Bulletin of the Seismological Society of America, Vol. 77, No. 4, pp. 1368–1381, August 1987

A FRACTAL APPROACH TO THE CLUSTERING OF EARTHQUAKES: APPLICATIONS TO THE SEISMICITY OF THE NEW HEBRIDES

By R. F. Smalley, Jr., J.-L. Chatelain, D. L. Turcotte, and R. Prévot

Abstract

The concept of fractals provides a means of testing whether clustering in time or space is a scale-invariant process. If the fraction x of the intervals of length τ containing earthquakes is related to the time interval by $x \sim \tau^{1-D}$, then fractal clustering is occurring with fractal dimension D ($0 < D < 1$). We have analyzed a catalog of earthquakes from the New Hebrides for the occurrence of temporal clusters that exhibit fractal behavior. Our studies have considered four distinct regions. The number of earthquakes considered in each region varies from 44 to 1,330. In all cases, significant deviations from random or Poisson behavior are found. The fractal dimensions found vary from 0.126 to 0.255. Our method introduces a new means of quantifying the clustering of earthquakes.

Introduction

Many studies have been performed on the distribution of earthquakes in time and space in order to better understand the earthquake generation process and possibly assist in earthquake prediction. If the occurrence of each earthquake was totally uncorrelated with other earthquakes, then the distributions of earthquakes would be a random process. The random distribution of point events is known as a Poisson process, and the mathematics of this process is well understood.

In studies of regional seismicity, distributions described as nests, swarms, and clusters are often observed. Even after the elimination of aftershock sequences, which are correlated to the main shock, the seismic distributions are not Poisson. Regional examples include southern California (Knopoff, 1964; Johnson *et al.*, 1984), central California (Udias and Rice, 1975) Nevada (Savage, 1972), New Mexico (Singh and Sanford, 1972), Fiji-Tonga-Kermadec region (Isacks *et al.*, 1967), New Zealand (Vere-Jones and Davies, 1966), and southern Italy (Bottari and Neri, 1983; De Natale and Zollo, 1986). A variety of statistical methods have been applied in order to quantify deviations from random occurrences (Shlien and Toksöz, 1970; Vere-Jones, 1970; Vere-Jones and Ozaki, 1982; Matsumura, 1984; Dziewonski and Prozorov, 1984). In general, these approaches have been empirical in nature.

Just as Poisson statistics model purely random processes, fractal statistics model processes that exhibit scale-invariant properties such as scale-invariant clustering. Mandelbrot (1967, 1975, 1982) developed the concepts of fractal geometry and fractal dimension, and applied them to the description of many natural features such as coastlines. A fractal curve is defined as a curve whose length or perimeter P is a function of the length of the measuring rod l such that

$$P \sim l^{1-D} \tag{1}$$

where D is the fractal dimension. Applying this definition to coastlines (where D is limited to the range $1 \leqq D \leqq 2$), typical values of D are found to be near 1.2. The fractal dimension measures the "ruggedness" of the coastlines, smaller D's represent smooth coastlines while larger D's represent rugged coastlines.

A fractal distribution of areas, such as islands or craters, is defined as a distribution where the number of objects, N, with a characteristic linear dimension

greater than r is given by

$$N \sim \frac{1}{r^D}. \tag{2}$$

Mandelbrot (1975) pointed out that the number-area relation for islands satisfies (2) with $D = 1.30$. Turcotte (1986) has shown that fragments of rock produced by weathering, explosions, and impacts often satisfy (2) over a wide range of scales. Aki (1981) has shown that the standard frequency-magnitude relation for earthquakes is equivalent to (2) with $r = A^{1/2}$ where A is the area of the fault break. The fractal dimension D is related to the b value of the distribution by $D = 2b$. A fractal approach to tectonics has been proposed by King (1983) and extended by Turcotte (1986/1987).

In this paper, the concept of fractal or scale-invariant clustering will be applied to earthquakes. Earthquake epicenters can be considered to be point events in space and time. To study fractal clustering, we must examine the distribution of events over a wide range of scales. In this paper, we will be primarily concerned with the temporal clustering of earthquakes because timing of events (event times are typically known to better than a minute) and the total length of time available for study (typically years) provide a wide range of scales for analysis. The approach can also be applied to spatial clustering. In this study, however, the region covered with continuous recording in the Cornell/ORSTOM network (approximately 2.5° by 2°) and the resolution of event locations (0.05° at best) limits the range of scales available for a spatial study (Chatelain *et al.*, 1986).

In order to quantify the clustering of earthquakes in the New Hebrides region, we will consider the two limits of random (Poisson) point events and fractal (scale-invariant) clustering of a finite set of point events. First, a discussion of the Poisson distribution will be given. This allows us to introduce the basic terminology and concepts. We then modify the fractal approach to the scale-invariant clustering of an infinite set of point events to a finite set. In particular, the relationship between the fractal dimension D and the degree of clustering will be discussed. The application to the New Hebrides data will illustrate the fractal approach to the temporal distribution of seismicity.

FUNDAMENTAL APPROACH

The basic problem we consider is the temporal distribution of N earthquakes occurring in a time interval τ_0. The number of earthquakes is determined by the area of the region considered and the magnitude cut-off. For practical considerations, the magnitude cut-off is made to insure that all earthquakes in the region are detected by the regional network. The interval τ_0 is the interval over which data is collected. The natural period for the data is is τ_0/N.

