RESUME

L'applicabilité de la géostatistique pour l'analyse de données d'écho-intégration est explorée à l'aide de deux ensembles de données provenant des eaux côtières du nord de la Norvège et du Vénézuéla. Les écarts aux conditions de stationnarité et les variations temporelles des structures spatiales sont examinées en portant une attention spéciale aux effets de : (1) l'utilisation des données brutes ou transformées en unités logarithmiques, (2) l'exclusion ou l'inclusion des séries de zéros, (3) différences jour/nuit dans les structures spatiales, (4) double échantillonnages et (5) différentes communautés de poisson. Tous ces facteurs affectèrent fortement la forme des variogrammes calculés, et conséquemment les estimations de biomasse qui en étaient reliées. L'utilisation de la géostatistique pour l'analyse des structures spatiales, l'estimation ou la cartographie de la biomasse en écho-intégration doit donc être effectuée avec précautions. Des recommandations sont suggérées et les aspects demandant plus de recherche sont soulevés.

ABSTRACT

The applicability of geostatistics to analyze fisheries acoustics echo-intégration data is explored with two typical data sets from coastal waters off northern Norway and Venezuela. Departures from stationarity conditions and temporal variations of the spatial structures are examined with special attention to the effects of : (1) using of raw or logtransformed data, (2) exluding of including the series of zeros, (3) day and night differences in the spatial structure (4) revisiting some sampled areas and (5) different communities of fishes. All these factors strongly affected the shape of the variogram computed, and consequently the biomass estimations also. Care should thus be taken when using geostatistics to analyze the spatial structures map or compute biomass estimations. Some recommandations are suggested and questions needing more explorations are pointed out.

INTRODUCTION

The echo-integration technique (Burczinski 1982) produces dense series of biomass estimates of pelagic or demersal fishes along routes surveyed by research vessels. The biomass is generally integrated over distance intevals of 1-5 nautical miles. Because of the continuous sampling serial autocorrelation is present in the data collected. This hinders data analysis with classical statistics, which require independence of the samples. Some methods have been suggested to minimize the effect of spatial autocorrelation to allow the use of classical statistics (Williamson 1982, Maclennan and Mackenzie 1988). A alternative methods using the autocorrelation structure through the Theory of regionalized variables (Matheron 1965, 1971) have been tried a few times in mid-1980s' (Gohin 1984, 1985 ; Laloe 1985 ; Guillard et al, 1987). but they have not stimulated a large interest until recently. The present paper explores the applicability of these geostatistical methods for spatial structures analysis and biomass estimation in fisheries acoustics.

Using geostatistics in fisheries acoustics presents two main potential interests. The first one is for analyzing the spatial structures of fish biomass that can be inferred from the characteristics of the structural functions (e.g David 1977 ; Sokal 1986) such as : the range of autocorrelation, the shape of the variogram, the sill level, the relative importance of the variability at various scales. The second one is for the estimation of local or global fish biomass in presence of spatial autocorrelation and for "optimal" "objective" mapping through kriging. This information can then be used for designing optimal sampling strategies.

The application of geostatistics to model spatial process relies on some basic assumptions about the studied variable. First, its structure is assumed to be stable in time, at least at the studied time scale. This might be true for mining deposits but it is not so for the distribution of fishes in a given environment, where variations of locations and activities are expected to occur over a wide range of time scales, hours, days, weeks etc... Second stationarity of the spatial process is assumed : this means that this spatial process must have some homogeneity and should be repeatable in space. Under the intrinsic hypothesis, the increment of the values of samples depend only on their relative spatial orientation, there is no trend or if there is one it can be modeled by the intrinsic random functions (David 1977 ; Clark 1979 ; Delhomme 1978). These stationarity conditions are far from the expectations from acoustics data. Complex spatial inhomogeneities are the rule in fish distribution data, which besides often include extremely high values (outliers), representing significant portions of the total biomass, or large proportion of total number of samples with no biomass. The present contribution explores how theses various assumptions are satisfied in two sets of data from tropical and arctic waters. Other specific objectives were to check the effects of : (1) using raw or log-transformed data, (2) excluding or not the series of zeros in the computation of the variogram (3) day and night differences in the spatial structures, (4) revisiting some sampled areas, (5) different communities of fishes on the spatial structures.

METHODS

The first data set has been collected in February-m... 1989 from northern Norway between latitudes 62 N and 64 N a... longitudes 4 E and 8 E. The echo (m of backscattering cross section per square nautical miles) of pelagic fishes were collected at intervals of 1? nautical miles along the cru... tract (fig. 1). This data set has been provided by K. Foo... (Institute of Marine Research. P.O. Box 1870, Nordnes, ... Bergen, Norway).

The second data set comes from a survey realized in August 1985 off the east coast of Venezuela between latitudes 10 N and 12 N and longitudes 65.5 W and 61.5 W. The acoustic system was composed of a sounder SIMRAD EK-400, 120 KHz, a digital echo-integrator AGENOR. The echos were integrated at intervals of 0.8 nautical miles and expressed in relative units, directly proportional to the backscattering cross-section per area unit. The daytime transects were revisited at night in order to cover the whole area twice, once by day and once by night (fig 10). Some areas visited by day or by night were revisited with additional transects, and the transects located in the Gulf of Cariaco were more closely spaced than elsewhere. characteristics of the echos were computed per stratum of ... min. lat. by 20 min. long . and analized by hierarchical clustering to identify the various homogeneous acoustic populations (Fig. 17 ; Gerlotto and Marchal 1987), which corresponded to different species communities as indicated ... pelagic trawl samples (n = 30).

The numerical analysis was done with a combination of the following spatial analysis packages. GEO_EAS (Environmental Protection Agency USA) was used for computing histograms and variograms and for kriging. SURFER (Golden Sofware, G... Colorado, USA) did the contouring, 3-dimensional plots and ... interpolation by the inverse square distance or by kriging using a variogram with linear model. Some variograms were also computed from an ORSTOM internal package.

RESULTS

NORTHERN NORWAY DATA SET

The echos from this area were characterized by a large proportion of zeros (60%). The distribution of positive echos over the whole area was lognormal but showed substantial spatial variations some areas having larger proportions of low or high echos indicating non-homogeneity (fig. 2). The time sequence plot of the echos (fig.3) were marked by 4 sharp peaks isolated from the rest of the series and serial autocorrelation at small scales was evident. Correlograms indicated that the correlation coefficient as well as the length of significant autocorrelation changed with the type of data pre-treatment used (Fig 4). This was also observed on the variograms (fig. 5), the range of autocorrelation, the levels of the nugget and the sill changing when the data were logtransformed or were excluding the zeros. The variograms computed per transects (fig.6) also presented large differences among them and were strongly affected by zeros. So were the variograms per strata (fig.7 and 8) indicating that the structure of echos was not homogeneous over the whole sampled area. The maps of echos obtained from various kriging options and from interpolation according to the inverse square distance method (fig.9) presented significant differences among them. Lognormal kriging with variograms including (ln x+1 Fig. 9) or excluding (ln x, fig 9) the zeros gave very different estimates. The estimates obtained when the zeros are considered in computing the variogram are systematically higher than the other options and seem erroneous. Lognormal kriging is very sensitive to the level of the sill of the variogram (Armstrong and Boufassa 1988) with may explain the high estimates obtained. The kriged maps from raw data gave similar results, the option including the zeros showing the highest peaks however. The linear kriging option of surfer (x lin. fig.10) produced a map slightly differing from the latter two, and which showed notable differences from the map obtained by the inverse square distance method (x 1/d . fig 10).

COASTAL VENEZUELA DATA SET

The sampling design in this region included significant proportions of revisited areas in both the day or the night surveys (fig. 10). Also contrasting with the other data set, the proportion of zeros was low (< 5%). The histograms of logtransformed daytime data per stratum (fig.11) evidenced significant spatial variations of the echo distribution. The logarithmic transformation failed to normalized the data in most cases. The corresponding histograms for the night survey (fig.12) were clearly different, the distributions being less skewed towards low echos and generally centered on similar means. Even though spatial variations were presents they were smaller than during the day. Variograms per strata (fig.13 and 14) also exhibited spatial variations and those of the daytime survey systematically showed higher sills and nugget as expected from the histograms. The effect of revisiting sampled areas on the variograms was tested for the Gulf of Cariaco data (fig. 15). This two-way sampling design greatly altered the variograms notably by raising the nugget and the variance at small scales. This was especially evident in the night variograms for the whole region sampled (fig. 16). The inclusion of even a small sub-area with such revisited data can drastically change the variogram as shown in fig. 16 when the revisited data of the Gulf of Cariaco are included in the computed variogram. Different community of fishes having their own acoustic signature (sensu Gerlotto and Marchal 1987, from cluster analysis of acoustic data) and species composition also appeared to be characterized by different variograms (fig. 17).

CONCLUSION

The above results clearly showed that typical fisheries acoustic data are not always well adapted for geostatistical analysis. Care should therefore be taken when using such methods. Deviations from stationarity conditions resulted in different variograms depending on (1) which area of the studied region was considered and (2) the type of data used, which could include or not the series of zeros and which could be transformed or not. Since the variogram is the basic tool in geostatistics, the variability presents some concerns : what data should be considered when computing the variogram and when kriging and what transformation should be used ? How biomass mapping and global estimation should be performed, globally or by fusion of homogeneous strata ? How to consider the outliers adequately ? Should lognormal kriging, relative kriging or other types of kriging be used ?

