International Council for the
Exploration of the Sea
ICES / CIEM

ICES CM 2000/K:25

On the clustering of fish schools at two scales and their relation with meso-scale physical structures

by

Pierre Petitgas

Line transect acoustic surveys provide digital echograms of fish schools. This enables to analyse structure in the occurrence of individual fish schools along the transect lines as if it were a spatiall point process. The paper characterizes two scales in the clustering of schools, clusters of schools (several kms) and clusters of clusters of schools (tens of kms), and analyses which structure can be related to physical characteristics of the water masses. Clustering of schools is analysed using the pair correlation function (an analog of the variogram but for point process) along the transects. Clusters of schools were modeled along transect lines by a correlation structure and a Matern point process. Clusters of clusters involved the regional 2D pattern on the map and were modeled with a trend surface. Physical structures considered are river plume, vertical water stratification and upwelling events. They are characterized using parameter outputs from a 3D hydrodynamic model. The trend was in best spatial coherence with the river plume. At the smaller scale, there was no environmental parameter to be related to the schools. It is hypothesized that the clustering of schools of a few kms is related to the behavioural schooling dynamics but that the meso scale aggregation of fish at tens of kms is related to the meso-scale physical structures and the food chain they support.

Keywords: Fish Schools, Aggregation, Clusters, Point Process

Pierre Petitgas: IFREMER, laboratoire Ecologie Halieutique, BP 21105, F- 44311 cedex 03, Nantes, France [tel: +33 240 374163, fax: +33 240 374075, e-mail: Pierre.Petitgas@ifremer.fr].

1. Introduction

Aggregation is a general biological characteristic (Taylor 1961). Aggregation occurs at different spatial scales and generates, in fisheries survey data, a few high densities, correlation between values at a scale of a few km, gradients or trends at a scale of tens of km. It is important to analyse the allocation of variance in the different spatial scales (ie characterise the spatial organisation) because such characterisation has 3 consequences: (i) the planning of survey strategy and effort, (ii) the understanding of the relation between the spatial structure and the underlying processes (relation between fish and fishermen, fish and environment). The objective of this paper is to characterize different scales in the clustering of schools, clusters of schools and clusters of clusters of schools, and analyse which structure can be related to physical characteristics of the water masses.

Pelagic fish aggregate in schools which themselves occur in clusters of schools (Cram and Hampton 1976, MacLennan and MacKenzie 1988, Swartzman 1994, Petitgas and Levenez 1996, Petitgas and Samb 1998, Mackinson et al. 1999, Petitgas et al. 2000). Numerical echosounders and image analysis softwares make it now possible to extract from digitally recorded echograms during acoustic surveys school data files where position and school parameters are recorded. Two approaches are possible for analysing the spatial structure of schools: one may consider the schools as discrete events of a spatial point process or one may consider the number of schools per unit sailed distance (Nm) as a continuous variable forming a density surface. In the latter case, geostatistics, GLM or GAM have been used. In the former case, Point process analysis can be used (ICES 1999). Both approaches should give coherent results. The interest in considering school occurrence as a point process is that one can work on all distances at all scales. In the density surface approach, residual variance within the Nm (ie support sample) will stay unaccounted which may limit the modeling.

The spatial distribution of fish has been related to environmental factors of various nature at different spatial scales. Trends in fish distribution as a function of covariates have been modeled using linear regression (Sullivan 1991, Guiblin and Rivoirard 1996), Generalized Linear Modeling (Stefánson 1996) or Generalized Additive Modeling (Swartzman 1994, Maravelias et al. 2000). Coherence and proximity in the spatial distribution of zooplankton and fish have been looked for (Swartzman 1999, Maravalias et al. 2000). Schneider (1989) evidences particular spatial scales at which occur predator-prey relationships. In this paper we focus on the relation between hydrology and schools after having characterized the spatial scales of the fish spatial organisation, following the spectral approach of Schneider (1989). Hydrodynamic meso-scale structures (front, upwelling, gyre, river plume) extent typically at a scale of tens of km and their dynamics are forcing events on the plankton production and food webs. Pelagic fish should respond at a similar scale in their spatial organisation. This is what we shall look for. Rather than use hydrological parameters (temperature or salinity) to characterise hydro-climat over the surveyed areas, we have estimated parameters of physical structures (upwelling, stratification, river plume) for the surveyed area using a 3D hydrodynamic model.

