moreover, estimation of lenght histograms in the same areas are required for scientific advice to fishery management.

In order to exploit the correlation structure in the data, a geostatistical estimation approach was applied to the available regionalized variables. Preliminary results of this work, applied to a selected zone, are presented and discussed.

1. INTRODUCTION

The objective of this paper is to illustrate a preliminary application of geostatistics to the estimation of fish abundance.

The available data are collected in groundfish surveys conducted by the INIP(National Institute for Fishery Research)every year, since 1979.

The results reported here regard the southern part of the 1986 October survey, for the species hake (<u>Merluccius merluccius</u>). The data set includes 49 bottom trawl stations, the location of which are approximatively shown in Fig. 1. For each haul, the water volume swept by the gear is constant - it is the sample support S, given by:

$$S = \Delta H \; v \; t \; D$$

where ΔH is the horizontal net opening between the wings (14.5m)
v is the tow speed (3 knots)
t is the duration of each tow (1/2 hour)
D is the vertical net opening (3.5 m).

M. Armstrong (ed.), Geostatistics, Vol. 1, 459–467.

In each haul station, the following biological variables are measured in the support S: total weight of the species, total number of individuals caugth and length of each individual. The length variable is divided into classes with a certain ecological and commercial meaning (<17 cm, 17/25 cm, 25/45 cm, > 45 cm).

