Fractional Brownian motion and Hurst coefficient in drinking water turbidity analysis

D A Prada1, Y A Herrera-Jaramillo2, J Ortega3, and J Gómez4
1 Universidad Pontificia Bolivariana, Bucaramanga, Colombia
2 Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia
3 Universidad Cuauhtémoc, Aguascalientes, México
4 Servicio Nacional de Aprendizaje, Bogotá, Colombia

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

Abstract. Brownian motion is a physical phenomenon studied by biologist Robert Brown through which it is possible to describe the random irregular motion of pollen particles suspended in a fluid using the probability of finding a particle at a specific time. In a fractional way, the Brownian movement describes the random fluctuation of continuous stochastic processes which can be characterized by the Hurst coefficient whose initial utility was the observation of the levels of the Nile River to predict the persistence of future phenomena and thus prepare for the drought or flood. Fractional Brownian motion has been used for study in various areas such as hydrodynamics. Water is a vital natural resource for the preservation of species and less than 1% of fresh water is available for human consumption. Turbidity is a measure of the degree of transparency that water loses, the greater the number of solids suspended in the water, the greater the degree of turbidity. With the data provided by the “Acueducto Metropolitano”, Bucaramanga, Colombia between January 2008 and August 2019, volatility and persistence of turbidity were established in the Colombia municipalities of Bucaramanga, Floridablanca and Girón. In the latter, persistent behavior and extreme volatilities are observed.

1. Introduction
Turbidity is considered a good measure of the quality of water because it quantifies the degree to which the water loses its transparency [1]. It is mainly caused by Phytoplankton, sediments from erosion, resuspended sediments from the bottom, waste discharge, algae growth and urban runoff. When some of these solids are present in water the light intensity through it is affected, therefore turbidity can be measured optically as a light-scattering property of water [2]. The more suspended solids the water has the more turbid it is.

The values of turbidity are used to establish the level of disinfection the water needs, the most convenient filtration rate, the coagulation process effectiveness, as well as to determine the drinking water quality. It is essential to remove the turbidity of water in order to effectively disinfect it for drinking purposes because turbidity can act as a shield to pathogens and its particles can harbor bacteria and viruses, in addition, it can affect the taste and odor of the water [3]. The water treatment process to remove turbidity includes coagulation, flocculation, sedimentation, filtration, and disinfection.

Although turbidity is not a direct indicator of health risk, numerous studies show a strong relationship between the elimination of turbidity and the elimination of protozoa. Turbidity particles provide a "refuge" to microbes by reducing their exposure to the attack of disinfectants. It has been considered
that microbial binding to particulate material helps the survival of microbes. Fortunately, traditional water treatment processes have the ability to effectively eliminate turbidity when operated correctly [3]. The nephelometric method for measuring turbidity is based on a comparison between the intensity of the light dispersed by a sample under defined conditions and the intensity of light dispersed by a standard reference suspension under the same conditions. The formation polymer is used as the primary standard reference suspension and the turbidity of a specified concentration of formation suspension is defined as having 4000 nephelometric turbidity units (NTU) [4]. Turbidity measures are prone to interference from debris, sediments and water bubbles.

In Colombia there is a maximum acceptable NTU value of two (2) according to the criteria established in Decree 1575 of 2007 and Resolution 2115 of 2007 of the “Ministerio de Protección Social y Vivienda y Desarrollo Territorial” [1], entity in charge of regulating the quality of drinking water in this country. This value is high compared to the ideal level set by the World Health Organization (WHO) that is below 1 NTU. Turbidity below 4 NTU can only be detected by instruments, however, at 4 NTU and above, a milky white, muddy, red-brown, or black suspension may be visible and may reduce the acceptability of drinking water.

The “Acueducto Metropolitano, Bucaramanga (AMB)”, Colombia, is the company that provides water and sewerage services to the city of Bucaramanga and the municipalities of Floridablanca and Girón. The water sources of AMB, Colombia, are the Suratá river, the Tona river and the Frio river, which have a treatment capacity of 2000, 1400 and 600 liters per second, respectively. The extraction system used in the Suratá river, Colombia, is the pumping system, while in the other two is used the gravity one.

