Physical applications: fractional Brownian movement applied to the particle dispersion

Y Herrera¹, D A Prada², J Ortega³, A Sierra³, and A Acevedo³
¹ Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia
² Universidad Pontificia Bolivariana, Bucaramanga, Colombia
³ Universidad Cuauhtémoc, Aguascalientes, México

E-mail: yherrera743@unab.edu.co, duwamg.prada@upb.edu.co

Abstract. Fractional Brownian motion allows you to see the correlation between past data increments in a shift with respect to future increments. Through this important physical property of fractional Brownian motion, it was possible to perform an analysis of the impact of suspended particles, also known as particulate matter, on the environment, which is a worldwide concern. The particulate matter produced in a coal mine is large and can be disastrous for neighboring communities, resulting in a decrease in their health and quality of life in general. In this work we analyze a time series of the values of total suspended particles registered by the “La Jagua” monitoring station located 5 km from the coal mine complex in the municipality of “La Jagua” de Ibirico in Colombia, we do this by using the fractional Brownian motion calculated from Hurst's coefficient to determine if there is persistence of the presence of this particulate material in the future. Initially a normality test is carried out, then the rescaled range and Hurst coefficient are computed with which it is possible to determine the increasing persistence and volatility of the series.

1. Introduction
A sustainable environment is the desire of every community, no matter if it is in a developed or in a third-world country. Unfortunately, the voracious race to increase wealth by industries is an adverse factor that hinders the reaching the goal of having that desirable environment. Colombia is no exception to this, and even less when it has the largest lung of the world: the Colombian Amazon. Year after year the armed groups, landowners, natural disasters (i.e. fires) make it much more difficult for this ecosystem to stay at peace. Due to pollution, we observe that there are places on the planet where people should use mask, they cannot perform outdoor exercises and the probability of respiratory-associated diseases is high, which evidently shows a low quality of life in that appearance [1].

In this paper we analyze a part of Colombia that is even richer thanks to its natural resources, a region that has one of the largest open-air coal mines in the world. Being more specific, the mine is located at the Municipality of “La Jagua de Ibirico”, in the center of the department of Cesar, in the foothills of the Eastern Rangers, 125 kilometers from the city of Valledupar, its capital. The municipality of “La Jagua de Ibirico” borders to the north with the municipality of Becerril, to the south with Chiriguáná, to the east by the Bolivarian Republic of Venezuela, and to the west by the municipality of El Paso. These four neighboring municipalities make up the mining corridor in the central subregion of Cesar, its territorial area is 728.93 km² and has a population of 46,722 inhabitants. This municipality ranks second income-wise in the Cesar department after Valledupar, the capital according to the 2016-2019
Development Plan of the Department [2]. In this document it also stated that the income increases of the 2012-2015 with respect to that of the 2008-2011 period was 130%, going from COP $348,651 million to COP$800 million. Even though the rate of income vs. population is high for this department, there is no prosperity in the levels of unsatisfied basic needs, and this is basically due to a problem of corruption. Hence, this municipality is recipient of lots of problems in addition to the money from the exploitation of the mine. In this paper we analyze the levels of environmental pollution.

The municipality of “La Jagua de Ibirico” is in the mining corridor of “Cesar”, it has a total suspended particles (TSP) monitoring station through a network installed since 2007 by the regional corporation of Cesar, “Corpocezar”. The suspended particles (SP) also known as particulate matter (PM) can be defined as any mixture of microscopic or very small compounds in the form of liquids and solids suspended in the air [3]. This mixture varies significantly in size, shape and composition, depending mainly on its origin. The particles are mainly composed of metals, organic compounds, material of biological origin, ions, reactive gases and the structure itself of the particles, normally formed by elemental carbon, the effect of the PM is represented in: visibility reduction, effect of smoke and dust on human health, chronic respiratory diseases, asbestosis, lead poisoning, contamination with dirt of houses and clothes, vegetation destruction, climate change, and many more [4].

