ANNALI DI GEOFISICA, VOL. 42, N. 5, October 1999

Fractal dimension confidence interval estimation of epicentral distributions

Luciana De Luca [1], **Stanislaw Lasocki** [2], **Dario Luzio** [1] **and Massimo Vitale** [1]
[1] *Dipartimento di Chimica e Fisica della Terra (CFTA), Università di Palermo, Italy*
[2] *Institute of Geophysics, University of Mining and Metallurgy, Krakow, Poland*

Abstract
Estimates of the fractal dimension of hypocentral distributions require evaluating the range of independent variables in which fractal parameters exhibit a power law. Systematic and accidental errors are produced mainly by the subjective selection of this range, the insufficiency of data sets as well as by hypocenter mislocations. Therefore it is very important to determine the confidence intervals which are associated with fractal dimension estimates. The effects of various sources of errors are studied using different geometric clusters of epicenters, which have been synthetically generated using a multicluster algorithm with different hierarchical levels, so as to reproduce some characteristics of the patterns typical of real epicenter distributions. Subsequently, groups of differently sized subsets of synthetic epicenters were obtained by randomly sampling each distribution. Confidence intervals of fractal dimensions were thus calculated using all the estimates obtained for the various subsets. This procedure was also tested on real seismic data, consisting of epicentral distributions in three Sicilian areas and five clusters of mining-induced seismic events (Wujek coal mine, Poland). In that analysis both correlation dimensions and their confidence intervals were taken into account.

Key words *correlation dimension – confidence interval – seismicity*

1. Introduction

Statistical fractality has long been studied in connection with seismicity patterns (Turcotte, 1992; Xie, 1993). This feature corresponds to a scale invariance of the spatial and temporal distribution of earthquakes. The Gutenberg-Richter law is associated, according to Aki (1981), with the scale invariance of the seismic phenomenon with respect to fracture sizes in faulted areas. Therefore, it seems reasonable to try to differentiate seismic activity generated according to different seismogenetic processes by evaluating their fractal dimensions in the domains stated above. A complete fractal characterization of homogeneous seismic areas must include estimates of the fractal dimension confidence intervals in order to evaluate if fractal dimension estimates of epicentral distributions located in distinct areas are significantly different.

So far, in the literature the uncertainty associated with fractal dimension estimates has been represented by that of the slope of the regression line. Here, on the contrary, significant samplings of the fractal dimension estimate distributions are obtained from sets of synthetic epicenters characterized by different geometric parameters, in order to give a more reliable variance of those estimates and thus find out if they are biased.

Mailing address: Prof. Dario Luzio, Dipartimento di Chimica e Fisica della Terra (CFTA), Via Mariano Stabile 110, 90159 Palermo, Italy; e-mail: luzio@unipa.it

In estimating the confidence interval of the fractal dimension the following causes of errors must be taken into account. First of all one has to consider the effect of the finiteness of the samples. Moreover, the evaluation of the range in which fractal parameters exhibit a power-law behavior is not, in general, based on an objective criterion. Typically that increases the variance of the estimates and such an increase is certainly uncorrelated with the uncertainty of the regression line slope. Finally, a further source of errors in the estimates resides in hypocenter or epicenter mislocations. Systematic effects on the correlation dimension estimate arising from random noise present in the data have been studied by Möller *et al.* (1989). They observed in sets of 20 000 values obtained from the Roessler attractors an overestimation of the actual correlation dimension, proportional to the ratio between the random noise variance and the square of the selected average length scale.

The depth of seismic events was neglected in this analysis since it is known that this experimental parameter is not usually determined with the same accuracy as the epicentral coordinates, and thus could bring about a considerable deterioration of the fractal dimension estimates. It should be noted in this context that it could prove interesting to investigate how the lack of accuracy in determining depth can influence the quality of fractal dimension estimates.

Here the correlation dimension D_2 has been taken into account. It was defined using the correlation integral method (Grassberger and Procaccia, 1983) as the slope, in a log-log diagram, of the function relating the correlation sum $C_2(r)$ to the counting radius r. The correlation integral is

$$C_2(r) = \frac{1}{N(N-1)} \sum_{j=1}^{N} \sum_{k \neq j} \Theta(r - |\boldsymbol{x}_j - \boldsymbol{x}_k|) \approx r^{D_2} \tag{1.1}$$

where Θ is the Heaviside function, N is the number of points and $|\boldsymbol{x}_j - \boldsymbol{x}_k|$ the distance between pairs (j, k) of points. The slope is calculated by least squares in the function *linearity range*.

Synthetic epicentral distributions were created by a multicluster generation with different hierarchical levels, in order to reproduce some of the geometrical features which are evident in real epicentral distributions.

2. Analysis of synthetic data

Synthetic sets of points were produced in three stages. The first consists of a *multicluster generation* of large sets (source samples) of points $P(\xi, \eta)$ on a plane. In the second the points were randomly sampled, so that the subsets extracted from them can represent, in terms of numerosity, experimental epicenter distributions. In the final stage the points are orthogonally projected onto a differently oriented plane with respect to the generating one. The generated distribution proves to be increasingly elongated in the direction of the intersection between the two planes, as the dihedral angle θ between them is increased. A similar elongation is commonly found in many epicentral distributions.

The source samples generated must reveal the tendency of the point set to form clusters. They must be large enough to be considered sufficient to provide reliable estimates of the geometric parameters which are being investigated, in particular the resulting variance of the estimates of D_2 should be very small with respect to the resolution power required for this parameter for the purpose of the investigation.

As to the first stage, a multicluster generation with L hierarchical levels was defined. Each cluster of points of the i-th level was centered at a point of the $(i-1)$-th level. Though a theoretical proof of the possibility to provide the generated sets with scale-invariance properties by the procedure here used has not been given, they proved to behave like statistical fractals.

Fractal geometry is a tool for the reproduction of real very complex objects by the iterative application of simple rules. These latter reproduce or characterize particular physical processes which act in the same repetitive way on many different scales. The multicluster generation carried out can thus be seen as a good representation of a seismogenetic phenomenon, which is supposed to feature a statistical scale invariance.

At level i the points (r_k, φ_k) are generated around each point of level $i-1$, with azimuth φ_k uniformly distributed in the interval $[0, 2\pi]$ and radial distance r_k following the cumulative distribution law

$$F(r_k) = \left(\frac{r_k}{R} \right)^b . \qquad (2.1)$$

The set of values r_k is in practice obtained by generating uniformly distributed values of F in the interval [0, 1], and finding through eq. (2.1) the corresponding r_k in the interval $[0, R]$.

The exponent b represents the reciprocal of the clustering index k which can be differentiated for each hierarchical level. The maximum clustering degree is achieved when $b = 0$, as the points distribution collapses to a single point; when $b = 2$ a two-dimensional uniform distribution is generated, while in the case $b > 2$ the surface points density becomes an increasing function of r.

This scheme was followed to generate a number of source samples having different size and complexity. The latter is related to the possibility to vary the number of hierarchical levels and to differentiate b in each of them.

