ICES 1989

Digitalization sponsored by Thünen-Institut

C.M. 1989/K:37 Sess U

APPLYING GEOSTATISTICS TO THE ESTIMATION OF A POPULATION OF BIVALVES

by

M. Armstrong*, D. Renard* and P. Berthou**

* Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau
**Direction des Ressources Vivantes, IFREMER, Brest, France

ABSTRACT

This case study was carried out on the measurements of the number and biomass of a certain type of bivalve, the spisule *Spisula ovalis*, which is found in the sand in shallow waters (in this case from 6 to 17m in depth) near the Ile d'Yeu on the Atlantic coast of France. The sampling procedure consisted of finding the limits of the sand bank and of estimating the overall biomass from random sampling using a Hamon grab. This takes the sediments in an area of $0.25m^2$ and up to 20 to 25 cms in depth. The sediment was then sieved to recover the bivalves which were measured, weighed and classed according to age.

Geostatistics is a set of spatial statistical techniques that were originally developed for estimating and simulating ore grades in the mining industry. Nowadays it is used in many other fields ranging from the oil industry to soil science, forestry and agriculture.

The aim of this study is to show how geostatistics can be used to quantify the characteristics of the spatial distribution of the spisules and to estimate the total quantity. The geostatistical technique, kriging, was used to estimate the insitu reserves. Both variables studied (biomass and number of spisules) show a large class of zero values (27 values out of 67). This raises one important question: should the zeros be included in the study or not? Or do they merely serve to delimit the periphery of the area? In order to answer this question the kriged estimates were made with and without the zeros.

In addition to kriging the variables, some more advanced geostatistical techniques including disjunctive kriging and conditional simulations were used to assess the recoverable reserves, that is, to estimate the quantity above a cutoff value.

GEOSTATISTICAL CASE STUDY ON SPISULES AT THE ILE D'YEU

By M. Armstrong*, D. Renard* and P. Berthou**

OBJECTIVES

The primary objectives of this study are twofold: firstly, to show how geostatistics can be used to quantify the characteristics of the spatial distribution of a certain type of bivalve, the spisule *Spisula ovalis*, which is found in the sand in shallow waters (in this case from 6 to 17m in depth) near the Ile d'Yeu on the Atlantic coast of France, and secondly to use kriging to estimate the total quantity in the area where the shells exist. In addition to this, some more advanced geostatistical techniques, disjunctive kriging and conditional simulations, were used to assess the recoverable reserves. It is assumed that the readers are familiar with geostatistics (e.g. variogram analysis and kriging). These are presented in the available textbooks such as Journel and Huijbregts (1978), and David (1977).

INTRODUCTION

This case study was carried out on the measurements of the number of spisules and their biomass, made during April 1988. The sampling procedure consisted of finding the limits of the sand bank and of estimating the overall biomass from random sampling using a Hamon grab. This takes the sediments in an area of $0.25m^2$ and up to 20 to 25 cms in depth. The sediment was then sieved to recover the bivalves which were measured, weighed and classed according to age.

Figures 1 and 2 show the number and biomass of spisules found at each of the 67 sample locations. These two figures show that the molluscs exist in a clearly defined area (the region where nonzero values are found) and that further outward there are no more.

The histograms of the data are shown in Figures 3 and 4. Both show a large class of small values most of which are strict zeros (27 values out of 67). Figures 3b and 4b present the histograms after eliminating the zero values. This raises one important question: should the zeros be included in the study or not? Or do they merely serve to delimit the periphery of the area? In order to answer this question the study was carried in duplicate, with and without the zeros.

BASIC STATISTICS

The basic statistics (mean, variance and correlation coefficient) were calculated for both variables with and without the zero values. Table 1. The mean values are much lower when the zeros are included as is the variance. So the choice of whether to include them in the calculation of the total resources is crucial. The scatter diagram (Figure 5) of biomass against number of molluscs is typical of highly correlated variables (here the correlation coefficient is 0.95).

* Centre de Geostatistique, Ecole des Mines de Paris, 77305 Fontainebleau, France
** Direction des Ressources Vivantes, IFREMER, Plouzane, Brest, France

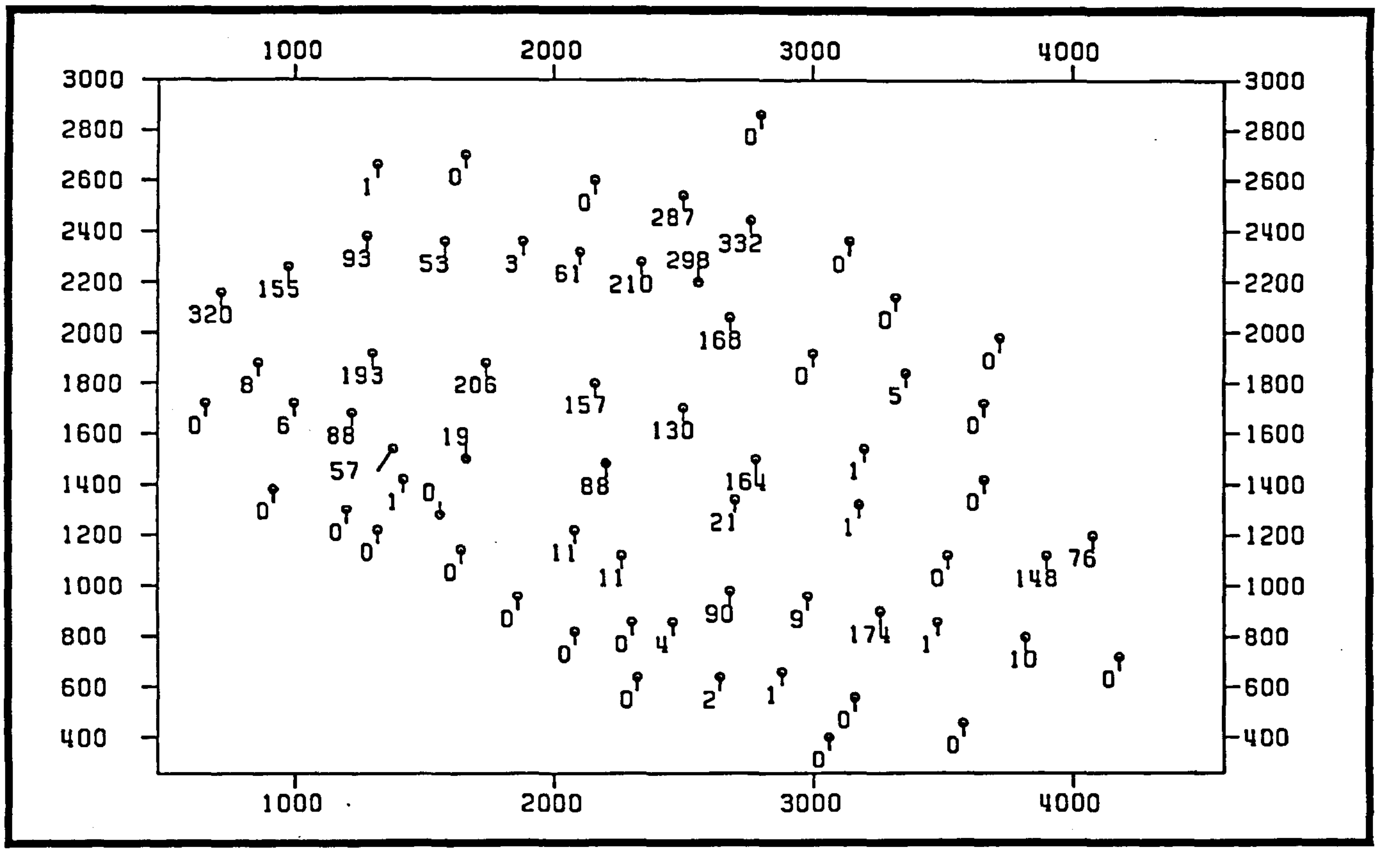


Figure 1. Number of Spisules at each location

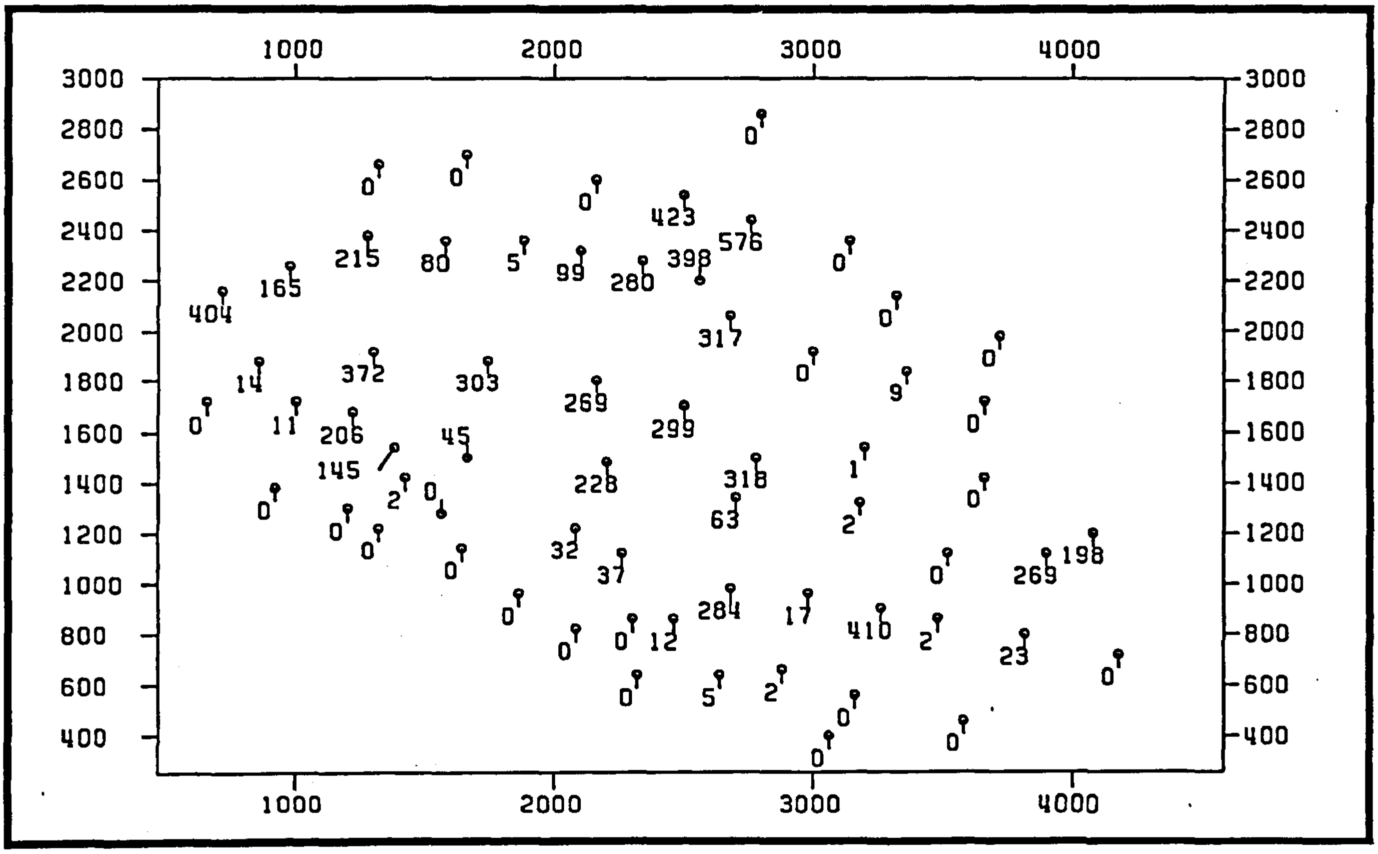


Figure 2. Biomass at each location.