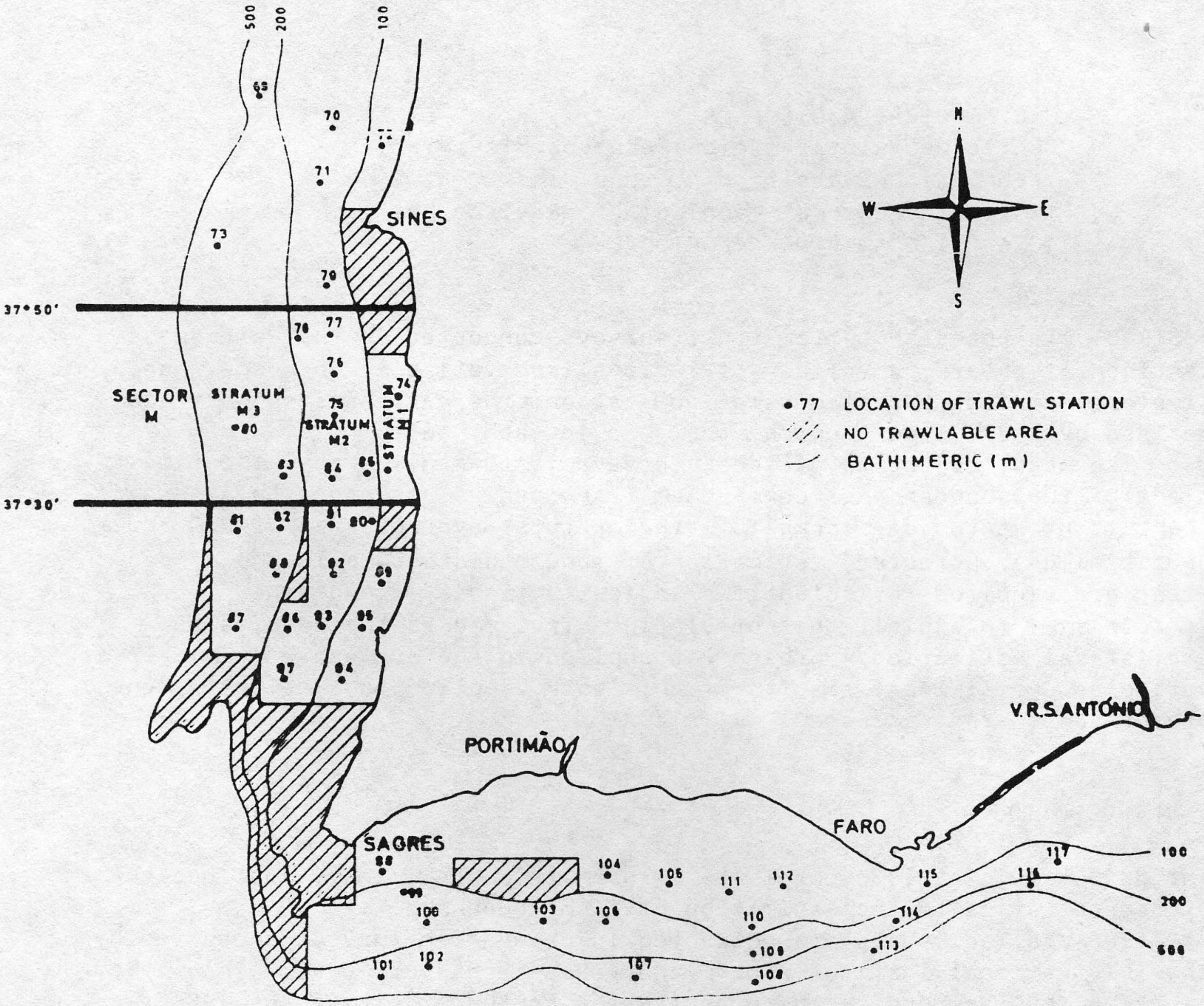

Figure 1. Location of the sampling stations (69-117) and selected zone (SECTOR M = M1 U M2 U M3) to be estimated.

Since the time delay between the first and the last haul of this part of the survey is small (7 days) compared with the period inter--surveys (several months), all stations of this data set can be considered to be sampled at the same time, and the necessary condition to undertake a spatial geostatistical study is met: usual variograms and cross--variograms for the biological variables can be calculated, disregarding the temporal sequence of stations. Also, the estimation results for a selected zone (depicted in Fig. 1 as SECTOR M, divided into strata M1, M2, M3) are valid for the entire period of time represented by this specific survey. Obviously, this approach can not account directly for the fish mobility during the sampling period. This feature is included into the nugget effect, and can be considered sampling error.

2. VARIOGRAMS AND CROSS-VARIOGRAMS

Assuming that the leading factor controlling the spatial variability of fish abundance is the orientation, referred to the shore line, of the vector $\vec{h}$ (Cf. eq. [1]), the $\vec{NS}$ variogram for stations 69-97 (Cf.Fig.1) was averaged with the $\vec{EW}$ variogram for stations 98-117. Hence, all variograms and cross variograms are computed "along the shore".

Relative "along the shore" variograms and cross-variograms are calculated using [1] , for the whole data set.

$$\gamma_{ij}(\vec{h})/\sigma_{ij} = \frac{1}{2} E\left[(Z_i(x+\vec{h}) - Z_i(x)).(Z_j(x+\vec{h}) - Z_j(x))\right]/\sigma_{ij} \quad [1]$$

where σ_{ij} is the variance-covariance matrix of variables

$\vec{h}$ is a vector along the shore (maximum 30')

$Z_i(x), Z_j(x)$ are variables i and j at location x.

The experimental curves obtained for the relevant variables (weigth, total number of individuals, absolute frequency for the classes < 17 cm , 17/25 cm, 25/45 cm) are shown in Fig. 2 (results for the frequency of individuals greater then 45 cm are not reliable, and it was assumed that they follow the general structure).

It is clear from Fig. 2 that all pairs of variables depict the same structural pattern. A global model (Cf. [2]) was fitted to the experimental curves

$$\gamma_{ij}(\vec{h})/\sigma_{ij} = .547 + .006\ \vec{h}^{1.6} \quad [2]$$

Since accurate variograms can not be calculated perpendicularly to the shore line, due to lack of data, the anisotropy problem was handled suppressing each observation point at a time and providing an estimate $\overset{*}{Z}$ in that point by the remaining N-1 measurements, based on different anisotropy ratios. For the zone to be estimated (Cf. SECTOR M, Fig. 1), the best cross-validation results (MRE≃0 and MRSE≃1, Cf. TABLE 1) were found for the anisotropy ratio of 1:2 (the EW component of vector $\vec{h}$ is twice the NS one).