Particulate matter of aerodynamic diameter less than 10 µm (PM10) and less than 2.5 µm (PM2.5) are considered extremely dangerous, especially the latter because it can affect human health when inhaled [5]. The allowed maximum media of this particulate matter are 50 µm/m³ in a year and 100 µm/m³ in a day, whereas for the PM2.5 they are 25 µm/m³ in a year or 50 µm/m³ in a day [6]; and for the TSP the maximum media is 100 µm/m³ in a year or 300 µm/m³ in a day [7].

In this paper we analyze the TSP values in the “La Jagua” monitoring station via Brownian motion theory. In the study on the processes of fertilization of flowers, R Brown in 1821, observed that the pollen particles presented a rapid irregular oscillatory movement suspended in a fluid. This phenomenon was defined as a Brownian motion. For the year 1905, the kinetic theory is positioned, and it asserts that Brownian motion is caused by a bombardment of the fluid molecules on the particle.

According to Einstein, Brownian motion is described through the probability p(r, t) of finding a particle at position r at time t, which satisfies the macroscopic diffusion equation [8]. Today, beyond the study of the dynamics of a particle in a fluid, Brownian motion has a very wide range of different applications in areas such as physics, hydrodynamics, polymer dynamics, seismology for the analysis of vibrations, generation of pseudo-random sequences, among others [8].

2. Mathematical model

A time series is a collection of data recorded in the order they are obtained in time, with time measured in a uniform scale [9]. The time series applications abound in all the fields of knowledge because they allow to model the behavior of a quantity in time.

In this paper we analyze the time series of the total suspended particles (TSP) measured by the “La Jagua” monitoring station, nearest to the “La Jagua’s” coal mines. The dataset we used corresponds to the daily TSP values (in µm/m³) from July 3, 2007 until December 28, 2017 and are 1080 records in total. The data was obtained from the official website of the “Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)” (http://sisaire.ideam.gov.co/ideam-sisaire-web/consultas.xhtml) in July 21, 2020.

The graph of the given series and its residuals are displayed in Figure 1. As can be noticed from Figure 1(a), the TSP values do not have a defined trend and are increasing in the last years which causes distress and despair, even more, the 300 µm/m³ a day is violated several times. We can also appreciate that the TSP value is chaotic in its behavior. In addition, by analyzing Figure 1(b) we can imply that the time series is a white noise, that is, it is a Brownian motion [10]. To prove that this time series is a Brownian motion we need to perform a normality test; first, we plot the histogram of residuals (Figure 2). Figure 2 shows that the density of the residuals of the TSP values is far from being normally distributed; then, we proceed to plot the quantile vs quantile (Figure 3). From Figure 3 is evident that
the residuals of the TSP values do not follow a normal distribution, this can be stated with a confidence of 95% as provided by the opensource software R (region in gray).

Figure 1. TSP time series and its residuals (a) TSP time series, (b) residuals of the TSP time series.

Figure 2. Histogram of the log of residuals of the time series.

Figure 3. Quantile-quantile plot of the time series.
Figure 2 and Figure 3 suggest strongly that the time series in question does not follow a normal distribution, that is, the PST is a Brownian motion, but we have to make sure this is the case. Therefore, we proceed to perform the Kolmogorov-Smirnov normality test K-S. The K-S test [11], is utilized to determine if a dataset corresponds to a specific distribution. In particular, the test is used to determine if the residuals of a time series are normally distributed. The K-S test is based on the empirical distribution function EDF, which is defined as follows:

Let $z_1, z_2, ..., z_n$ be a collection of ordered data points from the time series $\{x_i\}_{i=1}^N$, i.e., $z_i \leq z_{i+1}$, for $1 \leq i \leq n$; and let $n_i$ the number of $z_i$'s that are less than $z_i$. Then $EDF(i) = n_i/N$. The K-S test is then based on the maximum distance between the collections of points, $\{(i, z_i)\}_{i=1}^N$ and $\{(i, n_i)\}_{i=1}^N$. More precisely, it is stated as follows:

- $H_0$: The data follow a specified distribution
- $H_A$: The data do not follow the specified distribution
- Test statistic: The Kolmogorov-Smirnov test statistic is defined by the Equation (1) given in [11].

$$D = \max_{1 \leq i \leq N} \left(F(z_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right),$$  \hspace{1cm} (1)

where $F$ is the continuous theoretical cumulative distribution of the distribution being tested and max is the maximum of the two given values.

