

# Fractal analysis of persistence in fluctuation of levels of the Magdalena River

D A Prada<sup>1</sup>, D J Serrano<sup>1</sup>, M S Rueda<sup>1</sup>, L Duarte<sup>1</sup>, A Acevedo<sup>1</sup>, S C Prada<sup>2</sup>  
and J Gómez<sup>3</sup>

<sup>1</sup> Universidad Pontificia Bolivariana, Bucaramanga, Colombia

<sup>2</sup> Universidad Industrial de Santander, Bucaramanga, Colombia

<sup>3</sup> Servicio Nacional de Aprendizaje SENA, Bogotá D.C., Colombia

duwamg.prada@upb.edu.co

**Abstract.** The measures of water level in a river contributes to understand the dynamic of the discharge in a specific point of the catchment. With this kind of information, it is possible to predict and study future flooding events. The official entity CORMAGDALENA with support of IDEAM report every two days the water level along the Magdalena River. The Hurst coefficient  $H$  generates a new statistic methodology, which is based on the tendencies of data series, determining the persistence a dataset and consequently its fractality. This research applies this methodology to a dataset of time, registered by CORMAGDALENA in the municipality of Barrancabermeja, in the Department of Santander. The time has been taken between September 29<sup>th</sup> of 2011 and March 31<sup>st</sup> of 2017. According to the value of  $H$  we can say if the series of behaviours taken has persistence in the case that  $H$  is greater than 0.5, if  $H$  is less than 0.5 we will say that the series is anti-persistent. With these values of  $H$  it is possible to calculate the fractal dimension associated to the compartment and thus to determine the volatility for the analysis of the risk in terms of the fluctuation of the Magdalena River.

## 1. Introduction

There are some high complexity systems in the nature, in particular the water basins, because there are some Variables like time, space, weather, geology, vegetation and others are facts linked to their formation process, whose evolution is directly linked with the combination of those forces and weather facts. Due to the dynamic of discharge in specific points of a water basin, the study of water fluctuation levels becomes important, for example in a river, to predict and consider upcoming flood events. The flow of rivers also depends directly of the speed it has, so if the speed is low, the flow will be regular, if the speed increase then it becomes irregular, creating swirls and swirls between swirls. By the Navier-stokes equations the hydrodynamic describes those nonlinear movements initially because the newton laws for fluids. The famous hydrologist Harold Edwin Hurst (1880-1978), studied the fluctuation levels of the Nile River in order to predict fluctuations in order to be prepared for floods and take precautions against drought.

## 2. Magdalena river

According to the time in the year the level of water in the Magdalena River may change. This kind of phenomena is directly related with periods of drought and floods. Whereby, people can suffer lack of



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

food or damage in the households. The hydrologist Harold Hurst studied the fluctuations of the Nile River with the aim of taking advantage of the reservoir capacities and the measures in drought epochs [1] [2] [3]. For this purpose, Hurst created a statistical methodology which consists of analysing the persistence of a time series [5] according to the duration of its cycles and determining if that time series is fractal [6].

The Magdalena River is one of the most important rivers in Colombia with an average flow is 6700 m<sup>3</sup>/s. In fact, the country is considered the fourth in the world in water richness. About 90% of the territory presents rains superiors to 2000 millimetres (mm), with an average evaporation of 1150 mm, and volume of runoff of 2112 km<sup>3</sup> [7]. The river starts in the Colombian Mountains in the department of Huila, and flows into the Caribbean Sea in the Atlantic department. Its length is approximately 1600 km, being 886 navigable. The narrowest part is 2.20 meters, and the widest is 1073 meters [7].

CORMAGDALENA is the entity in charge of the river and is supported by the IDEAM. It registers every two days the level of the Magdalena River, whose measurement stations are in the municipalities of Puerto Berrio, Barrancabermeja, San Pablo, Las Varas, El Banco and Calamar [8]. The data is used to generate time series of depth, with 984 days from September 29, 2011 to March 31, 2017. The fluctuations of the Magdalena river are observed in the municipality of Barrancabermeja, Santander. The analysis is performed to observe the persistence of the time series. Furthermore, it is required to execute a Jarque-Bera test as an examination of normality. Then, it is estimated the fractality in the time series and then the analysis of the rescaled range. After that, it is done the calculation of the Hurst coefficient, and estimated the fractal dimension associated to the fluctuation levels in the Magdalena river, Barrancabermeja.