2. Material and Methods

Schools

We have worked on the 1997 french pelagic acoustic survey of IFREMER performed by the R/V Thalassa in the bay of Biscay and designed for direct monitoring of pelagic fish abundance (anchovy, sardine, horse macquerel and macquerel). The survey targeted at the anchovy population and covered the area where the specie concentrates for spawning in spring. The survey time period was 8-29 may 1997. The survey coverage was made of parallele E-W acoustic transects sailed at 10 knots, regularly spaced every 5Nm, from the spanish border (43°30N) to the isle of Ré (46°30N) (Fig.1). Oportunistic pelagic trawl hauls were undertaken to identify echotraces to species and gather biological parameters of the fish. The echosounder used (Ossian) had a frequency of 38kHz and a nominal beam angle of 5.7°x5°. The echograms recorded were analysed in the laboratory using the image analysis software Movies developped at IFREMER (Weill et al. 1993). The processing threshold was set to -55dB. Schools were not identified to species for the present study. They were collapsed on the vertical and latitude and longitude of the school centres were estimated. School echotraces that were smaller in length than 2 beam widths at depth were not considered as such echotraces have been proved to be too small to allow estimation of their dimensions with the equipment settings (ICES 1999). These small schools represented 10% of the total nb of school echotraces and 7% of the total acoustic energy of the school echotraces. Basic statistics on the schools are in Tables 1-3.

Hydrology

In the area studied in spring, there are two major physical structures, the Gironde river plume and a coastal upwelling, while a thermal stratification is developping on the plateau in spring (Allain et al. 1999). The dynamics of the structures can be tracked using the hydrodynamic model developped at IFREMER (Lazure and Jegou 1998). The model extends from coast to the 200m isobath, it has a 5x5km mesh size, 10 vertical layers (sigma coordinates) and 900s time lag. The model simulates the evolution of temperature, salinity, currents, at depth, forced by tides, winds, surface-atmosphere thermal exchanges, river discharges. The model has been validated by field measurements of temperature, salinity and currents and satellite images of temperature (Lazure and Jegou 1998). The model was run to produce maps of salinity, temperature, currents at depth, every 3 days for the period 1-29 may 1997. The river plume extent on the plateau is traced by surface salinity lower than 34psu. Average values are given every 3 days. The upwelling off the coast of Les Landes is traced by a positive vertical current. At each time step of the model run and each grid node, the vertical current is estimated as the current resulting from the continuity equation. To estimate net transport, the values were summed every 3 days. The stratification index of the water column is estimated as the energy necessary for mixing homogeneously the column (Allain et al. 1999). Average values are given every 3 days. We estimated for the period 1-29 may, one map for each parameter to be coupled to the spatial distribution of the schools. We estimated the union of spatial extents of the river plume over the period: the area where at least once, salinity was < 34psu. We estimated the union of spatial extents of the upwelling: the area where at least once, vertical current was $>0.1 ms^{-1}$. We estimated the average of the stratification index for the period at each grid node.

Correlation structure analysis

In acoustic surveys, because samples are collected continuously along the ships sailing track, the data are collected with a major anisotropy which influences the data analysis. School occurrence along the sailed lines can be studied as a one dimensional point process (MacLennan and MacKenzie 1988, Petitgas and Samb 1998, Petitgas et al. 2000). A point process analysis in 2D is here impossible as this approach requires mapped point patterns where all process events have been recorded (as on a sonar image). The analysis in 2D is best performed by the density surface approach, thus working on a school count per Nm.

Correlation structure in the school occurrence process along the acoustic survey transects was analysed using variogram and pair-correlation function. The variogram was computed on the number of schools per Nm (ie on density values). We used the estimate proposed by Matheron (1971):

$\hat{\gamma}_\theta(r) = \frac{1}{2\, n_\theta(r)} \sum_i (z_{x_i} - z_{x_{i+r}})^2$, where z_{xi} is the school count at location x_i, r is the distance between locations x_i and x_{i+r}, $n_\theta(r)$ is the number of pairs of point samples r apart in the direction θ. The pair correlation function is the analogue for Point Process of the variogram for random functions. It can be defined as $P(r) = \lambda^2 g(r) dxdy$, where *P(r)* is the probability that a segment of length *r* has one process event at each of its extremities, λ is the intensity of the point process and *g(r)* is the pair-correlation function. When there is no correlation, the variogram equals the variance for all distance, the pair-correlation function equals 1 for all distances. The pair correlation function curve reads like a classical correlation function. We used the estimate for *g(r)* proposed by Stoyan and Stoyan (1994). Along a 1D- transect line, it writes:

$\hat{g}(r) = \frac{1}{\lambda^2} \sum_i \sum_{j \neq i} \frac{e_h(r - d_{ij})}{l_{\mathrm{TR}} - d_{ij}}$, where r is the distance at which the function is computed, *e* is the epanecnikov kernel of width *h*, d_{ij} the distance between school *i* and school *j* on transect *TR*, l_{TR} is the length of transect TR and λ the intensity. An average estimate for all transects has been computed considering one single intensity (total nb of schools/sum of transect lengths) and summing all the ratio values for all the d_{ij} from the different transects. The variogram was estimated along and across transects, the pair correlation function was estimated along transects.