In order to analyze the available data, we divide the interval τ_0 into a series of smaller intervals τ defined by

$$\tau = \frac{\tau_0}{n}, \quad n = 2, 3, 4, \cdots. \tag{3}$$

Our measure of clustering will be the fraction $x(\tau)$ of the intervals in which an earthquake occurs as a function of the interval length τ.

As a specific example, consider a uniform (equally spaced in time) series of earthquakes. For this example,

$$x = \begin{cases} \dfrac{\tau N}{\tau_0} \text{ if } \tau < \dfrac{\tau_0}{N} \\ 1 \quad \text{if } \tau > \dfrac{\tau_0}{N} \end{cases} \tag{4}$$

If the number of earthquakes is greater than the number of intervals, then an earthquake occurs in every interval and $x = 1$. If the number of earthquakes N is less than the number of intervals, n, then $x = N/n$.

Poisson Distribution

If the distribution of N earthquakes is totally random, it is a Poisson distribution. For this case, there is distribution of intervals between earthquakes given by (Cox and Lewis, 1966, p. 22)

$$\Pr(\tau_e) = \frac{\tau_0}{N} \exp\left(-\frac{N\tau_e}{\tau_0}\right) \tag{5}$$

where $\Pr(\tau_e)\, d\tau_e$ is the probability that an earthquake will occur after an interval between τ_e and $\tau_e + d\tau_e$. Clearly, a fraction of the earthquakes will have an interval less than τ_0/N, and a fraction will have longer intervals.

We wish, however, to express the Poisson distribution in terms of the fraction x of intervals of length τ that are occupied by randomly occurring earthquakes. This is the classic problem of the random distribution of N balls distributed into $n = \tau_0/\tau$ boxes. The probability $\Pr(x)$ that the fraction of boxes occupied lies between x and $x + dx$ is given by (Feller, 1968, p. 102)

$$\Pr(x) = \binom{\tau_0/\tau}{\tau_0(1-x)/\tau} \sum_{\nu=0}^{\nu=\tau_0 x/\tau} (-1)^{\nu} \binom{\tau_0 x/\tau}{\nu} \left(x - \frac{\tau\nu}{\tau_0}\right)^N \tag{6}$$

where the binomial coefficient is defined by

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}. \tag{7}$$

For earthquake distributions, N and τ_0/τ are generally large. In this limit, $\Pr(x)$ from (6) has a strong maximum at a specific value of x, x_0.

Fractal Approach to Clustering

The fractal approach to clustering is based upon the Cantor set model (Figure 1). To construct a Cantor set, start with the closed interval [0, 1]. The middle third is then removed leaving two closed intervals each with a length $\frac{1}{3}$. Continue by deleting the middle third of these two intervals, leaving four closed intervals of length $\frac{1}{9}$. The process of removing the open middle third of the remaining closed intervals is repeated an infinite number of times and defines the Cantor set. This set exhibits scale-invariant clustering that can be quantified by the fractal dimension.

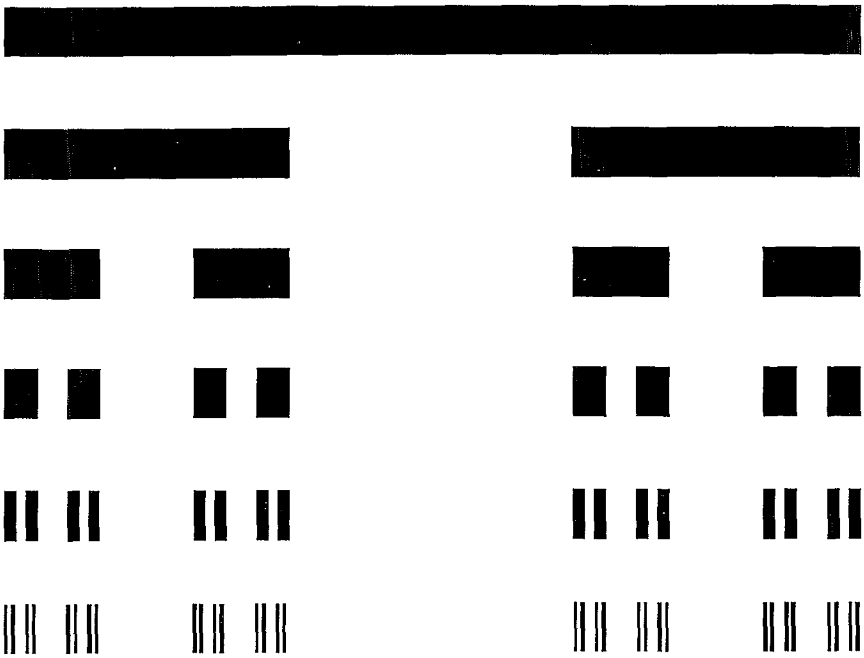

FIG. 1. Illustration of a Cantor set. At each step, the middle third of each solid line is removed. The process of generating thinner and thinner lines is referred to as curdling. The infinite set of lines generated is known as a Cantor dust.

Returning to the definition of the fractal set (2), we write

$$N_i \sim \frac{1}{r_i^D} \tag{8}$$

where N_i is the number of objects with size r_i and for the Cantor set $N_{i+1}/N_i = 2$ and $r_{i+1}/r_i = \frac{1}{3}$. Thus,

$$D = -\frac{\ln(N_{i+1}/N_i)}{\ln(r_{i+1}/r_i)} = \frac{\ln 2}{\ln 3} = 0.6309. \tag{9}$$

The process of generating thinner and thinner lines is referred to as curdling by Mandelbrot (1982). The infinite set of objects generated by this process is known as a Cantor dust.