Two large source samples of $\sim 10^6$ points were generated using a different number of hierarchical levels. From each population eight groups of two hundred subsets including from 100 up to 7000 points were extracted so as to evaluate the effect of the sample size and of the number of hierarchical levels on the correlation dimension estimates. The mean value of $\tilde{D}_2$ and its standard deviation as well as the sample standard deviation were calculated for every group.

Figure 1a,b shows, *e.g.*, two small subsets of synthetic epicenters extracted from two different source samples. They evidence how their geometric patterns reproduce some typical clustering features of experimental epicentral distributions.

To study the effect on the estimates of an azimuth dependence of the points density, produced by their projection from the generating plane onto a differently oriented one, three different rotation angles ($\theta = 0°$, $\theta = 45°$, $\theta = 90°$) were considered, for sets of points generated by three hierarchical levels.

In order to carry out a statistically significant test a very large number of sets of points had to be analyzed. For this purpose, it was necessary to automatize the calculation of the correlation dimension, by determining the linearity range in a log-log data representation by means of an optimization procedure.

The algorithm used here consists of progressively enlarging a $\log r$ interval centered on an intermediate point between two extreme values of r, determined from the coordinates of the set points.

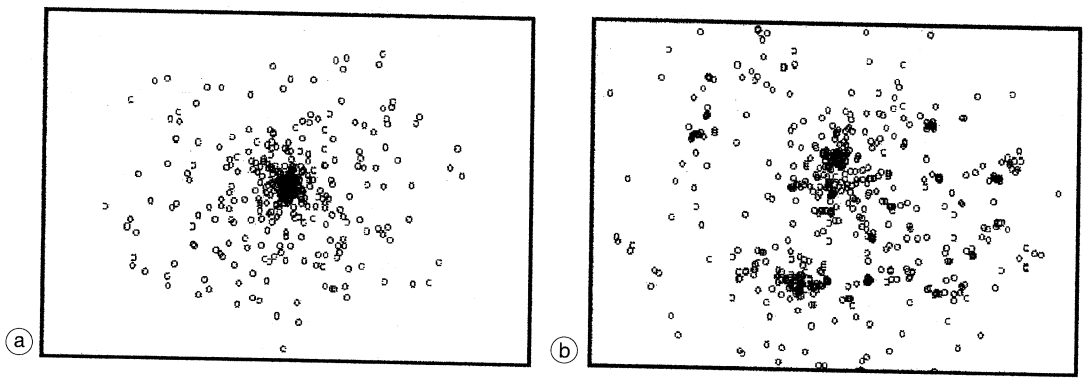

Fig. 1a,b. Two sets of synthetic epicenters: a) one hierarchical level: $b_1 = 0.2$; $\theta = 0°$; b) five hierarchical levels: $b_1 = 0.5$; $b_2 = 0.4$; $b_3 = 0.3$; $b_4 = 0.2$; $b_5 = 0.1$; $\theta = 0°$.

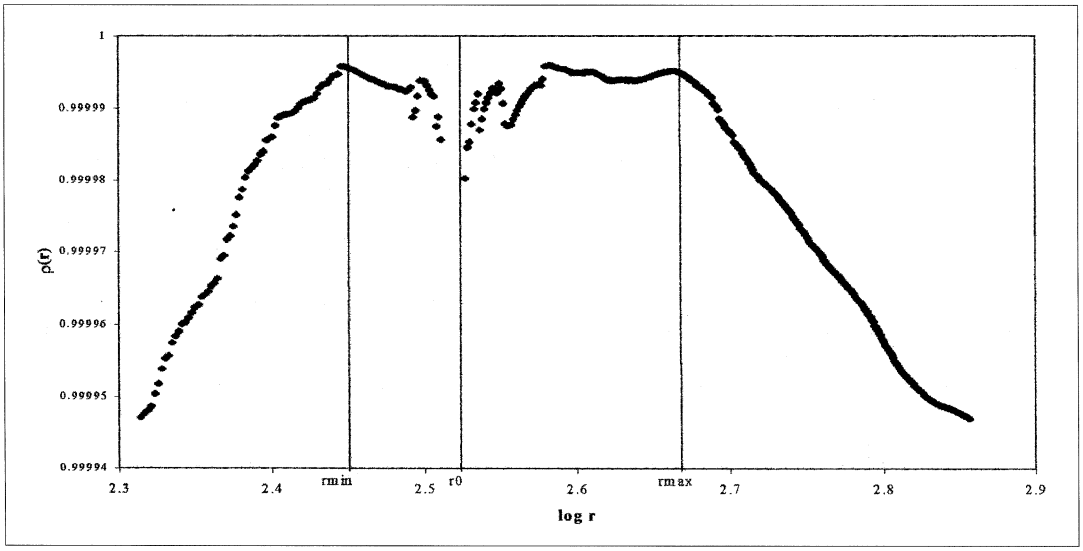

Fig. 2. Trend of the correlation coefficient for a 3000-point set. The estimated linearity range and the intermediate point ($r0$) are indicated by the three vertical lines.

A defining criterion for these values, which proved effective for the location of the intermediate point within the linearity range, was based on the choice of the maximum r as half the smaller between the x and y ranges and the minimum r as one quarter the side of the average square occupied by each point, $r_{min} = \sqrt{(x_{max}-x_{min})(y_{max}-y_{min})/N}$, where N is the number of points. These limits should lie respectively in the depopulation and the saturation zones, so as to make it possible to disclose the whole linearity range in the subsequent analysis.

The intermediate point was defined as the logarithm of the geometric mean between the two limiting values of r.

For each point added to its right or left the correlation coefficient $\rho(r)$ between $\log C_2(r)$ and $\log r$ in the new interval is calculated, until a decrease of $\rho(r)$ is observed consecutively for more than a number η of points. The value of η must be large enough to allow possible $\rho(r)$ relative minima to be included in the linearity range, but small enough to render unlikely to find an increase in $\rho(r)$ in the nonlinearity zones; that would in fact bring about an undesired widening of the interval sought and a consequent underestimation of D_2. A typical plot of $\rho(r)$ is shown in fig. 2.

The function $\log C_2(r)$ *versus* $\log r$ was sampled in the fixed range of r by setting a constant step δ either on a linear or on a logarithmic scale. Estimates obtained using both criteria showed a remarkable consistency within their confidence intervals, no systematic effects in their trends with respect to each other and the tendency of their differences to decrease as the subset size increases. The first criterion was used for all the synthetic tests shown in this paper.

3. Results of the tests

The synthetic tests carried out show a systematic effect of the sample size on the estimated value of the correlation dimension, as well as a dependence of it on the different parameters involved in the generating procedure, *i.e.* the number of hierarchical levels, the clustering index and the number of points for each level.

The complex dependence on the generating parameters was not modeled here. The correlation laws $\tilde{D}_2(N)$ *versus* N expressed in fig. 3 indicate that the estimates have on average biases greater than 5% for a size lower than 3000 points and greater than 10% for a size lower than 1000 points, while samples with size greater than 8000 points are needed to obtain a bias lower than 1%.

The asymptotic values (when the subset size approaches infinity) are about 1.65 and 1.59 for two source samples of 606 060 and 521 050 points respectively. The first was generated by three hierarchical levels with $b = 0.5$ and number of points in each cluster of the first, second and third level equal respectively to 60, 100 and 100; the second source sample was generated by four hierarchical levels with $b = 0.5$ and number of points equal respectively to 50, 20, 20 and 25.