Grouping schools in Clusters of schools

The cumulative distribution of the distance to the next school along the acoustic lines is always skew and is well modeled by a Weibull distribution. This in itself evidences clustering (Petitgas et al. 2000). Any clustering algorithm is based on the distance between elements. The question is how to define the "distance threshold", ie the distance at which two schools are considered too far from each other to belong to the same cluster. This distance lies somewhere in the inflexion part of the cumulative distribution of the distance to the next school. A procedure to determine the "distance threshold" was proposed by Petitgas and Samb (1998) which was improved and made multicriteria within the EC FAIR program Cluster (Petitgas et al. 2000). The procedure is essentially the following: for a given "distance threshold", the schools are aggregated in clusters along the transects, the number of schools per cluster is linearly regressed on the cluster lengths (clusters need have at least 2 schools) and the R-square of this regression is reported, the number of clusters and of single schools (clusters made of one school) are reported, the non-homogeneity of school occurrence within the clusters is tested and the number of homogeneous clusters is reported. This is done for different "distance threshold" and curves are produced of the R-square, the number of clusters, singles and percent of homogeneous clusters, each as a function of distance treshold. The distance threshold chosen to aggregate schools is the one which maximises by eye all criterias: high R-square, not too many clusters nor singles, high percent of homogeneous clusters. This empirical multicriteria procedure enables to aggregate schools in homogeneous Poisson clusters. Once the clusters are defined, cluster length along the transect is estimated as the distance between the two extreme schools belonging to the cluster. The distance between clusters is estimated as the distance between the last school of a cluster and the first school of the next cluster along the transect.

Trend structure

A 2D trend surface was fitted on the nb of schools per Nm using a Generalised Additive Modelling (Hastie and Tibshirani 1995) as a function of an interaction between latitude and longitude. The function gam() in Splus (Mathsoft Inc.) was used: gam(y~lo(lat,long,0.25), family=Poisson, link=log) where the smoother is loess with 0.25 degrees of freedom. Different trend surfaces have been tried with and without interactions, with and without environmental parameters. The previous model gave a trend surface which fitted best the data.

Simulations - scales in the aggregation pattern.

Using the pair correlation on the point events and the variogram on the nb of schools per Nm, 2 scales in the aggregation of schools were evidenced (see section 3 Results). A Point process for each scale was simulated and the pair correlation functions between model and data were compared. We considered two Point Process models, a Matern process for the small scale structure and an inhomogeneous Poisson process for the long scale structure. We simulated occurrence of schools along the transects using each model. The models were simulated along the surveyed transects and the pair- correlation function was estimated for each simulation. 10 simulations were considered and the average of the pair correlation functions was computed.

Stoyan and Stoyan (1994) propose a statistical procedure to infer parameter values of a Matern process. Different parameter values are tried untill the best fit is obtained between pair correlation functions for the model and the data. A Matern process is a particular cluster process where parent points are Poisson distributed with intensity ρ and where daughter points are uniformly and independently distributed within distance R from each parent point. The number of daughter points per cluster (around a parent point) follows a Poisson distribution with parameter μ. Values for the parameters ρ, R and μ were taken to be the values estimated by the multicriteria clustering algorithm : nb of clusters over total surveyed length, average cluster length over 2, average nb of schools per cluster. Then, these values were changed to obtain best fit between model and data pair correlation functions.

An inhomogeneous Poisson process is a Poisson process where the intensity varies in space. We considered the spatial variation to be the response surface of the nb of school per Nm fitted by means of a GAM using covariates Latitude and longitude. Within the Nm segments along the transects, the schools were distributed uniformely.

3. Results

Fig.1 shows the spatial occurrence of the schools along the transects. Fig.2 shows the pair correlation function. Two correlation ranges are clearly noticeable, one near 7km and a larger one near 35km. On Fig.1 in front of the mouth of the Gironde, long school clusters are visible which generate the long range structure. Figs3 show along and across transect variograms for the nb of schools per Nm. The along transect variogram shows the same correlation ranges as the pair correlation function, one small range structure near 7km and a longer range structure near 35km. The across transect variogram is not informative and does not show any anisotropy.

