Comparative analysis of seismic persistence of Hindu Kush nests (Afghanistan) and Los Santos (Colombia) using fractal dimension

D A Prada¹, M P Sanabria¹, A F Torres¹, M A Álvarez² and J Gómez³
¹Universidad Pontificia Bolivariana, Bucaramanga, Colombia
²Universidad de Puerto Rico, Mayagüez, Puerto Rico
³Universidad Industrial de Santander, Bucaramanga, Colombia
E-mail: duwamg.prada@upb.edu.co

Abstract. The study of persistence in time series in seismic events in two of the most important nests such as Hindu Kush in Afghanistan and Los Santos Santander in Colombia generate great interest due to its high presence of telluric activity. The data were taken from the global seismological network. Using the Jarque-Bera test the presence of gaussian distribution was analyzed, and because the distribution in the series was asymmetric, without presence of mesocurtisity, the Hurst coefficient was calculated using the rescaled range method, with which it was found the fractal dimension associated to these time series and under what is possible to determine the persistence, antipersistence and volatility in these phenomena.

1. Introduction
Fractal geometry as a branch of mathematics allows research on self-similarity and dynamic systems. This branch uses notions of dimension and orbits in various disciplines such as geology, geophysics, medicine, engineering, biology, among others. Earthquakes are geology events, and the complex phenomenology exhibited by them is correlated to the deformation and sudden rupture of parts of the earth's crust because of the external forces acting from plate tectonic motions [1]. Some geophysical processes are characterized by self-similarity correlations because the dynamics are based on the interaction of many components over a wide range of time or space scales [2].

The study of seismic events is important, not only for geologists and geophysicists, but the community in general whom are affected by the natural phenomena. There are places on the planet where such events occur more frequently than others. One area of greatest tectonic activity is Pamir Hindu Kush, in the Afghanistan region [3]. This nest is located at the western syntax of the Himalaya, within the broad deformation belt, area formed by the collision of the Indian and Eurasian lithospheric plates. It is located approximately between the latitudinal boundaries 35-39°N and longitudinal boundaries 69-75°E [4]. Another geographic region considered as a seismic nest is found in the municipality of Los Santos, situated at 30km from the capital of the department of Santander in Colombia. The depth of the seismic events that occur in this zone is in a range of 150 – 170km in average. And, according to the National Seismological Network of Colombia RSNC, 50% of the telluric movements recorded in Colombia occur in this municipality; however due, to the depth of the hypocenter are categorized as a low-risk event.
In the period between June 1, 1993, and February 17, 2016, 2502 seismic events were presented in the Hindu Kush zone and Los Santos 3964 seismic events, as recorded by the seismological network world. Each point shows evidence of a telluric event in the area; the color determines depth: violet $0 - 33\, km$, blue $33 - 70\, km$, green $70 - 150\, km$, yellow $150 - 300\, km$, orange $300 - 500\, km$, and red of $500 - 800\, km$. The size of the circle determines its magnitude with respect to the Richter scale [1].

2. Mathematical method
Fractal theory applied to objects allow to construct the notion of self-similarity, dynamic systems, chaos, orbits, and dimension, both topological and Hausdorff, a set of fractal properties that have become useful tools in different disciplines such as mathematics, science, art, etc. [5-7]. Besides the self-similarity, the fractal objects have an idea out of the common; the fractal dimension. The dimension that is assigned by convention to certain geometrical and physical objects is associated with a number infinite variable, for example, a cube is assigned a triple defined directly by the thickness, width, and height thereof, and then the size of this object is three. This type of dimension is known as the topological dimension.

The fractal dimension, as its name suggests, is a fractional dimension and is determined by a rational number. Long-range power-law correlations are traditionally measured by a scaling parameter or fractal dimension ($D$). If the time series is self-similar and self-affine, the parameter $D$ is related to the Hurst exponent ($H$) through the expression $D = 2 - H$ [8,9]. Thus, the Hurst exponent is a measure of the long-range correlation in time series data and allows to distinguish the persistence (correlation), anti-persistence (anti-correlation) or randomness of the data [10]. The original estimation of the Hurst exponent was first performed in hydrology by Harold Edwin Hurst in 1951 [11], by introducing an empirical relationship called the Rescaled-Range (R/S). Posteriorly, this relationship became the start point to establish the Classical R/S (CR/S) method developed by Mandelbrot and Wallis into the context of the fractal geometry [10,12,13].

Although the CR/S is one of the most popular methods to calculate the Hurst exponent, it has shown some serious limitations to study long-range correlation when the time series is not large enough [13,14]. In the series of seismic events it is necessary to use the modified method of calculating the ratio between the rescaled range and the standard deviation, due to the number of data about CR/S might lead to an overestimation of the Hurst exponent, especially if seismic time series are constructed with less than 5000 data points as happened in this work. On the other hand, a study previously performed using data of Southern México [15] has detected the existence of cycles in the interevent interval sequence and this fact has been considered as a sign of trends in the associated seismic time series.