The underestimation of D_2, evidenced in all tests carried out, must probably be attributed to a slope decrease which can be observed in the plot $\log C_2(r)$ *versus* $\log r$ both for low and high values of r, these effects being known respectively as *depopulation* and *saturation* (Nerenberg and Essex, 1990). Both of them appear for every limited data set and affect the estimates more heavily as the number of points in the data set decreases, since the saturation and depopulation effects tend to become evident also in the intermediate zone.

The effects on the bias of $\tilde{D}_2$ due to the different orientation between the generating and the analysis planes was observed for rotation angles of 0°, 45° and 90°. In the first two cases nearly coincident trends of $\tilde{D}_2$ *versus* N were obtained (fig. 4). In the third case samples made up by few hundreds points allow estimates of D_2 with a negligible bias, their asymptotic value being lower than one, which is the actual embedding dimension of the set.

An empirical dependence law for D_2 on the parameter b was investigated by extracting groups of 7000-point subsets from three-level source samples with different values of b kept

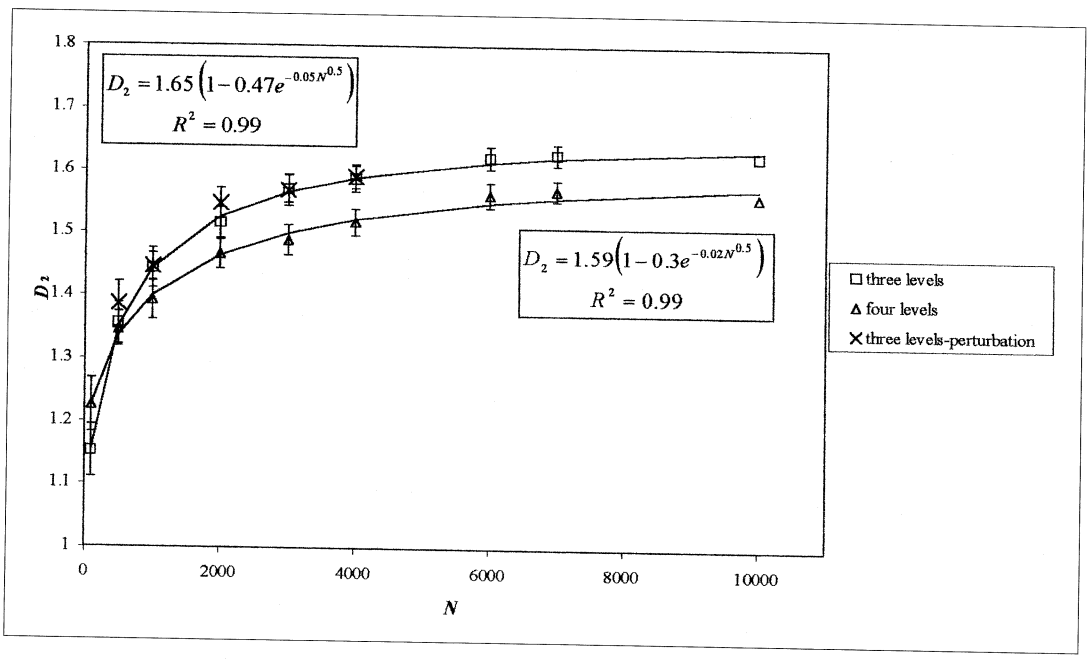

Fig. 3. Dependence of $\langle \tilde{D}_2 \rangle$ on size for subsets extracted from source samples generated using different hierarchical levels. The equation on the top is referred to the three-level generation, the other to the four-level generation.

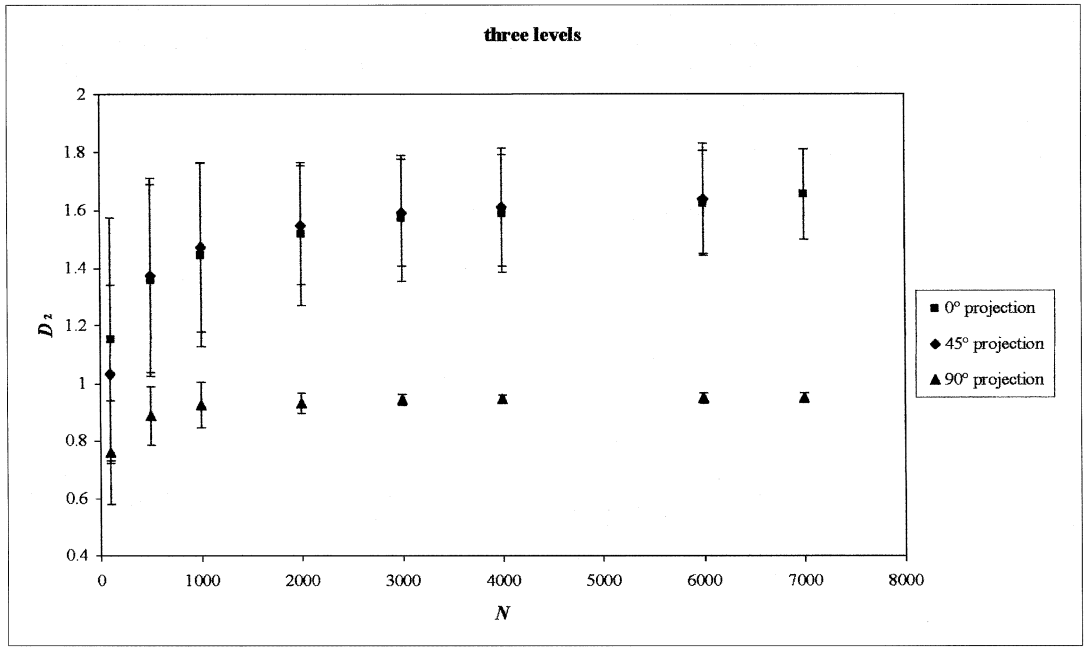

Fig. 4. Dependence of $\tilde{D}_2$ on the rotation angle between the generating and analysis planes.

constant in all generating levels. The relationship used to fit the experimental data is expressed as

$$D_2 = 2\,(1 - e^{-\alpha 2^{\beta}})^{-1}(1 - e^{-\alpha b^{\beta}}) \qquad (3.1)$$

with parameters α and β to be statistically estimated.

Equation (3.1) is such that D_2 is equal to 0 when $b = 0$, *i.e.* the point distribution collapses into a center, and equal to 2 when $b = 2$, *i.e.* the point distribution becomes uniform. Figure 5 shows the best-fit curve, in which $\alpha = 3.8$ and $\beta = 1.2$. An evident narrowing of the 95%-confidence intervals is observed while b increases.

Some tests were carried out to study the dependence of the confidence interval of $\tilde{D}_2$ on sample size. Since the standard deviation of $\tilde{D}_2$ also depends on the genetic process of the experimental data for any fixed sample size, an analysis was conducted to find a correlation between it and some parameters connected with the experimental data distribution. Finally the effect was investigated of data accidental errors on the bias and the standard deviation of $\tilde{D}_2$.