In this article, we analyze the relationship between rainfall and increased turbidity in the last ten (10) years measured monthly, in the city of Bucaramanga, and the municipalities of Floridablanca and Girón, in the department of Santander, Colombia, by means of an analysis of the time series of the data of this time interval. Due to the nature of the phenomenon, alternative methods based on the analysis of the standardized route or the rescaled range method will be used, a statistical test that is maintained to quantify the dynamics of the time series by determining the fractional Brownian motion from the calculation of the Hurst coefficient.

2. Mathematical method

Time series analysis is a tool that permits to observe the behavior of different phenomena to be able to predict the trend of the same. A time series is a sequence of observations, measured at certain intervals of time, ordered chronologically and evenly spaced in a uniform way, and whose main objective is their analysis to make forecasts [1]. Different techniques are used with respect to time series, for instance, the one of stationary type of weak condition, for which the mean and variance of the analyzed data are required to be constant over time, and the amount of data is important for the self-regressive processes based on loss of degrees of freedom [2]. In the classic analyses of the time series, three main components can be observed: Trend that is defined as a long-term change in the ratio at medium level, seasonality that shows some periodicity in a given interval of time, and random which does not follow any pattern of behavior [1]. With regard to time series involving natural phenomena, in particular hydrological series, its importance turns up evident because it is a vital resource and therefore the analysis can be developed in different ways, in general, depending on the particular system in question, either the generation of univariate or multivariate, annual or seasonal synthetic series [3] may be required. Among the most commonly used mathematical models in this type of series, the best known is the integrated auto-regressive integrated moving average (ARIMA) process, which is based on the implicit assumption of the linearity of the system that generates the variables trajectory [4].

Robert Brown in 1821 carried out a study on the irregular oscillatory motion of pollen particles suspended in a fluid, which would later be defined as Brownian motion. For the year 1905, the kinetic theory that the Brownian motion is caused by a bombardment of the fluid molecules in the particle is positioned. 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 [5,6]. Today,
beyond the study of the dynamics of a particle in a fluid, Brownian motion has a wide range of different applications in areas such as hydrodynamics, polymer dynamics, seismology for the analysis of vibrations, the generation of pseudo-random sequences, among others [5]. The fractional Brownian motion (FBM), past increments in displacement are correlated with future increments, at least for the first steps of the process, hence the system has memory. Hurst exponent for a time series provides a measure of whether it is a pure white noise random process or has underlying trends. Dynamic process, that might naively characterized with purely white noise, sometimes turn out to exhibit Hurst exponent statistics for long memory process, i.e., colored noise. A long memory process is a process where past events have a decaying effect on future ones. But those are forgotten as time moves forward [7]. The FBM is characterized by a parameter, the so called Hurst parameter H [8].

The famous hydrologist Harold Edwin Hurst studied the fluctuations of the heights of the Nile River for long periods of time to be able to project reserve capacities and thus take precautionary measures in times of drought. For this purpose, Hurst created a statistical methodology which consists of analysing the persistence of a time series [9]. 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 [10]. The standardized path analysis, also called rescaled range analysis (R/S), is a statistical test used to quantify the dynamics of a time series [11], this methodology was developed by Hurst, and obtains the H coefficient, or exponent of scaling, which can take values between zero and one [12].