To apply this method, it is necessary to ensure that the data does not present a Gaussian distribution or that the data are not normalized. The study of these phenomena in front of their distribution may present asymmetry and homocedasticity, due to their leptokurtic, mesokurtic or platykurtic behavior. In order to determine if the data series presents Gaussian distribution, the Jarque-Bera test was applied. The Jarque-Bera test [16] is performed to corroborate that the time series is not linear, that is to say, that it presents fractalness. For the data analyzed, the asymmetry value was 1.00 and -0.11 for kurtosis. Then the following hypotheses are posed: $H_0$ (null hypothesis): Mean=0; Variance=1, i.e., the normal distribution; $H_1$: Mean≠0; Variance≠1, i.e., the distribution is not normal. To find the value of the Jarque-Bera test (JB), the following equation:

$$JB = n \left[ \frac{S^2}{6} + \frac{(K-3)^2}{24} \right]$$  \hspace{1cm} (1)

Where $S$ represents the asymmetry and $K$ the kurtosis. The Jarque-Bera coefficient is greater than 6, so the series does not show normality.

Specifically, a value $0 < H < 0.5$ corresponds to anti-correlated data (anti-persistence behaviour), a value $0.5 < H < 1$ corresponds to correlated data (persistence behaviour) and the value $H = 0.5$ corresponds to random data (uncorrelated behaviour) [10]. The definition of $H$ can be extended to values
larger than 1 [12]. In this way, the case $H = 1.5$ corresponds to Brownian motion, the case $H = 2$ corresponds to brown noise and the case $H > 2$ corresponds to black noise [17]. Shown below CR/S method the time series described in [18] under consideration $X: \{x_i\}$ is composed by $N$ values. The full-time series is divided into windows of size $M$. The number of windows is defined by $s = N/M$ and therefore there are $s$ windows of data $Y_j$, with $j = 1, 2, ..., s$. Defining the vector $k = (j - 1)M + 1, (j - 1)M + 2, (j - 1)M + 3, ..., (j - 1)M + M$, the average over each window is calculated as:

$$\bar{Y}_j = \frac{1}{M} \sum_k x_k$$

(2)

The profile or sequence of partial summations $Z_j: \{z_n\}$, with $n = 1, 2, ..., M$ is defined as the cumulative summation minus the average of the corresponding window

$$z_n = \sum_k^n \left( x_k - \bar{Y}_j \right)$$

(3)

The range $R_j$ of the window is defined as the maximum minus the minimum data point of the profile

$$R_j = \max\{Z_j\} - \min\{Z_j\}$$

(4)

The standard deviation of each window $\sigma_j$ is given as:

$$\sigma_j = \left[ \frac{1}{M} \sum_k (x_k - \bar{Y}_j)^2 \right]^{1/2}$$

(5)

The rescaled range is described by the quantity $(R/S)_M$, which is defined as

$$(R/S)_M = \text{mean}\{R_j/\sigma_j\}.$$  

(6)

For the case in which a stochastic process associated to the data sequence under study is rescaled over a certain domain $M \in \{M_{\text{min}}, M_{\text{max}}\}$, the R/S statistics follows the power law

$$(R/S)_M = aM^H$$

(7)

Herein, $a$ is a constant and $H$ is the Hurst exponent which represents a fractal measure of the long-range correlations in the analysed data. However, it is necessary to use the improved method of the rescaling range, in which autocovariance and its autocorrelation are used. The equation (5) of the standard deviation changes for this calculation by the equation:

$$\sigma_j(q) = \sqrt{\sigma_j^2 + 2 \sum_{k=1}^{q} \omega_k(q) \gamma_k} = 1 - \frac{k}{q+1}$$

(8)

Where $\sigma_j^2$ is the sample variance and $\gamma_k$ the sample autocovariance. The parameter $q$ represents the lags in the weighted autocovariances.

3. Results and discussion

The time series of seismic events occurred between June 1, 1993, and February 17, 2016, in the Hindu Kush Afghanistan and Los Santos Colombia zones, presented 2502 and 3964 events respectively, as recorded by the global seismological network. Each group of data was divided into as many groups as
divisors greater than 100. Table 1 shows the group of the first 139 data recorded in the Hindu Kush area from June 6, 1993, to August 27, 1996.

Table 1. Data of the standard deviation, data number, and rescaled range.

| Place       | Standard deviation | Number of data | Rescaled range | Max      | Min      | Average   |
|-------------|--------------------|----------------|----------------|----------|----------|-----------|
| Hindu Kush  | 0.630957685        | 139            | 13.22446       | 3.934E-13| 4.097122 |           |
| Los Santos  | 0.423422389        | 283            | 20.18763       | 3.117E-13| 4.139929 |           |

With each of the series the same procedure is performed, in Table 2 and Table 3 the necessary data for the calculation of the Hurts coefficient for this time series are shown, in addition, we compare the natural logarithm of the data of each of the subgroups generated versus the natural logarithm of the ratio between the rescaled range and standard deviation.