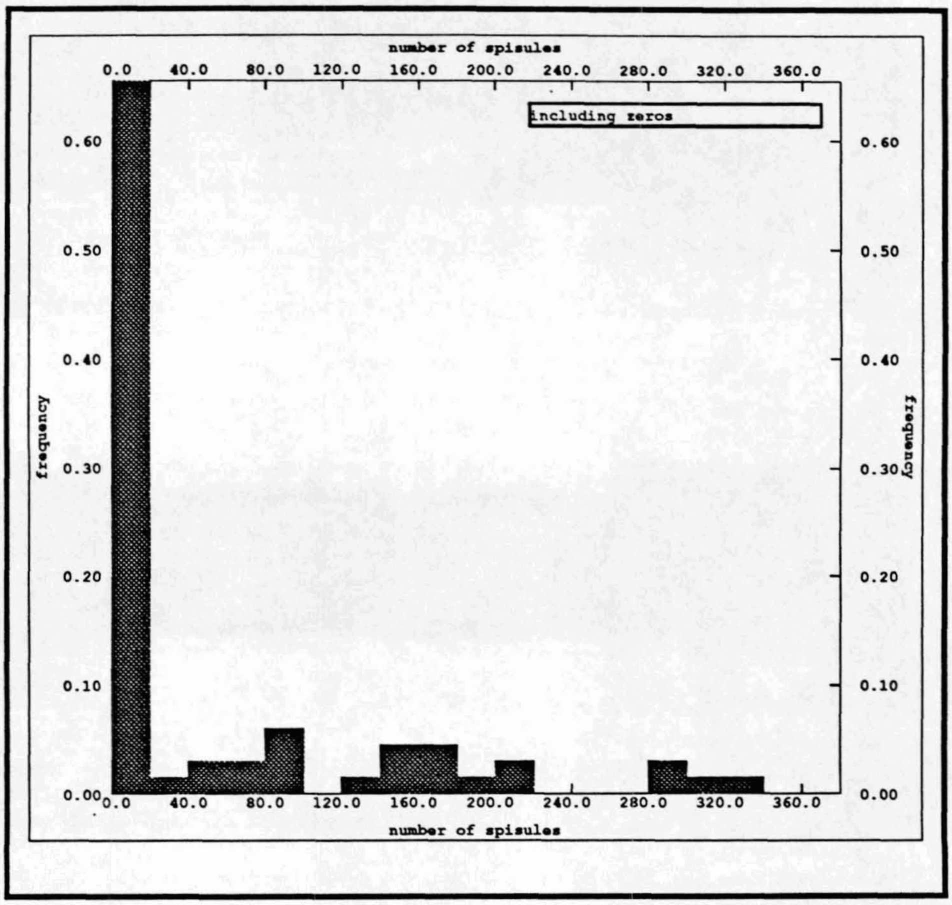


Figure 3a. Histogram of Number of Spisules including zeros

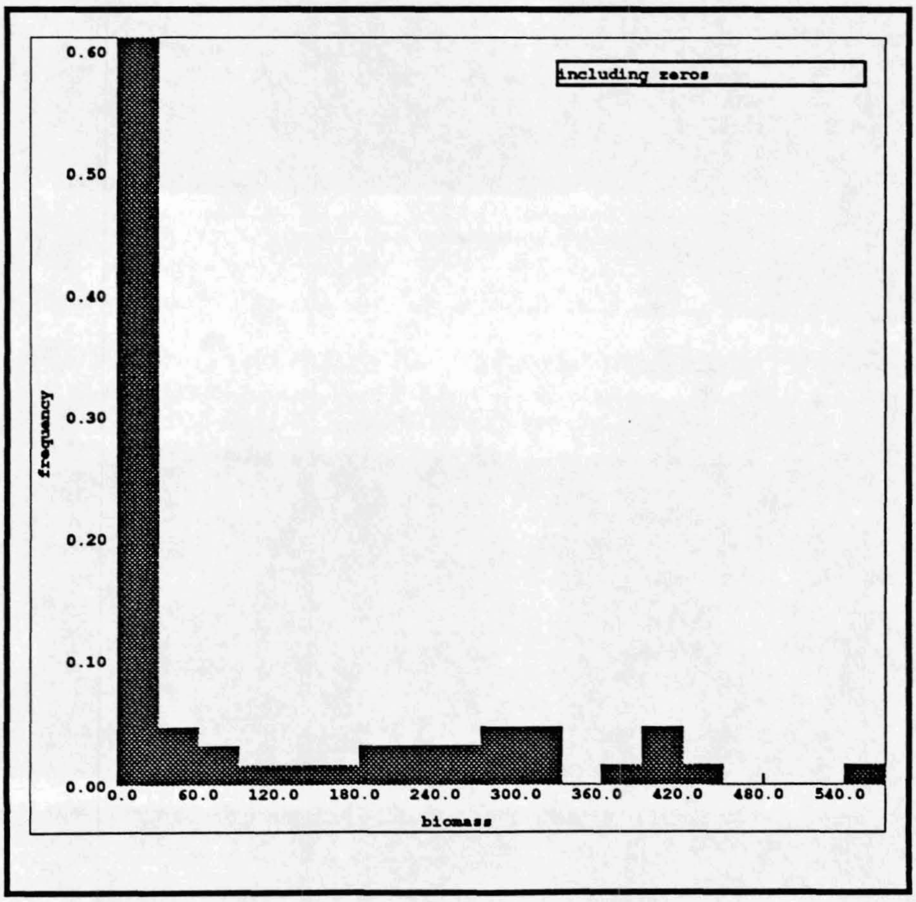


Figure 4a. Histogram of Biomass including zeros

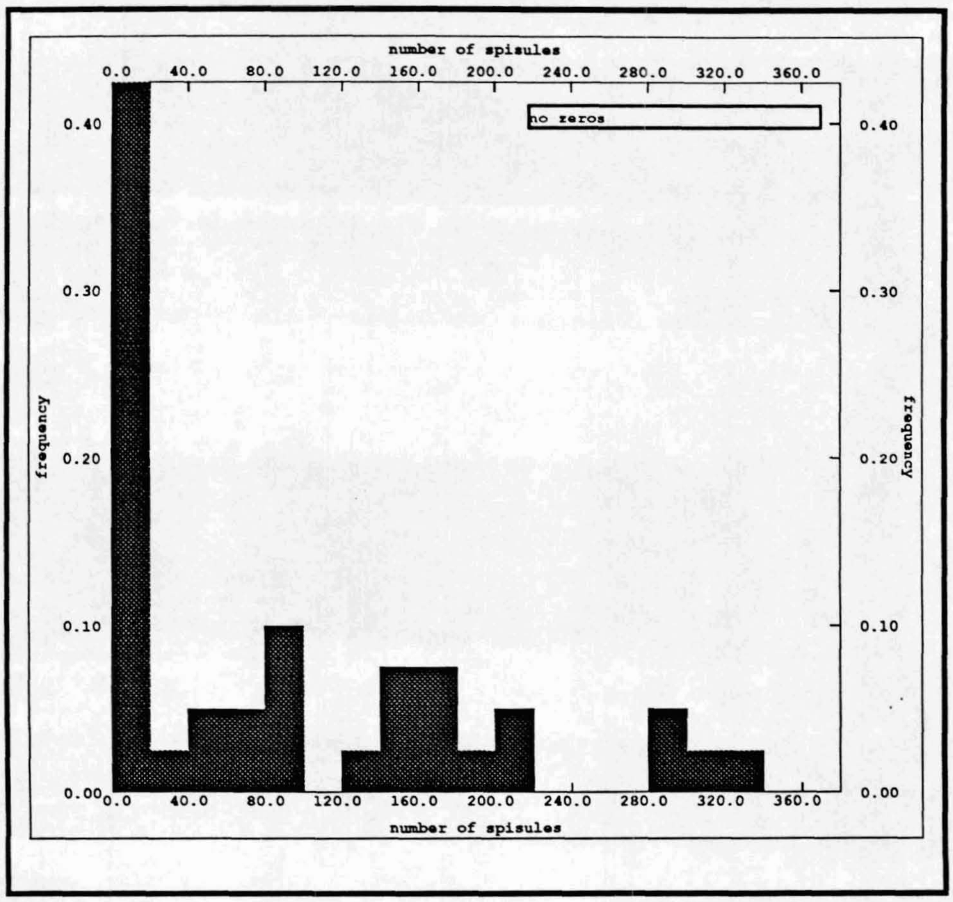


Figure 3b. Histogram of Number of Spisules excluding zeros

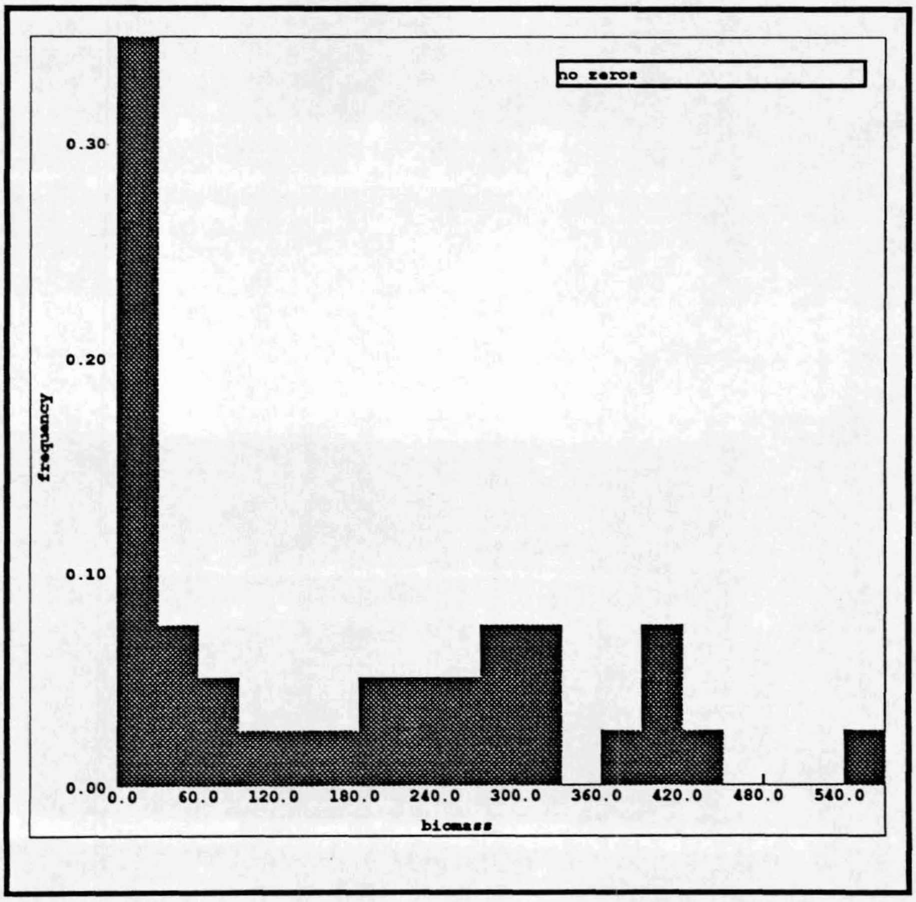


Figure 4b. Histogram of Biomass excluding zeros

Basic Statistics	Mean Value	Variance
Biomass (including zeros)	97.66	21501.00
Number of Molluscs (including zeros)	54.69	7899.44
Biomass (excluding zeros)	163.58	25232.55
Number of Molluscs (excluding zeros)	91.58	9854.85

Table 1: Basic Statistics (N.B. the measurements refer to an area of 0.25m2).

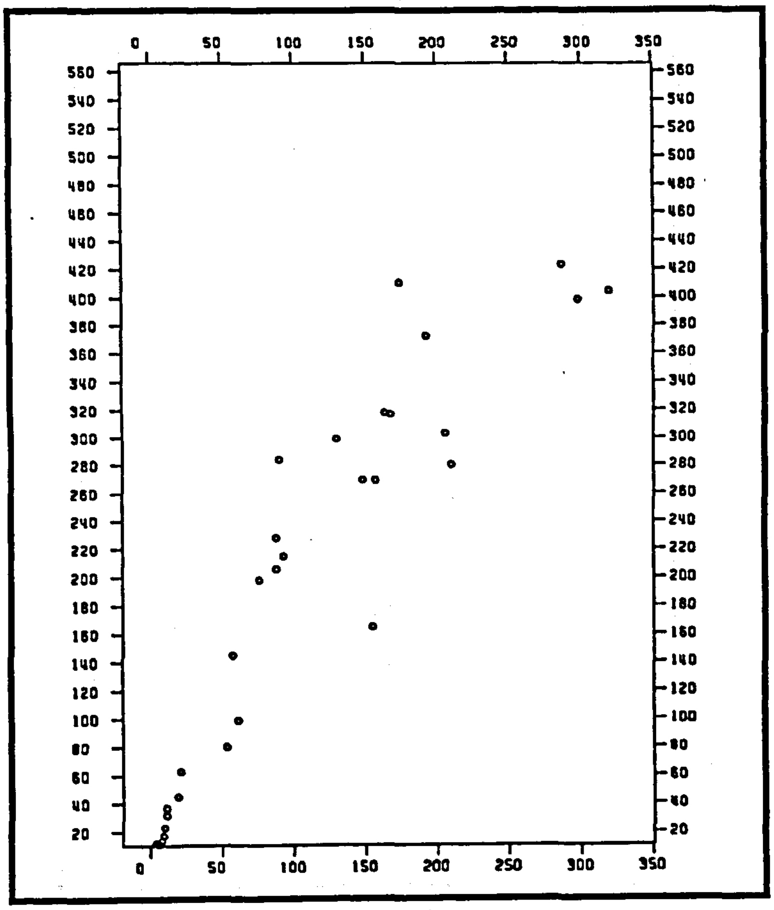


Figure 5. Scatter Diagram of Biomass (on the Y axis) versus Number of Spisules (on the X axis)

VARIOGRAM ANALYSIS

The basic tool in geostatistics for quantifying the spatial correlation between samples a certain distance h apart is the variogram. The experimental variograms were calculated for both variables with and without the zeros. These were calculated for a lag of 400m in 2 perpendicular directions 60^0 N of E and 60^0 W of N since these are the directions along and across the area. Given the limited number of samples available it would be illusory to calculate variograms in four directions (which is the standard procedure). Figures 6, 7 8 and 9.

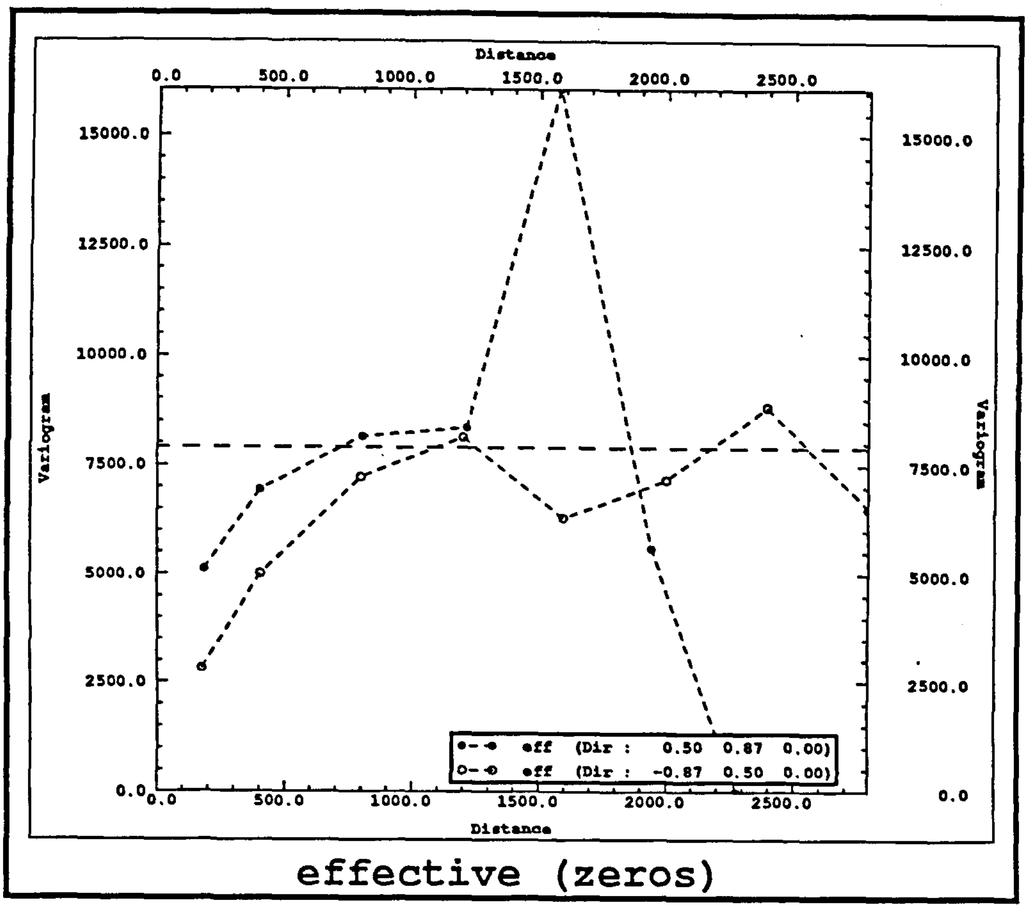


Figure 6. Experimental Variograms of the Number of Spisules including zeros.