The 95%-confidence interval of $\langle \tilde{D}_2 \rangle$, represented in fig. 3, shows a decreasing trend as the subset size increases. A quantitative relation to describe this was obtained extracting from a three-level source distribution sets of two hundred samples of different size. Figure 6 shows that, with 0.95 probability, the standard deviation of $\tilde{D}_2$ ranges from about 0.5 for small sample sizes to about 0.15 for 7000-point subsets. A significant correlation between the 95%-confidence interval S_s and the subset size N was found, expressed by the empirical law

$$S_s = 0.54\, e^{-0.044 N^{0.37}}. \qquad (3.2)$$

This significant loss of efficiency of the estimator $\tilde{D}_2$ as N decreases will be clearly understood observing fig. 7. The graphics show two plots of the function $\log C_2(r)$ *versus* $\log r$ which refer to

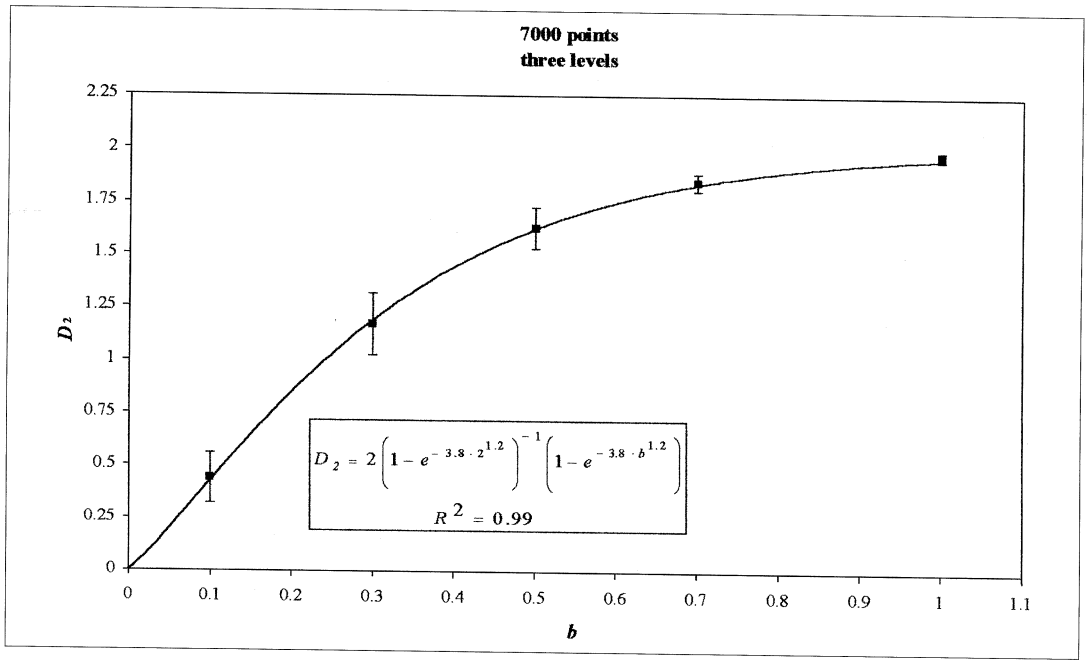

Fig. 5. Representation of $\tilde{D}_2$ *versus* b.

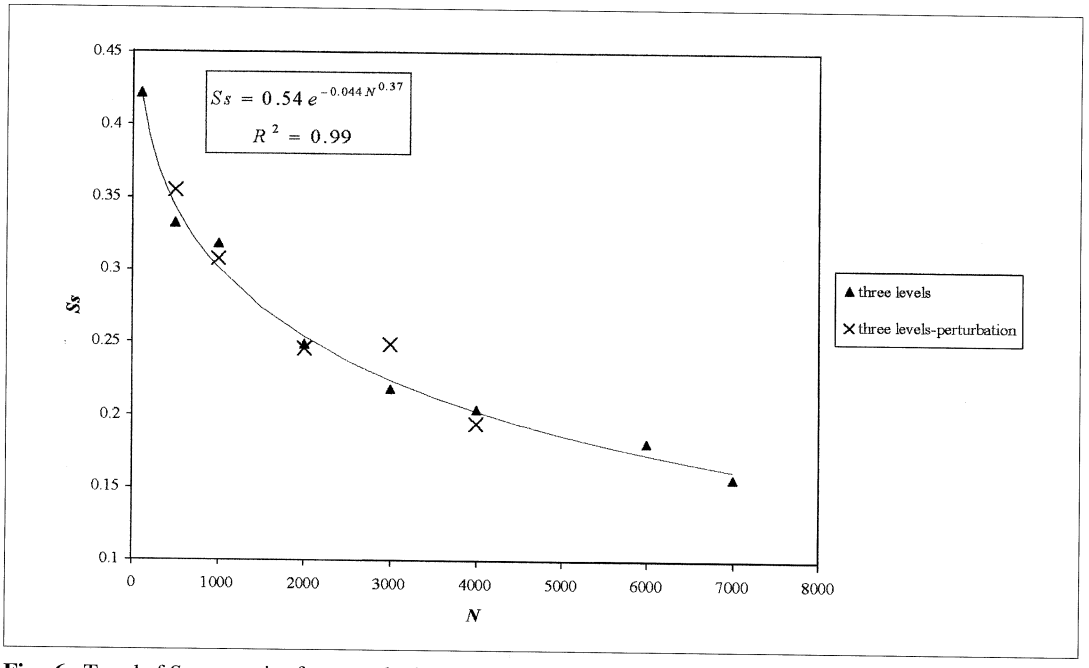

Fig. 6. Trend of S_s *versus* size for perturbed and unperturbed data of subsets extracted from a three-level generated source sample.

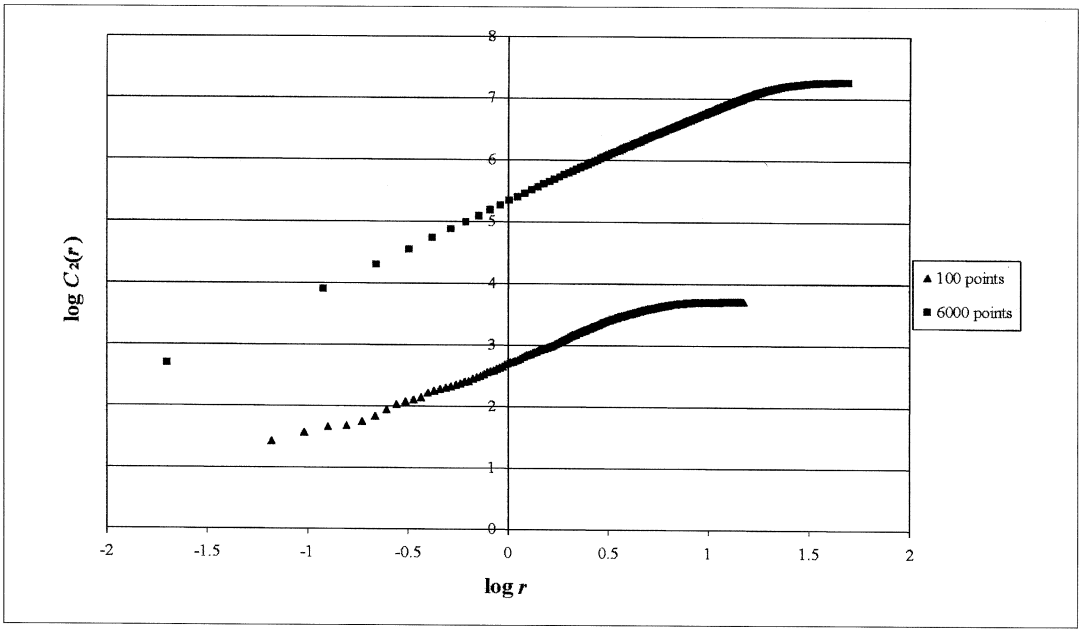

Fig. 7. Plots of $\log C_2(r)$ *versus* $\log r$ for a 100-point sample and a 6000-point sample.