The fraction x of the steps of length r that include dust is given by

$$x_i \sim r_i^{1-D}. \tag{10}$$

For the Cantor set $x_{i+1}/x_i = \frac{2}{3}$ and the fractal dimension $D = \ln 2/\ln 3 = 0.6309$ can be obtained from either (8) or (10).

The Cantor set illustrated in Figure 1 is both scale-invariant, highly ordered, and deterministic. A scale-invariant random set is easily generated by randomly removing the first, middle, or last third of each closed interval instead of the middle third. This process is illustrated in Figure 2. The result is a scale-invariant fractal clustering that is also nondeterministic and has a fractal dimension $D = \ln 2/\ln 3$

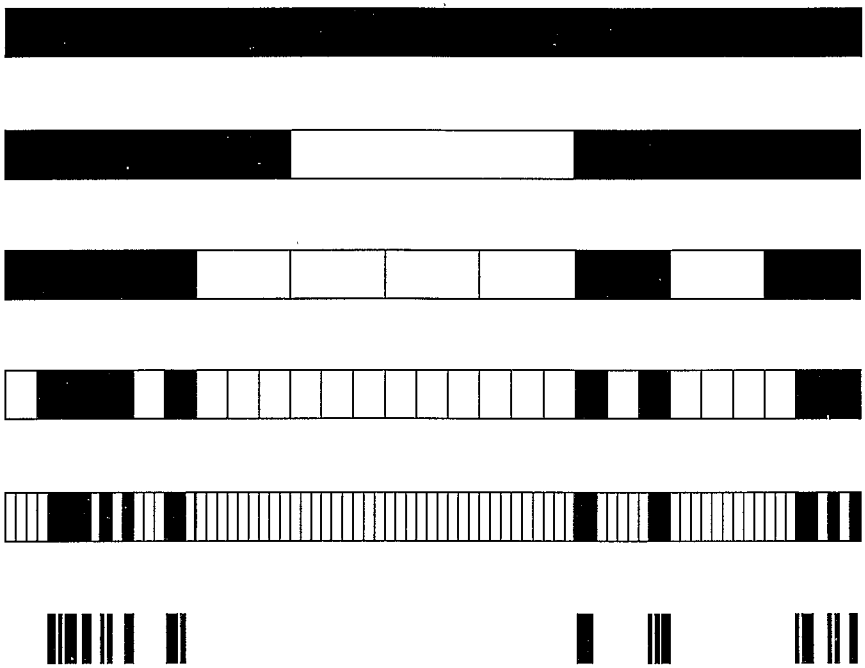

FIG. 2. Illustration of the first six orders of a random Cantor set. At each step, a random third of each solid line is removed.

= 0.6309. For actual applications of clustering, the number of events is finite. The number of events associated with a Cantor set of order m is 2^m. Our approach to the fractal analysis of clustering is illustrated in Figure 3, showing the results for a random Cantor set of order 9. First, we rescale the length so that the width of an element of the truncated random Cantor set is unity giving a total length of 3^9; containing 2^9 elements. To measure the fractal dimension by the "box method," we take intervals of length $r = 2^n$ and determine the fraction x that include at least one element. This fraction is given as a function of the interval length by the open squares in Figure 3. The best-fit straight line has a slope of 0.368 so that $x \sim r^{0.368}$ and $D = 0.632$. The deviation from the exact value of 0.6309 is due to the fact that we have taken intervals of length 2^n rather than the intrinsic interval size of 3^n and therefore obtain an estimate of the fractal dimension. If the same number of lines is uniformly distributed across the interval (no clustering), the probability of finding a line within an interval from (4) is given by the solid diamonds in Figure 3. In this case, the slope is unity for $r < (\frac{3}{2})^9$ and zero for $r > (\frac{3}{2})^9$. Thus, $D = 0$ for $r < (\frac{3}{2})^9$, i.e., a set of isolated points, and $D = 1$ for $r > (\frac{3}{2})^9$, i.e., a line.

INTERPRETATION OF D

In the Cantor set illustrated in Figure 1, the first step was to remove the middle third of the solid line. To illustrate the relation between clustering and the fractal dimension, D, the first two iterations in the construction of a series of Cantor-like sets with increasing D are shown in Figure 4. In this series, the closed unit interval [0, 1] is divided into nine equal intervals of length, $r = \frac{1}{9}$, and at each iteration in the construction of the set, M of these smaller intervals are removed (note that we must now be careful in the choice of open or closed ends on the intervals). In Figure

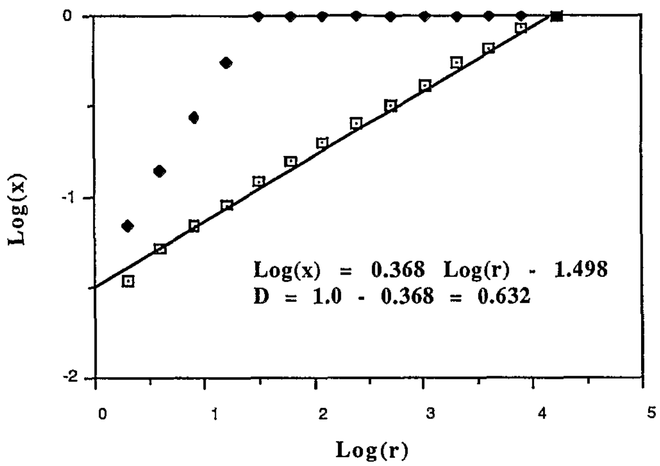

FIG. 3. The fraction x of steps of length r that include solid lines for a ninth-order random Cantor set are given by the open squares. The unit length is the length of the shortest line, the original line length is 3^9. The solid diamonds are for a uniform distribution of the same number of lines as given by (4). The solid line corresponds to (10) with $D = 0.632$.

