Verifying the Dependence of Fractal Coefficients on Different Spatial Distributions

Dragomir Gospodinov [a,b], Elisaveta Marekova [a], Alexander Marinov [a]

[a] Plovdiv University "Paisii Hilendarski", 24, Tsar Asen Str., Plovdiv, Bulgaria, drago_pld@yahoo.com

[b] Geophysical Institute of Bulgarian Academy of Sciences, Akad. G. Bonchev Str., bl.3, Sofia, Bulgaria

Abstract. A fractal distribution requires that the number of objects larger than a specific size *r* has a power-law dependence on the size $N(r) = C/r^D \propto r^{-D}$ where *D* is the fractal dimension. Usually the correlation integral is calculated to estimate the correlation fractal dimension of epicentres. A 'box-counting' procedure could also be applied giving the 'capacity' fractal dimension. The fractal dimension can be an integer and then it is equivalent to a Euclidean dimension (it is zero of a point, one of a segment, of a square is two and of a cube is three). In general the fractal dimension is not an integer but a fractional dimension and there comes the origin of the term 'fractal'. The use of a power-law to statistically describe a set of events or phenomena reveals the lack of a characteristic length scale, that is fractal objects are scale invariant.

Scaling invariance and chaotic behavior constitute the base of a lot of natural hazards phenomena. Many studies of earthquakes reveal that their occurrence exhibits scale-invariant properties, so the fractal dimension can characterize them. It has first been confirmed that both aftershock rate decay in time and earthquake size distribution follow a power law. Recently many other earthquake distributions have been found to be scale-invariant. The spatial distribution of both regional seismicity and aftershocks show some fractal features. Earthquake spatial distributions are considered fractal, but indirectly. There are two possible models, which result in fractal earthquake distributions. The first model considers that a fractal distribution of faults leads to a fractal distribution of earthquakes, because each earthquake is characteristic of the fault on which it occurs. The second assumes that each fault has a fractal distribution of earthquakes. Observations strongly favour the first hypothesis.

The fractal coefficients analysis provides some important advantages in examining earthquake spatial distribution, which are:

- Simple way to quantify scale-invariant distributions of complex objects or phenomena by a small number of parameters.
- It is becoming evident that the applicability of fractal distributions to geological problems could have a more fundamental basis. Chaotic behaviour could underlay the geotectonic processes and the applicable statistics could often be fractal.

The application of fractal distribution analysis has, however, some specific aspects. It is usually difficult to present an adequate interpretation of the obtained values of fractal coefficients for earthquake epicenter or hypocenter distributions. That is why in this paper we aimed at other goals – to verify how a fractal coefficient depends on different spatial distributions. We simulated earthquake spatial data by generating randomly points first in a 3D space - cube, then in a parallelepiped, diminishing one of its sides. We then continued this procedure in 2D and 1D space. For each simulated data set we calculated the points' fractal coefficient (correlation fractal dimension of epicentres) and then checked for correlation between the coefficients values and the type of spatial distribution.

In that way one can obtain a set of standard fractal coefficients' values for varying spatial distributions. These then can be used when real earthquake data is analyzed by comparing the real data coefficients values to the standard fractal coefficients. Such an approach can help in interpreting the fractal analysis results through different types of spatial distributions.

Keywords: fractal, spatial distribution, earthquake, Monte Carlo simulations, uniform distribution
PACS: 91.30._f

INTRODUCTION

The results from the spatio-temporal analysis of seismicity provide insight both on seismic hazard estimation and on understanding the physical mechanisms underlying earthquake occurrence. The probabilistic seismic hazard assessment (PSHA) in dense-populated geographical regions and subsequently the design of the strategic objectives are primary based on the knowledge of the seismicity parameters of the seismogenic sources which can generate ground motion amplitudes above the minimum level considered risky at the specific site. The PSHA usually deals

CP1203, 7th *International Conference of the Balkan Physical Union,* edited by A. Angelopoulos and T. Fildisis

with a Poissonian earthquakes distribution in time and uniform distribution in space. This assumption implies the necessity of a catalog containing only independent events.