Study the continuous changes in the fluctuations in the levels of watersheds, it is of great importance because it allows you to adjust predictions about behaviors that can lead to floods or droughts. The first to recognize the characteristics of invariance of the drainage network to establish the laws of scale was Horton, and denoted as: Relationship of fork ( $R_B$ ), a ratio of length ( $R_L$ ) and ( $R_A$ ) [9].  $R_L R_A$  Natural objects that have a complex spatial structure can be studied as fractals. Fractal geometry to describe irregular patterns and fragmented, which are repeated at different scales, usually in isotropic [10]. Fractal geometry as a branch of mathematics has been allowed to carry out research in relation to the similarity and dynamic systems, using notions of dimension and orbits in various disciplines such as medicine, engineering, psychology, music, biology, among others. In the particular case of drainage networks two fractal dimensions can be calculated, a fractal dimension ( $d$ ), which describes the sinuosity of a current and the fractal dimension of the entire network ( $D$ ), related to the characteristics of branching or compactness of the system [11,12]. With the analysis data between the length of the currents and the area of drainage basins, he found that the Mandelbrot fractal dimension of water courses ( $d$ ) was equal to 1.2, while the fractal dimension of the entire system ( $D$ ) was equal or close to 2 [13].

### 3. Mathematical method

The Convention dimension assigned to certain geometric and physical objects, it is associated with an infinitive variable number, for example, to the cubic number is assigned the 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 [13, 15].

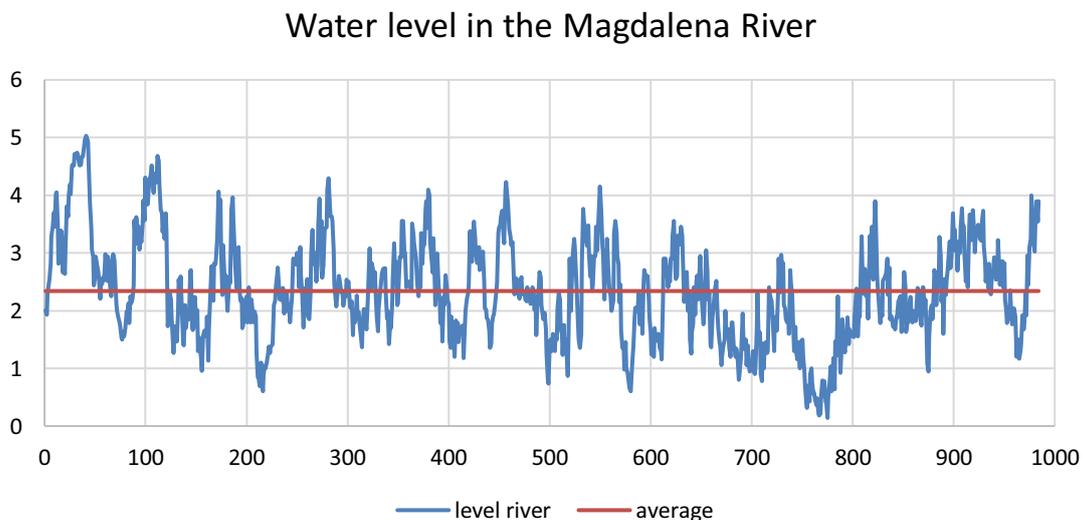
The fractal dimension, as its name suggests is a fractional dimension and is determined by a rational number. Long range power law correlations are traditionally measured by a scale parameter or fractal dimension ( $D$ ). If the time series is self-similar and self-related the parameter  $d$  is related to the Hurst exponent ( $H$ ) through the expression  $D=2-H$  [16, 17]. Therefore, the Hurst's exponent is a measure of long range correlation in time series data and allows distinguishing persistence (correlation), persistence

(anti-correlation) or the randomness of data [18]. The original estimate of Hurst's exponent was first made in hydrology by Harold Edwin Hurst in 1951 [19], by introducing an empirical relationship called rescale range (R/S). Subsequently, this relationship became the starting point to establish the classical method R/S (CR/S) developed by Mandelbrot and Wallis in the fractal geometry context [18, 20, 21]. Although CR/S is one of the most popular methods to calculate the Hurst exponent, it has shown some serious limitations for studying the long-range correlation when the time series is not large enough [21, 22].