Other problems with the method are related to the dynamic behaviour of fishes : their day/night cycle of activity and distribution pattern, their displacements, and the species composition of the communities. The day and night survey off Venezuela clearly showed that the distribution patterns of fish echos changed with the day/night cycle. To maximize stationarity day and night data should not be mixed, because they corresponds to different spatial organizations of fishes. This finding imposes some constrains to sampling strategies, by limiting the effective of work per day and the size of the surveyed area. Night echos were more homogeneous at all scales and over the whole region than those during the day which favours the use of night data for geostatistical analysis.

Since fishes are not sessile organisms but they continously move in their environment, this violates basic condition of geostatitics. Revisiting previously sampled locations should be avoided if the revisited data are to be considered simultaneously with the original data for the geostatistcal analysis. Such strategy would result in significant increases of the variance, particularly at small mean scales, and give erratic variograms, as was shown in the survey off Venezuela. The use of both revisited and one-way data requires strategies of data analysis differing from the standard procedure. This observation stresses the importance of well ordered transects designs in the time-space domain, i.e. the area covered should be swept in order during the survey, with noroutes back already sampled area or too close to them. For example long transects tightly spaced could generate a high nugget on the variogram if the fishes seen on one transect are not seen on the adjacent transects because they have moved away. The transects design must therefore be adapted to the relative speed of survey compared to the fish displacement speed. Zig-zag transect design seems well-adapted to these sampling contrains because the samples are well ordered in time-space ; the closer they are in space, the closer they are in time and vice-versa.

It seems that different fish communities have different spatial organisation, from the variograms computed for the "accoustic populations" off Venezuela. To maximize stationarity, it would therefore be appropriate to stratify the region according to the different fish communities and then to study the strata separately.

Finally we have not considered the problem of anisotropy, because the data were not ideal for such an anlysis, but this must be done. Is anisotropy important ? What controls anisotropy ? Bottom topography ? Water masses ? Distance from the coast ? These are questions that should be addressed.

ACKNOWLEDGMENTS

We are grateful to K. Foote for making available the data set of northern Norway.

REFERENCES

Armstrong, M. and A. Boufassa. Comparing the robustness of ordinary kriging and lognormal kriging : outlier resistance. Math. Geol. 20 (4) : 447-457.

Burczynski, J. 1982. Introduction to the use of sonar systems for estimating fish biomass. FAO Fish. Tech. Pap. 191 Rev. 1. 89 p.

Clark, I. 1979. Practical geostatics. Elsevier. 129 P.

David, M. 1977. Geostatiscal or reserve estimation. Elsevier. 364 p.

Delhomme, J.P. Applications de la théorie des variables régionalisées dans les sciences de l'eau. Bulletin du B.R.G.M. (2) III, 4-1978, 341-375.

Gerlotto, F., and E. Marchal. 1987. The concept of acoustic populations : its use for analyzing the results of acoustic cruises. Int. Symp. Fish. Acoustics, June 22-26, 1987, Seattle, 30 p.

Gohin. F. 1985. Geostatistics applied to fish distribution as derived from acoustic surveys. ICES/FAST Working Group Meet., 22-24 May, Tromso, 6 pp.

Guillard. J., D. Gerdeaux et J-M. Chautru. 1987. The use of geostatistics for abundance estimation by echointegration in lakes : the example of lake Annecy. Int. Symp. Fish. Acoustics, June 22-26, Seattle, 17 pp.

Laloe, F. 1985. Contribution à l'étude de la variance d'estimateurs de biomasse de poissons obtenus par echointegration, Oceanogr. Trop., 20 (2) : 161-169.

Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris, 305 p.

Matheron, G. 1971. The theory of regionalized variables and its applications. Les cahiers du CMM, fasc. 5, E.N.S.M.P. 211 p.

Maclennan, D.N., and I.G. Mackenzie 1988. Precision of acoustic fish stock estimates. Can. J. Fish. Aquat. Sci. 45 : 605-616.

FIGURE CAPTIONS

Figure 1. Map of the area sampled in northern Norway, with samples locations (crosses) along the cruise tract and gross contours (100, 2100, 4100) of echo (m per naut. mi.). Scale in km.

Figure 2. Histograms of the natural logarithm of echos in northern Norway data set for the whole area sampled and for the sub-areas marked by an X on the map.

Figure 3. Time-sequence plot of echos from northern Norway data set.

Figure 4. Correlograms of echos from the northern Norway raw (x) and logtransformed (ln) data including or excluding the zeros. Lag = time sequence.

Figure 5. Variograms (semi-variance vs distance in km) of northern Norway raw (x) and logtransformed (ln) data including or excluding the zeros.

Figure 6. Variograms (semi-variance vs distance in km) of three different transects (pointed on the map) in northern Norway logtransformed (ln) data including or excluding the zeros.

Figure 7. Variograms (semi-variance vs distance in km) per stratum (X on the map) in northern Norway logtransformed (ln x+1) data including the zeros.

Figure 8. Variograms (semi-variance vs distance in km) per stratum (X on the map) in northern Norway logtransformed (ln x) data excluding the zeros.

Figure 9. Three-dimensional maps of kriged estimates and of interpolated values according to the inverse square distance for northern Norway raw (x) and logtransformed (ln) data. The grid mesh size was 5 X 5 km, the search circle radius was 15 km, searching by quadrants a maximum of 8 points. The top 4 maps used the corresponding variograms presented on fig. 5. The zeros were always considered for interpolation at the kriging step, even though the variograms of x > 0 and ln x used did not include them. The bottom 2 maps were obtained from SURFER, by kriging using a linear variogram (x lin) or by interpolation according to the inverse square distance method.

Figure 10. Maps of the sampled area off Venezuela by day and by night, with the sample locations (dots) along the cruise tract. The grid rectangles are 10 min. lat. by 20 min. long.

Figure 11. Histograms of the natural logarithm of echos during daytime from the Venezuela data set for the strata marked by an X on the inserted map.

Figure 12. Histograms of the natural logarithm of echos during daytime from the Venezuela data set for the strata marked by an X on the inserted map.

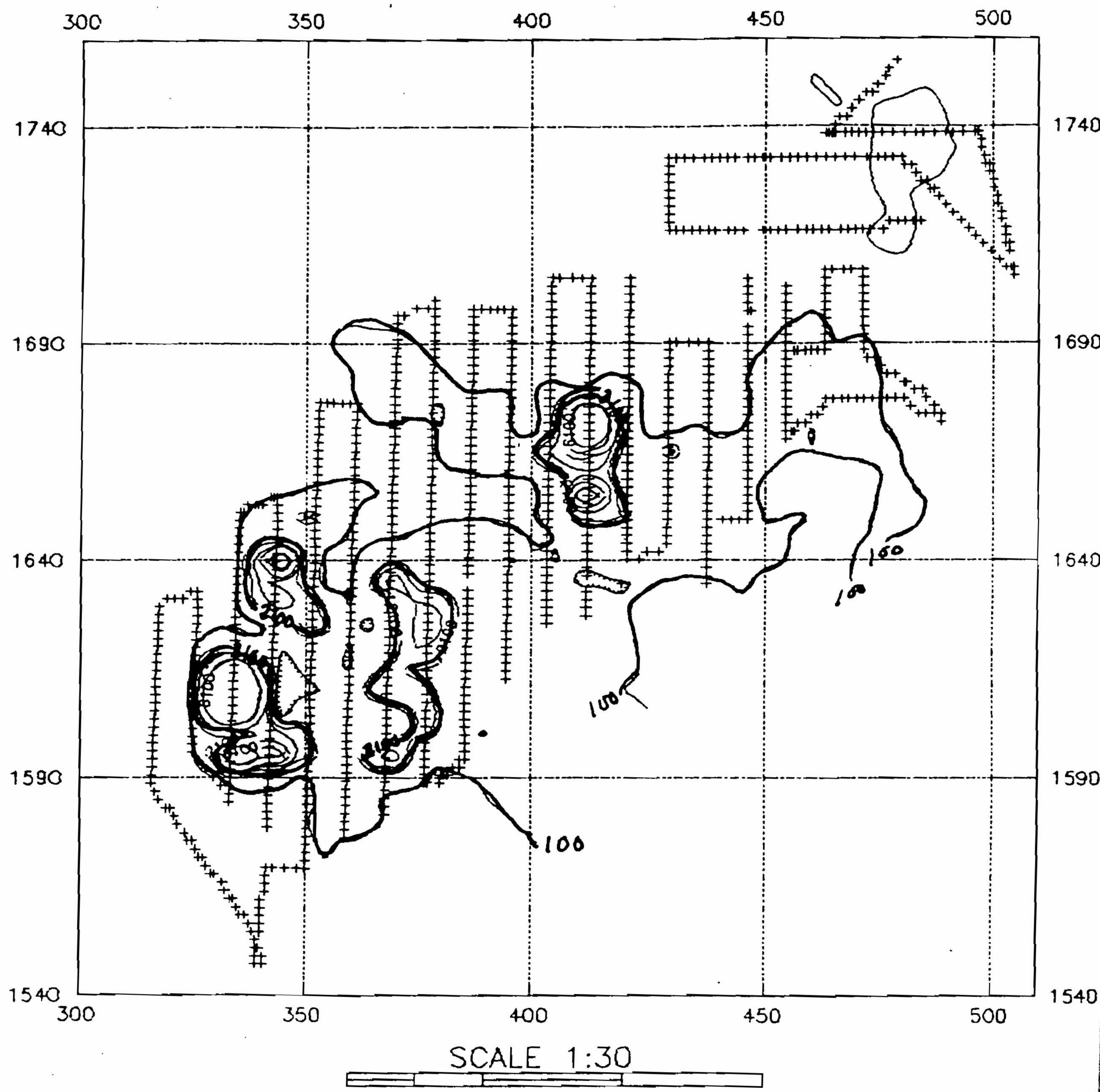
Figure 13. Variograms of the natural logarithm of echos during daytime from the Venezuela data set for the strata marked by an X on the inserted map. Note : Y scale = 0 - 20.