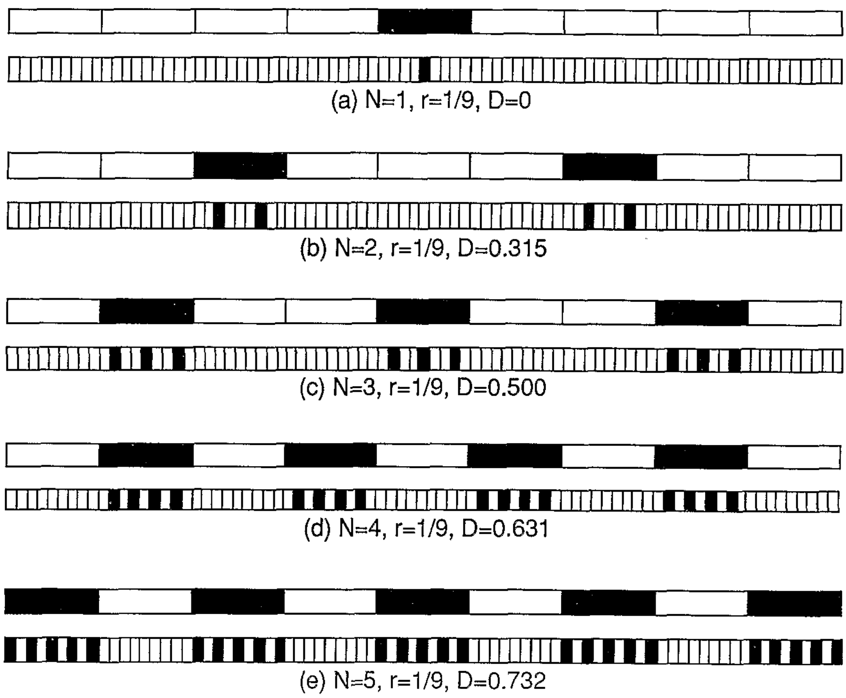

FIG. 4. Illustration showing relationship between clustering and fractal dimension. Clustered fractal sets with different values of D are constructed in an iterative process similar to that used for the Cantor set of Figure 1. The original solid line is divided into nine parts, and from one (a) to five (e) closed segments are retained at each iteration (e.g., in the first iteration in (a), the intervals [0, 1/9], [1/9, 2/9], [2/9, 3/9], [3/9, 4/9], [5/9, 6/9], [6/9, 7/9], [7/9, 8/9], [8/9, 1] are removed, leaving the closed interval [4/9, 5/9]). The corresponding values of D from (8) are also given.

4a, $M = 8$ intervals are removed, $N = 1 = 9 - M$ and from (8), $D = 0$. As the process is continued, only a point would remain and the fractal dimension, $D = 0$, of a point is equal to the usual Euclidean dimension, $E = 0$, of a point. In Figure 4b, $M = 7$ intervals are removed, $N = 2$ and from (8) $D = 0.315$. Sets with $M = 6$, 5, and 4, intervals removed, or $N = 3$, 4, and 5, are shown in Figure 4 (c to e) and have $D = 0.500$, 0.631, and 0.732, respectively. The fractal dimension describes the strength of the clustering: the more isolated the clusters the smaller the value of D.

Application to New Hebrides Seismicity

In order to test the applicability of the fractal approach to clustering, we will consider the temporal variation of seismicity in several regions near Efate Island in the New Hebrides Island Arc (map in Figure 5), an area with a high level of