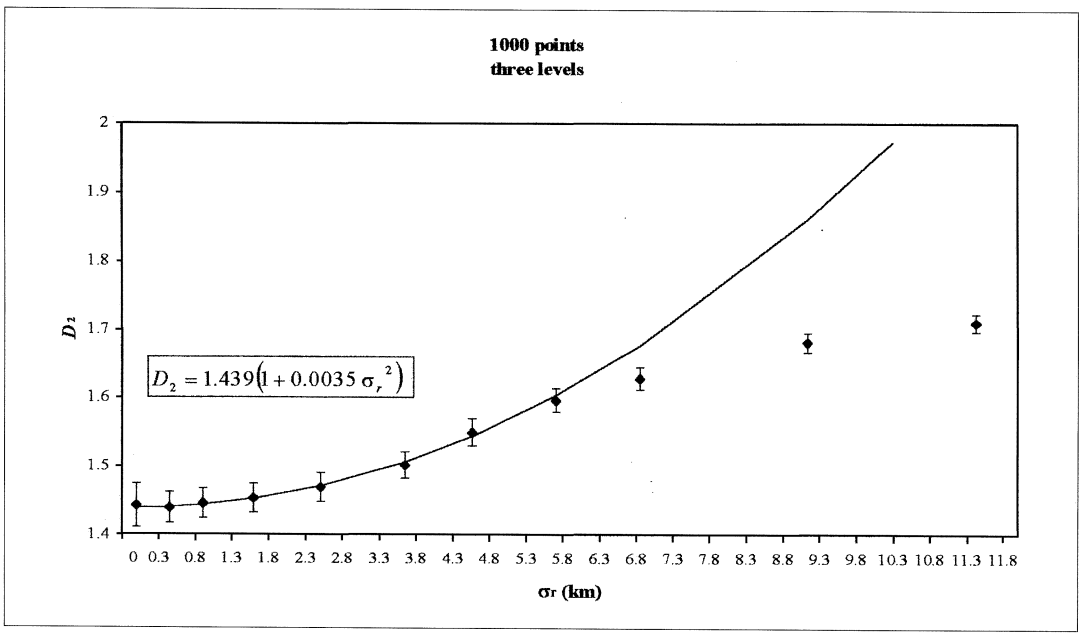

Fig. 8. Dependence of $\tilde{D}_2$ on the perturbation standard deviation.

100-point and 6000-point samples. They stress how the choice of the linearity range becomes critical in the case of small-sized samples. Evident irregularities affect the tail of the function near the depopulation zone in the 100-point sample; conversely, an almost perfect linearity is visible in the other plot.

The 95% probability level confidence intervals prove to be almost independent of the rotation angle between the generating and analysis planes, up to $\theta \cong 90°$, where they decrease by a factor ranging from 0.4 down to 0.1 when the subset size rises from 100 up to 7000 points (fig. 4).

Another source of error on $\tilde{D}_2$ is given by the inaccuracy of experimental points coordinates. In order to evaluate the effect of such errors on the estimates, the planar coordinates of the analyzed sets of synthetic epicenters were randomly perturbed. The perturbations $\Delta \boldsymbol{r}$ were uniformly generated in a considerably larger range than typical epicenter mislocations of local or regional seismic networks.

In a first test samples of 500, 1000, 2000, 3000, 4000 points, characterized by average interdistances between close points equal to 3.1, 2.2, 1.6, 1.3, 1.1 km, were perturbed with errors, uniformly distributed both in azimuth and amplitude, having a standard deviation $\sigma_r = 0.9$ km. No particular effect on the standard deviation of $\tilde{D}_2$ was observed after the perturbation, as shown in fig. 6. In fig. 3 it is possible to note how the perturbations produced a slight tendency to overestimate D_2.

The effect of different values of σ_r was also tested on $\tilde{D}_2$ (fig. 8). The σ_r were fixed from ~ 0.4 km up to ~ 11.4 km for groups of subsets containing 1000 points and $b = 0.5$ in each of the three hierarchical generating levels. The average interdistance between close points in the several sets is approximately equal to 2.2 km. From the plot D_2 *versus* σ_r, shown in fig. 8, it can be observed that $\tilde{D}_2$ is incremented of a bias ~ 15% in the case of the largest perturbation applied, and could possibly tend to 2 as $\sigma_r \to \infty$.

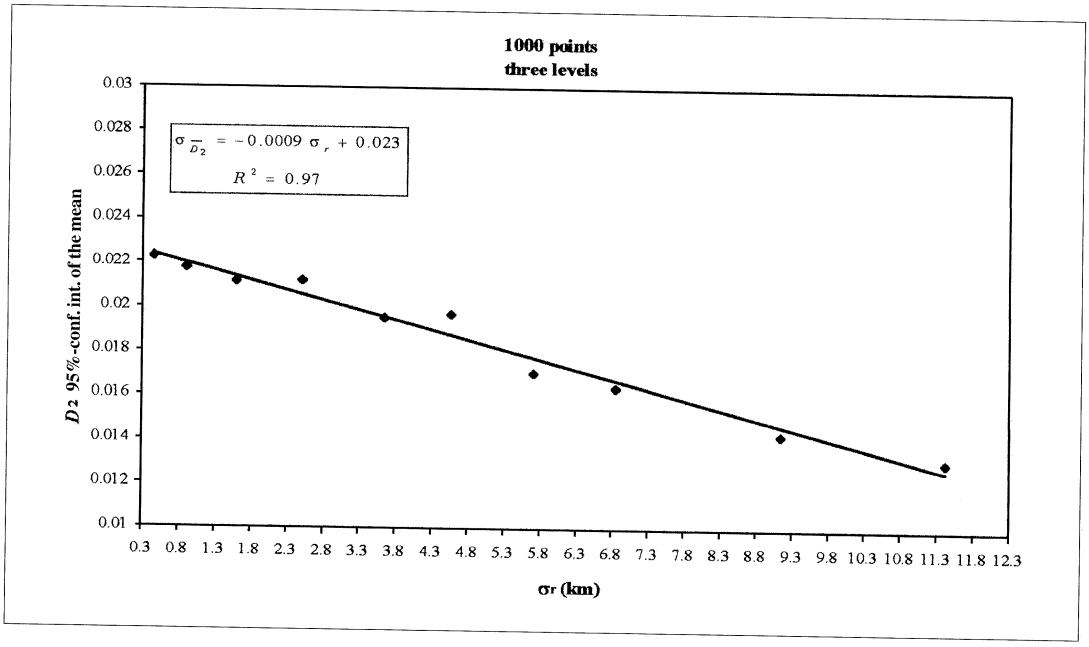

Fig. 9. Dependence of 95%-confidence interval of $\langle \tilde{D}_2 \rangle$ on the perturbation standard deviation.