Figure 14. Variograms of the natural logarithm of echos during the night from the Venezuela data set for the strata marked by an X on the inserted map. Note : Y scale = 0 - 10.

Figure 15. Day and night variograms of the natural logarithm of echos from the Venezuela data set for the Gulf of Cariaco sampled one-way or revisited (two-way). Note the different scales of semi-variance.

Figure 16. Variograms of raw (x) or logtransformed (ln) echos during the night from the Venezuela data for the whole region sampled one-way or revisited (top), or for the whole region sampled one-way but including or not the revisited data of the Gulf of Cariaco (bottom). Note the different scales of semi-variance.

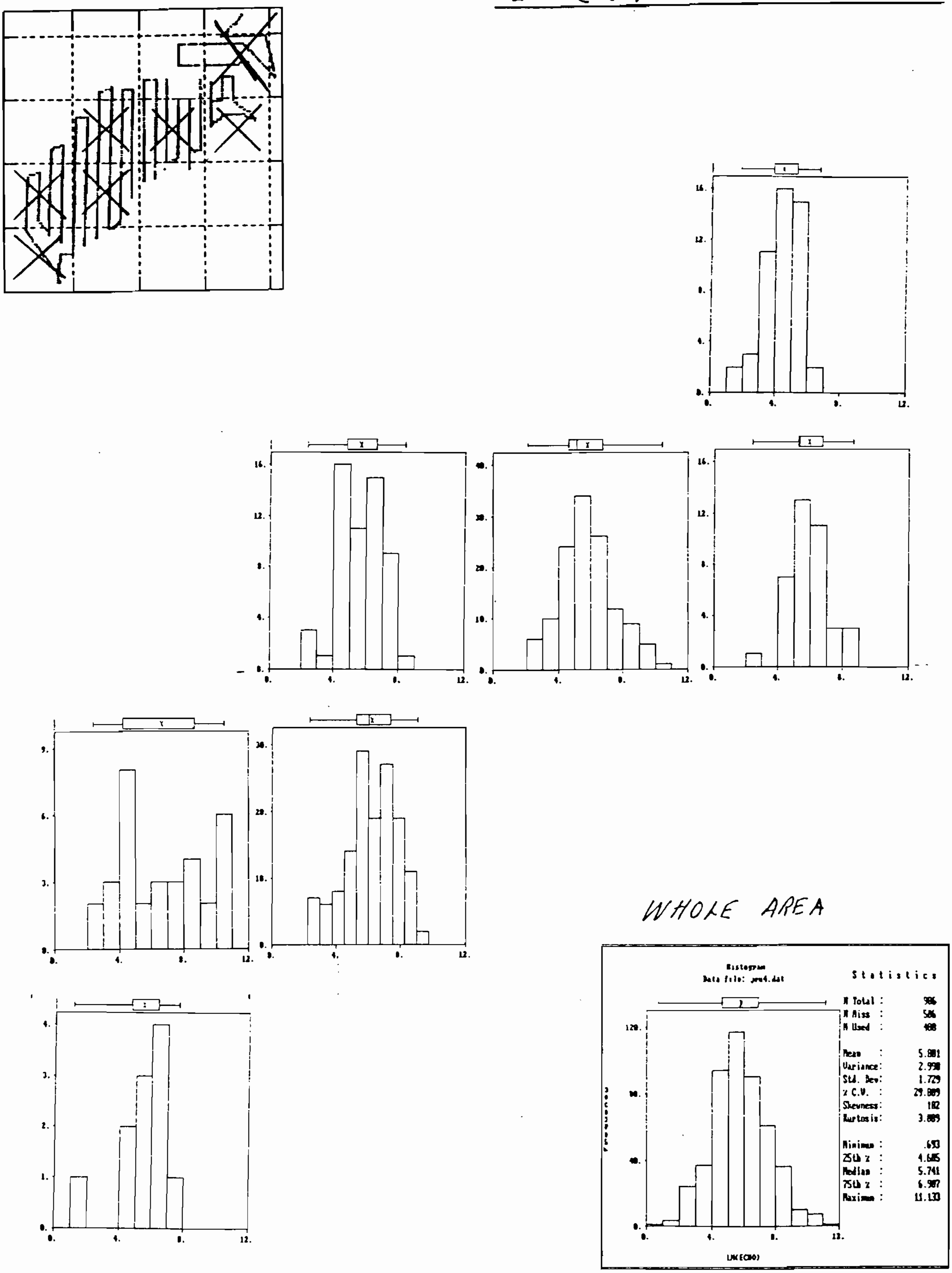
Figure 17. Variograms of the natural logarithm of echos during the night from the Venezuela data set for the groups stratat presenting similar acoustic signatures and species composition identified on the inserted map resulting from the cluster analysis. Note : the different scales of semi-variance.



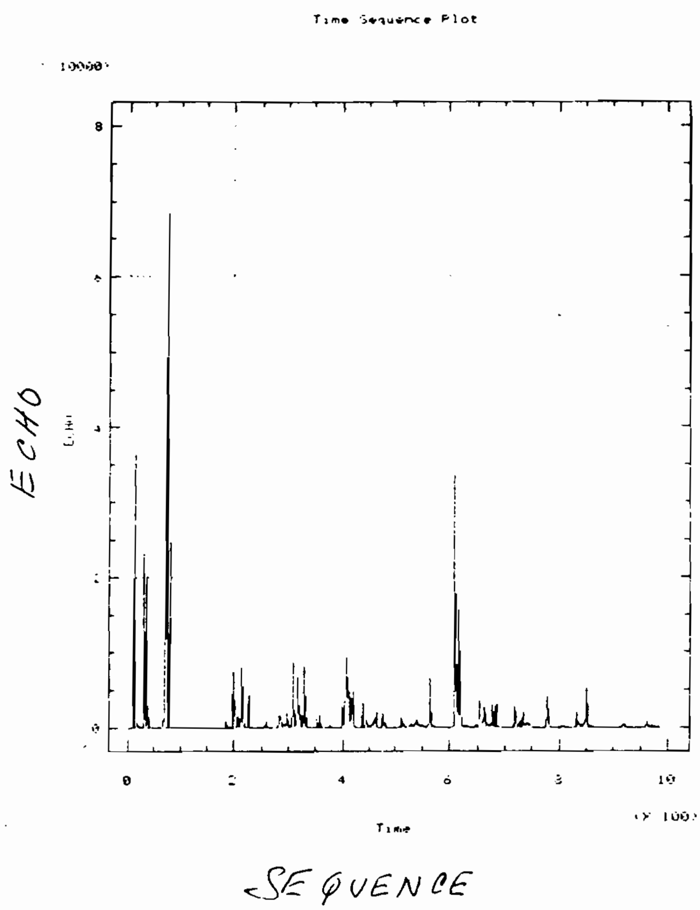
Simard and Gerlotto. FIG. 1.

SET 4, NORWAY

LN (X) ZEROS EXCLUDED



Simard and Gerlotto. FIG. 2.



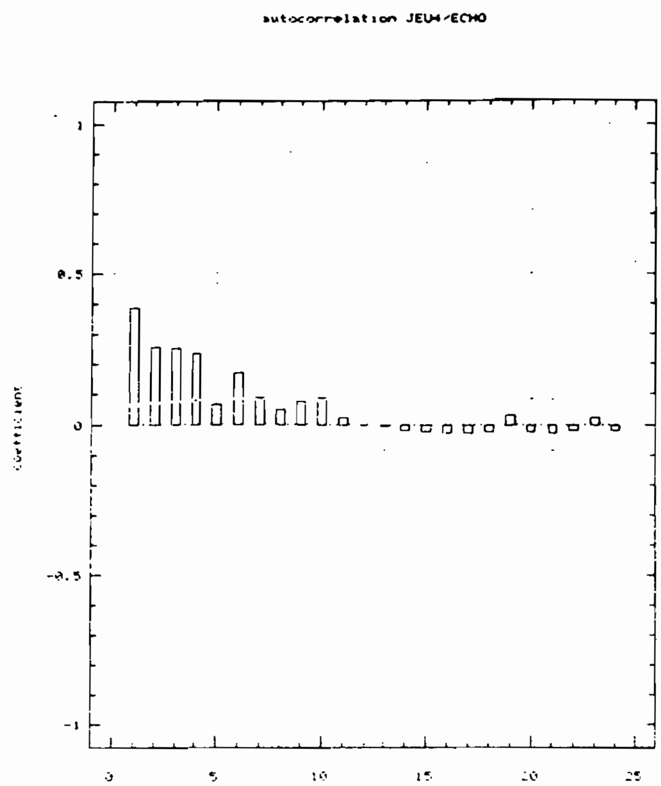
Simard and Gerlotto. FIG. 3.

