Australian Journal of Science and Technology

ISSN Number (2208-6404)
Volume 5; Issue 3; September 2021

Original Article

Estimation of correlation dimension from earthquake data using grassberger and procaccia algorithm

Folasade Lucia Aderemi*, Bolaji Opeyemi Oladele

Department of Physics, Faculty of Science, University of Ibadan, Ibadan, Nigeria

ABSTRACT

Correlation dimension is a measure of the dimensionality of the space occupied by a set of points usually referred to as a type of fractal dimension, it is an important parameter for understanding the complexity of a system and a tool for investigating the non-linear and chaotic characteristics of a system. The correlation dimension can, therefore, be used to distinguish between true stochastic processes and deterministic chaos. This research work is aimed at investigating the dynamics characteristics behavior of Earthquakes' occurrence by estimating the correlation dimension using the Grassberger and Procaccia algorithm. The data used for this study were obtained from Advanced National Seismic System, the data covers the major global seismic regions. The data were processed, the correlation integral and with their respective associated radius and the correlation, dimension was extracted using Grassberger and Procaccia algorithm with code written in Python 3. The results show that the values of the correlation dimension for the three regions considered are fractional and the correlation exponent tends to converge which is symptomatic of a deterministic chaos.

Keywords: Chaos, deterministic system, fractal dimension, non-linear dynamics, seismicity

Submitted: 29-06-2021, **Accepted:** 02-07-2021, **Published:** 30-09-2021

INTRODUCTION

Correlation dimension is a measure of the dimensionality of the space occupied by a set of random points usually referred to as a type of fractal dimension.[1] It is obtained from the correlations between random points on an attractor. The correlation dimension (D_2) for a set of points is defined as:

$$D_2 = \lim_{r\to\infty} \frac{\log \sum_i p_i^2(r)}{\log r}$$

where $p_i^2(r)$ is the probability of finding a pair of points in a box of size r.

If the coverage of the attractor is uniform, we have, $\mathrm{Pi} = \frac{1}{\mathrm{M(r)}}$,

where M(r) is the number of cells needed to cover the attractor.

The correlation integral or correlation sum C(r) is given as

$$C(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{\substack{i,j \\ (1\le i<j\le N)}} H(r - |Y_i - Y_j|)$$

where N is the number of points

H is the Heaviside step function; with H(u) = 1 for u > 0, and H(u) = 0 for u ≤ 0

where u = r–|Yi–Yj|, r is the radius of a sphere centred on Yi or Yj (attractor),

Yi stands for a point on which we center our measuring device (e.g. a box, a sphere or circle).

Yj are other point on the trajectory, for each centre point, the absolute distance between yi and yj is |yi-yj|.

Grassberger and Procaccia established that for small values of r and for sufficiently large numbers of data points N, the probability of having a pair of points in a box of size r is the same as the probability of having a pair of points with separation distance less than r.

For small r, the correlation sum grows according to a scaling relation;

Address for correspondence: Folasade Lucia Aderemi, Department of Physics, Faculty of Science, University of Ibadan, Ibadan, Nigeria. E-mail: folaremi@gmail.com

$$C(r) \approx r^v$$

where v is the Correlation Exponent, this scaling relation is only valid if r is small and N is large.

Rearranging the scaling relation and taking logarithms of both sides, shows that D_2 may be approximated by log (C(r)) divided by log(r), which is usually approximated as the slope of the straight line scaling region in a plot of log(C(r)) against log(r).[2] Practically, the estimation of v is done by plotting the correlation integral C(r) values obtained for different lengths (r) against the corresponding length (r) on a log-log graph and the slope is deduced. Literarily, v is equal to D2 but, the value of v can further be processed by estimating v for different embedding dimensions. This can be achieved by plotting the values of v against its corresponding embedding dimension, and v is expected to converge at a point. The values at which v converges is taken as the real correlation dimension. However, it is often useful to represent the attractor in a higher dimensional space than absolutely necessary, in order to reduce systematic errors. These errors result from a strongly non-uniform coverage of the attractor, provided this non-uniformity is not so strong as to make v not equal to D.[1]