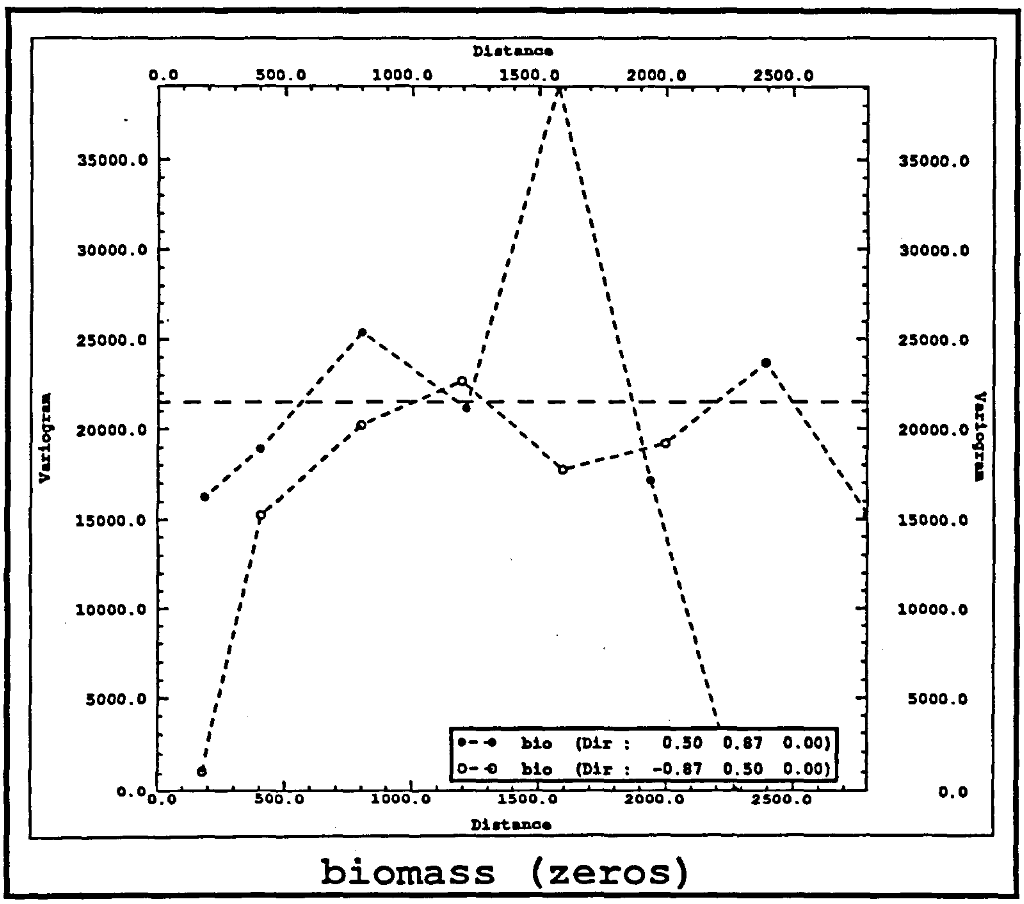


Figure 7. Experimental Variograms of the Biomass including zeros.

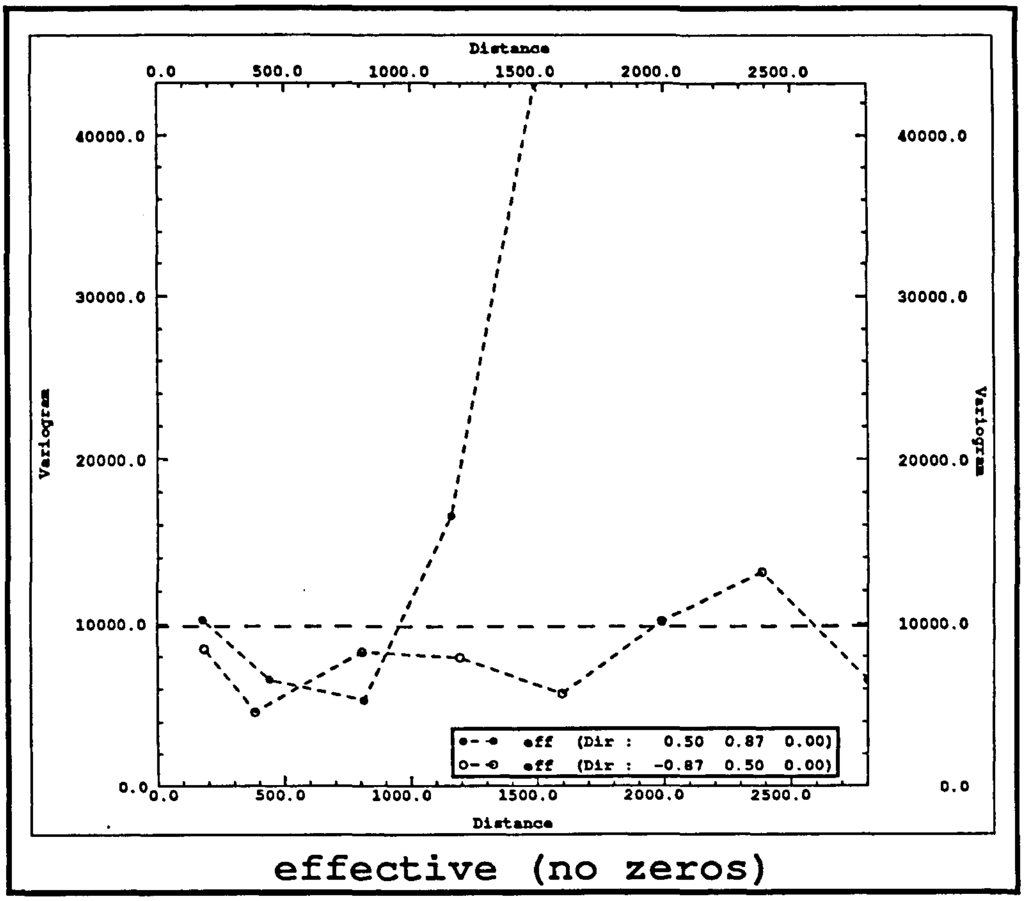


Figure 8. Experimental Variograms of the Number of Spisules excluding zeros.

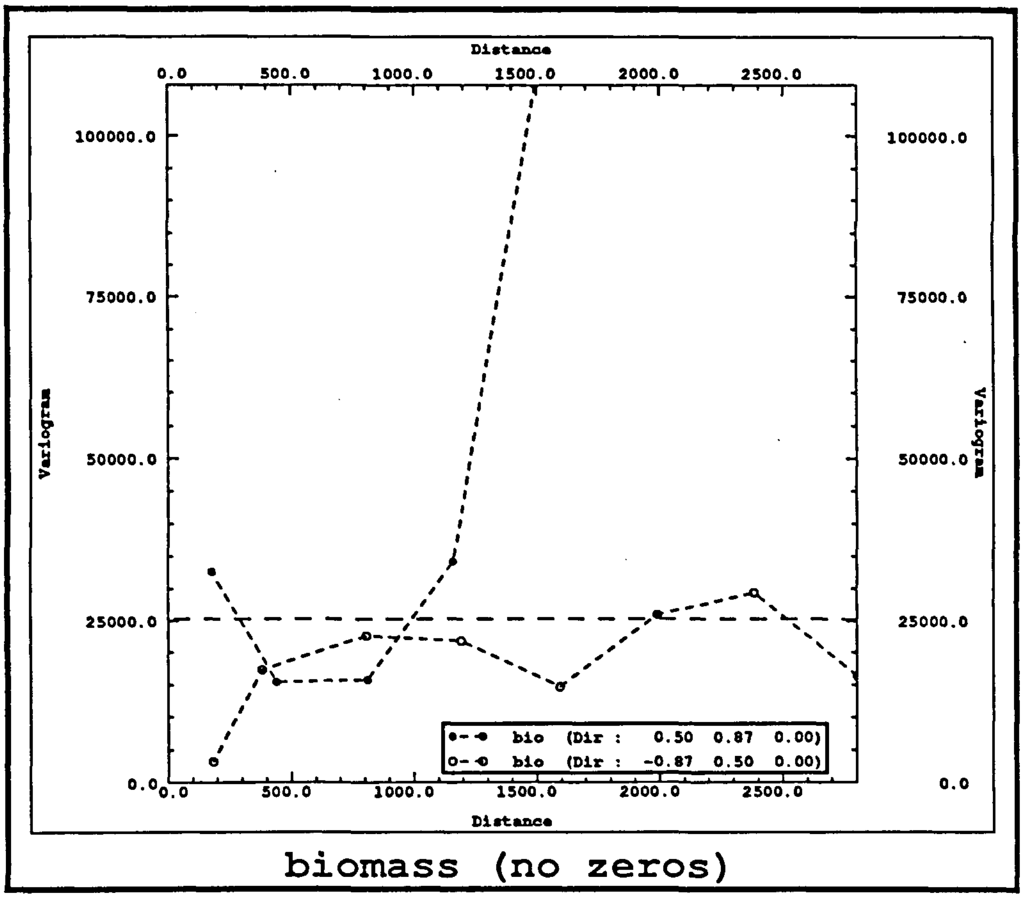


Figure 9. Experimental Variograms of the Biomass excluding zeros.

Variograms including the zeros:

The experimental variograms for the number of shells and for their biomass including zeros (Figures 6 and 7) behave differently for the two directions. They rise more rapidly in the direction across the deposit than along it. The variograms in the direction across the deposit become erratic for distances above 1200m. Since this is more than half the diameter across the deposit, this effect need not be taken into account when fitting a mathematical model to the variogram. Consequently the variogram model will be chosen to rise up from 0 at a distance h =0 up to a limit value called the sill which will be taken to be the value of the variance (shown by a dotted line). So the variables are assumed to be statistically stationary. Exponential variograms models were fitted in both cases. A model of geometric anisotropy was used to take account of the directional differences. The long axis of the ellipse of anisotropy was in the direction 600 W of N and the ratio between the long axis and the small one is 2.2 for both variables. The parameters of these are shown in Table 2.

	Number of Shells	Biomass
Nugget Effect C_0	0.	0.
Sill C_1	8000.	22000.
Scale parameter a	440m	400m

Table 2: Parameters for the Exponential Models for the case where zero values are included.

These models are not the only ones that could be fitted to the experimental variograms. To illustrate this a second model consisting of the sum of two sphericals with sills of 10500. and 11000. and ranges of 500m and 1250m, and a zero nugget effect gave an equally good fit for biomass. The anisotropy is the same as before. Figures 10, 11a and b show these three fitted models. As these two models for the biomass variogram are almost identical when plotted (in particular their behaviour near the origin is linear with no nugget effect) they will give virtually identical results in any subsequent use.

Variograms Excluding the zeros:

The experimental variograms for both variables when zeros are excluded (Figures 8 and 9) show two important features:– a sharp rise after 1000m in the direction across the sandbank and a rather large value for the first variogram class in some cases.

The first effect is due to the presence of a drift or trend in the values. So the variable is not stationary in this direction for distances above 1000m but this is not important in practice because points more than 1000m away will not usually be taken into account during the subsequent estimation procedure or if they are used, they will have a very low weighting factor.

The second effect (the high value for the first class) is due to the very small numbers of pairs of points (only 1!) this distance apart. Table 3 below shows the numbers of pairs of points used to calculate the variogram for each distance class for both directions. These figures show that whereas the variogram along the length of the area is meaningful up to the sixth class (up to 2000m) the other one is not. In particular in both cases the first distance class only contains one couple and so this variogram point need not be taken into consideration when fitting the model.

After eliminating this variogram point we are faced with the problem of choosing the value for the nugget effect. It would have been helpful to have had some additional closely spaced samples (e.g. in the form of a cross) to aid in choosing the nugget effect. These samples should of course be located in a typical area and not preferentially in a rich or poor one.

After taking account of these points, a single isotropic exponential variogram model with no nugget effect was fitted in both cases. The parameter values are shown below in Table 4. As before the choice of a model with a sill indicates that we assume stationarity. Figures 12a and 13 show the fitted models.