Table 2. Related data from the four groups depending on the range and standard deviation. Calculation of the natural logarithm of the data number and the natural logarithm of the quotient between the rescaled range and the standard deviation. Hindu Kush.

| Number of data | Rescaled range | Standard deviation | ln(num)   | ln(R/S)  |
|----------------|----------------|--------------------|-----------|----------|
| 139            | 13.22446043    | 0.630957685        | 4.934473933| 3.042584656|
| 278            | 25.49064748    | 0.596857382        | 5.627621114| 3.754388704|
| 417            | 37.37266187    | 0.56846170         | 6.033086222| 4.185436354|
| 834            | 49.42829736    | 0.580429930        | 6.726233402| 4.444509273|
| 1251           | 99.99904077    | 0.598812745        | 7.131698510| 5.117966936|
| 2502           | 383.97274180   | 0.723505287        | 7.824845691| 6.274218991|

Table 3. Related data from the four groups depending on the range and standard deviation. Calculation of the natural logarithm of the data number and the natural logarithm of the quotient between the rescaled range and the standard deviation. Los Santos.

| Number of data | Rescaled range | Standard deviation | ln(num)   | ln(R/S)  |
|----------------|----------------|--------------------|-----------|----------|
| 283            | 20.18763251    | 0.423422389        | 5.645446898| 3.864455207|
| 566            | 66.10477032    | 0.506678775        | 6.338594078| 4.871118969|
| 1983           | 316.59671880   | 0.656226596        | 7.592366129| 6.178877913|
| 3962           | 460.88414940   | 0.644585530        | 8.284504227| 6.572294466|

With the results shown in Table 2 and Table 3, the data of the last two columns are taken, the column of the natural logarithm of the data number and the column of the natural logarithm of the rescaled range over the standard deviation of each subseries. With the data of these two columns, a linear regression is elaborated whose abscissa is the natural logarithm of the data number, and the ordinate corresponds to the natural logarithm of the rescaled range over the standard deviation. Figure 1(a) shows the graph of the linear regression and the scatter data. The value of the slope of said linear equation is 1.0427 with a correlation coefficient $R^2$ of 0.9566, which represents the associated fractal dimension of the data corresponding to the Hindu Kush seismic events. Figure 1(b) shows the graph of the linear regression and the dispersion data corresponding to the data of the seismic events in Los Santos, the value of the slope of said linear equation is 1.092, which represents the fractal dimension associated to these dates. The correlation coefficient $R^2$ is 0.9797.
4. Conclusions

From the data recorded by the world seismological network between June 1, 1993, and February 17, 2016, in the areas of Hindu Kush Afghanistan and Los Santos Colombia, which presented 2502 and 3964 events respectively, the Hurts coefficient was $H = 1.0427 > 0.5$ for Hindu Kush with a fractal dimension of $D = 0.9573$ and for Los Santos the coefficient was $H = 1.0292 > 0.5$ with a fractal dimension of $D = 0.9768$, which indicates that the time series presents low fractal dimension and for this reason the series present persistence associated with a volatility of 47.86% for Hindu Kush and a volatility of 48.54% for Los Santos. These series did not present Gaussian distribution because the test of Jarque-Bera was greater of 6 for each one.

References

[1] Michas G, Vallianatos F and Sammonds P 2013 Non. Proc. Geo. 20 713–724
[2] Ashkenazy Y, Havlin S, Ivanov P, Peng C, Schulte-Frohlinde V and Stanley H E 2003 Phys. A 323 19–41
[3] Bai L and Zhang T 2015 Tecto. 638 177–184
[4] Sarkar I and Sanyal S 2003 J. Asi. Ear. Sci. 23 449–459
[5] Mandelbrot B 2006 Los objetos fractales. Forma, azar y dimensión (España: Metatemas)
[6] Falconer K 2006 Fractal geometry: Mathematical foundations and applications (India: Wiley)
[7] Barnsley M 2012 Fractals Everywhere (United States: Dover Publications)
[8] Feder J 1988 Fractals (New York: Plenum Press)
[9] Chen C, Lee Y and Chang Y 2008 Phys. A 387 4643–4648
[10] Mandelbrot B and Wallis J 1968 Water Resour. Res. 4 909–918
[11] Hurst H E 1951 Long-term storage capacity of reservoirs Trans. Am. Soc. Civ. Eng. 116(1) 770–799
[12] Mandelbrot B and Wallis J 1969 Wat. Res. Res. 5 321–340
[13] Mandelbrot B and Wallis J 1969 Wat. Res. Res. 5 967–988
[14] Sánchez M, Trinidad J and García J 2008 Phys. A 387 5543–5551
[15] Álvarez J, Echeverria J C, Ortiz A and Hernández E 2012 Geo. 54 1–12
[16] Andrew W 1991 Econ. 59 1279-1313
[17] Jimenez A 2011 Phys. A 390 2146–2154
[18] Martin L, Aranda N and Quimbay C 2015 Phys. A 421 124–133
[19] Gujarati D and Porter D 2010 Econometria (Mexico: Mc Graw Hill)