Correlation dimension has been applied both theoretically and practically in various fields of science such as agriculture, medicine, atmospheric physics, economics, and mathematics. The popular Grassberger and Procassia algorithm for estimating correlation dimension was first applied.[1] It was used to distinguish between random noise and chaos, also to determine the order of Non-Linear Auto-regressive Model, and to test the residuals of linear models.[3] The correlation dimension was used in medicine to differentiate electroencephalographys arising from dementia due to Alzheimer's disease from that of vascular etiology[4] to differentiate healthy brain states from those with schizophrenia or brain tumors.[5,6] Correlation dimension to differentiate between abnormal and normal achromatic Visual evoked potentials,[7] to distinguish between normal retinas and pathological retinas of diabetic patient using the box-counting method.[8] Diabetic retinopathy is damage to the blood vessels of the retina in people with diabetes which is a leading cause of blindness in people with diabetes. It is an important parameter for understanding the complexity of a system and can be used to distinguish between true stochastic processes and deterministic chaos whether low-dimensional or high-dimensional.[1,9]

MATERIALS AND METHODS

The data used for this study was obtained from Advanced National Seismic System (ANSS). The Earthquake data between 1962 and 2012 was obtained for three different seismic active regions with a delta search parameter. The three regions are North America, South America and Asia. Large earthquake of at least magnitude 8.5 was made the reference point for each of the regions. The delta search parameters enabled us to obtain the data radially outward from our reference large earthquake, the radii were chosen such that the data is confined to the respective regions and the rupture areas of the respective referenced large Earthquake. The major elements identified before estimating the correlation dimension are, Phase space or state space, Attractor, Trajectory, Correlation integral, and Embedding dimension. Phase space is the graph or in more formal terms, phase space or state space is an abstract mathematical space in which coordinates represent the variables needed to specify the phase (or state) of a dynamical system. The solid earth represents our phase space where we have the epicentre or geolocations of the earthquake, this is achieved by plotting a graph of the latitude against longitude of each epicentre of each earthquake. The attractor is a dynamical system's set of stable conditions, the attractor in this case is the epicentre of the earthquakes located on our state space. The finite volume of phase space that the attractor occupies usually is quite small relative to the volume of the phase space itself, which justifies our representation when the volume of spaces occupied by the epicentres is compared with the entire surface area of the solid earth. The trajectory represents the path in which the attractor follows. Both the attractor and the trajectory can be visualized using software such as PHASER, XPPAUT, and WOLFRAM MATHEMATICA.

Estimation of the Correlation Integral

The correlation sum as given by[1] is

$$C(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} H\left(r - \left|x_i - x_j\right|\right) \qquad (1)$$

which implies

$$C(r) = \frac{\text{Total number of points within radius r}}{\text{Largest number of mathematically possible points}} \qquad (2)$$

Equation 2 can also be written as

$$C(r) = \frac{\text{Total number of points within radius r}}{N(N-1)} \qquad (3)$$

The x_i in Equation 1 stands for a point on which we cantered our measuring device; our x_i is the reference large earthquake of radius r. The first radius chosen is the distance between our large reference earthquake and the next earthquake that occurs both in space and time. x_j are other points on the trajectory. For each centre point, the absolute distance between x_i and x_j is $|x_i-x_j|$. the separation distance between every pair of points was calculated using the equation of separation distance between two vectors as used in the work of[10] where they represented a two-dimensional system with variables x and y separation distance by equation 4.