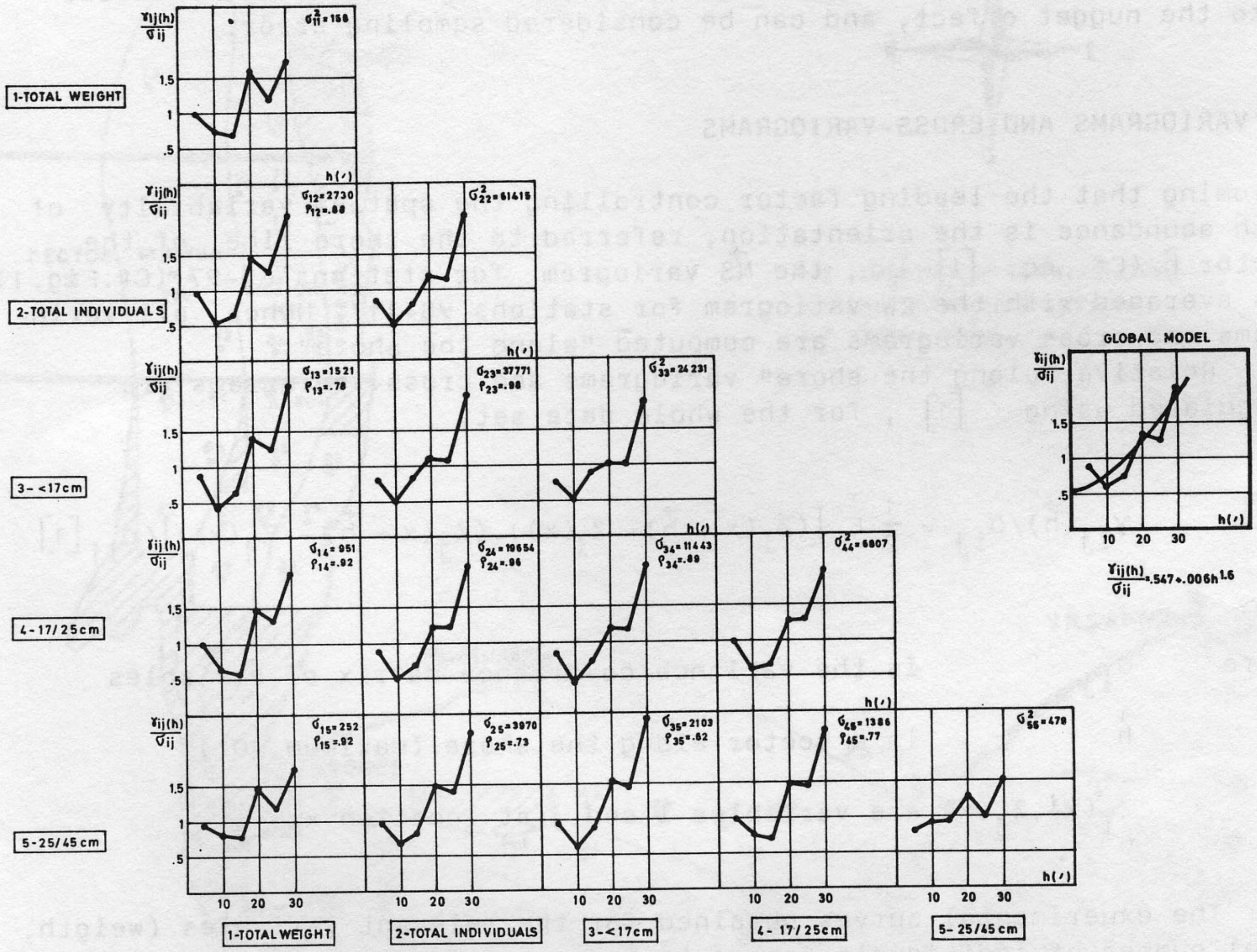

Figure 2. Relative variograms and cross-variograms along the shore.

TABLE I - Cross-validation criteria for SECTOR M

VARIABLES	WEIGHT	TOTAL INDIVIDUALS	ABSOLUTE FREQUENCY < 17 cm	ABSOLUTE FREQUENCY 17/25 cm	ABSOLUTE FREQUENCY 25/45 cm
$MRE = \frac{1}{N} \sum_{i=1}^{N} (Z^*_i - Z_i)/\sigma_{k_i}$	0.002	0.005	0.003	0.003	0.010
$MRSE = \frac{1}{N} \sum_{i=1}^{N} (Z^*_i - Z_i)^2/\sigma^2_{k_i}$	1.054	1.080	1.110	1.305	1.090

MODEL USED: $\gamma(h) = \sigma^2\,(.547 + .006\,h^{1.6})$

Anisotropy ratio = 1:2

N - no. of samples

$\sigma^2_{k_i}$ - kriging variance for sample i

MRE - Mean Relative Error

Z^*_i - estimated value for sample i

Z_i - real value for sample i

MRSE - Mean Relative Squared Error

3. GLOBAL ESTIMATION IN A SELECTED ZONE

In order to compare the geostatistical estimation method, with the one currently used by the INIP, the SECTOR M and strata M1, M2, M3 were selected (Cf. Fig. 1).

The current estimation method, denoted "stratified random" in TABLE II, is briefly outlined in the sequel (Cardador, 1983):

The average and variance of each stratum are calculated by the arithmetic mean ($\bar{Z}_\ell$) and variance (σ^2_ℓ) of the n_ℓ samples contained in the stratum.

The average in the Sector is given by:

$$\bar{Z} = \frac{1}{A} \sum_{\ell=1}^{3} \bar{Z}_\ell \, A_\ell$$

where A_ℓ is the area of stratum ℓ and A is the area of the Sector.