#### 4. Respect to the test of normality

For the analysis of chaotic or fractal time series, it is required to understand the data applying the linearity and normality test. The Jarque-Bera test is applied for the linearity analysis [23]. This is an asymptotic examination, or large samples based on the ordinary least squares OLS. The idea of persistence of a time series, understood as the tendency to repeat this series (long-term memory), is highly related to the notion of self-sufficiency. Self-splicing is a feature of fractal objects, under which it is possible to observe the irregularity or other properties that maintains the same structure independently of the scale at which it is analysed [4] [13].

The official entity CORMAGDALENA registered 984 data on the fluctuations of the Magdalena River in the municipality of Barrancabermeja between September 29, 2011 and March 31, 2017. The figure 1 shows the graph corresponding to the 984 fluctuations recorded. Likewise, a line represents the average value of the data.

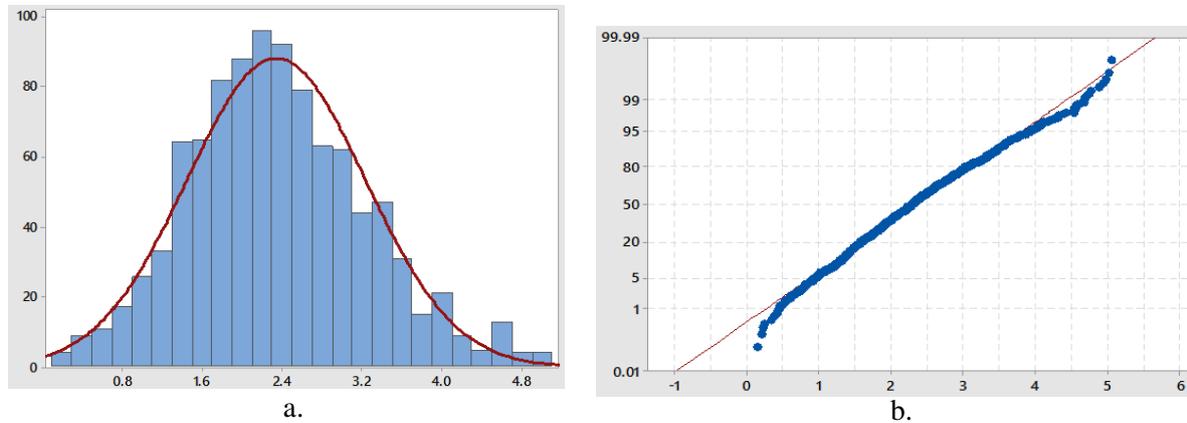


**Figure 1:** Data on the 984 fluctuations of the Magdalena river in the municipality of Barrancabermeja between September 29, 2011 and March 31, 2017

Thereupon, the Jarque-Bera test is performed to corroborate that the time series is not linear. In other words, it presents fractality. For the data analysed, the asymmetry value was 0.33 and 0.02 for kurtosis. Then the following hypotheses are posed:  $H_0$  (null hypothesis): Mean = 0; Variance = 1, which means normal distribution;  $H_1$ : Mean  $\neq$  0; Variance  $\neq$  1, it is not normal distribution. The following equation is used to find the value of the Jarque-Bera test (JB):

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

Where S represents the asymmetry and K the kurtosis. Therefore, the value of the test is 381.956, as the statistic  $JB > 6$ , it is rejected the null hypothesis. Thus, the series does not present any normality.



**Figure 2.** a. Histogram with normal curve of levels of the Magdalena River. b. Graph of probability of levels of the Magdalena River, proof Kolmogorov-Smirnov

Besides the auto-similarity, the fractal objects present another important characteristic, the fractal dimension. The Euclidean or topological dimension is one that assigns to a certain geometric and physical objects, a whole number, for example in a three-dimensional object that has thickness, width and height has three topological dimension. On the other hand, the fractal dimension assigns rational values, due to its irregularity, to certain geometric or physical objects. For example, the measurement of the perimeter of the coast of an island, the traditional measurement method does not allow to use a scale small enough, so that it can be measured in the same way, whereas with the fractal dimension it is possible to assign a value to such presence of irregularity of the contour of the same island. [13]

The Hurst Coefficient is one of the methods which it is possible to calculate the fractal dimension, associated to a time series for statistically self-similar data [24]. Using the accumulated sum and rescaled range method, the coefficient also allows measuring volatility, understand the maximum, and minimum peak values to which data are to be found in a time series with respect to the average value of the same, for the risk analysis of a time series [24]. The persistence of a time series depends on the value of the Hurst coefficient (H). If  $0 < H < 0.5$  the Hurst exponent has a high fractal dimension, it will exist for a time series with antipersistent behaviors. A larger exponent of H,  $0.5 < H < 1$  has a lower fractal dimension and thus a persistent analysed [24] [5] [25]. According to the statistical mechanics, if H is equal to 0.5, the series presents a random path [6].