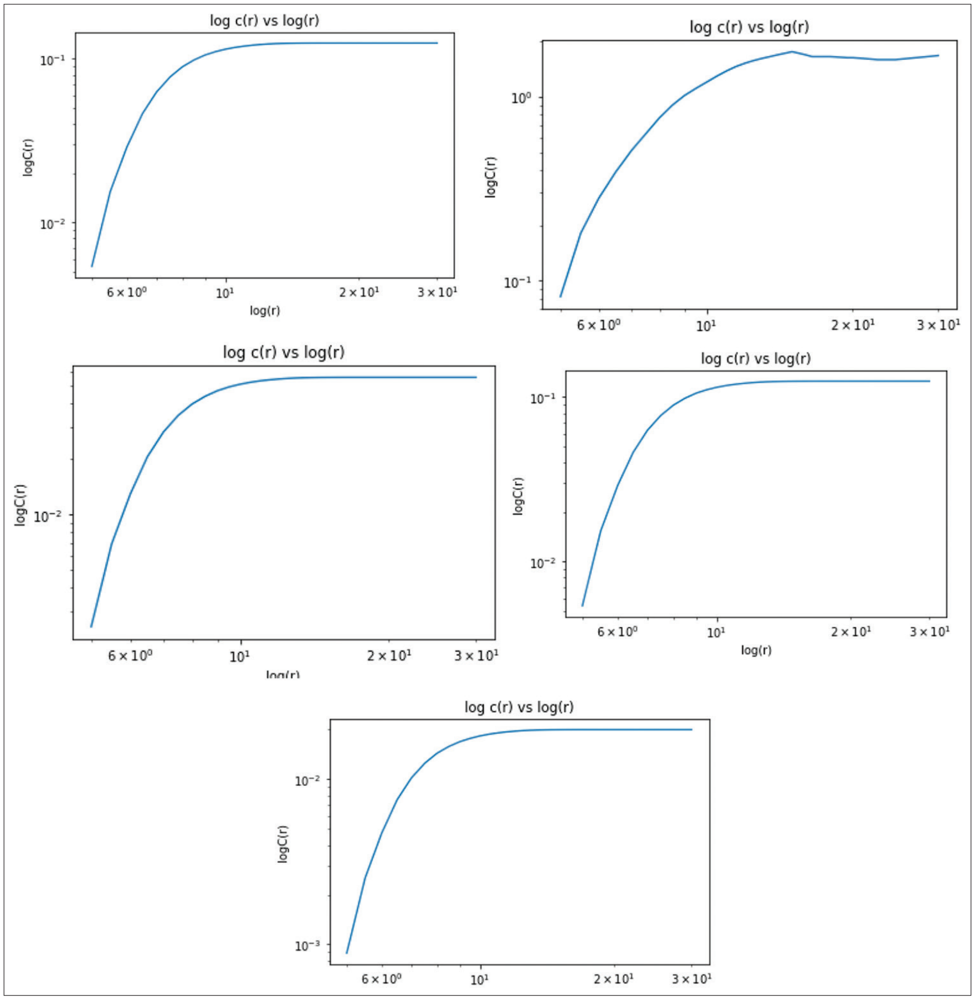

Figure 1: Graph of the correlation sum C(r) against the radius (r) for different embedding dimensions (Dm)

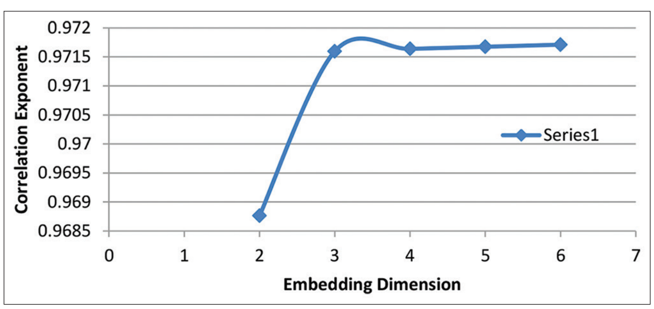

Figure 2: Correlation dimension curve for region 1

$$d_j = \sqrt{\left\{(x_a - x_b)^2 + (y_a - y_b)^2\right\}} \quad (4)$$

For a two dimensional system with variables x and y, the separation is defined by equation 4

Where a and b denote the two orbits. If a and b denote two orbits, d_j is defined as the separation between the jth pair of nearest neighbours. Since earthquake is a single observable, for the purpose of this research, x is defined as the latitude while y is defined as the longitude of epicenter of the earthquake. X_a and Y_a are defined as the latitude and longitude of a referenced earthquake, respectively, while X_b and Y_b are the latitude and longitude of the subsequent