Similar studies of temporal and spatial features of the seismic process also help to better understand the geotectonic settings of a region, to identify earthquake interaction, especially considering triggering seismicity and to determine the size of a strong event source zone etc. There are different approaches to examine hypocenter locations. One way of analyzing the spatial distribution of seismicity is to determine the fractal dimension (D-value). This D-value is an extension of the Euclidean dimension and measures the degree of clustering of earthquakes: in a 3D-space, D can be a decimal number and ranges from 0 (point) to 3.0 (uniform distribution in space). Within the fractal approach to studying the distribution of seismic event locations, different fractal dimension definitions and estimation algorithms are in use. Although one expects that for the same data set, values of different dimensions will be different, it is usually anticipated that the direction of fractal dimension changes among different data sets will be the same for every fractal dimension.

In this study we have chosen the correlation dimension to study the behavior of simulated data. The simulations were made after a uniform distribution by the application of Monte Carlo methods in different spatial volumes for 3D, 2D and 1D case. We estimated the correlation dimension for each simulated data set and then examined the coefficient variation with changes of the volumes size. The fact that uniform spatial data is analyzed reveals that the fractal dimension values will provide the maximum values for any real earthquake data coefficients. These upper limit dimension values can then be referred to by any researcher, who is doing analysis of earthquake scale-invariant properties.

DEFINITIONS OF FRACTAL COEFFICIENTS

In principle various fractal dimensions may be used as a quantitative measure of the degree of heterogeneity of seismic activity in fault systems. These in turn are controlled by the heterogeneity of the stress field and the pre-existing geological, mechanical or structural heterogeneity. However, fractal dimensions obtained by different methods generally reflect different aspects of scale invariance, and need not be equal or even positively correlated [1, 2]. Among various methods used to estimate the fractal dimension of hypocenter or epicenter distributions, the three most frequently used are:

- Box-counting method which gives estimates of the capacity dimension, defined by

$$D_c = \log N(r) / \log(1/r)\,, \tag{1}$$

Where $N(r)$ is the number of boxes (squares, cubes, etc.) of side r occupied by point sources. If the set of sources has a statistical fractal character, the log $N(r)$ vs. log $(1/r)$ relation can be approximated with the straight line in a certain range of r. The slope of the line assigns the value of capacity dimension [3, 4];

- Number-radius method which is used to evaluate the cluster dimension D_b. Let (x_i, y_i) denote the position of epicenter i. This method consists in counting the number of sources $N(r)$ located inside circles of radii r and centers located in the center of mass of the epicenter distribution, defined by the pair of expected values of the random variables X and Y. The slope of linear approximation of the log $N(r)$ vs. log (r) relation provides the estimate of cluster dimension [5];
- The correlation dimension estimate D_2 is defined through the two point correlation integral $C(r)$, which is presented as

$$C(r) = \frac{1}{N^2} \sum_{i \neq j} H(r - L)\,, \tag{2}$$

where N is the total number of points, $L = \left\| \overline{x_i} - \overline{x_j} \right\|$ is the distance between position vectors $\overline{x_i}$ *and* $\overline{x_j}$ of points and $H(z)$ is the Heviside function equating to 1.0 when $z \geq 0$ and to 0.0 when $z < 0$. If points spatial distribution exhibits scale-invariance then $C(r)$ must follow a power law

$$\lim_{r \to 0} C(r) \propto r^{D_2}\,. \tag{3}$$

And D_2 can be presented in the form

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r}, \tag{4}$$

which reveals that D_2 is the fractal dimension achieved by the integral correlation method as the slope of the straight line, which relates the number of pairs of sources whose mutual distances are smaller than a certain value of r, with r in a double logarithmic scale [6]. D_2 is often used to evaluate the structure of a point collection distributed in space and this fractal dimension is put to examination in the present study.