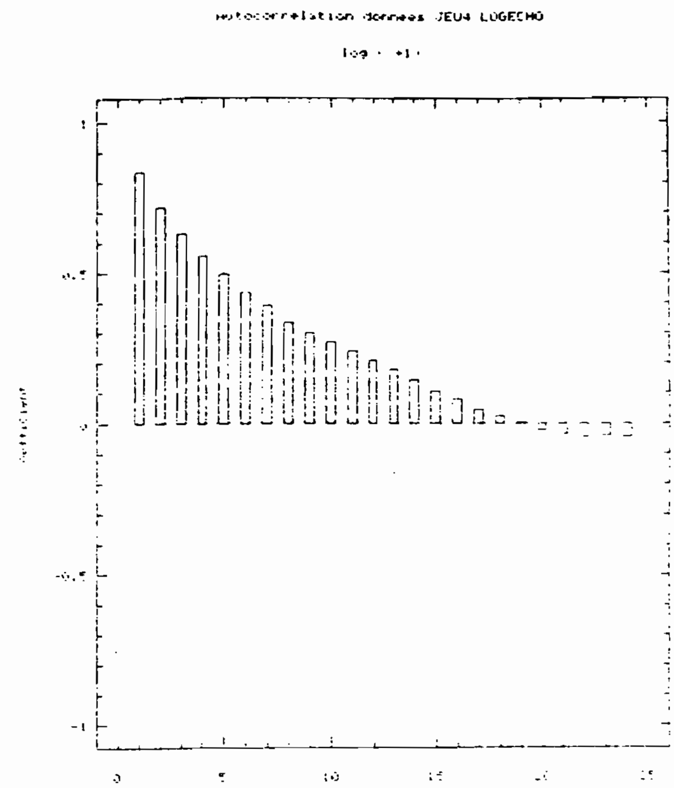
AUTOCORRELATION: DATA SET NO 4, NORWAY

ALL DATA

(X)

Ln (X+1)

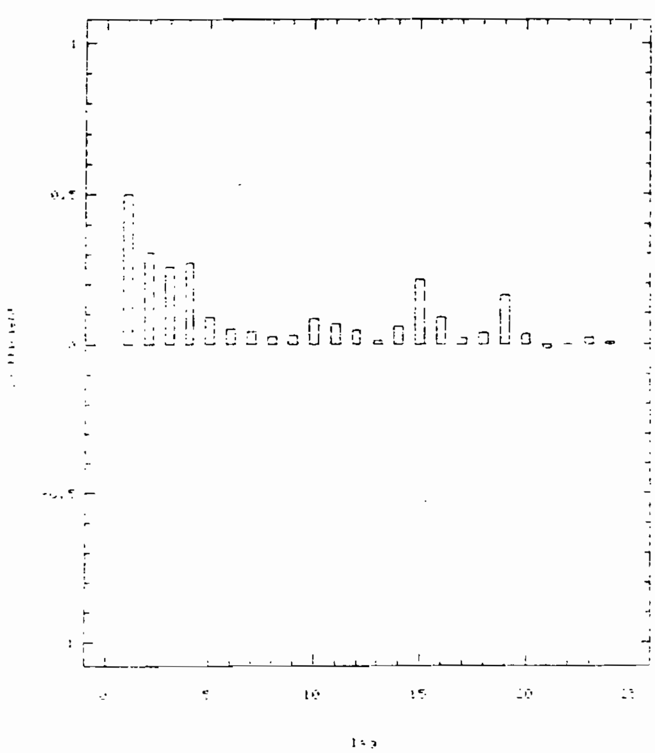


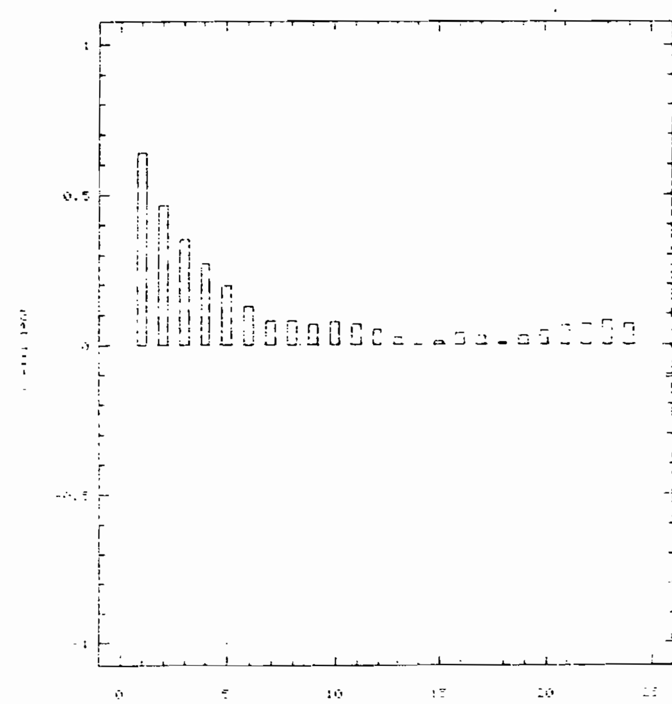


WITHOUT ZEROS

(X > 0)

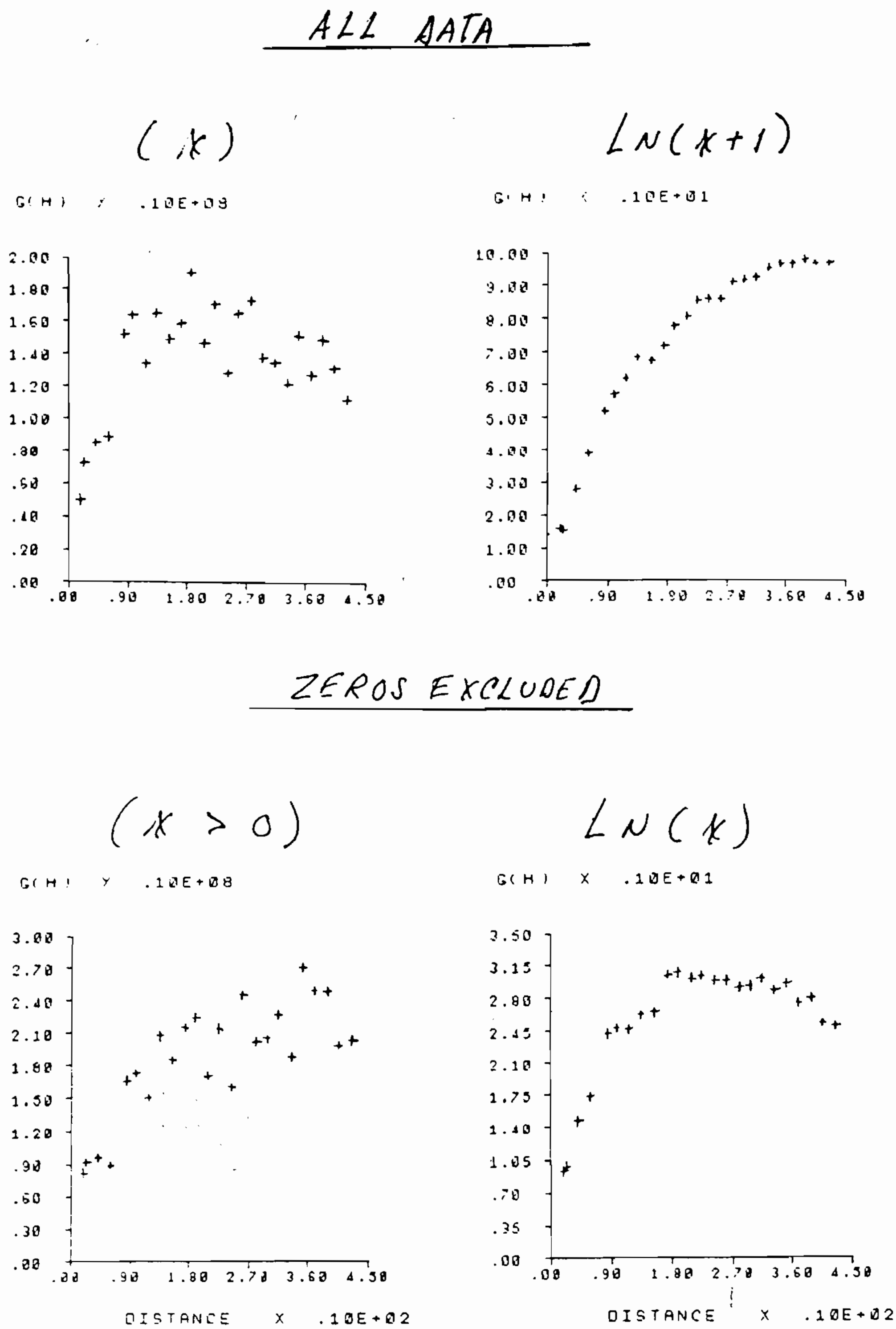
Ln (X)