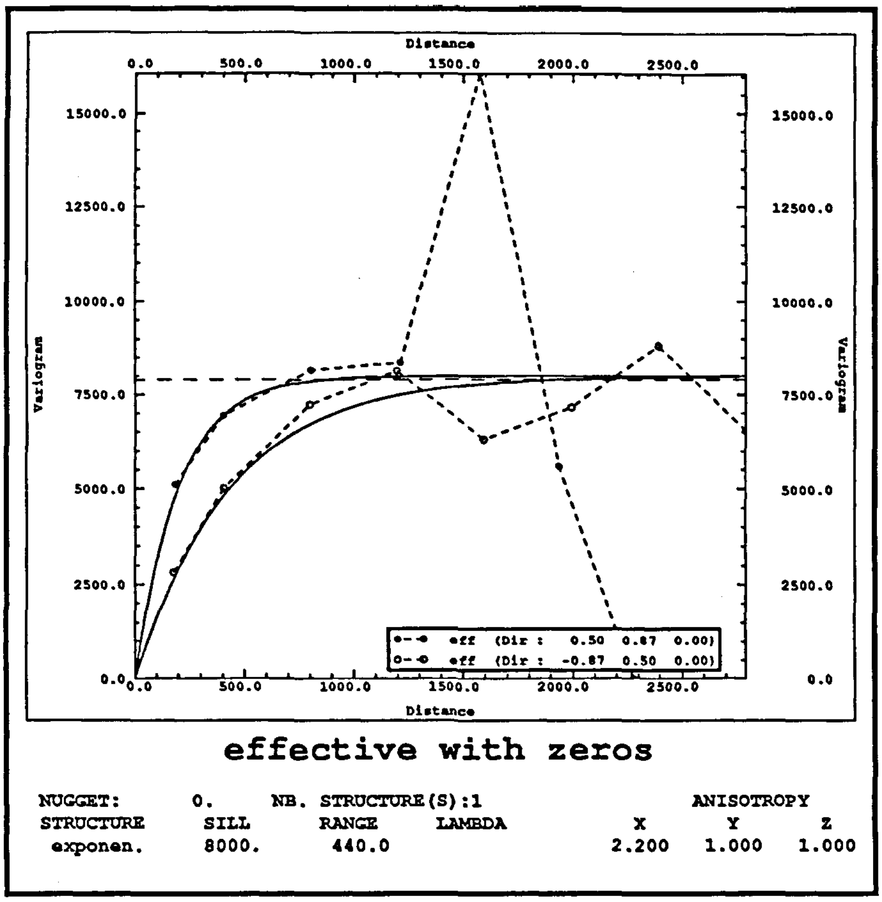


Figure 10. Fitted Exponential Variogram Model for the Number of Spisules including zeros.

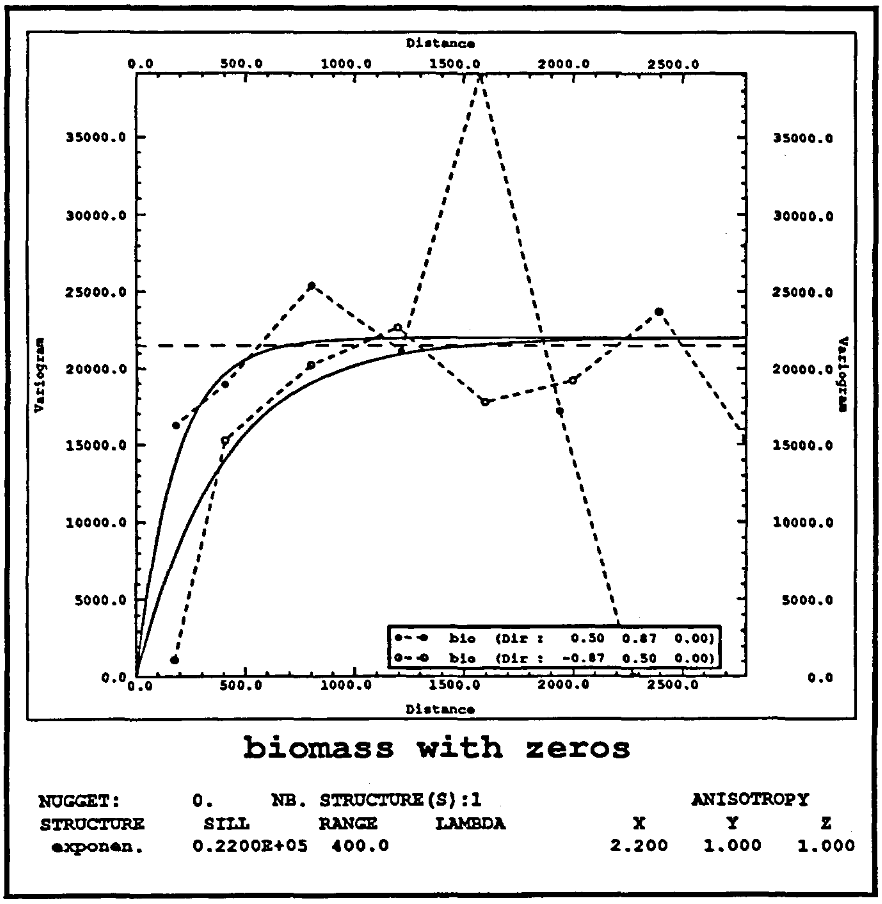
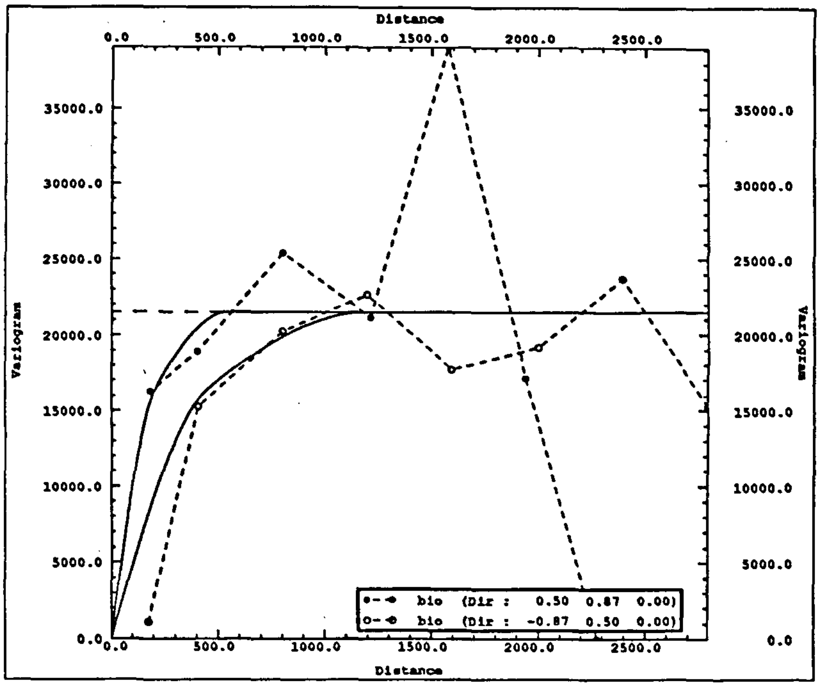


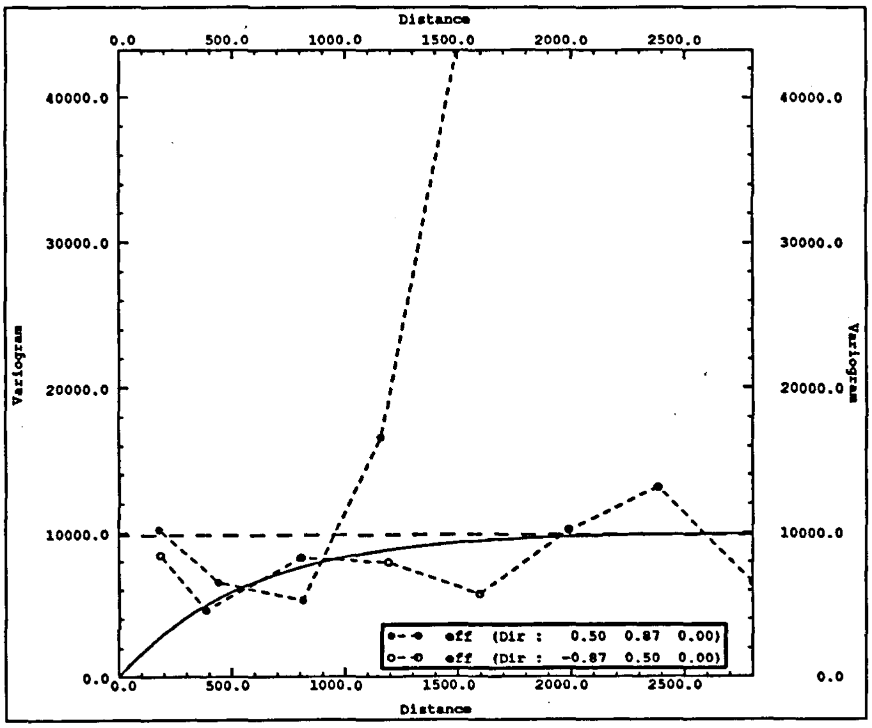
Figure 11a. Fitted Exponential Variogram Model for Biomass including zeros.



biomas - spherical model

NUGGET:	0.	NB. STRUCTURE(S):2		ANISOTROPY		
STRUCTURE	SILL	RANGE	LAMBDA	X	Y	Z
spheric.	0.1050E+05	450.0		2.200	1.000	1.000
spheric.	0.1100E+05	1200.		2.200	1.000	1.000

Figure 11b. Fitted Spherical Variogram Model for Biomass including zeros.



effective without zeros

NUGGET:	0.	NB. STRUCTURE(S):1		ANISOTROPY		
STRUCTURE	SILL	RANGE	LAMBDA	X	Y	Z
exponen.	0.1000E+05	560.0		1.000	1.000	1.000

Figure 12a. Fitted Exponential Variogram Model for the Number of Spisules excluding zeros.

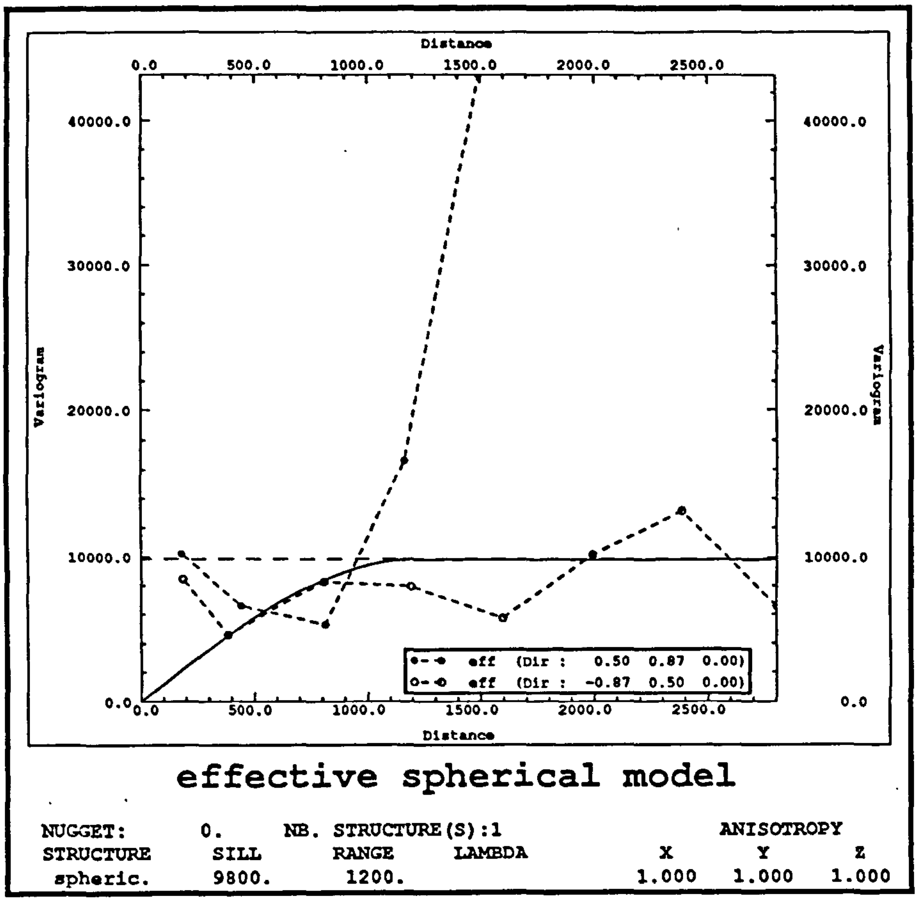


Figure 12b. Fitted Spherical Variogram Model for the Number of Spisules excluding zeros.