The variance in the Sector is given by:

$$VAR = \frac{1}{A^2} \sum_{\ell=1}^{3} \frac{\sigma^2_\ell \, A^2_\ell}{n_\ell}$$

Regarding the geostatistical estimation method, it is obvious from equation [2] that conditions of "intrinsic correlation" (Matheron, 1979) are met for this data set. This means that all variograms and cross-variograms are proportional to a basic scheme $\gamma(h) = .547 + .006\ h^{1.6}$, being, in this case, co-kriging equivalent to ordinary kriging. Moreover, since kriging weights are the same for all variables, the estimate of the total number of individuals is the sum of estimates for all classes of absolute frequency.

Results of estimation using both methods are given in TABLE II. The column "ERROR" denotes the coefficient of variation (σ^*/Z^*) in %.

	ESTIMATION METHOD	TOTAL WEIGHT Average (Kg)	TOTAL WEIGHT Error (%)	TOTAL INDIVIDUALS Average	TOTAL INDIVIDUALS Error (%)	< 17 cm Average	< 17 cm Error (%)	17/25 cm Average	17/25 cm Error (%)	25/45 cm Average	25/45 cm Error (%)	> 45 cm Average	> 45 cm Error (%)
						ABSOLUTE FREQUENCY OF INDIVIDUALS PER CLASS							
SECTOR M	KRIGING	15.9	21.9	173.5	23.0	77.7	24.9	61.7	26.4	32.5	16.4	1.6	48.7
	STRATIFIED RANDOM	19.5	36.4	183.1	45.4	68.4	31.2	76.2	64.0	36.8	23.7	1.7	102.4
STRATUM M1	KRIGING	12.3	35.3	144.4	37.2	68.6	31.0	47.5	34.4	27.3	24.4	1.0	95.0
	STRATIFIED RANDOM	3.8	94.7	31.5	70.7	1.5	100.0	12.5	92.0	17.5	94.3	0.0	-
STRATUM M2	KRIGING	14.7	24.1	168.1	19.2	80.2	18.6	55.8	16.9	30.7	17.8	1.4	55.8
	STRATIFIED RANDOM	15.1	22.5	210.4	20.2	120.0	23.5	59.8	28.4	29.3	16.9	1.3	74.5
STRATUM M3	KRIGING	18.2	23.5	186.8	28.8	78.3	29.3	70.8	25.2	35.8	18.3	2.0	48.3
	STRATIFIED RANDOM	27.4	53.4	202.5	79.8	40.0	95.0	106.5	96.2	52.5	33.3	3.5	100.0

TABLE II - Estimation Results for each estimation method (Sector M and Stratum M1, M2, M3)

4. LOCAL ESTIMATION OF LENGTH CLASSES

For the entire Sector M (Fig. 1), a local estimation in a small grid of points was performed for the absolute frequency of classes <17 cm, 17/25 cm, 25/45 cm and for the total of individuals.

In Fig. 3, the Kriged results are shown trough the quartiles of the estimated values