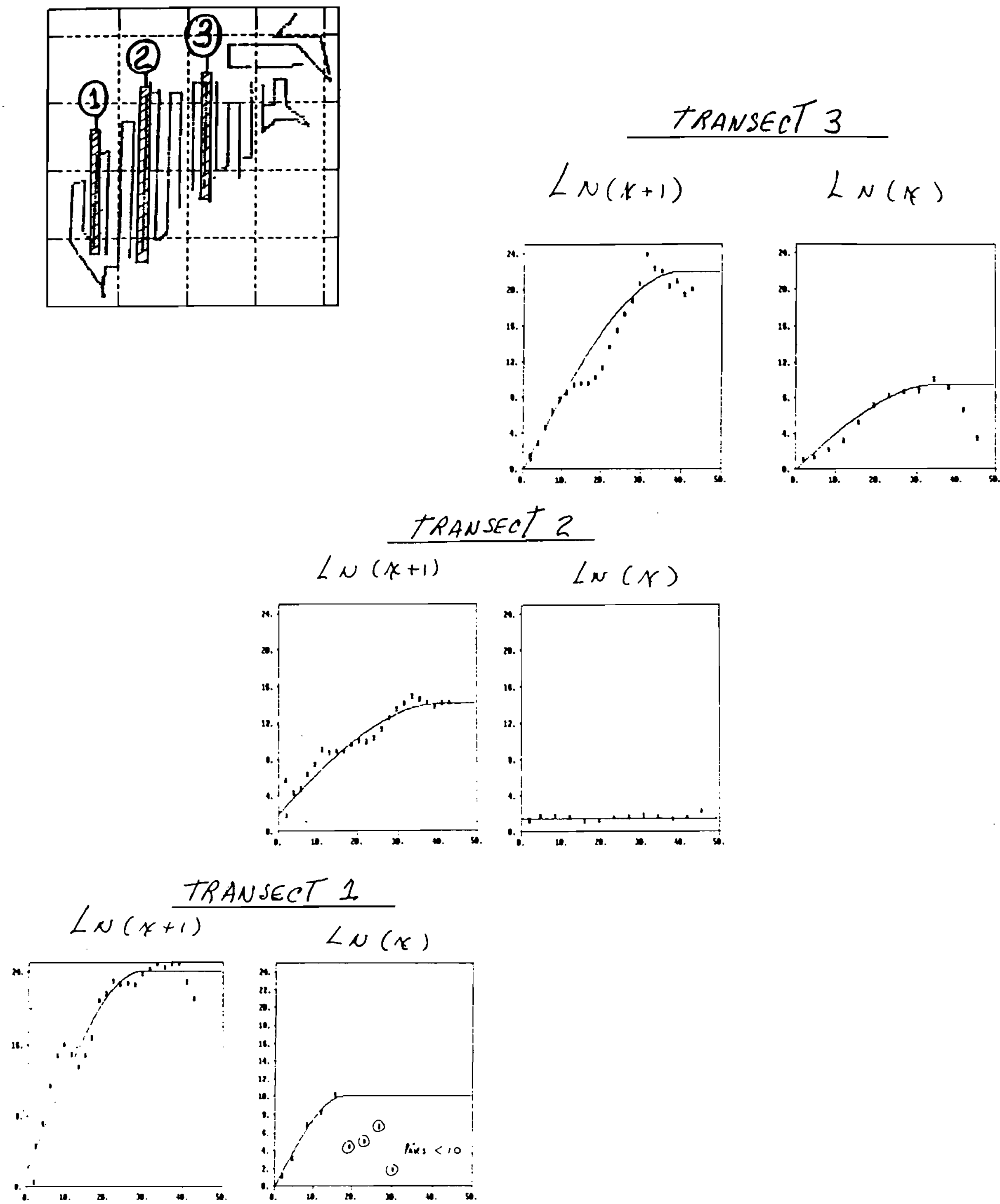
Simard and Gerlotto. FIG. 4.

SET 4, NORWAY



Simard and Gerlotto. FIG. 5.

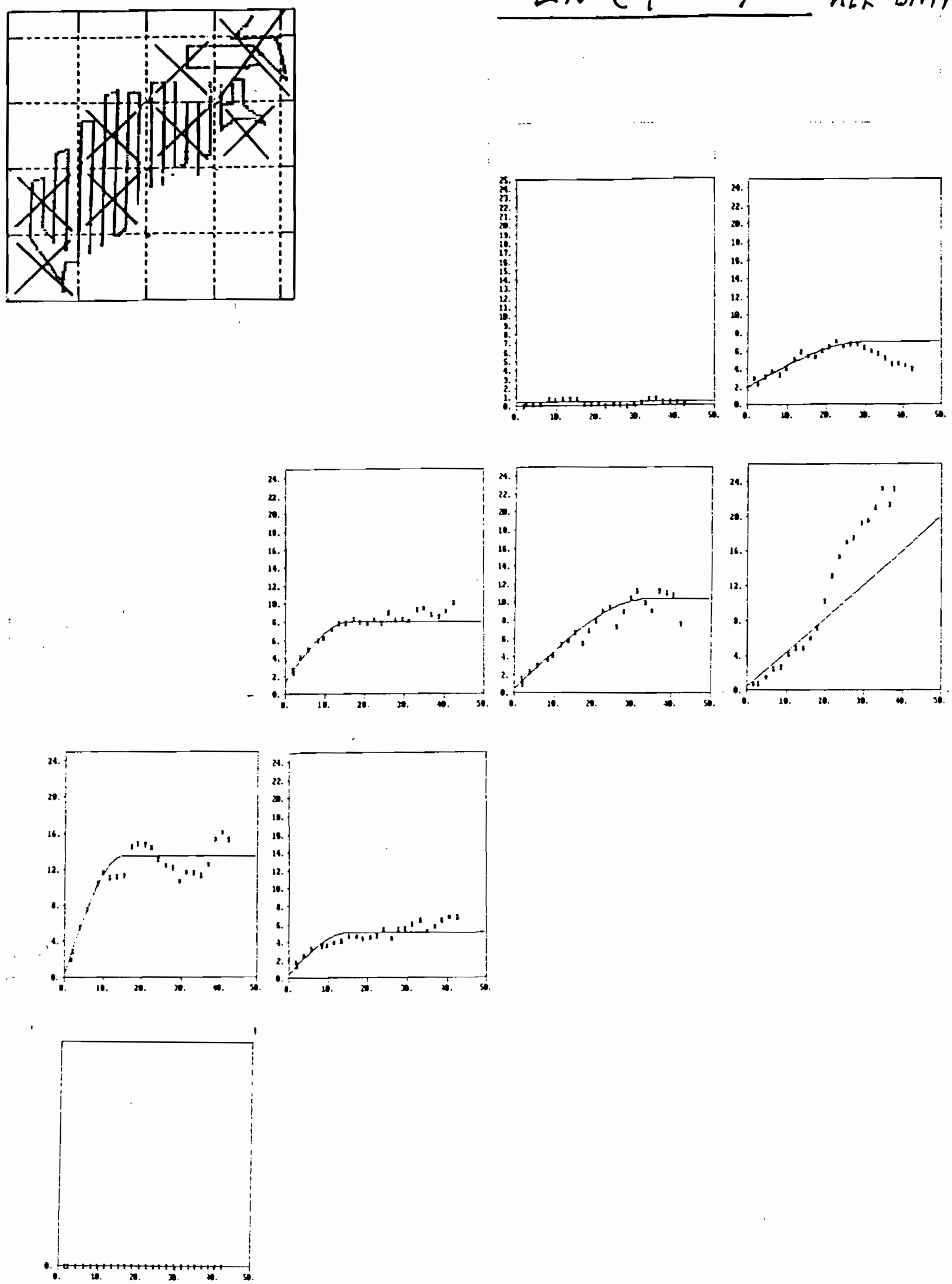
SET 4, NORWAY



Simard and Gerlotto. FIG. 6.