The mathematical procedure of the R/S analysis consists of the following steps [13]: First, for every integer 1 < n < N, where N is the number of total observations in the series, let < x >n and s(n) be the mean and the standard deviation of the subseries consisting of the n first xj’s, respectively. Second, we compute the following Equation (1).

\[ X(i, n) = \sum_{1 \leq j \leq i} (x_j - < x >_n). \]  

Third, we calculate the range associated with each of the given subseries by using Equation (2).

\[ R(n) = \max_{1 \leq i \leq n} X(i, n) - \min_{1 \leq i \leq n} X(i, n). \]  

Finally, we compute the generalization, i.e., the ratio between the range R(n) and the standard deviation s(n). The generalization follows an exponential law with the H coefficient as exponent, Equation (3).

\[ \frac{R(n)}{S(n)} \approx cn^H. \]  

The H coefficient corresponds to the slope obtained by the classic linear regression model adjusted by an ordinary least square’s estimator based on the logarithmic transformation of the generalization equation [13]. In [14], the time series ratings according to the Hurst coefficient are shown, for an exponent under 0 < H < 0.5, this implies that the series is anti-persistent (pink noise), i.e. more volatile, therefore more risk, this is a manifestation of non-similar trends in the future, if H > 0.5, the series is persistent, i.e. the series presents memory and similar behaviors in the future (black noise) and if H = 0.5, the series is random (white noise) which implies that it is a Brownian movement [15].

3. Results
By calculating the Hurst coefficient associated with the time series of the fluctuation of turbidity present in drinking water in each of the municipalities under study. According to the value of the Hurst coefficient, we observe whether the time series presents persistent or future memory fractional Brownian motion (black noise), higher volatility antipersistent with some risk (pink noise), or if it is non-fractional
Brownian motion (white noise). This phenomenon studied from physics allows characterizing time series under the idea of persistence, which indicates the care we must take with drinking water. From AMB 140 data were collected in a monthly time series from January 2008 to August 2019 with the aim of observing turbidity in treated drinking water distributed in Bucaramanga, Floridablanca, and Girón, Colombia. Every time series of turbidity of each municipality was subdivided into 4 temporary subseries. Table 1 shows the amount of data, range, standard deviation, natural logarithm of the amount of data, and the natural logarithm of the rescaled range (R/S) of each subseries of Bucaramanga, Colombia.

| Subgroup | Number data | Rescaled range | Standard deviation | Ln (Num) | Ln(R/S) |
|----------|-------------|----------------|--------------------|----------|---------|
| 1        | 35          | 2.78           | 0.29               | 3.56     | 2.25    |
| 2        | 70          | 4.23           | 0.27               | 4.25     | 2.77    |
| 3        | 105         | 3.97           | 0.29               | 4.65     | 2.61    |
| 4        | 140         | 4.00           | 0.27               | 4.94     | 2.71    |

Figure 1 shows the slope obtained by the classic linear regression model adjusted by an ordinary least square’s estimator based on the logarithmic transformation of the generalization equation (Equation (3)) for Bucaramanga, Colombia.

![Figure 1. Logarithmic linear regression of the data supplied in Table 1.](image)

Table 2 and Table 3 show the amount of data, range, standard deviation, natural logarithm of the amount of data, and the natural logarithm of the rescaled range (R/S) of each subseries of the municipality of Floridablanca and Girón, Colombia, respectively.

| Subgroup | Number data | Rescaled range | Standard deviation | Ln (Num) | Ln(R/S) |
|----------|-------------|----------------|--------------------|----------|---------|
| 1        | 35          | 2.30           | 0.31               | 3.56     | 1.99    |
| 2        | 70          | 4.80           | 0.38               | 4.25     | 2.55    |
| 3        | 105         | 8.46           | 0.37               | 4.65     | 3.13    |
| 4        | 140         | 12.71          | 0.37               | 4.94     | 3.53    |

| Subgroup | Number data | Rescaled range | Standard deviation | Ln (Num) | Ln(R/S) |
|----------|-------------|----------------|--------------------|----------|---------|
| 1        | 35          | 3.95           | 0.35               | 3.56     | 2.41    |
| 2        | 70          | 4.59           | 0.33               | 4.25     | 2.63    |
| 3        | 105         | 6.48           | 0.34               | 4.65     | 2.94    |
| 4        | 140         | 7.16           | 0.33               | 4.64     | 3.09    |
Figure 2 shows the slope obtained by the classic linear regression model adjusted by an ordinary least square’s estimator based on the logarithmic transformation of the generalization equation (Equation (3)) for the municipalities of Floridablanca, Colombia (Figure 2(a)), and Girón, Colombia (Figure 2(b)).