The following is CR/S method the time series described in [26] 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  $s \equiv N/M$  and therefore there are s windows of data  $Y_j$ , with  $j = 1, 2, \dots, s$ . Defining the vector  $k = (j - 1)M + 1, (j - 1)M + 2, (j - 1)M + 3, \dots, (j - 1)M + M$ , the average over each window is calculated as:

$$\bar{Y}_j = \frac{1}{M} \sum_k x_k \quad (2)$$

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

$$z_n = \sum_k^n \{x_k - \bar{y}_j\} \quad (3)$$

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

$$R_j \equiv \max\{Z_j\} - \min\{Z_j\} \quad (4)$$

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

$$\sigma_j = \left[ \frac{1}{M} \sum_k (x_k - \bar{y}_j)^2 \right]^{\frac{1}{2}} \quad (5)$$

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

$$(R/S)_M \equiv \text{mean}(R_j/\sigma_j) \quad (6)$$

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

$$(R/S)_M = aM^H \quad (7)$$

Herein,  $a$  is a constant and  $H$  is the Hurst exponent which represents a fractal measure of long-range correlations in the analysed released.

## 5. Results

Due to the series presents fractality, the series of 984 data is partitioned into 4 subgroups [2]. The first subgroup formed by the first 246 data included between the 29<sup>th</sup> of September 2011 and the 16<sup>th</sup> of March 2013. With these data the standard deviation was calculated, and the rescaled range. The group 2 consists of the first 492 fluctuations recorded, whose dates vary between September 29, 2011 and July 21, 2014. For the group 3, a number of 738 fluctuations that were recorded between September 29, 2011 and November 25, 2015 and finally, group 4 is formed by the 984 fluctuations recorded during the time between September 29, 2011 and March 31, 2017.

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

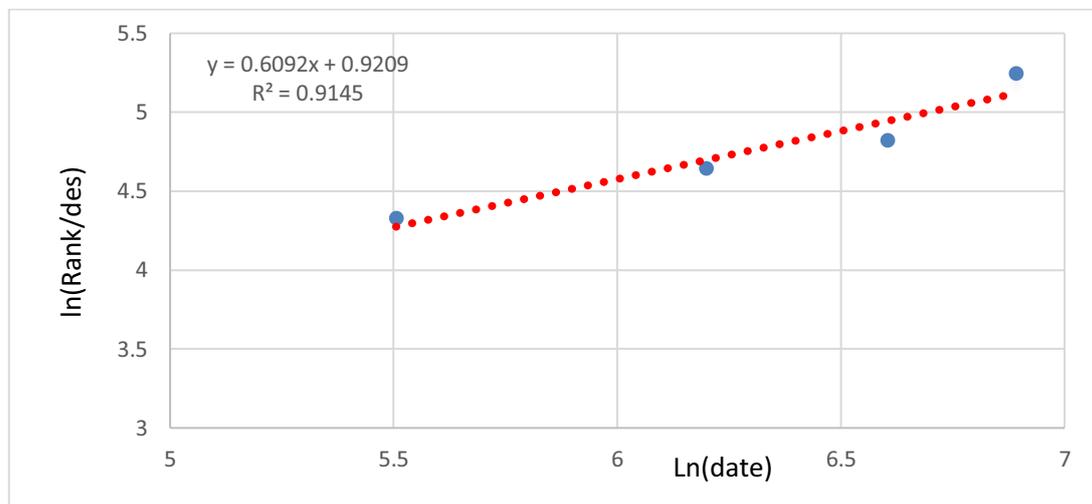
Subgroup	Number of released	Rank rescaled	Max	Min	Average
1	246	78.96	76.86	-2.10	2.66
2	492	90.61	84.95	-5.65	2.60
3	738	108.84	107.82	-1.02	2.41
4	984	169.21	128.71	-40.49	2.34

Table 2 shows the data needed for the calculation of the coefficient of hurts for this time-series, as well as 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

Subgroup	Number data	Rescaled range	Standard Deviation	Ln(Num)	Ln(R/S)
1	246	78.96	1.04	6.27	5.43
2	492	90.61	0.87	6.96	6.16
3	738	108.84	0.87	7.37	6.52
4	984	169.21	0.89	7.66	6.81

With this data we make a new graph in which 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, these data are related in Figure 3.