SET 4, NORWAY

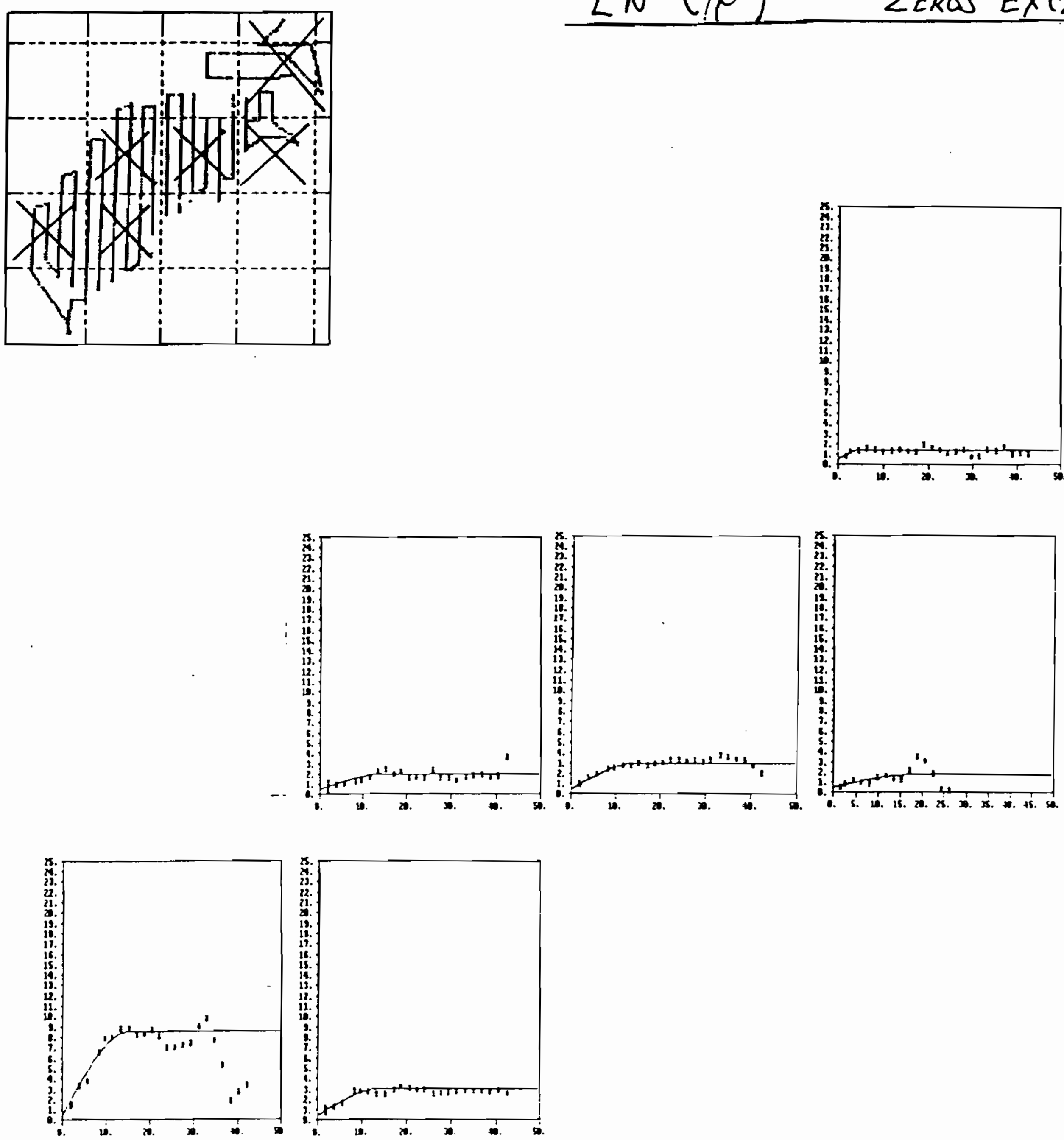
Ln (x+1) ALL DATA



Simard and Gerlotto. FIG. 7.

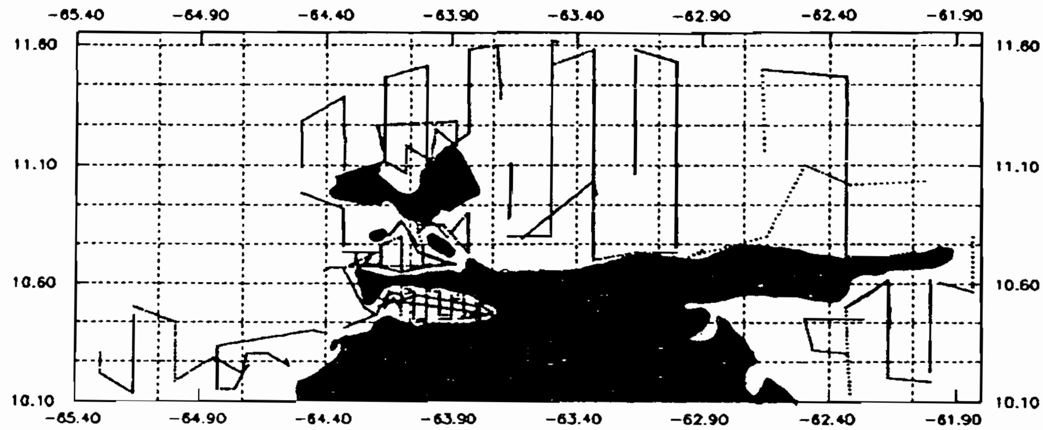
SET 4, NORWAY

LN (N) ZEROS EXCLUDED

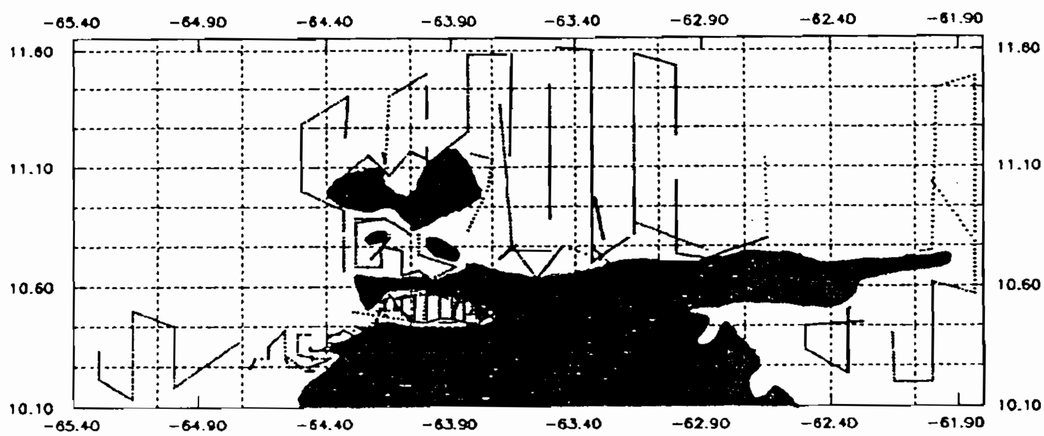


Simard and Gerlotto. FIG. 8.