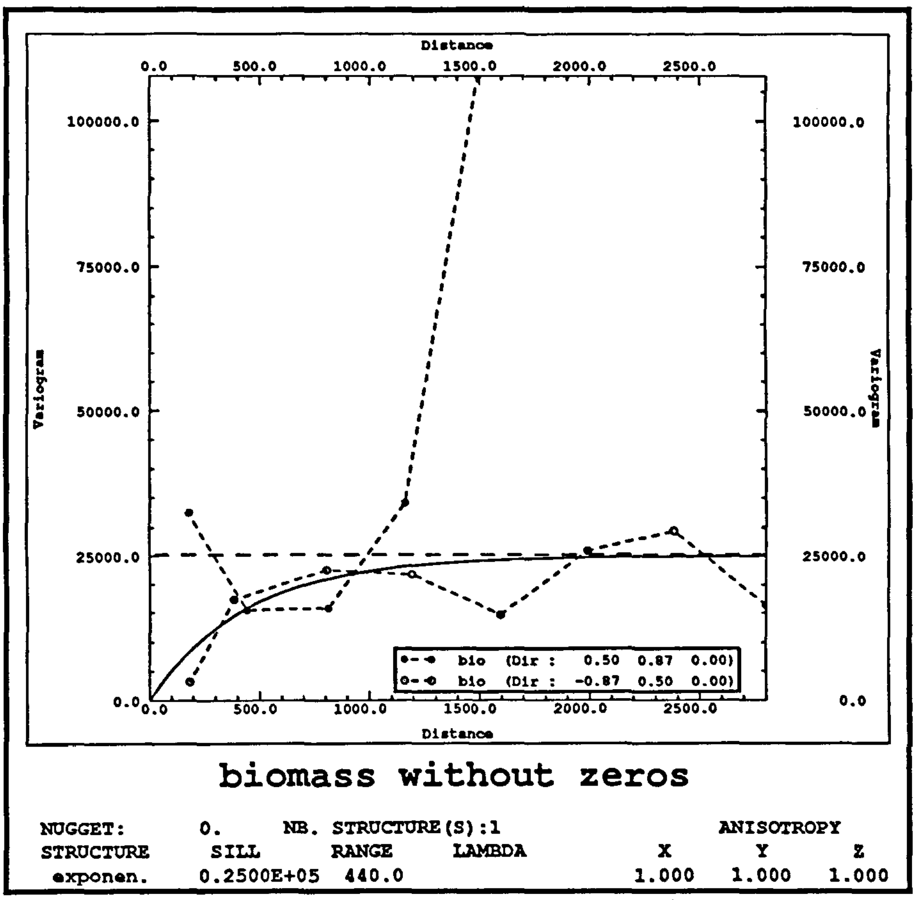


Figure 13. Fitted Exponential Variogram Model for Biomass excluding zeros.

Distance class	Across the breadth	Along the length
1	1	1
2	34	46
3	39	66
4	33	69
5	12	80
6	1	76
7	0	51
8	0	26

Table 3: Number of pairs of points used in variogram calculations

As before, these exponential models are not the only ones the can be fitted to the experimental variograms. For comparison purposes a second model consisting of a spherical model with a range of 1200m and a sill of 9800, and a zero nugget effect was fitted for the number of spisules (Figure 12b).

	Number of Shells	Biomass
Nugget Effect C_0	0.	0.
Sill C_1	10000.	22000.
Scale parameter a	560m	440m

Table 4: Parameters for the Exponential Variogram Model for the case where the zeros are excluded.

GLOBAL ESTIMATION

The next step is to estimate the total resources in this area. This means defining the limits of that area. Two different polygons were used for this. Figure 14 shows the two contours and also the samples (represented by a 0 when there were no spisules and by a + when there were). One contour is smoother and covers a larger area than the other (238 x 10^5 m2 compared to 176 x 10^5 m2), a decrease of 25.% . Clearly the difference in the surface ares will be reflected in estimated total quantities. The change in the contour will also affect the estimate of the mean because the area eliminated is the low grade border zone. Consequently the mean for the larger area will be lower than for the other.

The estimation technique, kriging, uses a weighted linear combination of the data values. Details of the method can be found in any geostatistical textbook. The estimates of the global reserves were made using the kriging program BLUEPACK working in a unique neighbourhood (because of the small number of points) and the exponential variogram models shown earlier in the tables. Tables 5 and 6 give the results. One advantage of geostatistics over other estimation methods is that it gives the estimation variance as well as the estimated value. As an example the corresponding standard deviation (i.e. the square of this variance) was 13.3 for a mean of 165.46 for the biomass for the case where the positive values were used inside the small polygon.

This leads to four sets of estimates for each variable. Which is right? Two sets of results can be eliminated immediately – those for the positive data inside the large polygon and similarly for all data inside the small one. In the first of these the large polygon covers many of the zero samples, and so removing these values before doing the estimates artificially extends the zone of influence of the positive ones. Clearly this is far too optimistic. The second case (kriging inside the small polygon which encloses the positive zone but including the zero values outside this in the estimation procedure) is less pernicious. The presence of the zeros outside tends to pull the estimates downwards along the boundary, which could be helpful.

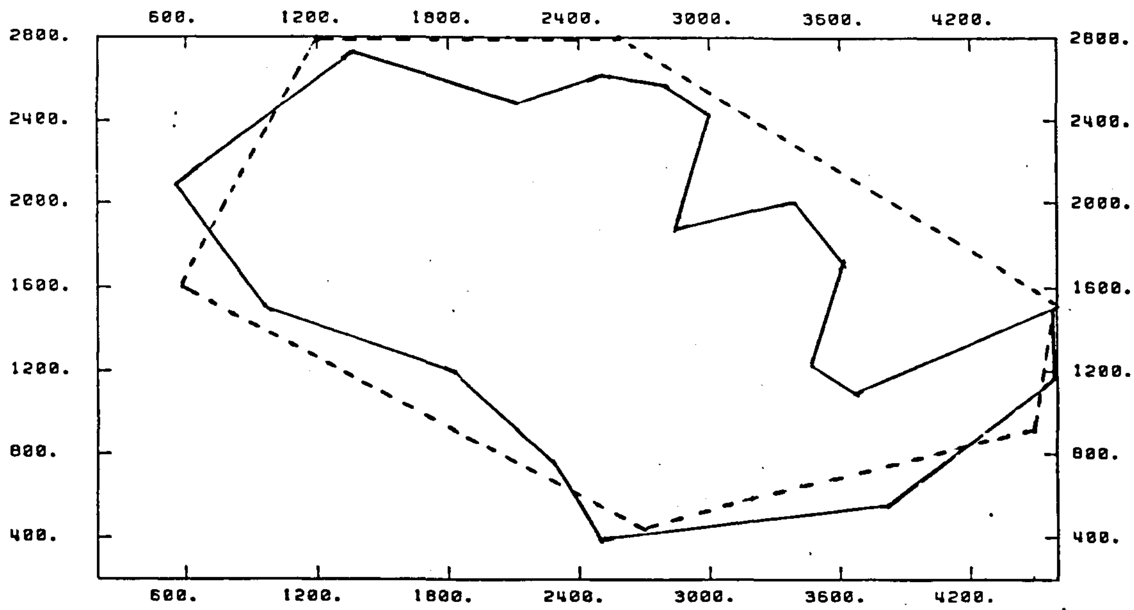


Figure 14. Shapes of the two polygonal contours used for kriging

NUMBER OF SHELLS.	Mean Value	Total Reserves
Large polygon, all data	65.0	1550×10^6
Large polygon, positive data	85.9	2050×10^6
Small polygon, all data	78.7	1380×10^6
Small polygon, positive data	90.9	1600×10^6

Table 5: Estimated global reserves: the number of individuals (N.B. the mean is over an area of $0.25m^2$)

BIOMASS	Mean Value	Total Reserves
Large polygon, all data	118.8	2830 tons
Large polygon, positive data	157.8	4560 tons
Small polygon, all data	142.31	2500 tons
Small polygon, positive data	165.46	2910 tons

Table 6: Estimated global reserves – the biomass. (N.B. the mean is over an area of $0.25m^2$)

The remaining two sets of estimates are intellectually consistent, and although the mean values are quite different the total reserves are similar (less than 3.5% difference). This shows a geostatistical sort of "conservation of matter". The total quantity of matter remains the same but is distributed differently in the two cases. When all the data are considered inside the large polygon the "estimated surface" is flatter on the edges and more peaked in the centre whereas in the other case there are "cliffs" along the border line between the area with molluscs and the outside. To illustrate this- the kriged estimates of the number of spisules were obtained at the nodes of a regular grid inside the appropriate polygon for the cases where the zeros are included and excluded respectively (Figures 15, 16a and b show these for 8 classes of values ranging from -50 to 350 by steps of 50. The negative estimates are horizontally striped whereas the positive ones are vertically striped. Figure 16a and 16b show two sets of estimates corresponding to the two different fitted models (exponential and spherical). The differences are very slight except along the edge. Please note that these three figures show the kriged estimates for point values. It is also possible to obtain estimates of the average values over blocks. These are of course smoother.

In many ways the choice between these two sets of results is more one of choosing between two different interpretations of the variable. If one believes that biological factors have lead to a barrier between the zone containing the molluscs and the exterieur, then the second case is appropriate. On the other hand if one believes that the number of molluscs merely tapers off, then the first case should be chosen. This choice cannot be made on geostatistical grounds; the biologist must make this choice from his scientific knowledge of the area and the mollusc. In this case because of the sedimentological characteristics of the area and the habitat of this species, it seems more appropriate to use only the positive values inside the small polygon.

Both geostatistical reserve estimates (2830 ± 640 tons and 2910 ± 470 tons) compare favourably with the estimate obtained using classical methods (2600 tons ± 600 tons).

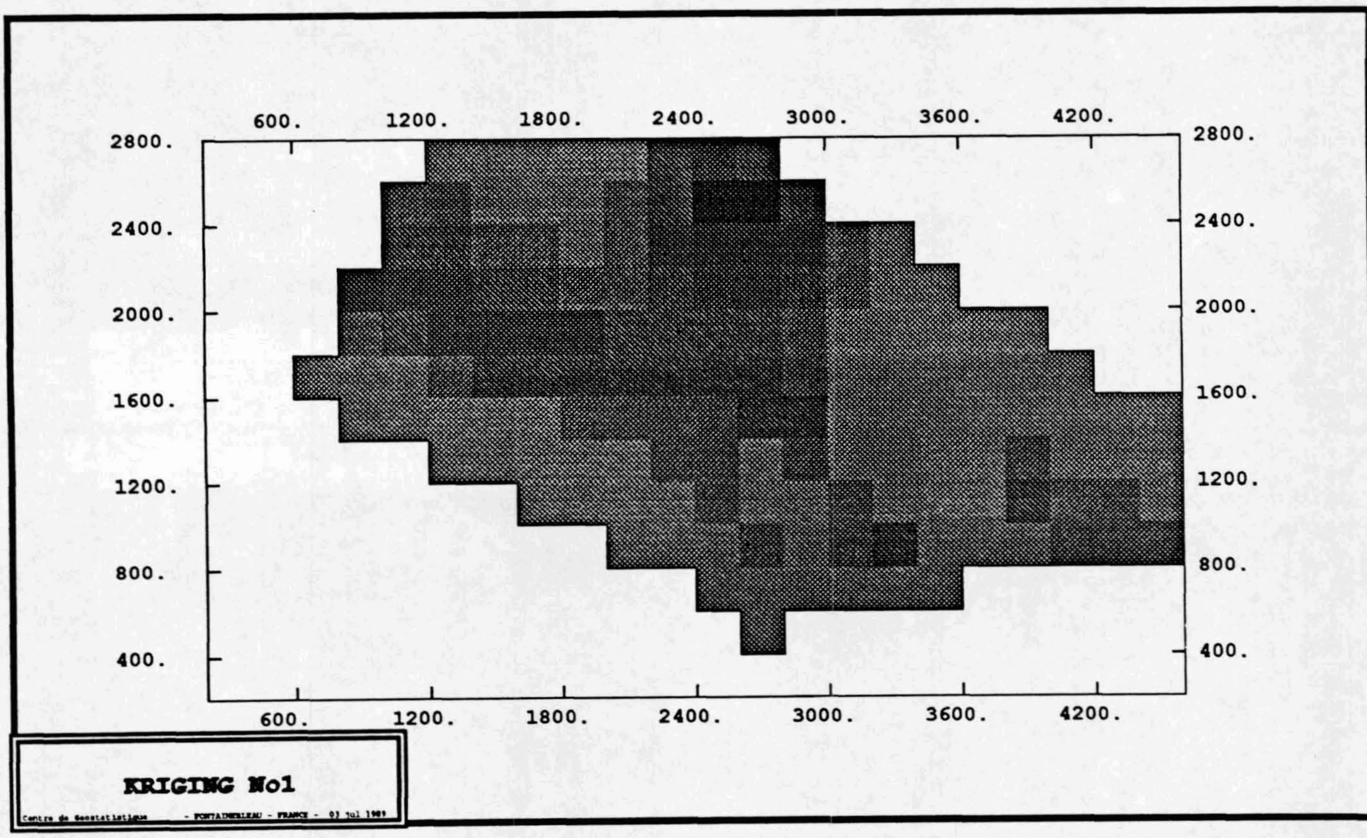


Figure 15: Kriged Results for the Number of Spisules inside the large polygon

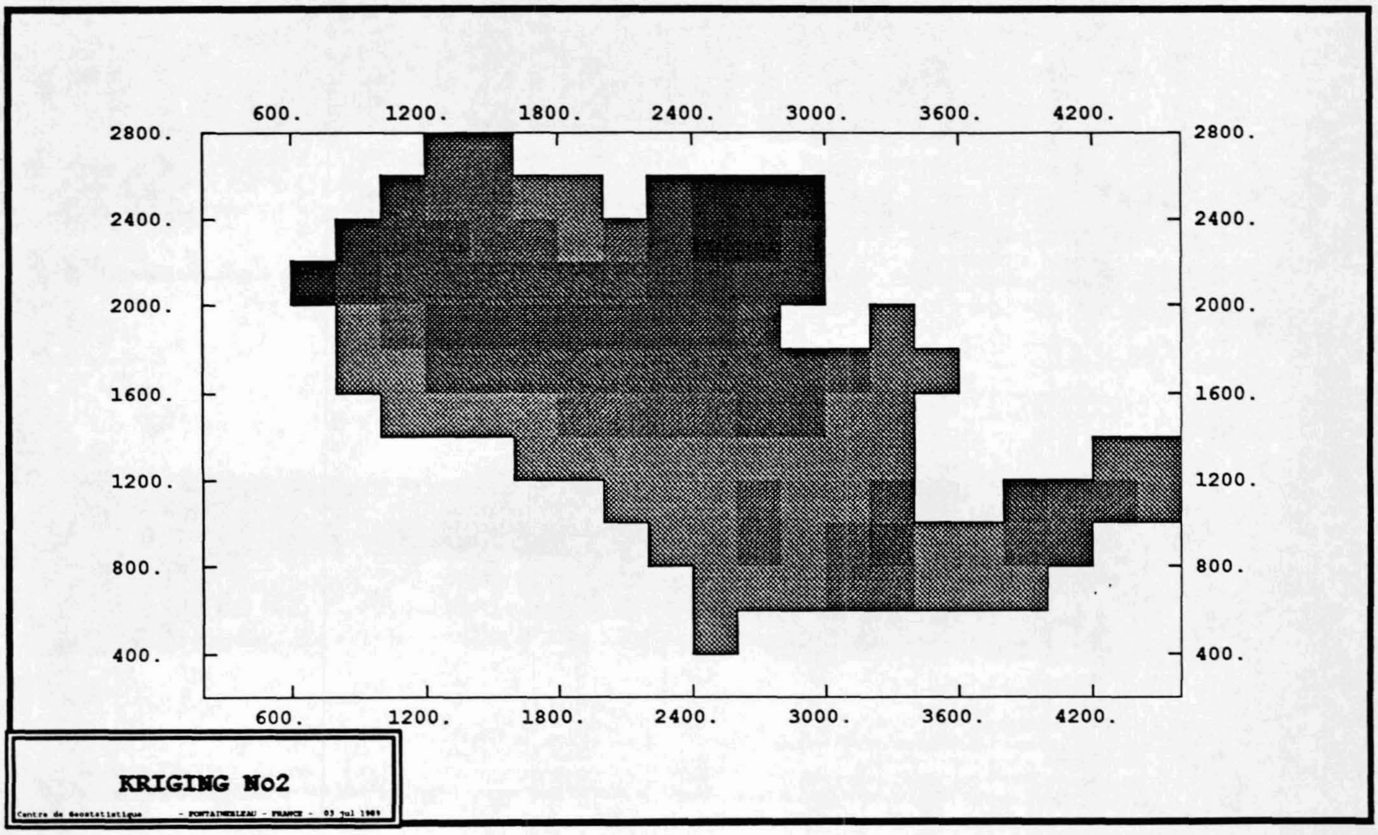


Figure 16a: Kriged Results for the Number of Spisules inside the small polygon using the exponential model