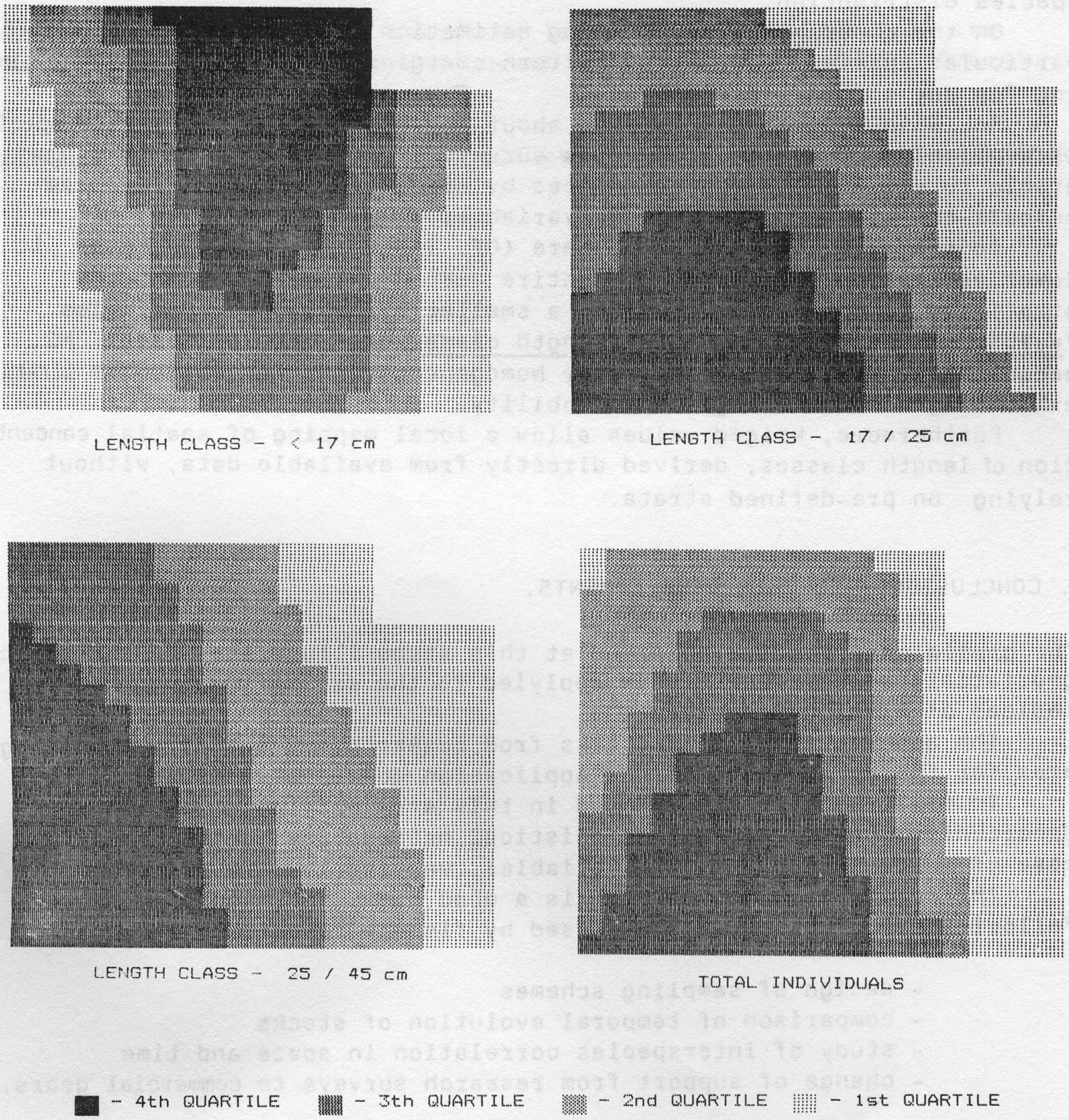

Figure 3. Kriged maps of length distribution within the selected Sector

5. DISCUSSION OF RESULTS

The basic assumption underlying the stratified random estimation method is the homogeneity of variables within strata. Hence, strata are designed so that variability is greater between than within strata, in order to account for the "patchiness" of the overall distribution.

This assumption requires an a priori knowledge of the environmental factors controlling the fish behaviour in space. Moreover, multispecies surveys are difficult to handle, since strata boundaries depend on target species distribution.

On the other hand, the kriging estimation method accounts for the particular spatial correlation pattern emerging from data, without claiming for any a priori assumption.

So, when a limited knowledge about the behaviour of a certain species is available, or when multispecies surveys are conducted, it is preferable to estimate abundance indices by kriging, since the variogram reveals the spatial structure of variables.

In the case study reported here (Cf. TABLE II), kriging leads to lower estimation errors for the entire SECTOR M, even though the stratified random method provides a smaller error for the variables "weight" and "frequency in the length class 25/45 cm" in stratum M2, because these variables are rather homogeneous in this particular stratum, exhibiting a narrow range of variability.

Furthermore, kriged values allow a local mapping of spatial concentration of length classes, derived directly from available data, without relying on pre-defined strata.

6. CONCLUSIONS. FURTHER DEVELOPMENTS.

The preliminary results obtained at this stage illustrate how the basic geostatistical approach can be applyied to the estimation of abundance indices of fishing resources.

The assessment of fish stocks from research surveys is a promising area for the enlargement of the application domain of geostatistics.

In fact, some problems found in this area provide a challenging opportunity for adjusting geostatistical methodology to the specific case of multiple regionalized variables, varying in time and space.

The geostatistical approach is a good basis for handling the following practical questions raised by fishery research:

- design of sampling schemes
- comparison of temporal evolution of stocks
- study of interspecies correlation in space and time
- change of support from research surveys to commercial gears.

7. ACKNOWLEDGEMENTS

We are indebted to the President of INIP for permission to publish these results, obtained in the course of a joint research work on application of computer to fishery data. The assistance of Fátima Cardador was indispensable in this work.

8. REFERENCES

Cardador, F. - "Indices of abundance from groundfish surveys in the Portuguese continental coast during 1979/82", ICES DOC/G:45, 1983.

Matheron, G. - "Recherche de simplification dans un problème de Cokrigeage" N-628 CGF, Nov. 1979, 19 p.