While estimation procedures for different fractal coefficients are fairly well defined, it is not easy to interpret the obtained coefficients' values. For 3D data for example (earthquake hypocenters) the maximum possible coefficient's value is 3 and if one obtains a smaller value for a real data set, this is considered as an indication of fractal properties for this data. The determined estimates, however, depend on the number of the analyzed data points and on the type of space, studied (cubic, parallelepiped space etc.). Uniform distribution of spatial points is considered to reveal opposite features to fractal behavior. The coefficients' estimates for such data, however, reveal the maximum possible values and any real data set with fractal properties will deliver a coefficient's value less than the one for uniform data. That is why we decided to estimate the coefficients' values for simulated data of uniform spatial distribution in different types of space volumes. These will yield the upper bounds of the fractal coefficient values for any real data set from a similar type of space. The purpose of this study is to examine the variability of the correlation dimension D_2, estimated for different simulated data sets, following a uniform space distribution. The simulation is done for 3D, 2D and 1D case.

DEPENDENCE OF FRACTAL COEFFICIENTS ON DIFFERENT SPATIAL DISTRIBUTIONS

As mentioned above, we consider simulated data sets of uniformly distributed points for 3D, 2D and 1D cases in different types of space volumes. For each data set we estimate the correlation dimension D_2 and after that we construct the relation between the coefficients' values and the type of space volume used.

In the case of an infinite fractal distribution, the resulting plot of $\log_{10}C(r)$ against $\log_{10}(r)$ will be a straight line whose gradient is the fractal dimension. In practice, however, it is found that for large values of 'r' the gradient is artificially low, whereas for small values of r the gradient is artificially high [7]. These two conditions have been called ''depopulation'' and ''saturation.'' Whereas it is common for an estimate of the fractal dimension to be made by fitting a straight line to a subjectively-chosen straight part of the curve, equations (5) provide formulae for determining the distances of depopulation and saturation, r_n and r_s as a function of the embedding dimension d, the number of data N, and the linear size R, of the hypercube encompassing a given data [8]:

$$r_n = R\left[\frac{1}{N}\right]^{1/d}; \qquad r_s = \frac{R}{2(d+1)}. \tag{5}$$

As discussed by it is often safe to start the scaling range at values of r as low as $r_n/3$, but in the case studied here we choose the more conservative approach of measuring a gradient from r_n.

As in this paper the fractal dimension was estimated as the correlation dimension, the objective is therefore to use the minimum number of events per subset that gives the most reliable estimate of the correlation dimensions. A proper scaling range is the one between r_n and r_s. For $r_s < r_n$, this implies that there is no proper scaling range for the data set. Otherwise, one may derive a straight segment of the correlation integral on a log-log plot that appears as a power law scaling, but is not a true reflection of the underlying distribution. Equations (5) are employed to compute the minimum number of events N that gives a scaling range (i.e., $r_n < r_s$) for the epicentral distribution. Formulae (5) have been used to determine the distances of depopulation and saturation, r_n and r_s for each data set and then to estimate the data fractal dimension.

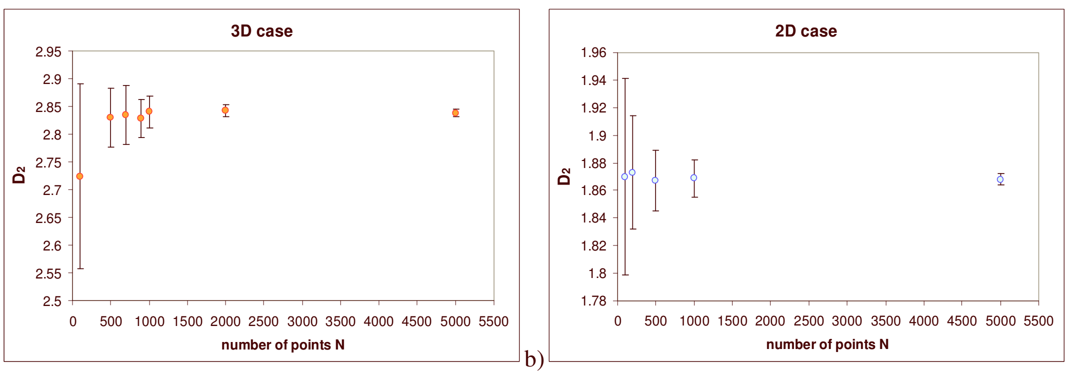
Simulation of Uniformly Distributed Spatial Data Sets