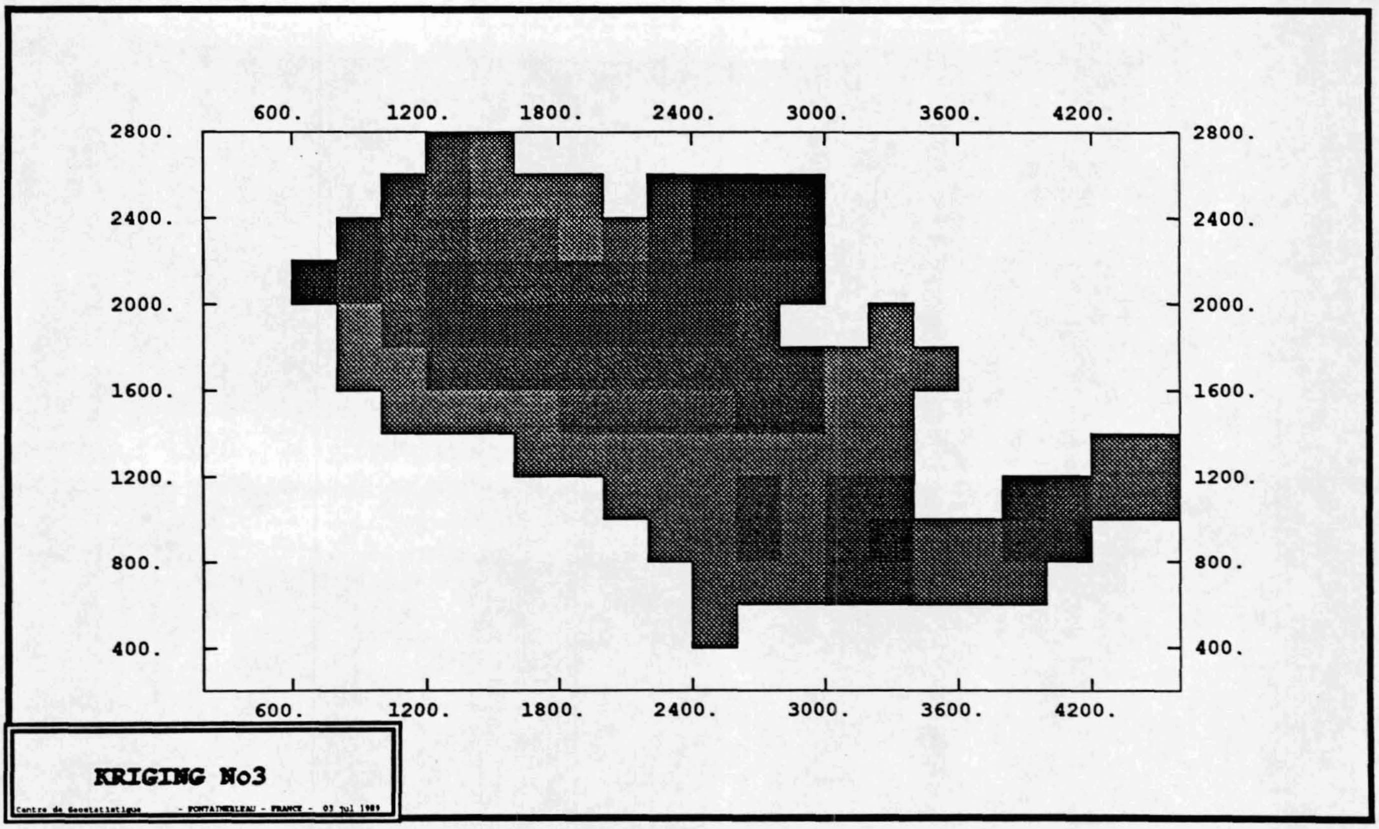


Figure 16b: Kriged Results for the Number of Spisules inside the small polygon using the spherical model

RECOVERABLE RESERVES

Having obtained estimates of the global insitu reserves and displayed the point estimates of the number of spisules, we can now proceed to calculate the quantity that can be recovered. By this we mean the number of blocks (exploitation areas) where the variable (number of spisules or biomass) exceeds a certain economically profitable cutoff. To reduce the workload, since the two variables are strongly correlated (0.95) we will only present the results for one of them, the number of spisules. The other would be very similar. Our objectives are to calculate the percentage of blocks of a given size (points, or 250m x 250m, or 500m x 500m for example) where the number of spisules is above the cutoff. We also want to know the cumulated quantity above the cutoff.

To illustrate this, suppose that we knew the true values of 10 blocks: 0, 1, 1, 3, 5, 10, 20, 50, 80, 150. Then if the cutoff is 15 (say), we recover 40% of the blocks with an average grade of (20 + 50 + 80 + 150)/4 = 75. Although only 40% of the area is exploited we nevertheless recover 300 out of a total of 320, i.e. 93.75% of the total. In mining geostatistics, the percentage of blocks recovered gives the percentage of ore recovered (called the tonnage), while the second corresponds to the metal recovered. For want of better names we will continue to use the terms "tonnage" and "metal" to designate these concepts.

When most people are asked to estimate the recoverable reserves, they naturally decide to estimate blocks of the appropriate size using linear kriging or some other estimation method and then to count up the tonnage and metal above the cutoff. Unfortunately the results are not accurate. Linear kriging (and other estimation methods) smooth the values too much and hence are not suitable for calculating recoverable reserves. Special techniques (such as disjunctive kriging and conditional simulations) are needed for this. The first step in both is to transform the initial distribution (an arbitrary one) to a standard normal one.

For mathematical convenience the transformation function (or anamorphosis as it is called) is expressed in terms of Hermites polynomials. In principal an infinite number of terms are required for the expansion. In practice we limit ourselves to a finite number usually 20 or 30. Figures 17a and 17b show the anamorphosis function for the number of spisules (excluding zeros), when 10, 20 or 30 terms are used. The first figure shows the curve within the range where the gaussian equivalent goes from -2. to +2. (corresponding to the central 90% of the normal distribution). The second figure extends the range much further. In theory the anamorphosis function must be nondecreasing (in order to back transform to the initial scale). In this case the skewness of the raw data distribution and the peak of identical small values make it difficult to get a good fit. Several tests using an expansion with 30 Hermite polynomials were carried out to check the quality of the fit. We calculated the theoretical histogram using the anamorphosis function and compared it to the experimental one based on the available 40 sample values. Figure 18. In addition two of the family of grade/tonnage curves were calculated theoretically and compared to their experimental equivalents. Figures 19a and 19b show the tonnage recovered (i.e. the area) and the metal recovered expressed as percentages of the total, as a function of the cutoff. This shows that the anamorphosis function fits well.

So we can now go on to calculate the conventional profit as a function of cutoff. This is defined as the quantity of metal above cutoff minus the tonnage above cutoff times the cutoff grade:

$$\text{Conventional Profit} = \text{Metal} - \text{Tonnage} \times \text{Cutoff}$$

In many ways this corresponds to the reserves that will actually be recovered by the fisherman. See Figure 21. In addition to this we calculated the metal vs tonnage curve for several different block sizes; firstly for points (i.e. the sample volume), then for blocks 250m x 250m, and finally for 500m x 500m blocks. From Figure 21, we see that if we exploit the richest 50% of the blocks (i.e. tonnage = 50%), we would recover 97%, or 92%, or 85% respectively of the total, depending on the size of the exploitation areas. This gives an indication of how much less selective the operation becomes as the block size is increased.