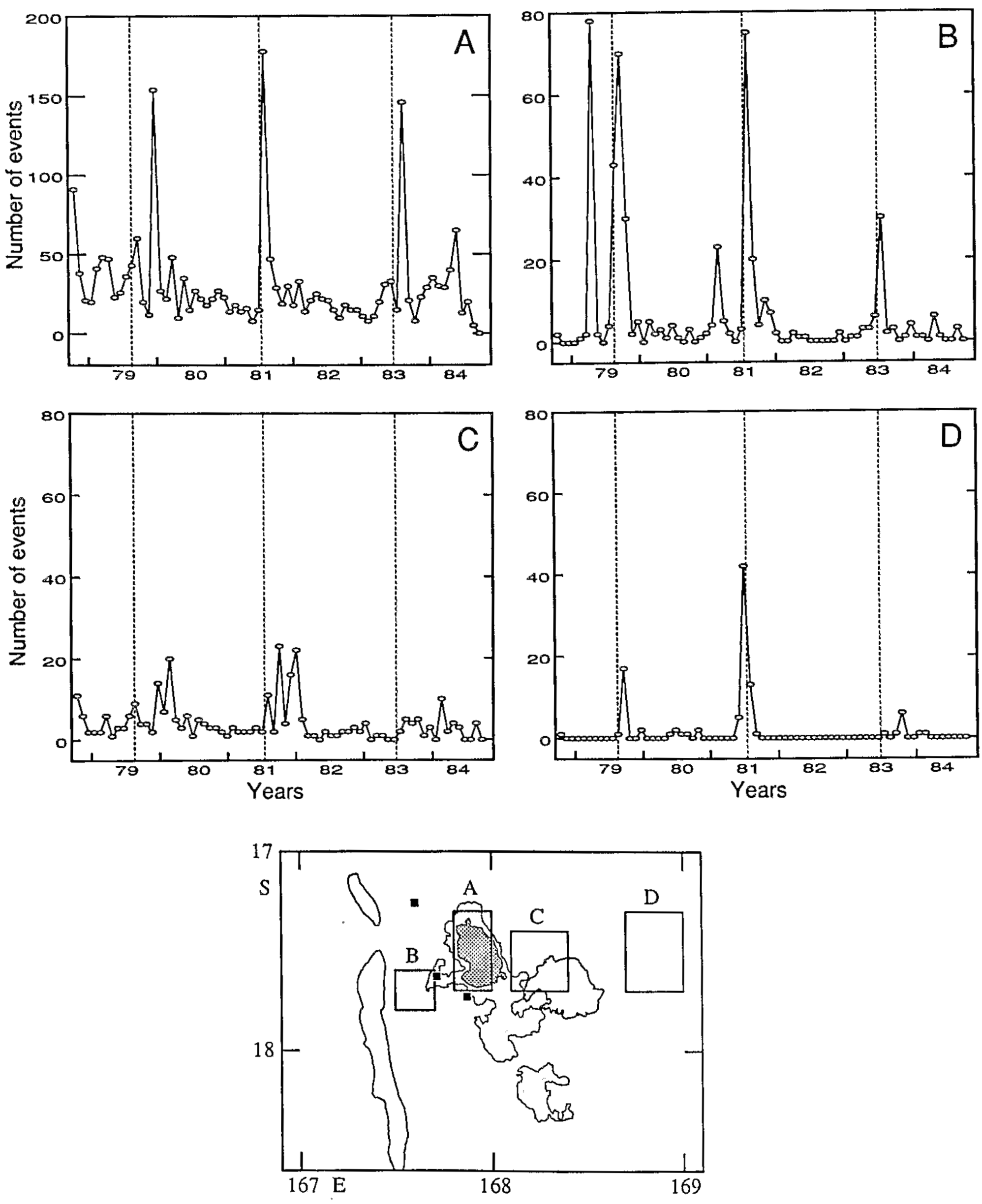

FIG. 5. Histograms of the number of earthquakes per month in four seismic regions in the New Hebrides island arc from Chatelain *et al.* (1986). The dashed lines indicate the occurrence of four major earthquakes.

seismicity in which a local network has provided a continuous record of events between mid 1978 and mid-1984. We have chosen the four zones in the Efate region defined by Chatelain *et al.* (1986) which contain clusters of earthquakes, as can be seen in the histograms of monthly seismic activity shown in Figure 5. The clusters,

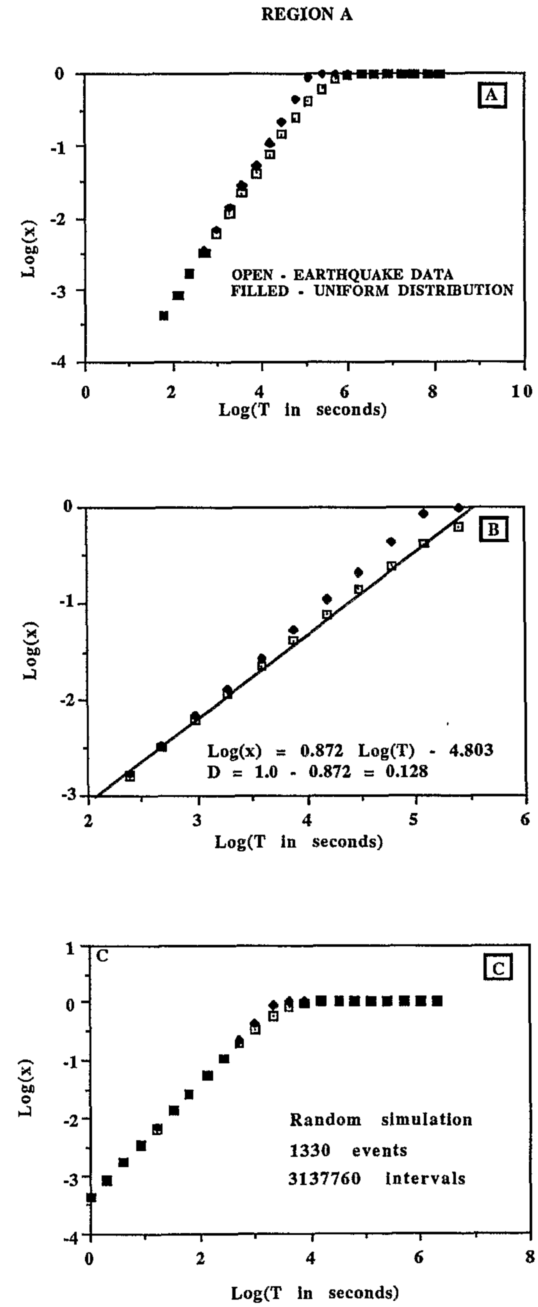

FIG. 6. (a) The open squares are the fraction x of time intervals of length t that contain earthquakes in region A. The solid diamonds are for a uniform distribution of 1,330 earthquakes as given by (4). (b) The best fit of the data from (a) with (10) in the range $4 < \tau < 4{,}096$ min. The derived fractal dimension is $D = 0.128$. (c) Open squares are a random simulation of 1,330 events for the Poisson process given by (6).