For the simulation of data we developed a standard Monte Carlo program in FORTRAN using well known random number generator *ran2* with long period (> 2×10^{18}), highly recommended by Press and Teukolski, who state that *ran2* provides perfect random numbers [9].

Reference of Fractal Coefficients to the Number of Simulated Points

We have first analyzed the dependence of the correlation dimension D_2 on the number of simulated points in a space volume for 3D and 2D.

3D case – we generated uniformly distributed points in a cube of dimensions 1x1x1. Several different cases are considered for number of simulated points N = 100, 500, 700, 1000, 2000, 5000, and for each case the correlation coefficient D_2 was estimated. The linear part of the relation (2) was chosen after formulae (5). We performed 20 simulations for each specified N and then calculated the D_2 average and standard deviation for these simulations. The dependence of the average fractal coefficient D_2 on the number of simulated points is graphically revealed on Figure 1a together with the corresponding error bounds. The results unclose that the coefficient values increase with the number of simulated points. The values rise, however, is quite small after several thousands of points and this motivated us to choose the upper limit of 5000 simulated points for our experiment.



a)

FIGURE 1. Relation between the correlation dimension D_2 values and the number of simulated points; a) 3D case; b) 2D case.

2D case – a similar procedure was executed for the 2D space volume. Different numbers of points (N= 100, 200, 500, 1000, 5000) were generated to follow a uniform distribution in a 2D space volume of dimensions 1x1. We used formulae (5) to choose a linear part of relation (2) and estimate the correlation coefficient D_2. A number of 20 simulations was performed for each specified N and then we calculated the D_2 average and standard deviation for these simulations. The relation of the average fractal coefficient D_2 on the number of simulated points is graphically presented on Figure 1b together with the corresponding error bounds. It can be seen that again the coefficient values increase with the number of simulated points but this tendency is much less evident in the 2D case compared to the 3D one. We again chose the upper limit of 5000 simulated points for our experiment.

Relation of Fractal Coefficients to Points' Spatial Distribution (3D case)

We simulated uniform spatial distribution of points in different 3D space volumes. We started with a cube of dimensions 1x1x1 and then diminished one of the cube's dimensions to examine other cases (1x1x0.5; 1x1x.25; 1x1x0.125). In each of these volumes we generated a set of 5000 uniformly distributed random points. Then the correlation integral was calculated through all possible distances between the generated events and formulae (5) were applied to identify a linear part of formula (2) and calculate the D_2 fractal dimension (an example of the correlation integral plot is shown on Figure 2a). Our purpose was to follow the variation of the D_2 coefficient with changing size of the 3D volume.

The results are plotted on Figure 2b and they point out that the D_2 fractal dimension decreases when one of the 3D volume sizes is lowering. Considering that we analyze a case for uniform distribution, the corresponding values of the D_2 coefficient for each volume provide the upper limits of the fractal coefficient values for any real earthquake hypocenter data. The results mean that for real hypocenter data we have to obtain smaller fractal coefficients' values, compared to the ones of uniform points distribution, in order to consider that the hypocenters follow scale-invariant behavior.