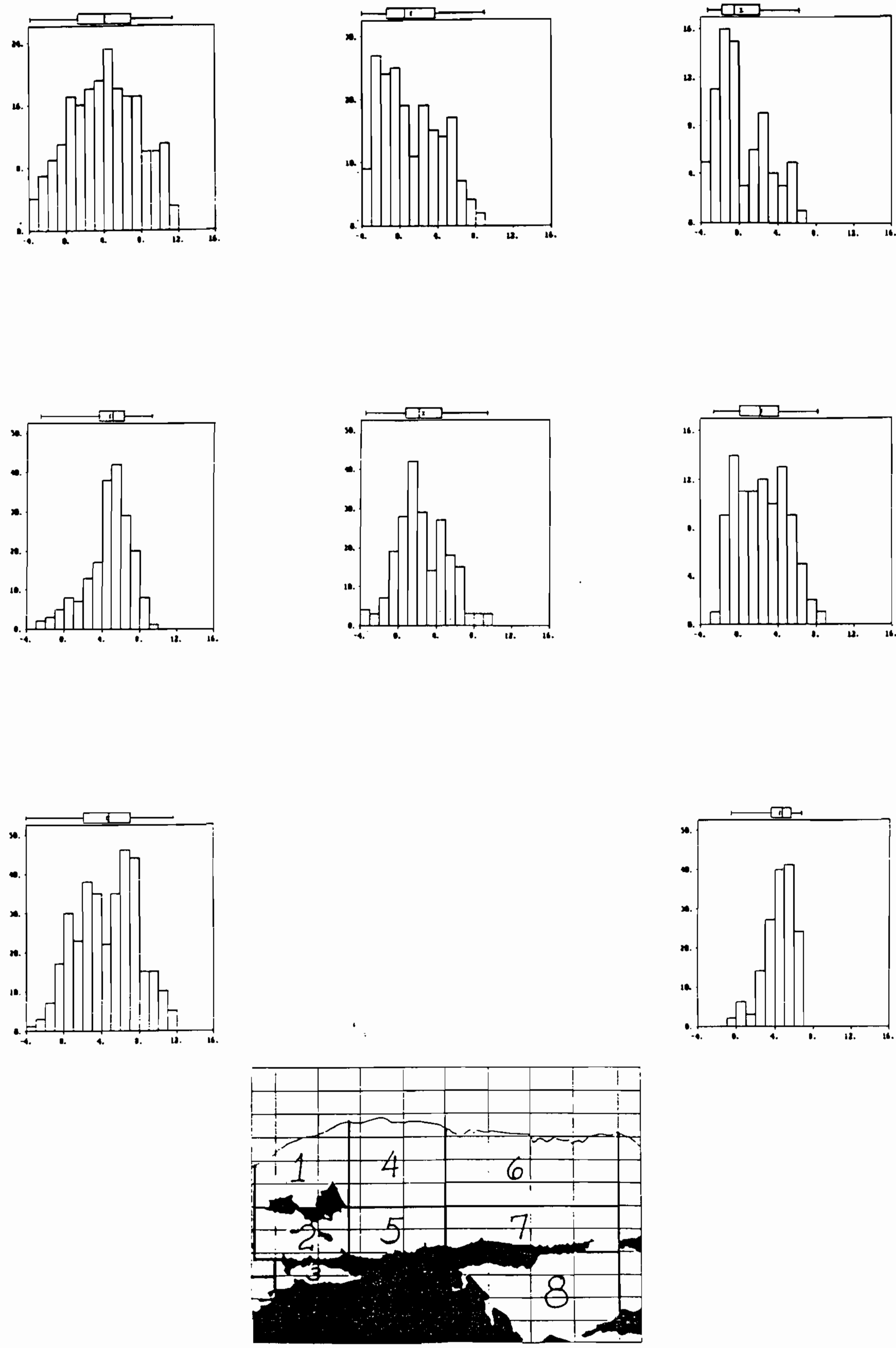
ECHOVEN NIGHT TRANSECTS



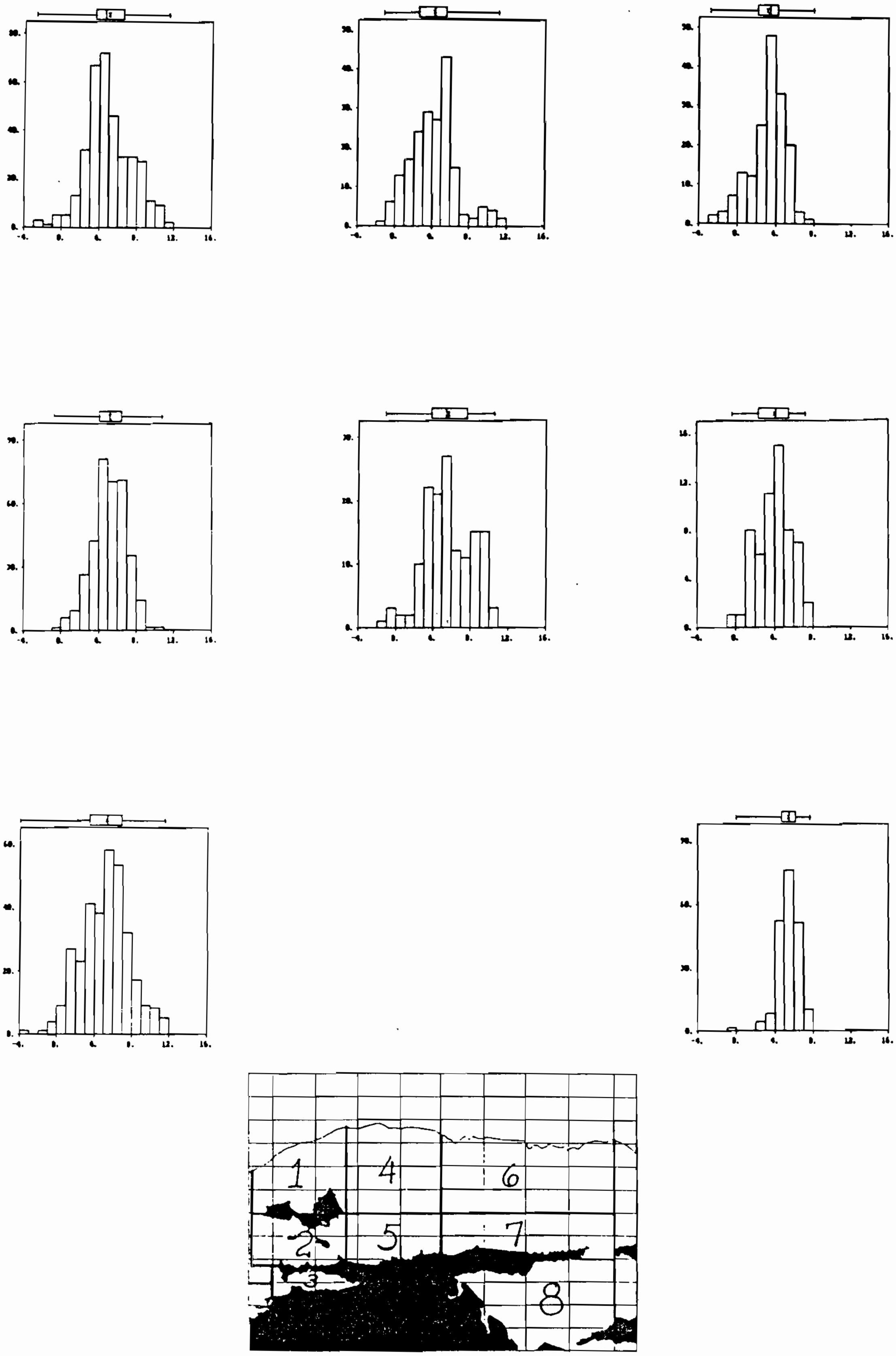
ECHOVEN DAYTIME TRANSECTS



Simard and Gerlotto. FIG. 10.