- Significance level: $\alpha$
- Critical values: The hypothesis $H_0$ is rejected if the test statistic, $D$, is greater than the critical value obtained from a table or a software program. In this work, we perform the K-S test in the open source software R.

In this work we use a significance level $\alpha = 0.05$, in which case, the critical value is $0.001572$ and the critical region is: Reject $H_0$ if $D > 0.001572$. After performing the K-S test in RStudio, we obtain that for this time series, $D = 0.057555$, therefore, we conclude the time series is not normally distributed since its $D$ value is in the reject region of the hypothesis test.

Since the time series does not present a normal distribution, it is possible to analyze its behavior in terms of its persistence and volatility with alternative methods useful in the area of physics, such as the fractional Brownian movement (FBM) that, unlike the ordinary Brownian movement (OBM), shows the correlation of the past data increments in a displacement with respect to the future increments, for which a memory system is observed. The FBM is characterized by a parameter, the so-called Hurst parameter $H$. An FBM with Hurst parameter $H > 1/2$ is called a persistent process, i.e., the increments of this process are positively correlated. On the other hand, the increments of an FBM with $H < 1/2$ constitute what is called an anti-persistent process, with increments being negatively correlated. For $H = 1/2$, an FBM corresponds to Brownian motion which has independent increments [12].

The Hurst coefficient of a time series is determined by partitioning the complete data, that is, subgroups are processed by accumulation of data. For each of this subsequences, find the average and the deviations of the data from that average. Using cumulative sums of the deviations, calculate the maximum and minimum of this process to determine the rescaled range associated with the data. With this value and the standard deviation, calculated in each subgroup, a logarithmic linear regression is performed where the slope is the Hurst coefficient. Hurst found that the a-dimensional ratio $R/S$ allows comparing the re-sizing of several temporal series and that such a resizing can be very well described by a law of power as follows in the Equation (2):

$$(R/S) = cN^H,$$  \hspace{1cm} (2)
where \( N \) is the time interval for the observations, \( H \) is the estimate for Hurst's Exponent as calculated from the R/S method and \( c \) is a constant. This Equation (2) and how the coefficient is calculated Hurst shown in more detail in [13,14]. Using Equation (2), it is possible to perform a logarithmic linear regression in which the slope represents the Hurst coefficient.

3. Results
To know the behavior of the time series associated with the total suspended particles (TSP) measured by the “La Jagua” monitoring station, nearest to the “La Jagua’s” coal mines TSP, the study of the fractional Brownian motion of the series will be carried out by computing the Hurst coefficient and the logarithmic linear regression of Equation (2), in order to determine if there is persistence or memory in the future. With the 1080 data collected from the SISAIRE platform between July 3, 2007 and December 28, 2017, four subgroups were made in which the rescaled range, the standard deviation of ordinary and logarithmic form, this grouping and its values were calculated. In each of the subgroups, shown in Table 1, the rescaled range was calculated, an indicator of the difference of the accumulated sums of the deviations of the data with respect to the average value of the number of data of each subgroup. With the rescaled range indicator and the standard deviation, the quotient between these data is calculated. Then Equation (2) is structured for each subgroup and then linear regression is performed to find the coefficient \( H \) associated with the fractional Brownian motion. The calculations are shown in Table 1, where the related data for the four groups are specified based on range, standard deviation and calculation of the natural logarithm of the number of data 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        | 270         | 2505.26        | 67.69             | 5.60     | 3.61    |
| 2        | 540         | 3794.14        | 61.31             | 6.29     | 4.12    |
| 3        | 810         | 7114.61        | 59.13             | 6.70     | 4.79    |
| 4        | 1080        | 9992.91        | 59.70             | 6.98     | 5.12    |