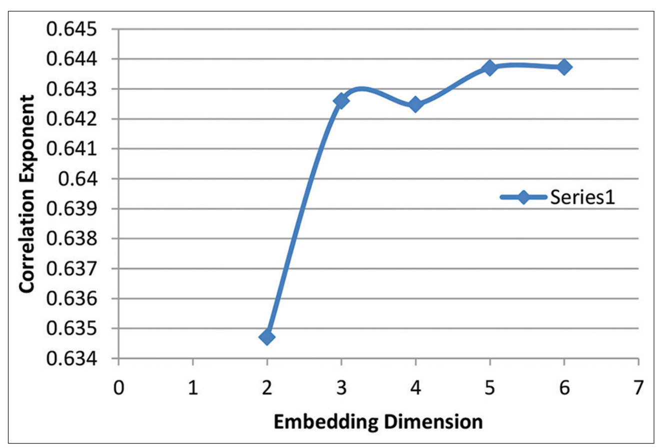

Figure 3: Correlation dimension curve for region 2

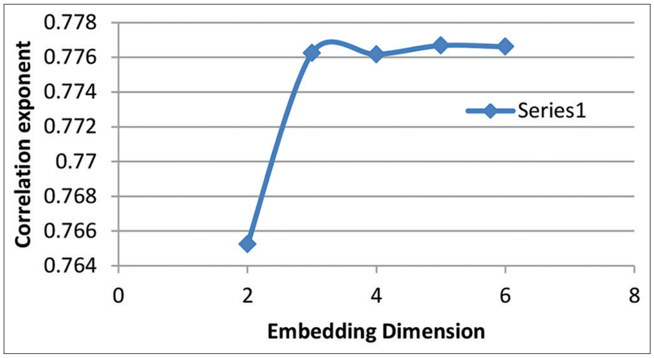

Figure 4: Correlation dimension curve for region 3

earthquakes, respectively. If the result obtained from r-|xi-xj| is negative, then the measured distance |xi-xj| is greater than ε. That means point X_j is beyond the circle of radius ε and therefore doesn't qualify for our count. On the other hand, if ε-|xi-xj| is positive, then |xi-xj| is smaller than ε, and the point X_j is within the circle. H the Heaviside function is an efficient way to label each qualifying point, that is, each point for which r-|xi-xj| is positive (>0). For all those cases the computer program was asked to assigns a value of 1 to the entire expression G(r-|xi-xj|). If, r-|xi-xj| is negative, it implies the point xj is beyond the radius of the circle. For those cases, the computer program was asked to assigns a value of 0 to G(r-|xi-xj|). The Heaviside function made it easy to earmark the points that qualify for the count. The next step is the counting of the pair of points inside the circle. It involves adding all the numbers pairs of points for the centre point x_j and the subsequent centre points x_i when each point is chosen inside the circle. this operation is done by the two summation signs in Equation 1. The two summation signs together simply mean that we go to the first center point xi. and sum the results of G(r-|xi-xj|) for all points (all Xj's), then do the same for the next centre point xi, and so on all the way through the total of N points. That gives the total counts of all the points that fall within the circle of the specified.

Normalization is achieved by dividing the total number of pairs of points by the total number of available points. The total number of available points is N(N-1). Hence, we multiply the total number counted (the numerator) by 1/[N(N-1)] but $1/N^2$ was used since N is very large, the 1 in N-1 becomes negligible. Then, N-1 simply becomes N. The notation "N(N-1)" (the denominator in the formula of correlation integral) therefore becomes N(N) or N^2. Multiplying the entire counts by $1/N^2$, with N^2 being the total number of available points or pairs on the trajectory hence normalises the equation. Having determined the correlation sum for our first radius of 0.1 km, the first radius was chosen to be the smallest distance between our large earthquake and the closest earthquake in time and space. The radius was increased by 1 km and the entire process of estimating the correlation integral was repeated for the new radius. Obviously the larger the radius the larger the number of points in the circle, that is, the new radius yields a larger total number of points to be counted, which gave a different result for the respective radius. The normalization constant N^2 depends only on the size of the basic dataset and so is constant regardless of the radius ε. Hence, the larger ε yields a larger correlation sum. We determined the correlation sum/integral with their respective associated radius, the result was put in a tabular form and a graph of the correlation integral versus the radius was plotted on a log-log scale and the value for the correlation exponent was deduced from the slope of the scaling region. The above method explained was achieved with the help of a written program on Python 3 programming language. Since placing a grid of boxes or spheres on the attractor is not feasible, we improvised by calculating the distances between every point (epicentres) on the attractor with reference to one another. The embedding dimension is the number of lagged values (xt, xt+m, etc.) Where t is 1,2,3.used in a pseudo phase space plot, usually for the purpose of reconstructing an attractor. The earthquake does not occur at regular time interval but with the aid of the search parameter with reference to the large earthquake the subsequent earthquakes follow in space and time. The time was lagged by 2 in order not to overstretch the trajectory until the embedding dimension of 5 was achieved. The process of the correlation sum was repeated for different embedding dimension. The computations produce a batch of radii (r) and their associated correlation sums C(r) for each of several embedding dimensions. The correlation sum obtained for each radius was then plotted against the radius respectively [Figure 1]; the slope of each graph is taken to be the correlation exponent. The correlation dimension obtained from each embedding dimension was then plotted against the dimension [Figures 2-4] which was used to characterise the dynamics of the occurrence of earthquake whether is chaotic or stochastic.