In the interval 0-5 km the experimental points can be fitted by a function

$$D_2 = 1.439\,(1 + 0.0035\,\sigma_r^2) \qquad (3.3)$$

which is very similar to that proposed by Möller *et al.* (1989).

The test also showed that the standard deviation of $\tilde{D}_2$ regularly decreases as σ_r increases (fig. 9). The same trend had been observed in the remaining tests, in which it could not be related to a simple increase of D_2, since generating parameters and sample size had been varied.

4. Estimate of confidence intervals of D_2 from sets of experimental data

The analysis described above to estimate the correlation dimension confidence interval of surface point distributions cannot be directly applied to experimental epicentral distributions because of the tiny number of seismic events generally present in seismic clusters. It is in fact impossible to extract an appropriate number of independent subsets from the data set having a sufficient size to allow a statistically significant estimation of D_2.

In order to define reliable estimators S_s, two different approaches were followed. The sample variances of the distributions of $\tilde{D}_2$, evaluated extracting several subsets of size N from very large sources of size $N_s (N_s > 1000N)$, were related in the first approach to the 95%-confidence intervals S_r of the regression line slope computed for each subset, while in the second approach they were related to the sample variances evaluated extracting the subsets of size N from a set of size N_{tot} typical of experimental data sets, thus not much larger than N. Obviously, in the second case the subsets of size N cannot be regarded as mutually independent.

In fig. 10 the estimates of several S_s obtained from homogeneous sets of two hundred samples characterized by different size and generating parameters are plotted against 67%-confi-

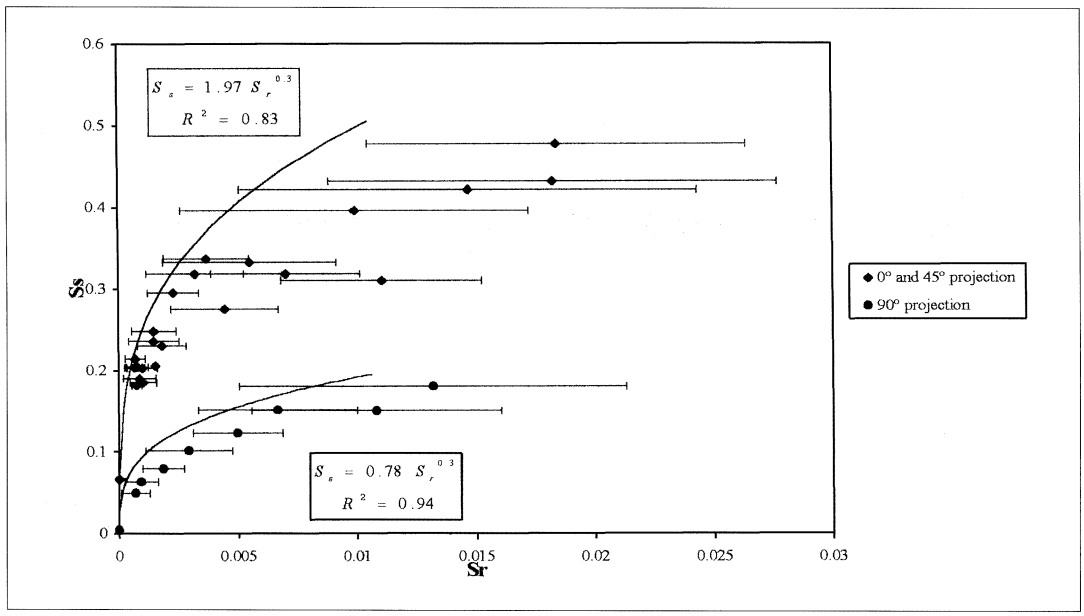

Fig. 10. Relation between S_s and S_r for subsamples extracted from source samples generated with three levels and various rotation angles between the generating and analysis planes.

dence intervals of S_r centered on the mean values S_{rm} of the two hundred estimates relative to the corresponding sets.

A nonlinear least squares regression between S_s and S_{rm} led to the relation

$$S_s = 1.2\, S_{rm}^{0.25} \tag{4.1}$$

obtained generating points directly on the analysis plane or on a 45° rotated plane, and to

$$S_s = S_{rm}^{0.4} \tag{4.2}$$

when the generating and the analysis planes are orthogonal.

The high amplitude of the error bars related to the estimates S_r suggests determining a conservative relation between the two parameters (method A).

This was defined by the regression law between S_s and the lower bounds of the 67%-confidence intervals of S_r. In this way, for a given S_r there is an ~84% probability that the actual 95%-confidence interval amplitude of $\tilde{D}_2$ is lower than the S_s corresponding to S_r.

The regression laws, obtained generating points directly on the analysis plane or on a 45° rotated plane, turned out almost coincident and are expressed as

$$S_s = 1.97\, S_r^{0.3}. \tag{4.3}$$

For the case of orthogonality between the generating and the analysis planes the relation found is

$$S_s = 0.78\, S_r^{0.3}. \tag{4.4}$$

A strong difference is evident between S_s and S_r, therefore it is misleading to use S_r as an estimate of the uncertainty of $\tilde{D}_2$, a practice commonly followed in many applied researches.

In the case in which $N_{tot} < 10N$, N being larger than the smallest acceptable sample size N_{min} to obtain a reliable estimate of D_2, the mutual dependence between the extracted subsets leads to an underestimation of the sample standard deviation of $\tilde{D}_2$, with respect to that calculated in the condition of complete independence.

It was proved that, given a set of N_{tot} experimental points, it is possible to estimate a corrected value of standard deviation of D_2 calculated from a given number $M \leq \binom{N_{tot}}{N}$ of samples, multiplying it times an underestimation coefficient (method B).

This underestimation effect was analyzed against the parameter $d = N/N_{tot}$, which provides a measure of the samples mutual dependence, d approaching 0 in the case of complete independence and 1 in the case of complete dependence.

A synthetic test was made extracting $M = 100$ samples of varied size N both from the source set, in order to realize the condition of full independence, and from a subset of it with size N_{tot}. Different combinations of subset sizes N and N_{tot} were used in order to have values of d equal to 0.1, 0.2, 0.4, 0.6, 0.7, 0.8 and 0.9. To verify that the underestimation coefficient $r = S_s/S_{sd}$, S_s and S_{sd} being respectively the standard deviations of $\tilde{D}_2$ when $d \to 0$ and when d assumes any of the other fixed values, mainly depends on d rather than the single values N and N_{tot}, each d was obtained setting N equal to 200, 600 and 1000 points.

The results of the test are shown in fig. 11. For $d < 0.8$ a good consistency between the three obtained estimates of r, the regularity of their trend *versus* d and the lack of a constant order between the estimates of r for the three different values of N makes significant a fit of all experimental data with the curve expressed by

$$r = 1.4\, d^2 + 0.1\, d + 1 \qquad d < 0.8. \tag{4.5}$$

The high dispersion of the three estimates and a likely sharp trend variation would require a greater number of synthetic tests to be carried out to extend the correlation interval to $d > 0.8$.