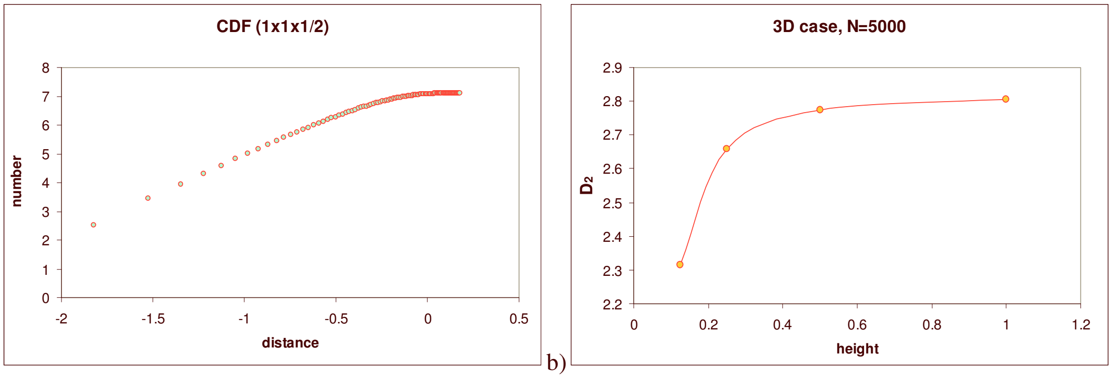


FIGURE 2. a) An example of the correlation integral for 3D (1x1x0.5 case); cumulative number of points versus the distance between them in a double-logarithmic scale; b) Relation between the D_2 fractal dimension and the different 3D volumes (1x1x1; 1x1x0.5; 1x1x.25; 1x1x0.125); the horizontal axis reveals the changing size of the 3D volumes axis.

Relation of Fractal Coefficients to Points' Spatial Distribution (2D case and 1D case)

All the above analysis was also carried out for 2D and 1D case. For the former we simulated uniform spatial distribution of points in different 2D space volumes (1x1; 1x0.5; 1x0.25; 1x0.125). A number of 5000 random points was generated in each volume after a uniform 2D distribution. Then all possible distances between the generated events were calculated and used to construct the correlation integral for each zone. Finally the D_2 fractal dimension was estimated and its values plotted as a function of the different 2D volumes (Figure 3a).

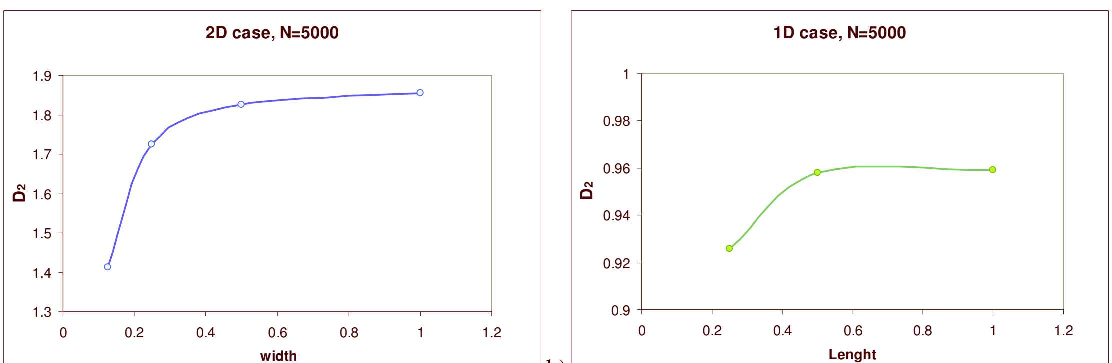


FIGURE 3. a) Relation between the D_2 fractal dimension and the different 2D volumes (1x1; 1x0.5; 1x0.25; 1x0.125); the horizontal axis reveals the changing size of the 2D volumes axis; b) The same as in 'a)' but for 1D case; the coefficient value is calculated for three different line segments L=1; 0.5; 0.25.