RESULTS AND DISCUSSION

In this study, the Grassberger and Procassia Algorithm technique of estimating Correlation dimension was employed to study the dynamical behaviour of the earthquake occurrence in three different regions across the world. The data analysed was limited to area of 0–200 km within the region and also from lower magnitude of 1 to the referenced large earthquake. This enabled us to study the occurrence of the earthquake as a single system which starts to build up stress until it finally releases the energy. We can determine whether or not the system is chaotic by the features and value of the correlation dimension, if the correlation exponent increases as the separation distance r is increased, it indicates absolute randomness. However if the correlation exponent tends to converge in a phase space this corresponds to chaos. It also indicates the amount of chaos in a system. Figures 2-4 show the results of the correlation dimension obtained for regions 1, 2 and 3 respectively. The values are fractions and each converges at a point for the three regions.

CONCLUSION

The correlation dimension for the three regions is fractions with 0.64 for region two being the lowest, followed by 0.77 for region three and 0.97 for region one The values of the correlation dimension indicate the quantity of chaos present in each region. The results show that as the separation distance increases, the correlation exponent increases but at a large separation distance r the correlation exponent converge at a value this is a symptom of a deterministic chaotic system.

ACKNOWLEDGMENTS

ANSS, Northern California Earthquake Data Centre, USA, for earthquake catalogue.

REFERENCES

1. Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. Physica D 1983;9:189-208.
2. Takens F. Detecting strange attractors in turbulence. In: Rand DA, Young LS, editors. Dynamical Systems and Turbulence, Lecture Notes in Mathematics. Vol. 898. Berlin, Germany: Springer-Verlag; 1981. p. 366-81.
3. Edrees MN. Mahmood applications of fractal dimension. Iraqi J Stat Sci 2006;10:54-73.
4. Jeong J, Chae JH, Kim SY, Han SH. Nonlinear dynamic analysis of the EEG in patients with Alzheimer's disease and vascular dementia. J Clin Neurophysiol 2001;18:58-67.
5. Kotini A, Anninos P. Detection of nonlinearity in schizophrenic patients using magneto encephalography. Brain Topogr 2002;15:107-13.
6. Antoniou PE, Anninos PA, Piperidou H, Adamopoulos A, Kotini A, Koukourakis M. Non-linear analysis of magneto encephalographic signals as a tool for assessing malignant lesions of the brain: First results. Brain Topogr 2004;17:117-23.
7. Boon MY, Henry BI, Suttle CM, Dain SJ. The correlation dimension: A useful objective measure of the transient visual evoked potential? J Vis 2008;8:6.1-21.
8. Uahabi KL, Atounti M. Applications of fractals in medicine. Ann Univ Craiova Math Comput Sci 2015;42:167-74.
9. David C, Theodore P. Using the correlation dimension to detect non-linear dynamics: Evidence from the Athens stock exchange. Finance Lett 2005;3:29-32.
10. Folasade LA. Olatunde IP. Computation of Lyapunov exponent for characterizing the dynamics of earthquake. Int J Recent Adv Phys 2016;5:9-15.