The test suggests that, though the greatest possible subset size is needed to give an estimate of D_2 with a minimum bias, there is a maximum number N of points which can be extracted from a set of experimental data such

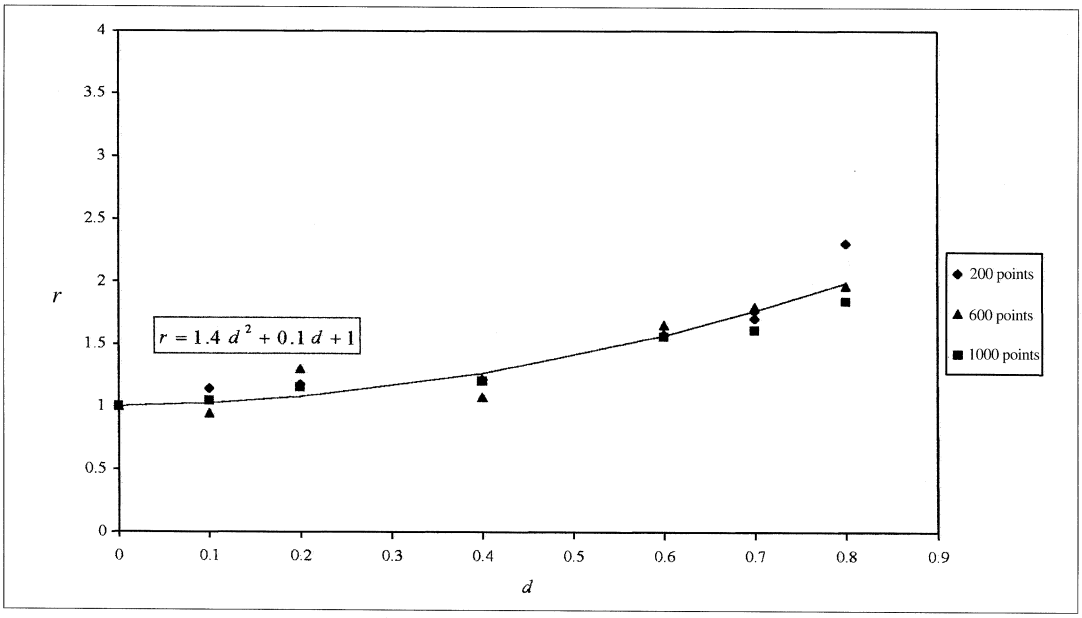

Fig. 11. Relation between the underestimation coefficient and parameter d.

Table I. Results of the comparison between 95%-confidence intervals of $\tilde{D}_2$ estimated by methods B and C; the $\bar{S}_s$ are the average values obtained on sets with each element characterized by a different d.

	Method C		Method B	
N	$\bar{S}_s$	$\sigma(S_s)$	$\bar{S}_s$	$\sigma(S_s)$
200	0.374	0.081	0.389	0.081
600	0.301	0.057	0.327	0.118
1000	0.271	0.028	0.283	0.058

that a reliable confidence interval of $\tilde{D}_2$ can be found by eq. (4.5).

To assess the congruence of the two approaches in estimating the standard deviation of $\tilde{D}_2$ and to supply a third method to that purpose (method C), from the M subsamples of each set of size N_{tot} the M uncertainties of the regression line slopes were estimated, and their mean value used in eq. (4.1) to calculate S_s. This was later compared with the corresponding sample confidence interval, computed for the same M subsamples by eq. (4.5).

Table I summarizes the results of the test, which showed a satisfying congruence between the two analysis techniques and a high efficiency of the two estimators of S_s. An optimum way of obtaining point estimates of S_s consists in an integrated application of methods B and C.

5. Experimental data

The first scope of the determination of $\tilde{D}_2$ and its relative confidence intervals from experimental data (natural and induced seismicity) is to find out, at least from a qualitative point of view, how the generating procedure affected the

results obtained from synthetic data, therefore to validate the method of simulation used.

The second scope is to test the possibility of differentiating, using these two parameters, the seismicity relative to volumes characterized by different seismogenetic conditions.

As regards natural seismicity, the analysis was carried out using four sequences of earthquakes (fig. 12) which occurred respectively:

– In the Aeolian Islands area (Sicily) between 1988 and 1992.

– In the area of Mount Etna (Sicily) between 1989 and 1991 (file 1, table II).

– In the area of Mount Etna between 1988 and 1992 (file 2, table II).

– In the area of Pollina (Madonie Mountains - Sicily) between 1993 and 1994.

Five clusters of mining-induced events (Cosentino *et al.*, 1997) which occurred in Wujek coal mine (Poland) in the period 1988-1993 have also been analyzed. These clusters are located in different areas of the mine: two groups, identified as 501_CS and 501_SW, are related to the exploitation carried out at mining level 501; three other groups, 510_NE, 510_NW and 510_SE, refer instead to a work which took place at the deeper level 510.

Table II shows the total number N_{tot} of seismic events, the number M of extracted subsets, the number N of events in each subset, the estimate $\tilde{D}_2$ for the whole data set, the mean value $\langle \tilde{D}_2 \rangle$ of the M estimates of D_2 and its 95%-confidence intervals resulting from methods B and C.

It is possible to observe that the correlation dimension for the epicentral distribution of the Aeolian Islands earthquakes is approximately 1.38, significantly lower than the values estimated for Mount Etna and the Pollina zone, which are, respectively, about 1.72 and 1.79.

These differences could hardly be imputed to the estimate bias connected with the lower number of events in the Aeolian Islands data set, as indicated by the very small differences between the estimates found for the Pollina and Etna zones by setting N respectively equal to 100 and 300. They proved in fact to be on average less than 0.1 times those found between the same areas and the Aeolian Islands.

In accordance with eq. (3.1), this probably indicates that the clustering of events in the case

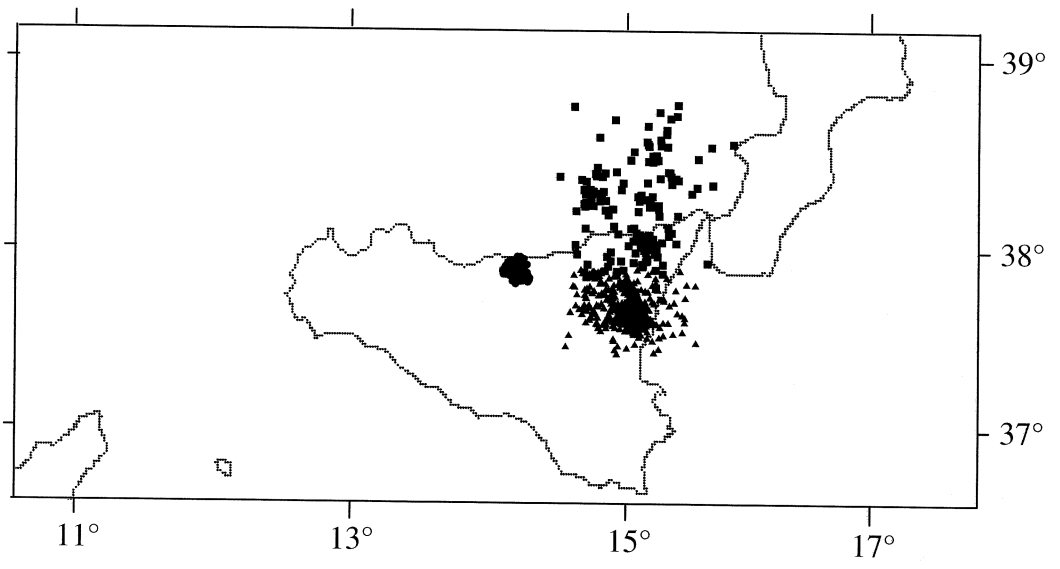

Fig. 12. Map of epicenters: circles = Pollina; triangles = Etna; squares = Aeolian Islands.