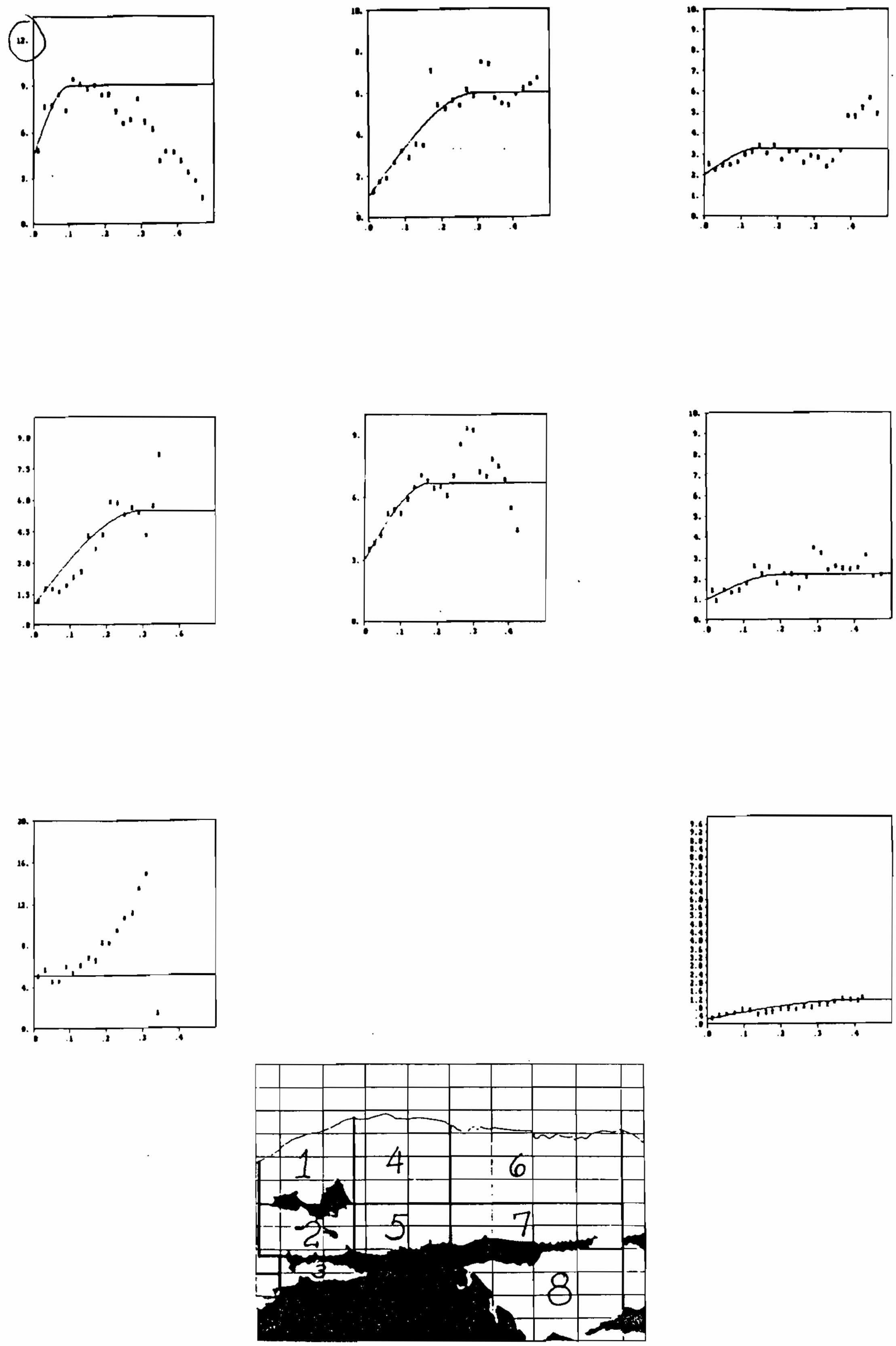


Simard and Gerlotto. FIG. 11

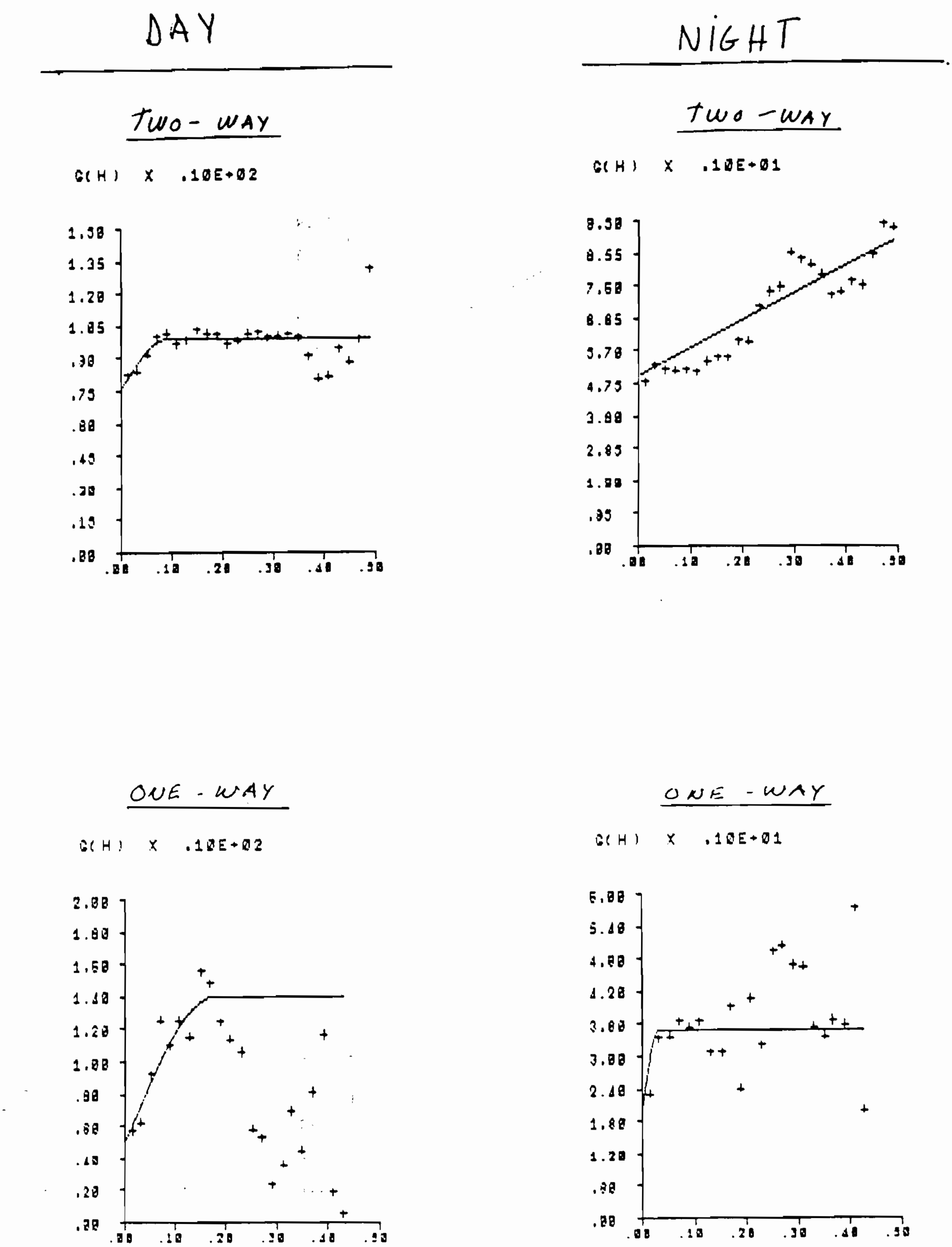


Simard and Gerlotto. FIG.

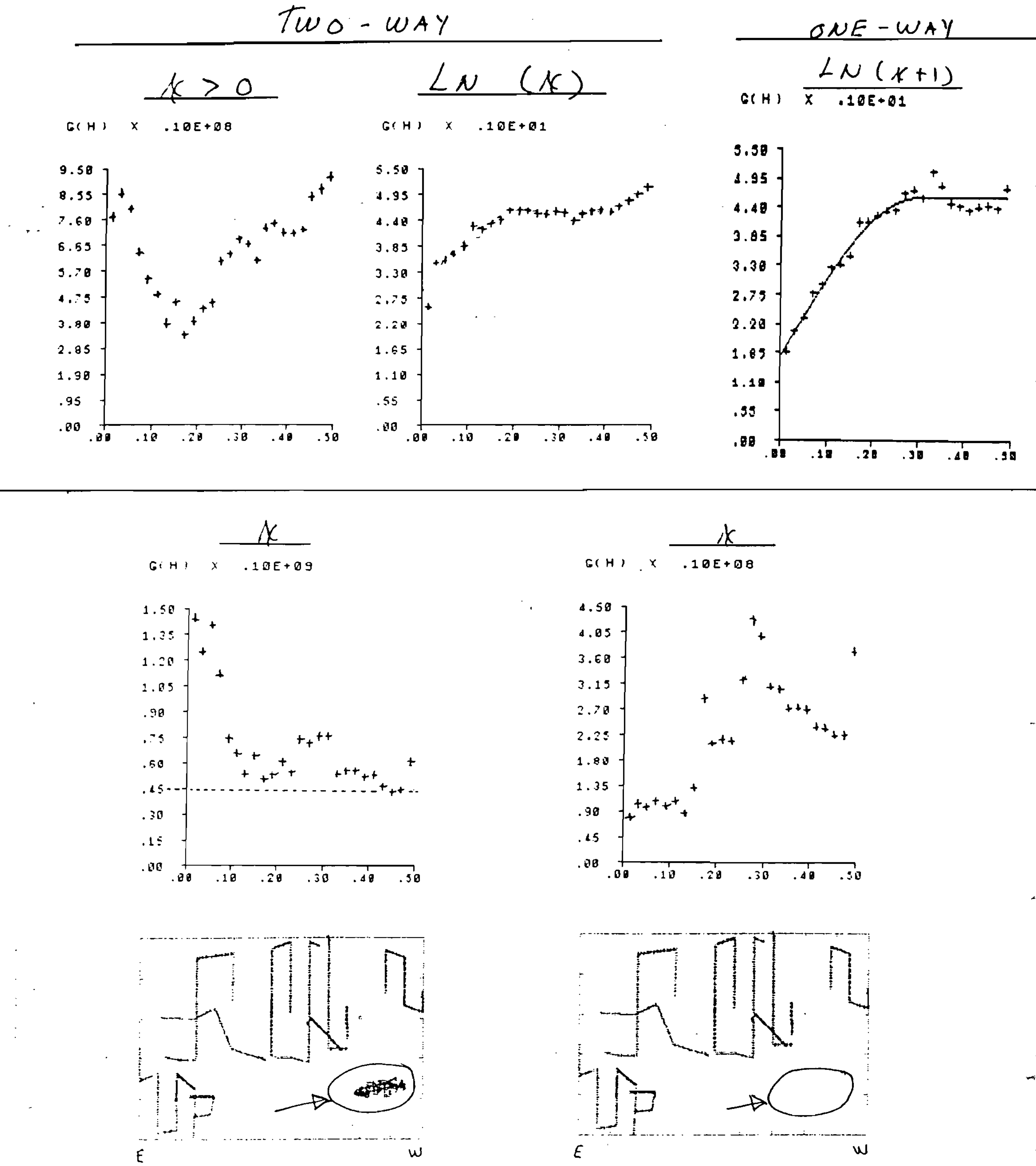
NIGHT



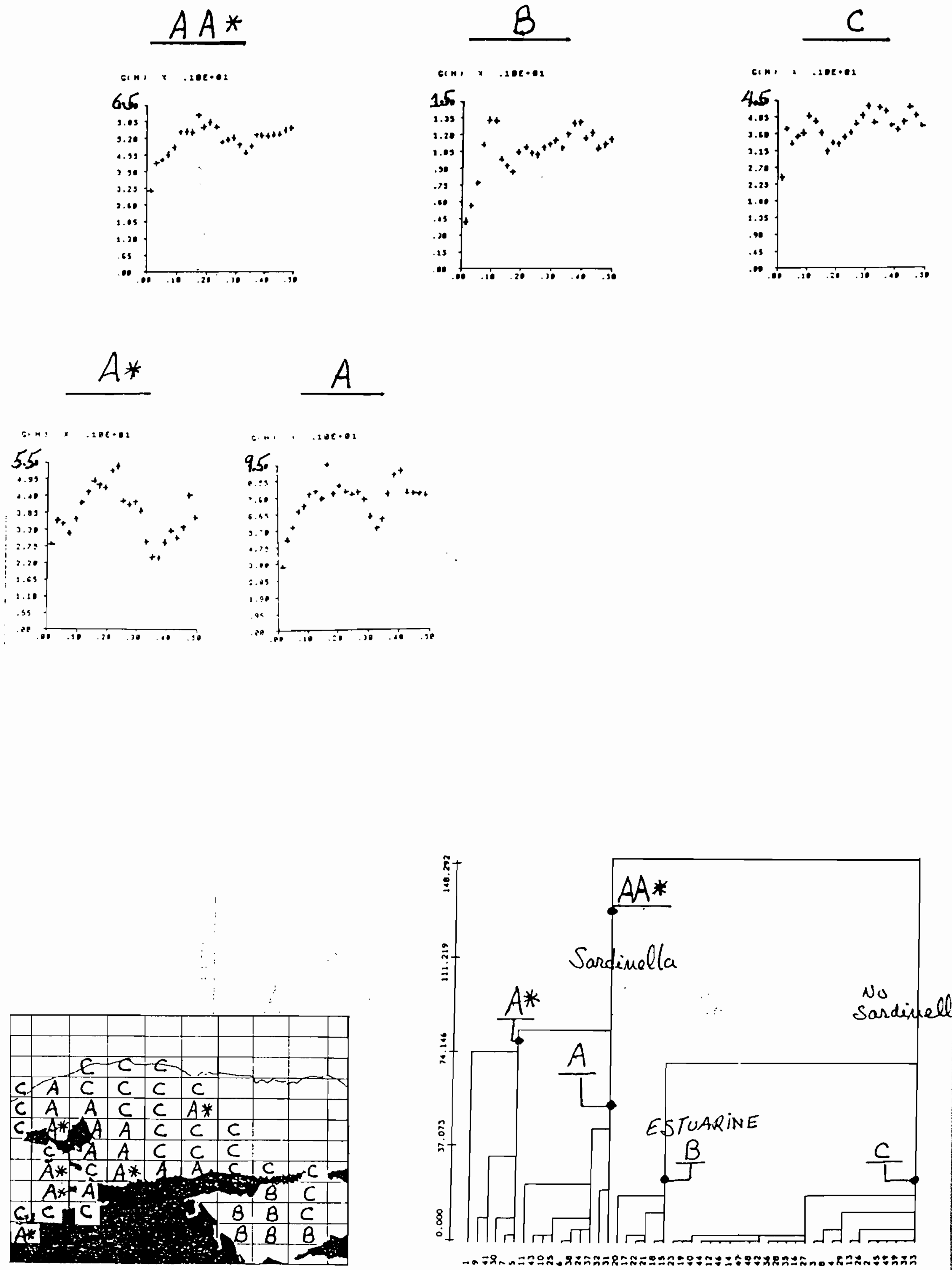
Simard and Gerlotto. FIG. 14.



Simard and Gerlotto. FIG. 15.



Simard and Gerlotto. FIG. 16.



Simard and Gerlotto. FIG.