The results are similar as for the 3D case, exposing that the D_2 fractal dimension decreases with the decrease of one of the 2D volumes size. This reveals that for real earthquake epicenter data (2D) one has to obtain smaller fractal coefficients' values, compared to the corresponding ones of uniform distribution, so that epicenter fractal features are considered to be identified.

Analogous procedures were applied for 1D case. We generated uniformly distributed points on segments of length L=1; 0.5; 0.25 the number of points being 5000 for each segment. The plot on Figure 3b reflects the change of the estimated D_2 fractal dimension for the different segments. Again one can observe that the coefficient's value gets smaller when the line's size decreases.

The 1D case can rarely be considered for real earthquake data. Sometimes, however, a number of events may occur on a single fault (especially aftershocks) and then the results from Figure 3b supply the upper limit values of fractal dimension if events on the fault exhibit scale-invariant properties.

CONCLUSIONS

In this paper we applied the Monte Carlo techniques to provide simulated data sets, following a uniform spatial distribution for 3D, 2D and 1D case in different types of space volumes. For each data set we estimate the correlation dimension D_2 and after that we construct the relation between the coefficients' values and the type of space volume used. We start for 3D case from a cube of dimensions 1x1x1 and decreasing one of its sizes to 0.5, 0.25, and 0.125. For the 2D case the start volume is quadratic 1x1and then we consider volumes 1x0.5, 1x0.25, and 1x0.125. And finally we examine three line segments of length L=1; 0.5; 0.25 for 1D case.

We first examined the relation between the correlation dimension D_2 values and the number of simulated points (Figure 1). The results point out that the coefficient values increase with the number of simulated points but this increase gets very small after several thousands of points and. This encouraged us to choose a number of 5000 simulated points for our experiment. So, for any of the above spatial volumes we generated 5000 uniformly distributed random points and calculated the fractal dimension D_2.

The plots on Figure 2b and Figure 3a, b expose that the D_2 values decrease when one of the volume sizes decreases. Considering that we analyze uniform distributions, the corresponding values of the D_2 coefficient for each volume provide the upper limits of the fractal coefficient values for any real earthquake data in a similar volume. So, the results from the present study provide a set of correlation dimension D_2 values for different spatial volumes. If one starts a fractal analysis of earthquake spatial data, he can consider the spatial volume his data occupies and refer to the corresponding maximum value of the fractal dimension for this volume. If for real data a smaller value of D_2 is obtained, this can point to possible fractal properties of the data.

ACKNOWLEDGMENTS

This study was supported by Project RS09-FF-011/15.04.2009 of Plovdiv University 'Paisii Hilendarski', Bulgaria.

REFERENCES

1. S. Lasocki, L. De Luca, *Pure appl. geophys*., 152, pp. 213–220, (1998)
2. C. Goltz, *Fractal and Chaotic Properties of Earthquakes*, Berlin, Springer-Verlag, pp. 187, (1998).
3. D. Turcotte, *Fractals and Chaos in Geology and Geophysics*, Cambridge, Cambridge University Press, pp. 398, (1997).
4. S. A. Preuss, "Some remarks on the numerical estimation of fractal dimension", in *Fractals in Earth Sciences*, eds. Ch. C. Bartone and P. R. La Pointe, New York, Plenum Press, 1995, pp. 65–76.
5. H. Xie, *Fractals in Rock Mechanics*, Rotterdam, A. A. Balkema Publishers, pp. 464, (1993).
6. P. Grassberger, I. Procaccia, *Phys. Rev. Lett*., 50, No. 5, pp. 346-349, (1983).
7. P. Mandal , B.K. Rastogi, *Pure appl. geophys.,* 162, pp. 53–72, (2005).
8. M. A. A. Narenberg and C. Essex, *Phys Rev A*, 42, pp. 7065–7074, (1990).
9. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN 77 - The Art of Scientific Computing, 2nd Ed.,* Cambridge, Cambridge University Press, (1992), pp. 272-273.