Next, the logarithmic linear regression is performed, taking as data in the coordinate the natural logarithm of the number of data for each subseries, and in the ordinate, we take the natural logarithm of the quotient between the rescaled range and the standard deviation corresponding to each subseries created, these data are related in Figure 3.

![Graph showing logarithmic linear regression](image)

Figure 4. Logarithmic linear regression of the data supplied in Table 1.
With the data provided in Table 1 and the logarithmic regression in Figure 4, we observe that the slope of the regression, which is the same Hurst coefficient, is 1.1062, which implies that the time series of the data recorded by SISAIRE with respect to a TSP have memory, that is, the behavior that we have analyzed between July 3, 2007 and December 28, 2018 will be replicated, if the same environmental policies of the environment continue to be clear. The volatility of the persistence of associated with the time series is 44.69%, that is to say that with respect to the mean of the data we have that measurements can be found in the environment of 100.65 μg/m³ to 159.58 μg/m³ particles.

4. Conclusion
According to the study carried out in the monitoring station of the municipality of “La Jagua de Ibirico”, the TSP results can be found in the environment between 100.65 and 158.58 per day, which implies that it is close to the maximum allowed Colombian norm of 300 in two days. On the other hand, the results show that the analyzed behavior tends to be repeated, which can cause that in a short time the allowed norm can be exceeded, which would cause damage to the environment and health. It would be recommended to carry out prevention activities to minimize the emission of total particulate matter into the environment.

References
[1] Prada D et al 2020 J. Phys. Conf. Ser. 1514 012016:1
[2] Asamblea del Departamento del Cesar 2016 Plan de Desarrollo del Municipio de la Jagua de Ibirico 2016-2019 (Colombia: Asamblea del Departamento del Cesar)
[3] Monn C, Braendli O, Schaeppi G, Schindler C, Ackermann-Liebrich U, Leuenberger P 1995 Atmospheric Environment 29(19) 2565
[4] Miller B 2011 Clean Coal Engineering Technology (Tokyo: Elsevier)
[5] Xing Y, Xu Y, Shi M, Lian Y 2016 Journal of Thoracic Disease 8(1) 69
[6] Hernández A 2013 Hoja Metodológica del Indicador Concentración Promedio Anual de Partículas Suspensas en el Aire – PST (Versión 1.00) (Colombia: Sistema de Indicadores Ambientales de Colombia. Colombia: Instituto de Hidrología, Meteorología y Estudios Ambientales)
[7] Ministerio del Medio Ambiente de Colombia 2017 Resolución 2254 (Bogotá: Ministerio del Medio Ambiente de Colombia)
[8] Giraldi H, Campos E 2015 ¿Es el movimiento Browniano un proceso estocástico o determinista? Boletín de la Sociedad Mexicana de Física 29(2) 103
[9] Shumway R, Stoffer D 2005 Time Series Analysis and its Applications (Germany: Springer Texts in Statistics)
[10] Mandelbrot B 2006 Los Objetos Fractales. Forma, Azar y Dimensión (España: Metatemas)
[11] Chakravarti M Laha G, Roy J 1967 Handbook of Methods of Applied Statistics, Volume I (New York: John Wiley and Sons)
[12] Yerlikaya F, Vardar C, Yolcu Y, Weber G 2013 Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method J. Compu. App. Math 259 843
[13] Martin L, Aranda N, Quimbaya C 2015 Phys. A 421 124
[14] Henrice E, Palma D, Conçalves A, Mesquita A 2017 Support to the identification of anomalies in an external neutron source using Hurst Exponents Progress in Nuclear Energy 99 119