REGION B

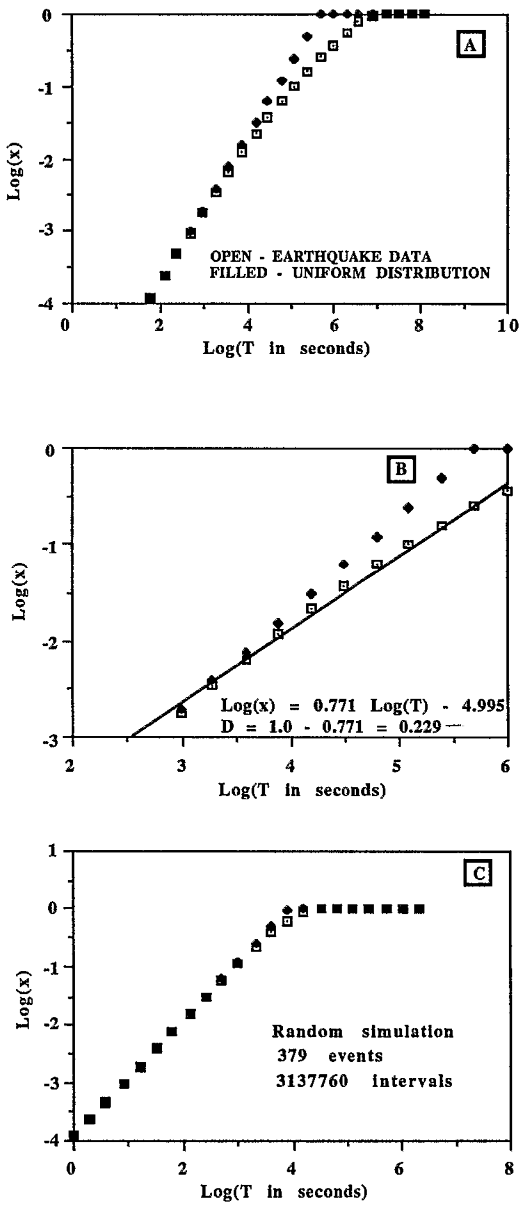

FIG. 7. Same as Figure 5 for region B.

which are observed in these zones, are related to the occurrence of moderate to large earthquakes in the up-dip part of the subducting plate (Chatelain *et al.*, 1986). In each region, log frequency versus magnitude plots were used to select the magnitude range in which the earthquake sample is complete.

Based on the time resolution, the number of events, and the total time of the data set, we have chosen the smallest time interval to be 1 min ($\tau = 60$ sec). This interval of time is increased by factors of 2 up to $2^{21} \sim 2 \times 10^6$ min. The total time interval for which data is available is about 3×10^6 min. For each zone, the fraction of intervals that include an earthquake, as a function of the interval size, is plotted, and the fractal dimension is determined from the slope.

For region A, 1,330 events occurred in the magnitude window. The results of the fractal analysis of the earthquake data are given in Figure 6A. For $\tau \leqq 4$ min, only single events occur in each time interval, and no clustering is observed. For $\tau \geqq$ 32,768 minutes = 22.8 days, every interval contains an earthquake. For values of τ between these limits, a fractal clustering is observed which gives a fractal dimension

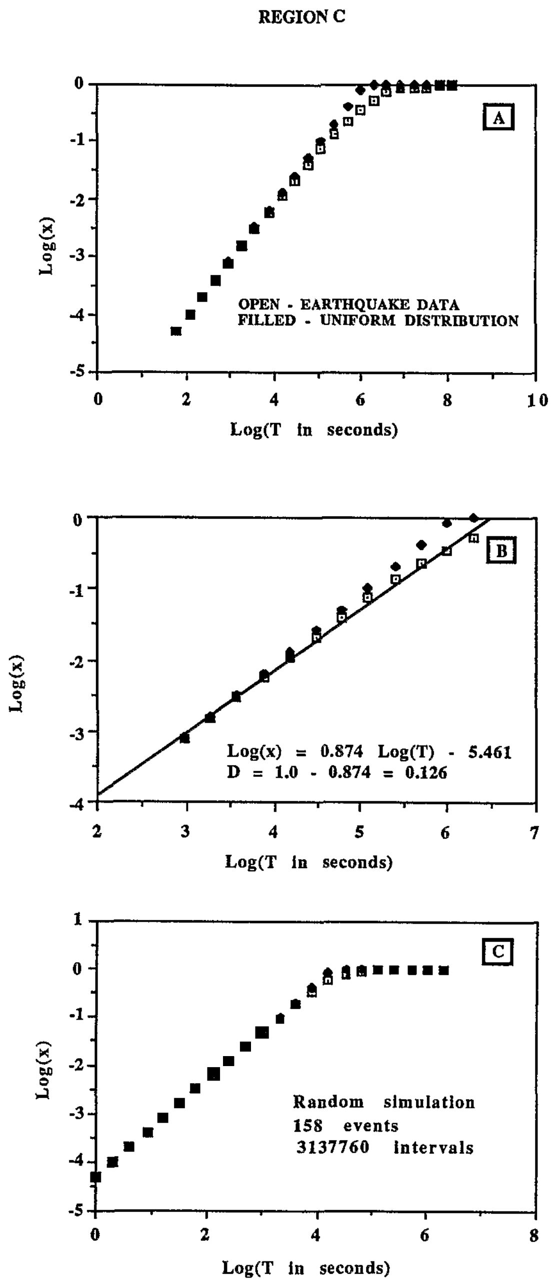

FIG. 8. Same as Figure 5 for region C.

of $D = 0.128$, corresponding to a slope of 0.872 (Figure 6B). The results of a random simulation of 1,330 events in the time interval studied is also plotted in Figure 6C and is significantly different from the earthquake data and close to the uniform distribution.