Fig. 4 shows the curves used in the multicriteria clustering algorithm. The threshold distance of 3.3km has been chosen for grouping schools in clusters of schools. Table 1 gives the estimated parameters for schools clusters formed by the empirical multicriteria clustering algorithm. Average cluster length (8.5km) and average inter-cluster distance (6.6km) are of the same order of magnitude as the small range structure (7km) on the variogram and pair correlation functions. The clustering algorithm gave access to the small scale structure.

Table 2 gives the values of the parameters used for the simulations of a Matern process along the transect lines. Pair correlation, empirical multicriteria clustering algorithm and Matern process agree well in the value of their parameters. Fig.5 (top left) shows one realization of a Matern cluster process along the surveyed transects. Fig.5 (bottom left) shows the pair correlation function estimated for this process when simulated on the transect lines. Clearly, the small scale structure (7-8km) was well reproduced but not the longer one (35km). Fig.5 (top right) shows one realisation of an inhomogeneous Poisson process that is simulated using a fitted GAM surface of the nb of schools per Nm. The R-square of the fit (ie, 1-deviance/null.deviance) was 40%. Fig.5 (botthom right) shows the pair correlation function estimated for this process when simulated on the transect lines. Clearly the long range structure (35km) is well reproduced but not the small one (7km). The small range structure (7km) is formed by clusters of schools. The long range structure (35km) is formed by the regional distribution of the school count that can be modeled with a trend surface.

On Fig.6, the spatial distribution of the individual schools surveyed has been superimposed on the maps of the hydrological parameters. The trend (or long range structure) in front of Gironde presents best spatial coherence with the river plume extent.

4. Conclusion

The work performed was essentially an analysis of scale in the spatial pattern of schools. Between the schools and the population, clusters of schools had been evidenced. The survey analysed here showed two scales in the clustering: clusters of schools and clusters of clusters. Clusters of schools were modeled along transect lines by a correlation structure and a Matern point process. Clusters of clusters involved the regional 2D pattern on the map and were modeled with a trend surface. The trend could be related to a major environmental structure, a river plume. It is thought that the spatial organisation described here is a general rule. At a regional scale, ie meso-scale, the school numbers vary as a function of some environmental parameters. At a smaller scale, ie a few kms, the school spatial organisation may depend more on school behaviour and inter-relations between schools. At the smaller scale, there was no environmental parameter to be related to the schools.

We propose the following conceptual 2 stage model for school occurrence in fisheries survey data: the first stage is a process of parent events which is an inhomogeneous Poisson process with a functional link with environmental parameters. It models the occurrence of clusters of schools (ie the overall spatial distribution of the biomass). The second stage process is a cluster process around the parent events (ie the spatial small scale organisation of the biomass). The first stage takes into account the trend and the second stage the clustering small scale behaviour of schools.

Ackowledgments
This study was founded by the European commission, contract FAIR-CT.96.1799, program Cluster.