**Figure 3.** Logarithmic linear regression of the data supplied in Table 2

Performing the linear adjustment, with the regression, the value of the Hurst coefficient is obtained as the slope of the resulting line.

## 6. Conclusions

Along the period studied, the 984 fluctuations recorded by the official entity CORMAGDALENA between September 29, 2011 and March 31, 2017, was verified that this series did not present linearity by the Jarque-Bera test through equation (1). This implies that the time series could be a fractal. Due to the fact, there is a certain degree of self-sufficiency. In other words, the fractal dimension can be assigned to this time series. The value of the Hurst coefficient, which was calculated using the accumulated sums, rescaling range and logarithmic linear regression, is equal to 0.6092,  $H > 0.5$ , which means, the time series is persistent or has long-term memory. Given the condition of the equation  $D + H = 2$ , where  $D$  is the fractal dimension and  $H$  is the Hurst coefficient, we have that the fractal dimension of the series is  $D = 1.3908$ . From the data above presented, it can be concluded that the series is persistent under a volatility of 69.54%, which implies an average risk in future behaviour.

## References

- [1] Braun E 2011 *Caos, fractales y cosas raras* (Mexico: Fondo de Cultura Económica) p 92
- [2] Rodriguez E 2012 *Lat. Amer. Not. Adm.* **3** pp 42-50
- [3] Hurst H 1951 *Trans. Am. Soc. Civ. Eng.* 116 pp 770-808
- [4] Mandelbrot B 2009 *La Geometría Fractal de la Naturaleza 2* (España: Metatemas)
- [5] Peters E 1991 *Chaos and Order in the Capital Markets* (EUA: John Wiley & Sons, Inc.)
- [6] Peters E 1994 *Fractal market analysis: applying chaos theory to investment and economics*, (New York: Willey)
- [7] E. B. Duffo, «Banco de la República, actividad cultural,» 1 Junio 2013. [En línea]. Available: <http://www.banrepultural.org/blaavirtual/revistas/credencial/junio-2013/el-rio-magdalena>. [Último acceso: 1 mayo 2017].
- [8] Corporación Autónoma del Río Grande de la Magdalena, Presidencia de la República, «CORMAGDALENA,» 5 mayo 2016. [En línea]. Available: <http://niveles.cormagdalena.gov.co/>. [Último acceso: 2 abril 2017].
- [9] Ramirez R 2017 *Min. and Geo.* **33** pp 279-292
- [10] Ramirez R, Rodriguez A and Ordaz A 2017 *Min. and Geo.* **33** pp 36-49
- [11] Barbera P and Rosso R 1989 *Wat. Res. Res.* **25** pp 735-741
- [12] Tarboton D 1996 *J. Hyd.* **187** pp 105-117
- [13] Mandelbrot B 2006 *Los objetos fractales; forma, azar y dimensión 6* (España: Metatemas)
- [14] Falconer K 2006 *Fractal geometry. Mathematical foundations and applications, second edition* (India: Wiley)
- [15] Barnsley M 2012 *Fractals Everywhere* (United States: Dover Publications)
- [16] Erdf J 1988 *Fractals* (New York: Plenum Press)
- [17] Chen C, Lee Y and Chang Y 2008 *Phys. A* **387** pp 4643-4648
- [18] Mandelbrot B and Wallis J 1968 *Wat. Res. Res.* **4** pp 909-918
- [19] Hurst H 1951 *Trans. Am. Soc. Civ. Eng.* **116** pp 770-808
- [20] Mandelbrot B and Wallis J 1969 *Wat. Res. Res.* **5** pp 321-340
- [21] Mandelbrot B and Wallis J 1969 *Wat. Res. Res.* **5** pp 967-988
- [22] Sanchez M, Trinidad J and García J 2008 *Phys. A* 387 pp 5543 -5551
- [23] Gujarati D, Porter D 2010 *Econometría 5* (México: McGraw Hill)
- [24] Mandelbrot B, Hudson R 2010 *Fractales y finanzas, una aproximación matemática a los mercados: arriesgar, perder y ganar* (España: Metatemas)
- [25] Mandelbrot B 1997 *Fractals and Scalind in Finance* (EUA: Springer)
- [26] Martin L, Aranda N and Quimbaya C 2015 *Phys. A* **421** pp 124 -133