Results for regions B, C, and D are given in Figures 7A, 8A, and 9A. For region B, 379 events satisfy the magnitude condition, for region C, 108 events satisfy the

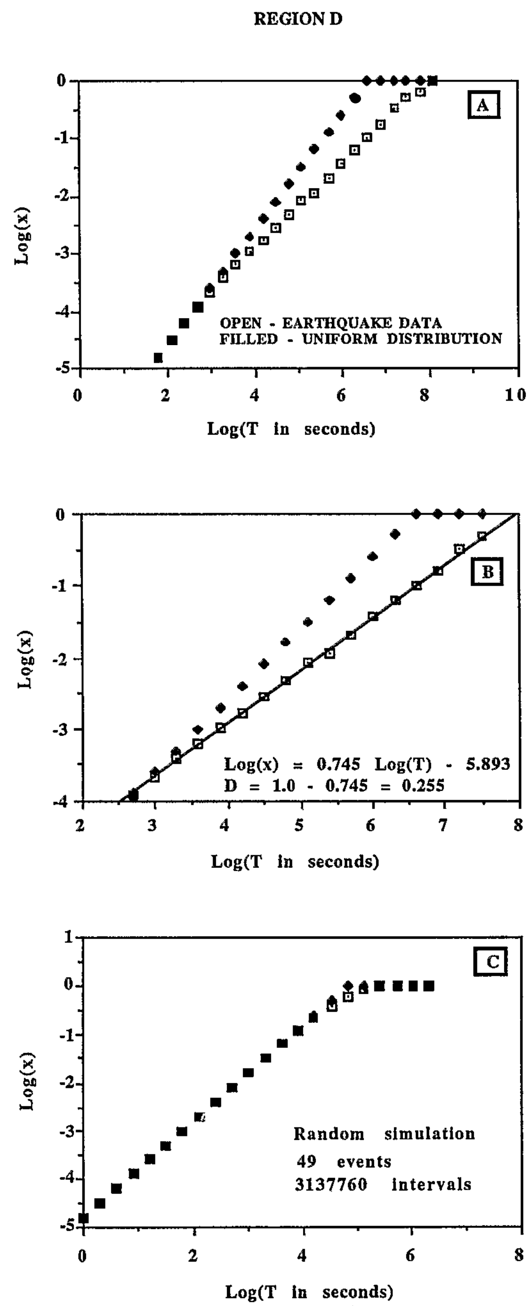

FIG. 9. Same as Figure 5 for region D.

magnitude condition, and for region D, 49 events satisfy the magnitude condition. The results for the time intervals with scale-invariant clustering, 16 min $\leqq \tau \leqq$ 16,786 min for region B, 16 min $\leqq \tau \leqq$ 32,768 min for region C, and 8 min $\leqq \tau \leqq$ 524,288 min for region D, are given in Figures 7B, 8B, and 9B. The best-fit straight line has a slope of $m = 0.771$ corresponding to a fractal dimension $D = 0.239$ in region B, $m = 0.874$, and $D = 0.126$ in region C, and $m = 0.745$ and $D = 0.255$ in region D. Results for random simulations using the number of events in each region are given in Figures 7C, 8C, and 9C.

Two aspects of our data correlations must be considered in terms of the fractal model. The first is the range of time scales over which a good fit to fractal model is obtained. Clearly, this is best established for region D as is illustrated in Figure 9B. An excellent fit to a straight line is obtained over 5 orders of magnitude. The well-defined earthquake swarms in this region exhibit a strong scale-invariant behavior. In the other regions, there is considerable curvature of the data. This is most easily interpreted as a combination of scale-invariant clustering and random events.

The second aspect of our data correlation is the fractal dimension obtained. For region D, the value $D = 0.255$ should be a good measure of the observed clustering. Thus, the clustering is relatively strong as illustrated in Figure 4B. In the other regions, while there is certainly a significant deviation from a purely Poisson process, a direct interpretation of the fractal dimension is difficult because of the superposition of random and scale-invariant processes.

Discussion and Conclusions

In this paper, we have presented a quantitative test of seismic clustering that is based on fractal or scale-invariant clustering. For the temporal distribution of seismicity, the fraction of time intervals containing an earthquake is plotted against the length of the interval. If a power law dependence is found, a fractal dimension for clustering is defined.

As a test of the method, we have considered four seismically active regions in the New Hebrides island arc. In all regions, our results deviate strongly from the results expected from a random or Poisson process. In each region, a fractal dimension is defined and values range from 0.126 to 0.255. The variation in fractal dimension does not correlate with the number of earthquakes considered. The fractal behavior of the region is best defined (minimum scatter) for the highest value of fractal dimension. A particularly well-defined fractal distribution is found for the back arc region D. This region did not have a major earthquake during the period of the study but the seismicity was strongly clustered.