References

Allain, G., Petitgas, P., Lazure, P., 1999, Environmental and stock effects on the recruitment of anchovy in the bay of Biscay, ICES CM 1999/Y:22
Cram, D., Hampton, J., 1976, A proposed aerial/acoustic strategy for pelagic fish stock assessment, *J. Cons. Int. Explor. Mer*, 37(1): 91-97
Guiblin, P., Rivoirard, R., 1996, Spatial distribution of length and age for Orkney-Shetland herring, ICES CM 1996/D:14
Hastie, T., Tibshirani, R., 1995, *Generalized additive models*, Chapman and Hall, 335p.
ICES, 1999, *Report on Echotrace Classification*, D. Reid. Ed., ICES Cooperative Research Report n° 238.
Lazure, P., Jegou, A.M., 1998, 3D modelling of seasonal evolution of Loire and Gironde plumes on Biscay Bay continental shelf, *Oceanologica Acta*, 21: 165-177.
Mackinson, S., Nøttestad, L., Guénette, S., Pitcher, T., Misund, O., Fernö, A., 1999, Cross-scale observations on distribution and behaviour dynamics of ocean feeding Norwegian spring-spawning herring, *ICES Journal of Marine Science*, 56: 613-626
Maravelias, C., Reid, D., Swartzman, G., 2000, Seabed substrate, water depth and zooplankton as determinants of the prespawning spatial aggregation of North Atlantic herring, *Marine Ecology Progress Series*, 195: 249-259
Matheron, G., 1971, *The theory of regionalized variables and their applications*, Les Cahiers du Centre de Morphologie Mathématiques, Fasc.5, Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique, Fontainebleau.
Petitgas, P., Lévénez, J.J., 1996, Spatial organisation of pelagic fish: echogram structure, spatio-temporal condition and biomass in senegalese waters, *ICES Journal of Marine Science*, 53: 147-154
Petitgas, P., Samb, B., 1998, On the clustered occurrence of fish schools along acoustic survey lines and its relation to population abundance, ICES CM 1998/J:5
Petitgas, P., Carrera, P., Georgakarakos, S., Reid, D., 1999, Third annual progress report of the program Cluster, contract FAIR.CT.1799, European Commission, DGXIV-C2.
Schneider, D., 1989, Idnetifying the spatial scale of density-dependent interaction of predators with schooling fish in the southern Labrador current, *Journal of Fish Biology*, 35 (Suppl.A): 109-115
Stefánsson, G., 1996, Analysis of ground fish survey abundance data: combining the GLM and delta approaches, *ICES Journal of Marine Science*, 53: 577-826
Stoyan, D., Stoyan, H., 1994, *Fractals, random shapes and point fields*, John Wiley and Sons, 389p.
Swartzman, G., Stuetzle, W., Kulman, K., Powojowski, M., 1994, Relating the spatial distribution of pollack schools in the Bering Sea to environmental factors, *ICES Journal of Marine Science*, 51: 481-492
Swartzman, G., Brodeur, R., Napp, J., Walsh, D., Hewitt, R., Demer, D., Hunt, G., Logerwell, E., 1999, Relating spatial distributions of acoustically determined patches of fish and plankton: data viewing, image analysis, and spatial proximity, *Canadian Journal of Fisheries and Aquatic Sciences*, 56 (Suppl. 1): 188-198.
Taylor, L., 1961, Aggregation, variance and the mean, *Nature*, 189: 732-735
Weill, A., Scalabrin, C., Diner, N., 1993, Movies B: an acoustic detection description software, application to schoal species classification, *Aquatic Living Resources*, 6: 255-267

Table 1: Parameters of school clusters as estimated using the multicriteria clustering algorithm

Total surveyed distance (km)………………………..1255.7
Nb of schools………………………………………..1487
Nb of clusters……………………………………….79
Nb of solitary schools…………………………………13
Average nb schools per km in the clusters…..…………2.66
Average along transect cluster length (km)…………….8.46
Average along transect between cluster distance (km)…6.56

Table 2: Parameters for the simulation of a Matern point process along the survey transects which reproduce the small scale structure of the data pair correlation function and which are close to those of Table 1

ρ : Nb of parent events per km…………………………………………………0.05
R : Maximum distance (km) between daughter and parent events in a cluster…...4.23
μ : Average nb of daughter events per cluster………………………………...22.49

Table 3: Basic statistics of the nb of schools per Nm

Nb of values……………………..678
Minimum…………………………..0
Maximum…………………………27
Average……………………………2.21
Variance………………………….12.50

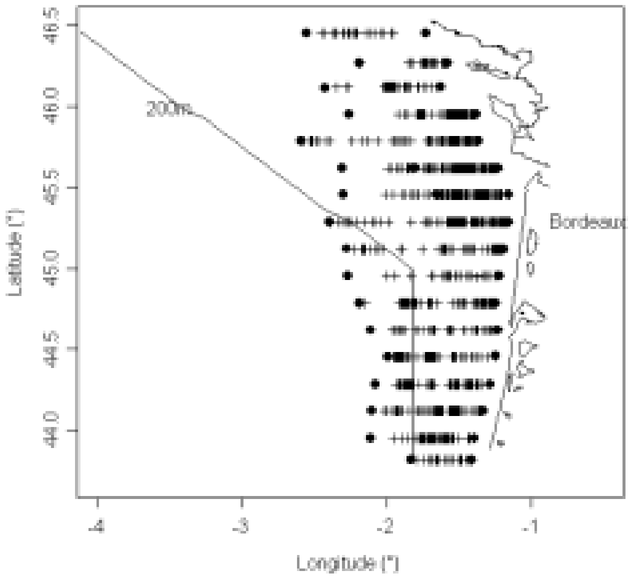

Fig. 1: Occurrence of individual schools along the transects of the acoustic survey Pegase97 of IFREMER in Biscay in may 1997. Crosses and school locations, black points are transect extremities.