4. Conclusions
The physical study of the random irregular movement of pollen particles suspended in a fluid, better known as Brownian movement, has allowed contributions to other areas of knowledge. Using the Hurst coefficient and fractional Brownian motion, it is possible to analyze time series to know the possible future behaviors of our object of study in relation to persistence or not, whose usefulness depends on the context. With regard to the fluctuation of the turbidity levels present in drinking water provided by the AMB, it was found that in Bucaramanga, Colombia, a regression model with equation and $y = 0.3012x + 1.275$ with a correlation coefficient $R^2$ of 0.6209 was presented, where the value of the coefficient of Hurst $H = 0.3012$, which is less than 0.5. Consequently, this series is pink noise, i.e., it is anti-persistent with volatility and therefore it has a greater risk of change. Regarding the analysis of the of the water turbidity in the municipalities of Floridablanca and Girón, Colombia, the models found have equations $y = 1.1077x - 2.0197$ with a correlation coefficient $R^2$ of 0.9788 and $y = 0.5544x + 0.3966$ with a correlation coefficient $R^2$ of 0.8812 respectively. The above indicates that for both Floridablanca and Girón, Colombia, the H coefficient is greater than 0.5, which implies that both time series are black noise, i.e., there is persistence for Floridablanca and for Girón, Colombia. This allows us to conclude that the level of turbidity in the municipality of Floridablanca, Colombia, will remain with 38% variability relative to the mean and for the municipality of Girón, Colombia, with 72.28% variability relative to the mean.

References
[1] Boyd C 2015 Water Quality: An Introduction (Switzerland: Springer International Publishing)
[2] Popek E 2018 Sampling and Analysis of Environmental Chemical Pollutants: A complete Guide (Cambridge: Elsevier)
[3] World Health Organization 2017 Guidelines For Drinking-Water Quality: Incorporating The First Addendum (Geneva: World Health Organization)
[4] American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF) 2012 Standard Method 2130: Turbidity. Standard Methods for the Examination of Water and Wastewater, 22nd edition (Washington: American Public Health Association/American Water Works Association/Water Environment Federation)
[5] Giraldi H, Campos E 2015 ¿Es el movimiento Browniano un proceso estocástico o determinista? Boletín de la Sociedad Mexicana de Física 29(2) 103
[6] Schlesinger M, Klafter J, Zumofen G 1999 Above, below and beyond Brownian motion American Journal of Physics 67(12) 1253
[7] Mancilla M, Leguto A, Riquelme B, Ponce P, Bortolato S, Korol A 2017 Hurst exponent: A Brownian approach to characterize the nonlinear behavior of red blood cells deformability Physica A: Statistical Mechanics and its Applications 488 1

[8] Yerlikaya F, Vardar C, Yolcu Y, Weber G 2013 Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method Journal of Computational and Applied Mathematics 259 843

[9] Peters E 1996 Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility (United States: John Wiley & Sons, Inc.)

[10] Mandelbrot B, Hudson R 2006 Fractales y Finanzas, una Aproximación Matemática a los Mercados: Arriesgar, Perder y Ganar (Barcelona: Tusquets Editores)

[11] Fernández B 1985 Análisis de la periodicidad en series hidrológicas Anales de la Universidad de Chile 8 635

[12] García I 2003 Análisis y predicción de la serie de tiempo del precio externo del café colombiano utilizando redes neuronales artificiales Universitas Scientiarum 8 45

[13] Hurst H 1951 Long-term storage capacity of reservoirs Trans. Am. Soc. Civ. Eng. 116 770

[14] Peters E 1994 Fractal Market Analysis: Applying Chaos Theory to Investment and Economics (New York: Willey)

[15] Mandelbrot B 2006 Los Objetos Fractales. Forma, Azar y Dimensión (Barcelona: Tusquets Editores)