In fact, it would be surprising if seismicity in a region was random. An increase in the state of stress in a region would be expected to cause clustering of seismicity. Earthquakes would be expected to occur with a variety of magnitudes. The dependence of frequency on magnitude generally satisfies the fractal or scale-invariant condition. Thus, it would not be surprising if the temporal occurrence of seismicity also satisfies a fractal or scale-invariant condition.

The results presented in this paper are clearly preliminary in nature. The primary purpose is to present the fractal approach to seismic clustering. Earthquakes represent a five-dimensional set of events: time, three space dimensions and magnitude. Clustering can be studied in any subset of these dimensions. One goal of this type of investigation would be to determine whether a change in the fractal dimension associated with the clustering of earthquakes may be a precursor for a major earthquake.

Acknowledgments

This research was supported by ORSTOM, Grant 81-16285 from the National Science Foundation, Contract 14-08-0001-19294 from the Department of Interior, and by Grant NGR-33-010-108 from the National Aeronautics and Space Administration. One of the authors (R. F. S.) held a NASA Graduate Student Research Training Grant NGT-33-010-108.

References

Aki, K. (1981). A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction*, D. W. Simpson and P. G. Richards, Editors, American Geophysical Union, Washington, D.C., 566–574.

Bottari, A. and G. Neri (1983). Some statistical properties of a sequence of historical Calabro-Peloritan earthquakes, *J. Geophys. Res.* **88,** 1209–1212.

Chatelain, J. L., B. L. Isacks, R. K. Cardwell, R. Prévot, and M. Bevis (1986). Patterns of seismicity associated with asperities in the central New Hebrides island arc, *J. Geophys. Res.* **91,** 12,497–12,519.

Cox, D. R. and P. A. W. Lewis (1966). *The Statistical Analysis of Series of Events*, Methuen, London, England.

De Natale, G. and A. Zollo (1986). Statistical analysis and clustering features of the Phlegraean Fields earthquake sequence (May 1983–May 1984), *Bull. Seism. Soc. Am.* **76,** 801–814.

Dziewonski, A. M. and A. G. Prozorov (1984). Self-similar determination of earthquake clustering, *Computa. Seism.* **16,** 7–16.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed., John Wiley, New York, 509 pp.

Isacks, B. L., L. R. Sykes, and J. Oliver (1967). Spatial and temporal clustering of deep and shallow earthquakes in the Fiji-Tonga-Kermadec region, *Bull. Seism. Soc. Am.* **57,** 935–958.

Johnson, C., V. I. Keilis-Borok, R. Lamore, and B. Minster (1984). Swarms of main shocks in southern California, *Computa. Seism.* **16,** 1–6.

King, G. (1983). The accommodation of large strains in the upper lithosphere of the earth and other solids by self-similar fault systems: the geometrical origin of *b*-value, *Pure Appl. Geophys.* **121,** 761–815.

Knopoff, L. (1964). The statistics of earthquakes in southern California, *Bull. Seism. Soc. Am.* **54,** 1871–1873.

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension, *Science* **156,** 636–638.

Mandelbrot, B. B. (1975). Stochastic models for the Earth's relief, the shape and the fractal dimension of the coastlines, and the number-area rule for islands, *Proc. Natl. Acad. Sci. USA* **72,** 3825–3828.

Mandelbrot, B. (1982). *The Fractal Geometry of Nature*, W. H. Freeman and Co., San Francisco, California.

Matsumura, S. (1984). A one-parameter expression of seismicity patterns in space and time, *Bull. Seism. Soc. Am.* **74,** 2559–2576.

Savage, W. U. (1972). Microearthquake clustering near Fairview Peak, Nevada and in the Nevada Seismic Zone, *J. Geophys. Res.* **77,** 7049–7056.

Shlien, S. and M. N. Toksöz (1970). A clustering model for earthquake occurrences, *Bull. Seism. Soc. Am.* **60,** 1765–1787.

Singh, S. and A. R. Sanford (1972). Statistical analysis of microearthquakes near Socorro, New Mexico, *Bull. Seism. Soc. Am.* **62,** 917–926.

Turcotte, D. L. (1986) Fractals and fragmentation, *J. Geophys. Res.* **91,** 1921–1926.

Turcotte, D. L. (1986/1987). A fractal model for crustal deformation, *Tectonophysics* **132,** 261–269.

Udias, A. and J. Rice (1975). Statistical analysis of microearthquake activity near San Andreas Geophysical Observatory, Hollister, California, *Bull. Seism. Soc. Am.* **65,** 809–827.

Vere-Jones, D. (1970). Stochastic models of earthquake occurrence, *J. R. Statist. Soc.* **32,** 1–45.

Vere-Jones, D. and R. B. Davies (1966). A statistical survey of earthquakes in the main seismic region of New Zealand. Part 2. Time series analyses, *N.Z. J. Geol. Geophys.* **9,** 251–284.

Vere-Jones, D. and T. Ozaki (1982). Some examples of statistical estimation applied to earthquake data. 1. Cyclic Poisson and self-exiting models, *Ann. Inst. Statist. Math.* **34,** 189–207.

Institute for the Study of the Continents
Cornell University
Ithaca, New York 14853-1504
Contribution No. 86 (R.F.S., D.L.T.)

Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM)
BPA 5, Noumea,
New Caldedonia (J.L.C., R.P.)

Manuscript received 19 August 1987

Present address for R. F. Smalley, Jr.: Center for Earthquake Research, Memphis State University, Memphis State University, Memphis, Tennessee 38152.