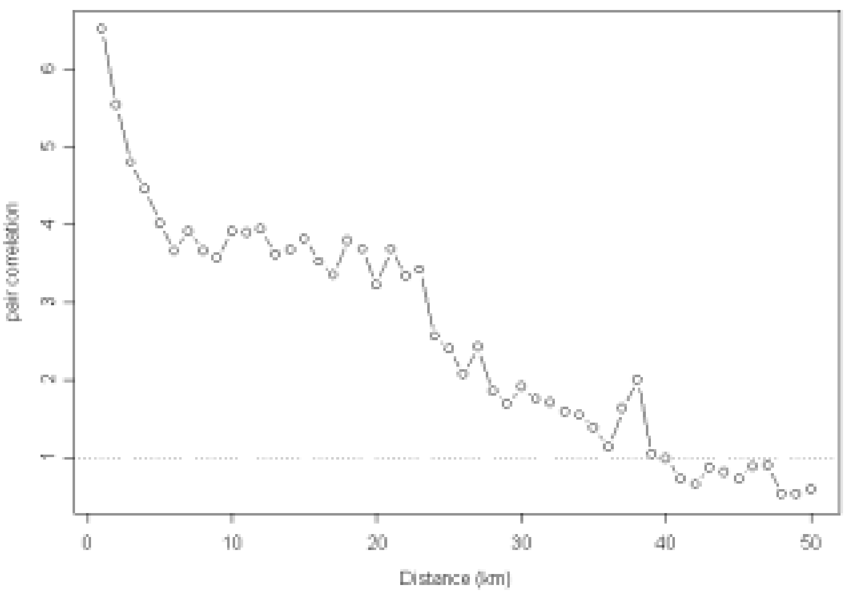

Fig.2: Along transect pair correlation function characterising the correlation structure in the occurrence of individual schools.

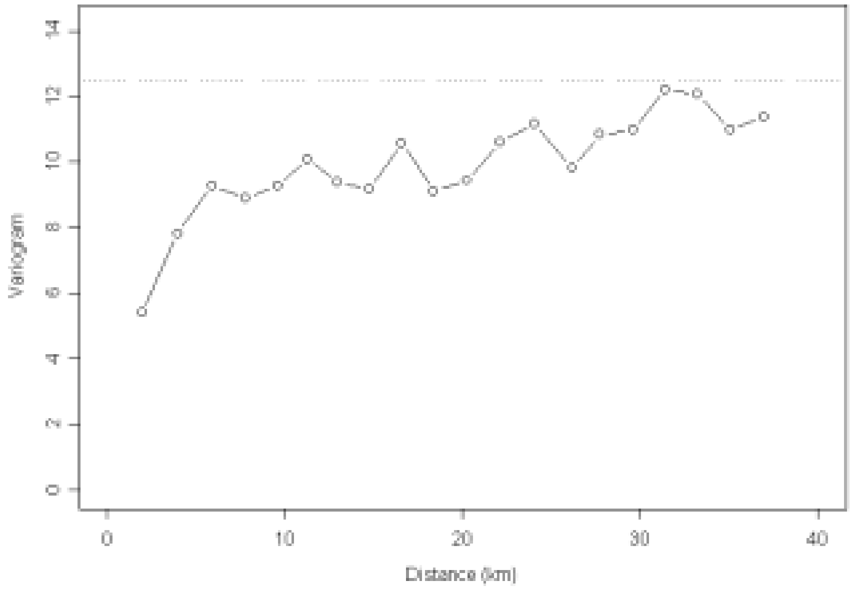

Fig. 3a: Along transect variogram of the nb of schools per nautical mile (Nm).

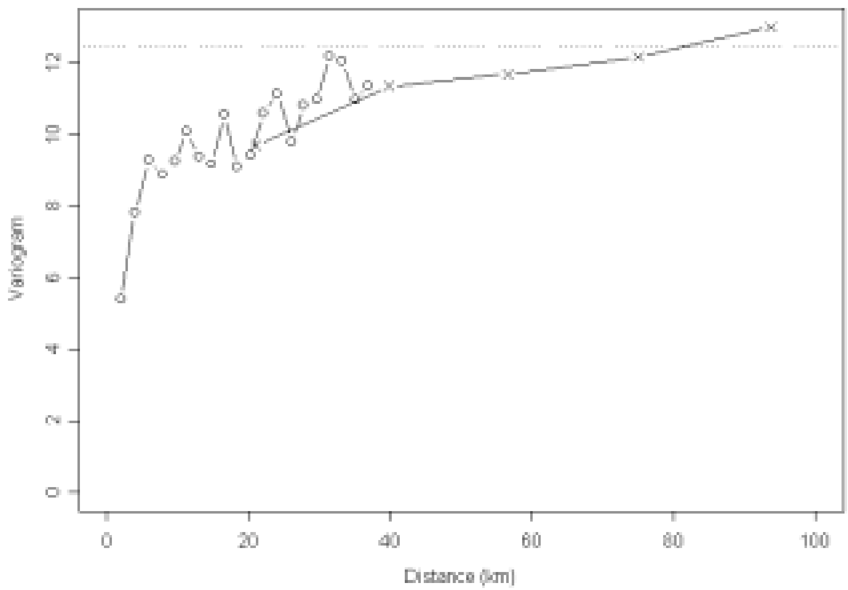

Fig 3b: Along (circles) and across (crosses) transect variograms for the nb of schools per nautical mile.