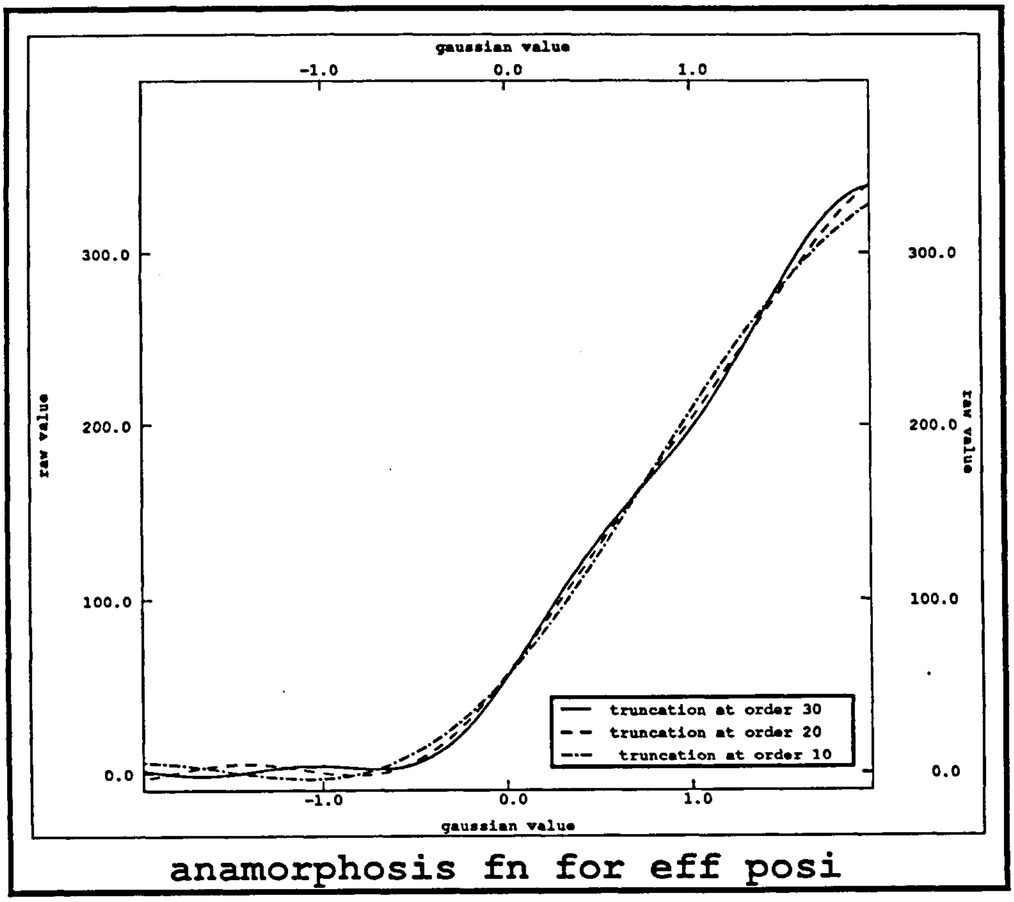


Figure 17a. Anamorphosis Function from -2.0 to +2.0

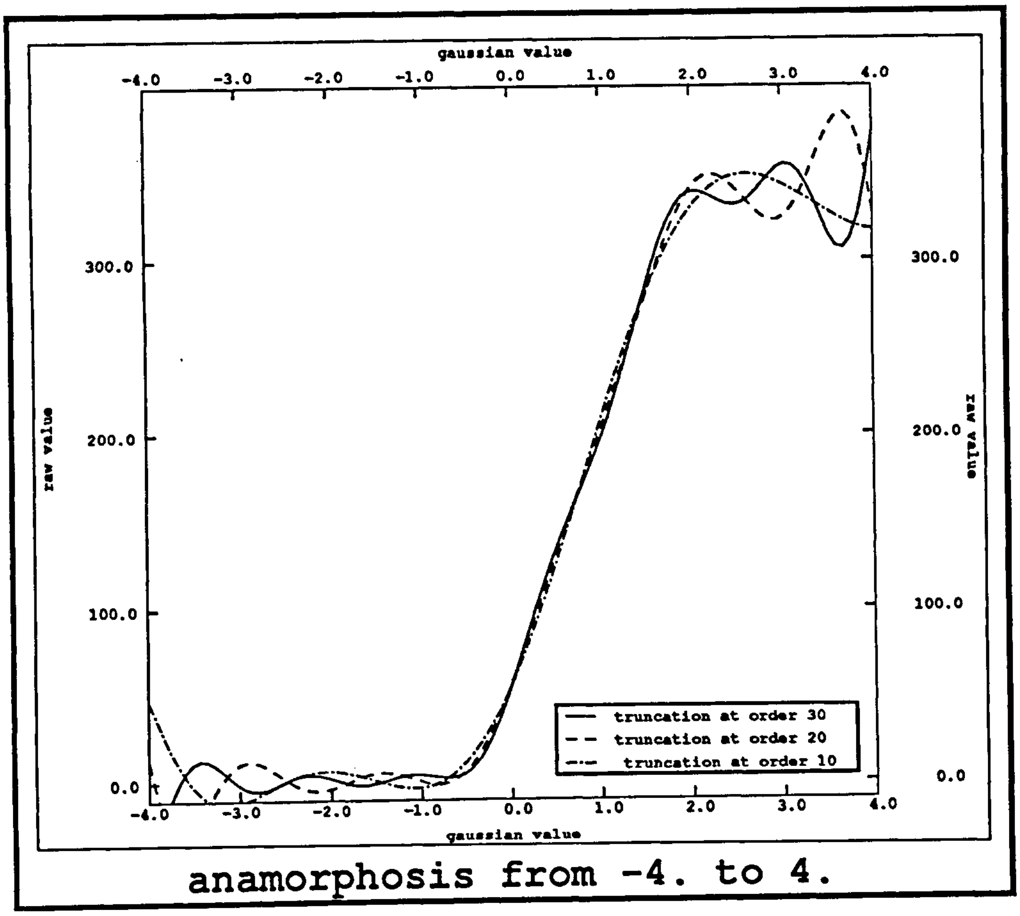


Figure 17b. Anamorphosis Function from -4.0 to +4.0

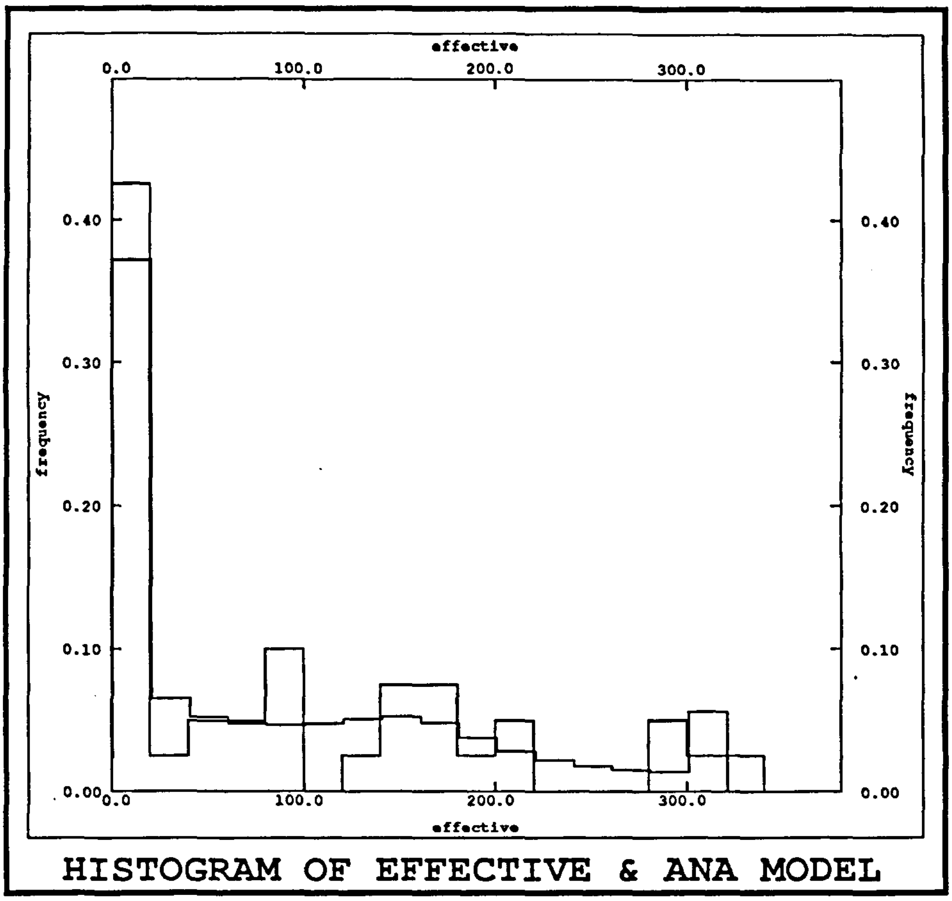


Figure 18. Comparison of Experimental and Theoretical Histograms

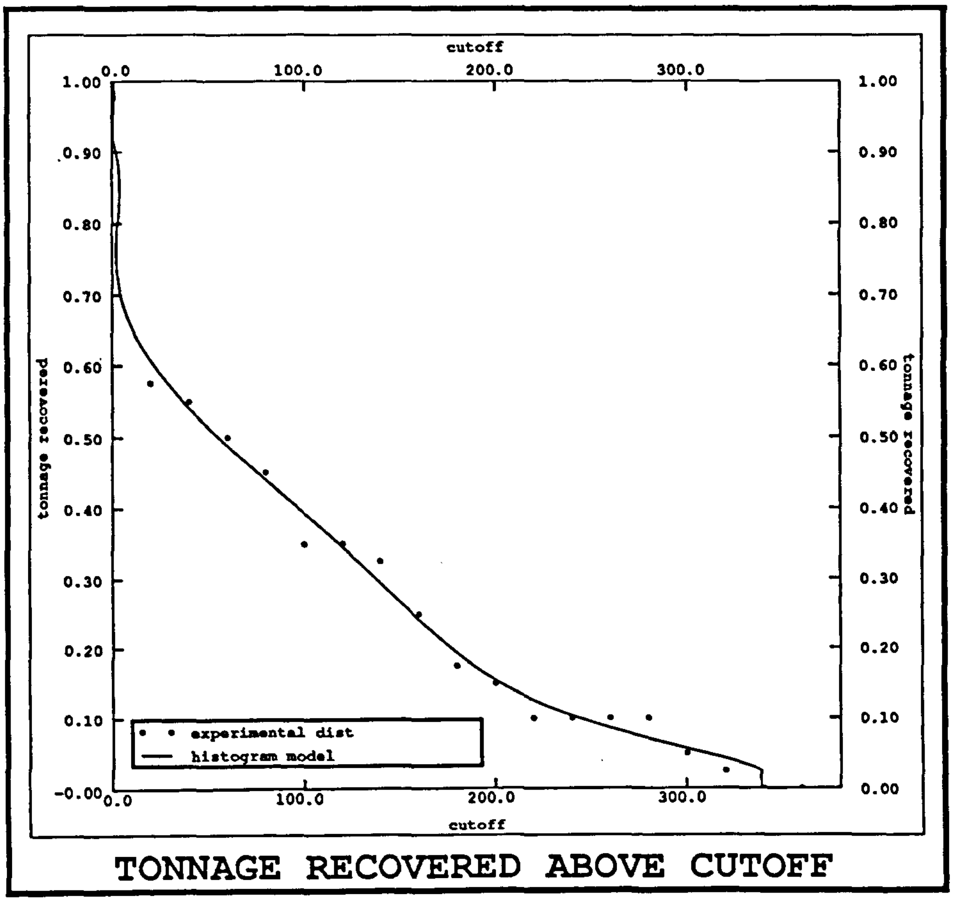


Figure 19a. Tonnage (or Area) above Cutoff expressed as a percentage as a function of Cutoff

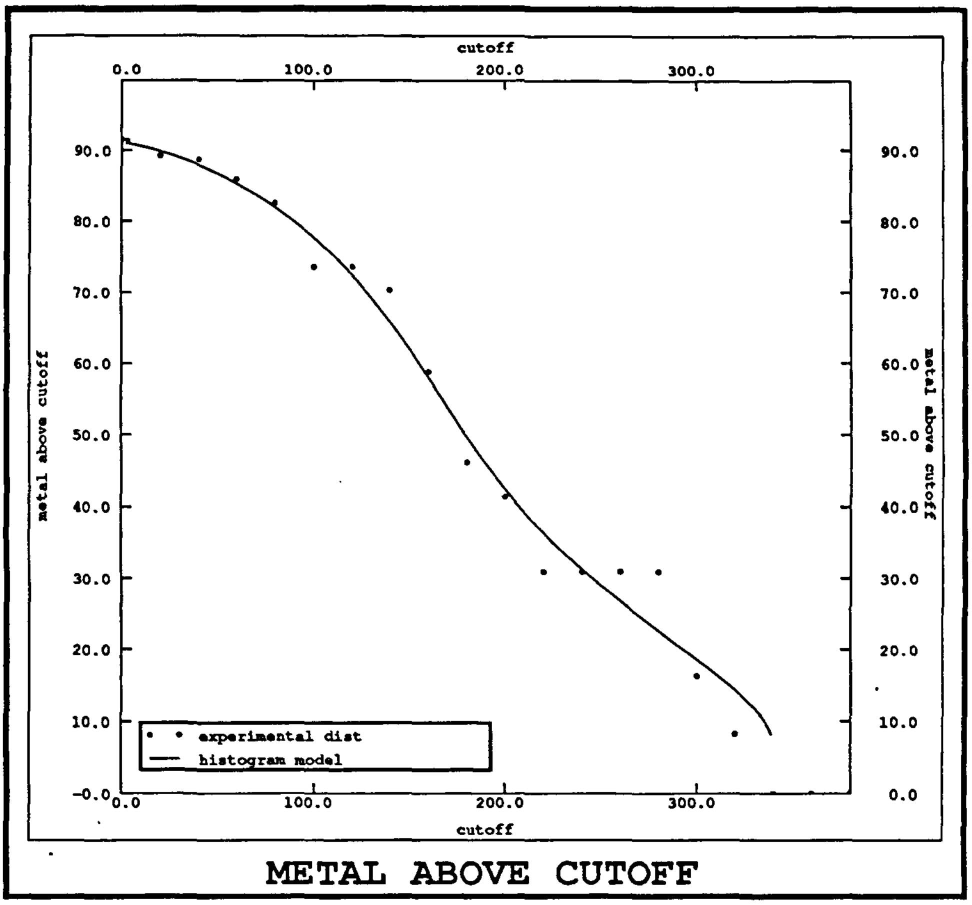


Figure 19b. Metal (or Quantity) above Cutoff expressed as a percentage as a function of Cutoff

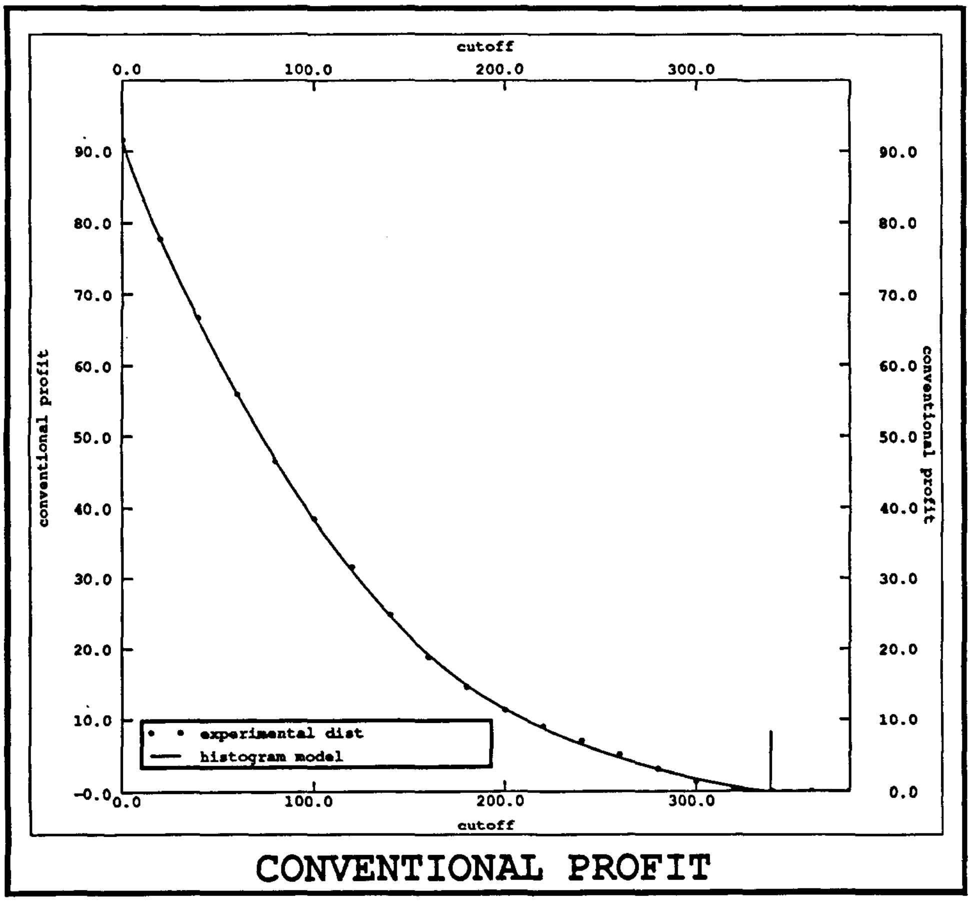


Figure 20. Conventional Profit Function as a function of Cutoff.