Table II. Results of the analysis on the experimental data.

Area	N_{tot}	M	N	d	$\tilde{D}_2$	$\langle \tilde{D}_2 \rangle$	Method C	Method B
Pollina	557	100	300	0.538	1.675	1.789	0.039	0.012
Etna File 1	607	100	300	0.494	1.780	1.734	0.038	0.016
Etna File 2	461	100	300	0.650	1.767	1.721	0.041	0.013
Aeolian Islands	170	100	100	0.588	1.380	1.382	0.042	0.021
501_CS	1572	100	500	0.318	1.616	1.601	0.035	0.009
501_SW	4816	100	1500	0.311	1.617	1.550	0.037	0.007
510_NE	360	100	100	0.277	1.430	1.402	0.045	0.025
510_NW	272	100	100	0.367	1.630	1.587	0.053	0.031
510_SE	1682	100	500	0.297	1.567	1.555	0.044	0.010

of the Aeolian Islands area is significantly more intense than that observed in areas characterized by volcanic seismicity or seismic swarms (Pollina).

Godano *et al.* (1996) analyzed a set of earthquakes which occurred in the Aeolian Islands, the Messina Strait and Mount Etna, evaluating the complete spectrum of the generalized fractal dimension (Grassberger and Procaccia, 1983) after splitting the whole set of events into five differently located clusters. The authors found a $\tilde{D}_2$ equal to ~1.2 in the zone of the Aeolian Islands, for a sample similarly sized to that used here and for events which occurred in about the same time interval, though ten years before.

The two estimates seem significantly different considering both their respective standard deviations and the almost coincident sample size. Their differences might be due to the presence of low-magnitude events ($M < 3$) in the analysis carried out by Godano *et al.* (1996), which determined an increase in clustering, typical of aftershock sequences. However, they could also be related to a non-stationariness of the phenomenon on the temporal scale of observation and to a better localization of seismic events in that analysis, since a local velocity model and data recorded by local seismological networks had been used.

Epicentral distributions Etna file 1 and Etna file 2 were referred to the same area, but the Etna file 2 earthquakes occurred in a longer period of time. The Etna file 1 earthquakes were also integrated with recordings made by the Calabrian Regional Seismic Network and localized using a different pseudo 3D-velocity model. The results indicate that the clustering properties evidenced by the two sets of epicenters do not significantly differ from each other.

Table II also reports the results of the analysis relative to seismic events induced by mining. Though a considerable homogeneity in the values $\tilde{D}_2$ is indicated by the analysis results, the estimate obtained for data set 510_NE is significantly lower than those obtained for the other data sets. Though this difference could be ascribed to an underestimation effect due to the lower N by which the test had been carried out, it can be observed that, since a lower fraction of high-magnitude events is present in the set, such a low value could be related to the actual geometric features of the distribution.

A similar statement can be made for data set 501_CS, for which the highest $\tilde{D}_2$ was obtained, probably connected with the higher fraction of high-magnitude events present in the set.

6. Conclusions

In order to carry out quantitative evaluations of the uncertainties which are associated with

the study of the fractal properties of epicentral distributions, multicluster point distributions generated on planes differently oriented with respect to the analysis plane were studied. These distributions have statistical fractal features and match seemingly with clustering characteristics of seismic epicenters.

Although the evaluations made so far do not yet have satisfying statistical significance, as the large number of parameters controlling the phenomenon requires processing a very large amount of data, the simulation and analysis techniques which have been described enable some useful conclusions to be drawn.

Some effects of the generating parameters on the estimation of $\tilde{D}_2$ were clearly observed, at least in a qualitative way.

An empirical law was determined, relating b to D_2 when the latter has the same value in each level of generation.

The effect of greater concentration along particular directions present in the data was also studied. This could be related to the average orientation of fault planes in a seismogenetically homogeneous volume.

The systematic effects on the estimates $\tilde{D}_2$ and on their uncertainties by the sample size and the random errors in points determination were analyzed in relation to different combinations of parameters b, θ of the generating algorithm.

The application of the same methodology of analysis to experimental data sets and the comparison of the results with those obtained from synthetic distributions have shown both the validity of the simulation carried out as well as the capability of the fractal parameters to characterize different kinds of seismic activity.

The possibility of integrating the characterization of seismogenetically homogeneous areas with the fractal dimension and its confidence interval seems to be proved, provided that experimental data are not affected by too large localization errors and that the sets are large enough to allow the extraction of subsamples of equal size, *i.e.* affected by a similar bias, from sets relative to different areas, such that significant estimations can be made.

Under these circumstances even small differences between the $\langle \tilde{D}_2 \rangle$ can be interpreted as an effective genetic heterogeneity. On the contrary, evaluation and interpretation carried out on the basis of estimates made using data sets containing only a limited number of points can lead to erroneous conclusions.

REFERENCES

AKI, K. (1981): A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction*, edited by D.W. SIMPSON and P.G. RICHARDS, American Geophysical Union, Washington, D.C, 566-574.

COSENTINO, P., L. DE LUCA, S. LASOCKI and D. LUZIO (1997): Evaluation of fractal dimension estimates: quantitative differentiation of seismicity clusters, in *Proceedings 4th International Symposium on Rockbursts and Seismicity in Mines*, edited by S.J. GIBOWICZ and S. LASOCKI (Balkema Rotterdam), 45-47.

GODANO, C., M.T. ALONZO and A. BOTTARI (1996): Multifractal analysis of the spatial distribution of earthquakes in Southern Italy, *Geophys. J. Int.*, **125**, 901-911.

GRASSBERGER, P. and J. PROCACCIA (1983): Measuring the strangeness of strange attractors, *Physica 9D*, 189-208.

MÖLLER, M., W. LANGE, F. MITSCHKE, N.B. ABRAHAM and U. HÜBNER (1989): Errors from digitizing and noise in estimating attractor dimensions, *Phys. Lett. A*, **138**, 176-182.

NERENBERG, M.A.H. and C. ESSEX (1990): Correlation dimension and systematic geometric effects, *Phys. Rev. A*, **42**, 7065-7074.

TURCOTTE, D.L. (1992): *Fractal and Chaos in Geology and Geophysics* (Cambridge University Press).

XIE, H. (1993): *Fractals in Rock Mechanics* (Balkema, Rotterdam).