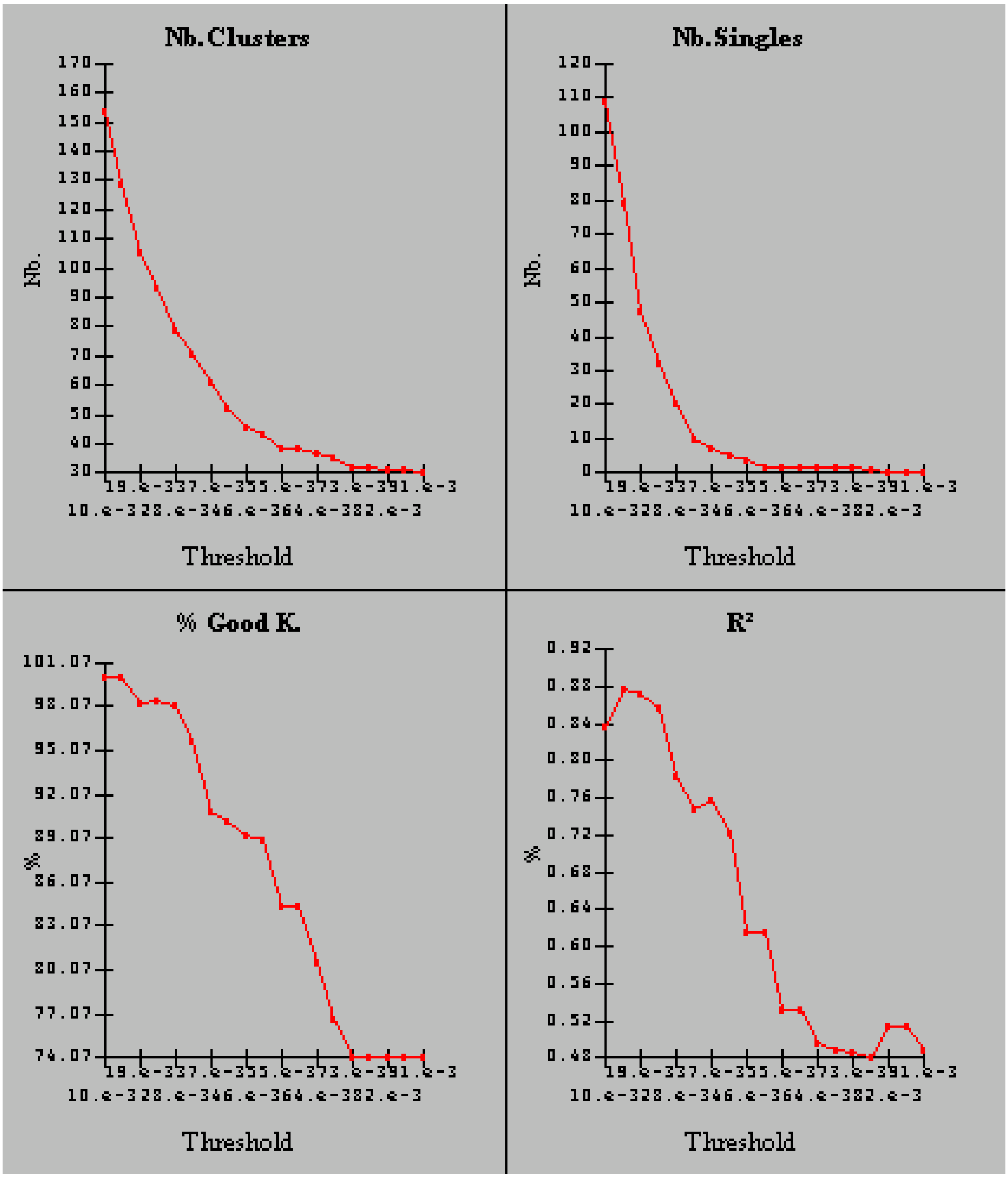

Fig. 4: Multicriteria curves for defining maximum distance between schools in a cluster. Threshold distance is in degrees. % Good is the percent of clusters with homogeneous distribution of schools inside. R2 is the R-square of the linear regression of the nb of schools per cluster on cluster lengths. Threshold distance chosen is 2km (ie between third and fourth points)