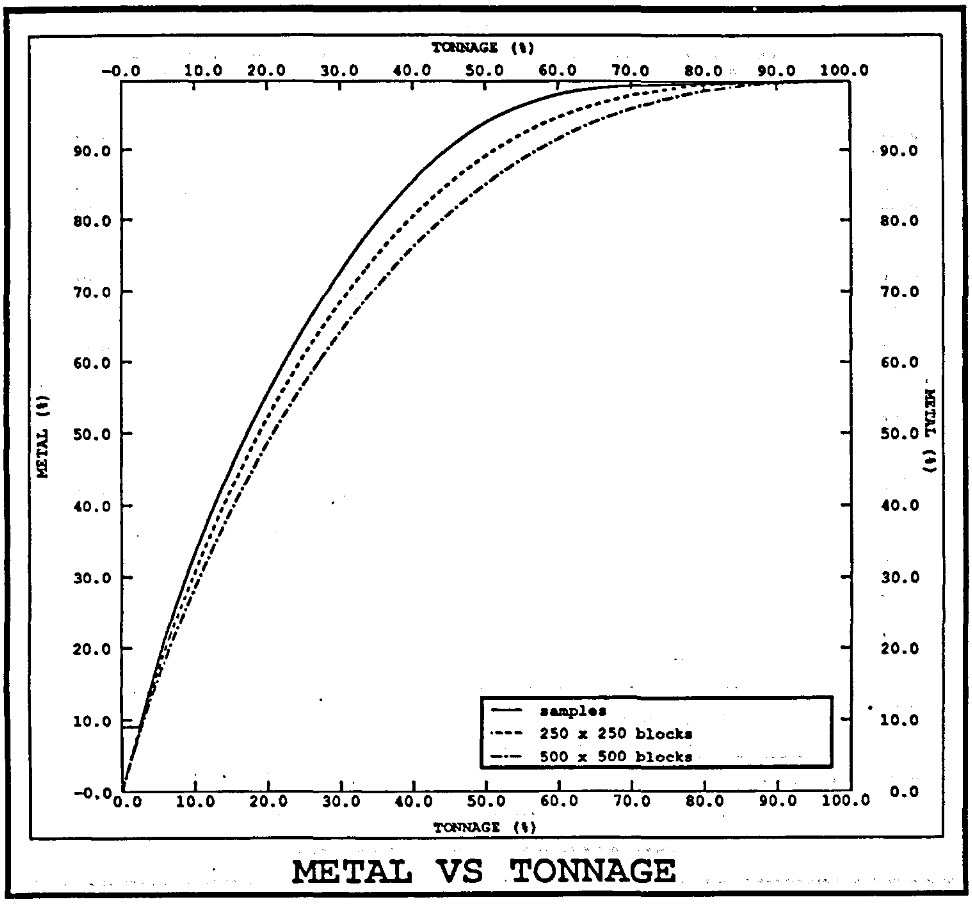


Figure 21. Metal versus Tonnage for three different Block Sizes

CONDITIONAL SIMULATION

Conditional simulations are designed to reproduce the spatial statistical characteristics of the data (i.e. the variogram and the histogram) and to go through the sample points (for a point simulation). The methodology is explained in most geostatistical textbooks. It consists in transforming the data to normality (as before), simulating a normal distribution with the appropriate characteristics, conditioning it to respect the data and then back-transforming. Clearly the variogram of the gaussian equivalents is not necessarily the same as that of the raw data. In this case the fitted model (Figure 22) for these was a spherical with no nugget effect, with a range of 1000m and a sill of 1.0 (equal to the variance of the standard normal). Tests were carried out to check that this matched with the model fitted earlier for the corresponding raw data. The simulation was carried out for 112 points on a regular 200m x 200m grid lying inside the small polygon. One such simulation is presented in Figure 23. It is instructive to compare this with the corresponding kriged maps (Figures 16a and b) which are, as expected, much smoother.

A total of 100 simulations were made and the global statistics for these were counted. Figure 24 shows the inverse cumulated histogram of the mean number of spisules per unit area obtained from each of the 100 simulations. This can be interpreted as giving the probability that the average number would exceed a specified value. Clearly the mean of this histogram is just the kriged average obtained earlier. This histogram gives more meaningful confidence intervals than those obtained as the kriged value plus (and minus) twice the standard deviation.

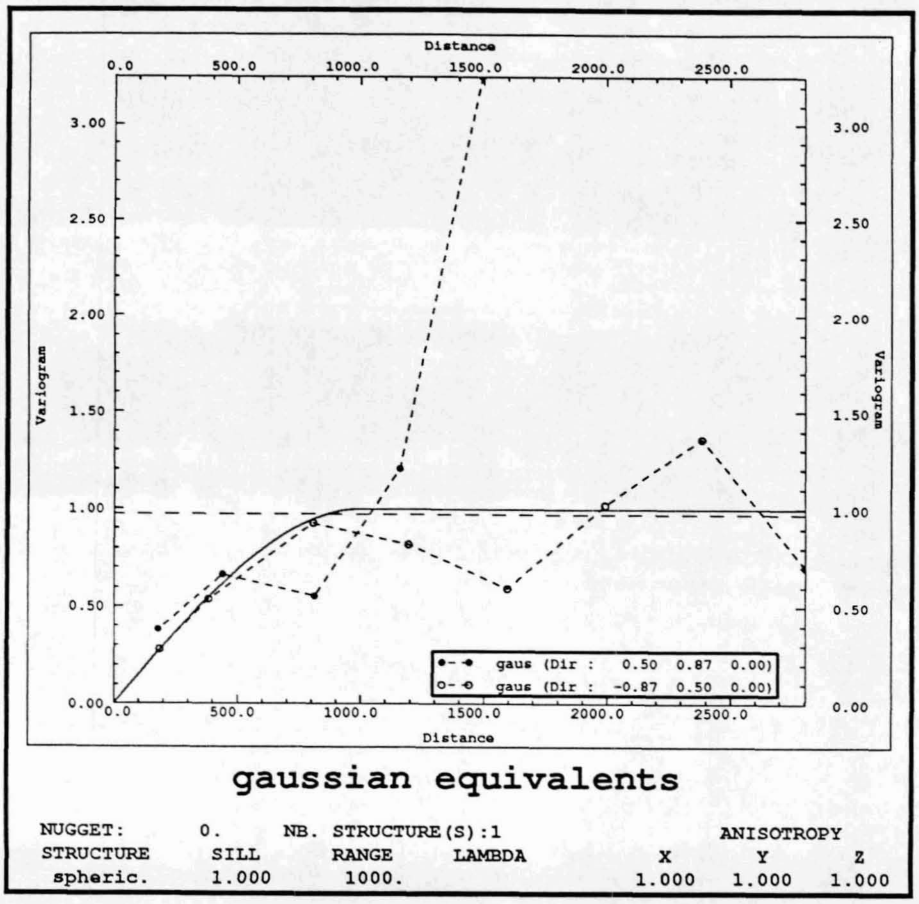


Figure 22. Fitted Variogram Model for the Gaussian Equivalents

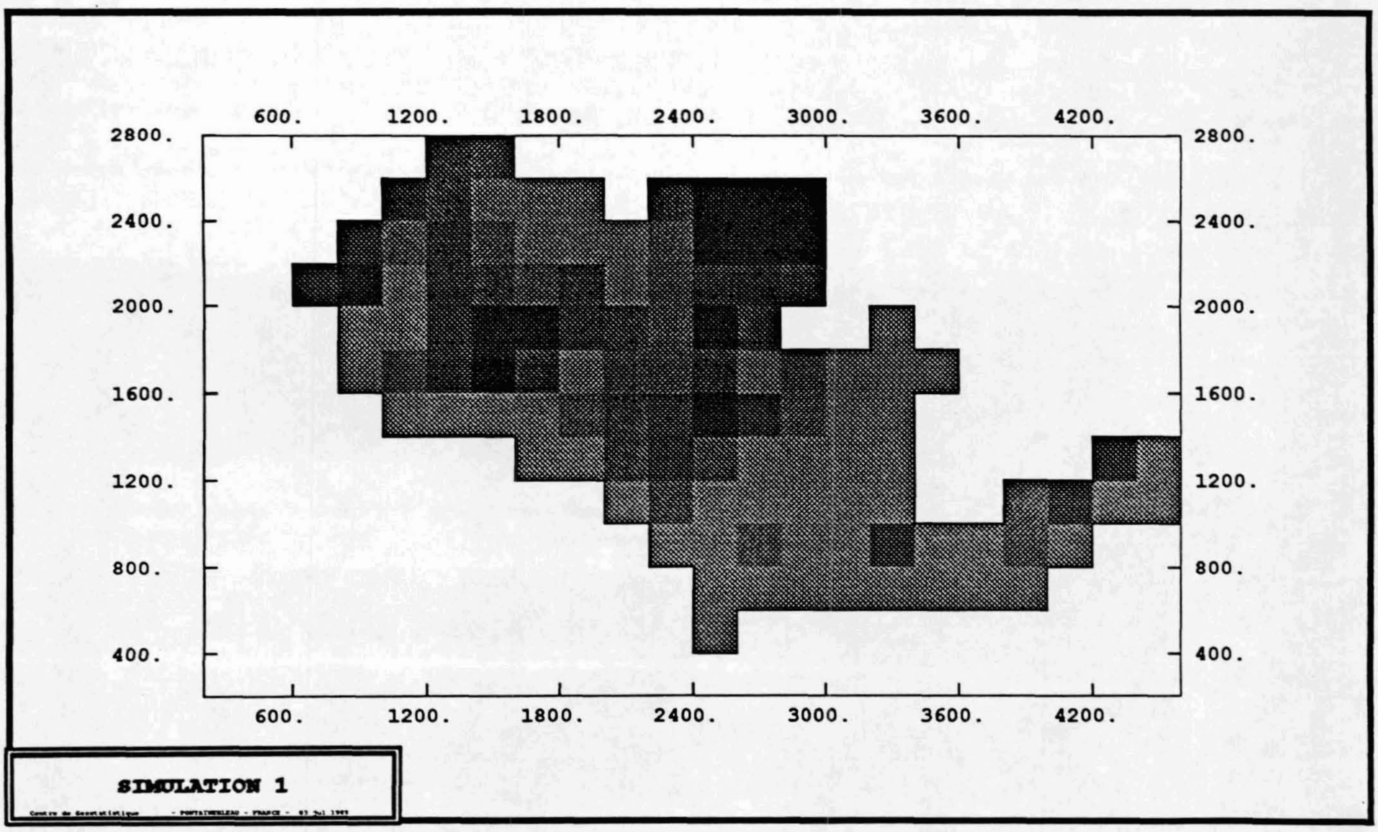


Figure 23. One Simulation of the Number of Spisules

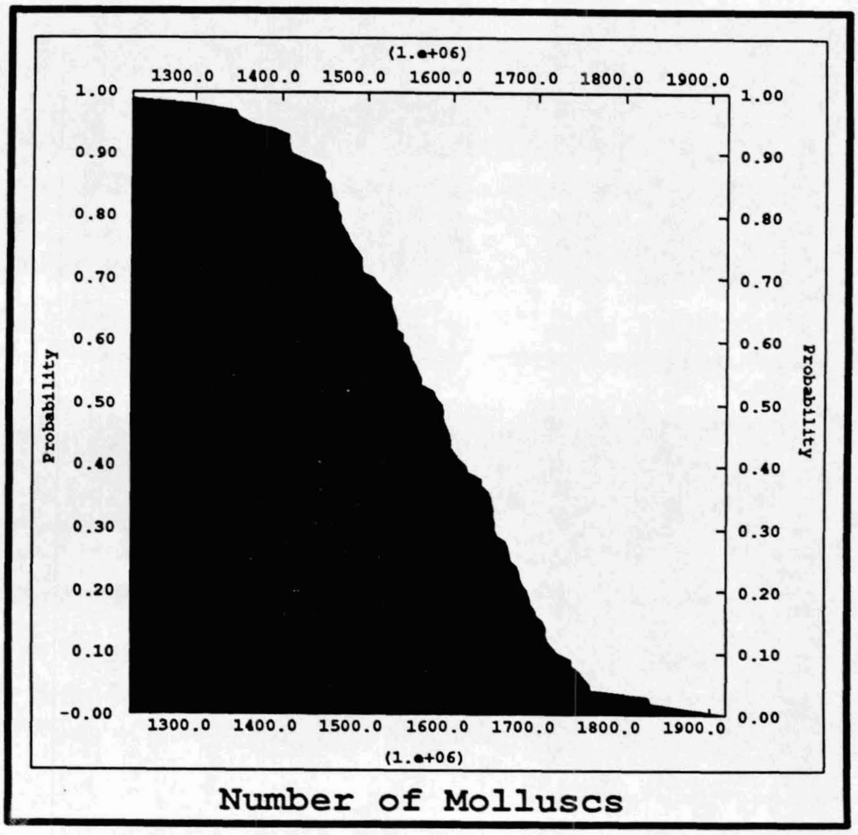


Figure 24. Histogram of the 100 mean values obtained from the simulations

DISCUSSION AND CONCLUSIONS

This preliminary report was designed to show how geostatistics can be used for assessing the reserves of sedentary species. The objective of this study was primarily pedagogical. In this case it was used firstly to calculate the total reserves and the average grades in a Spisule deposit near the Ile d'Yeu on the Atlantic coast of France. Two variables were studied the number of individuals and their biomass (in g/0.25m2). As the area containing the bivalves is clearly delimited by zero grades, we studied the impact of including or excluding these zero values. It was found that the total reserves were the same when the reserves were calculated using data including the zeros in the large polygon covering the whole area, as those obtained in the area containing positive values using only the corresponding data. These figures are in good agreement with previous estimates made using classical estimation methods. However the average grades were markedly different.

In addition to obtaining the estimates and the corresponding estimation variances using kriging, more advanced methods involving Hermite polynomials were used to estimate the recoverable reserves as a function of three possible sizes of exploitations blocks. This highlights the importance of the size of these blocks on the selectivity of the operation. Lastly 100 conditional simulations were carried out. One of these is presented to show how much more variable the number of spisules could reasonably be expected to be, compared to the corresponding kriged maps. The histogram showing the average number of spisules for each of the 100 simulations gives us an idea of the precision of the estimated value.