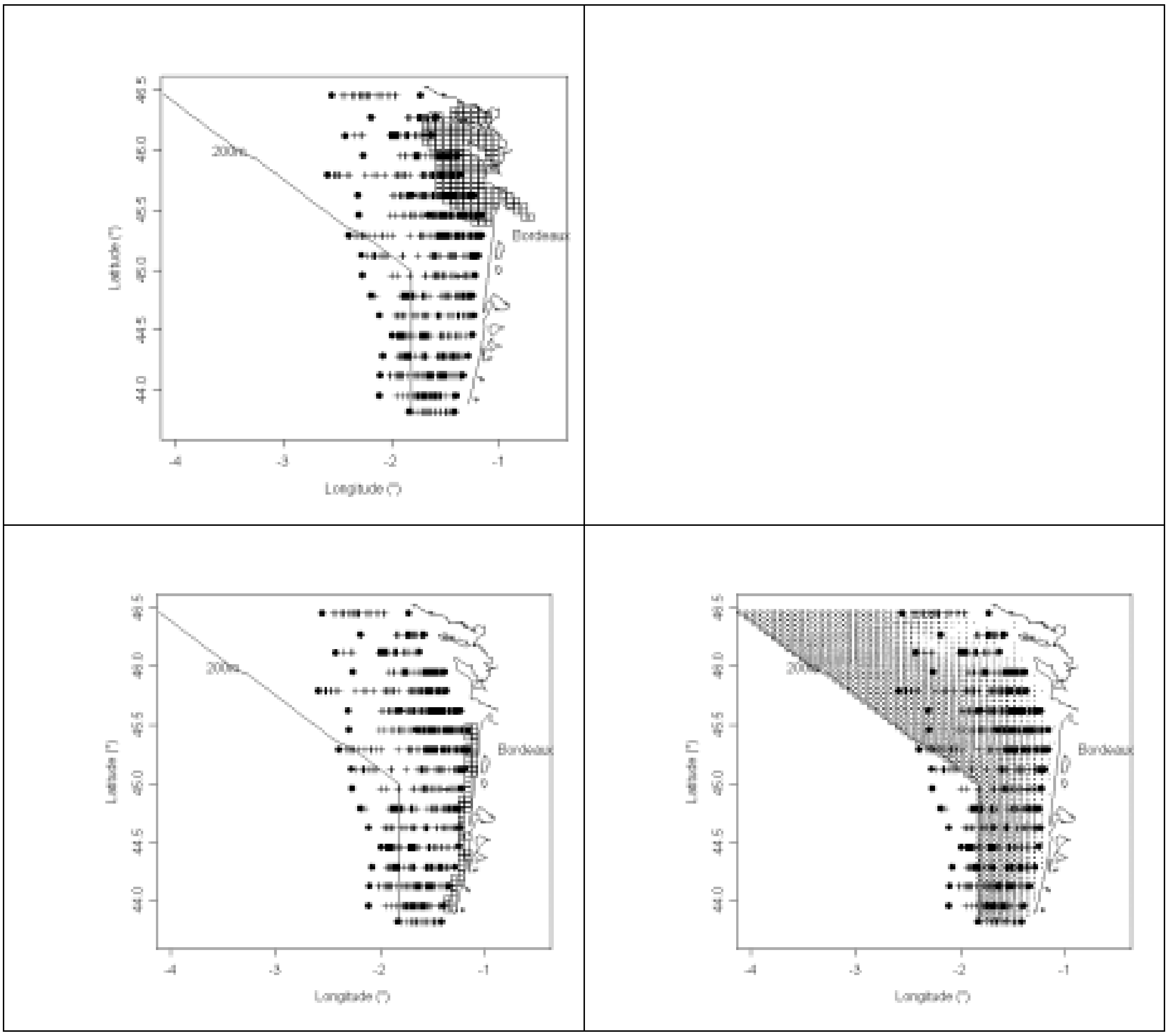

Fig.6 : Spatial distribution of schools in Biscay (may 1997) superimposed to the spatial distribution of hydrological parameters as infered with a 3D hydrodynamic model. Crosses are for the positions of individual schools, black points for the extremities of acoustic survey transects, empty squares and empty circles for the physical parameters.

- Top left figure represents the river plume extent (surface salinity < 34psu). The area is the union of all areas during the month of the survey.
- Bottom left figure represents the coastal upwelling extent (vertical speed from bottom to surface > 0). The area is the union of all areas during the month of the survey.
- Bottom right figure represents a stratification index of the water column (circles are proportional to the index value). Values are averages during the month of the survey.