Fractal analysis for estimating electricity demand through the application of Hurst's re-staging exponent

F Mesa¹, D M Devia¹, and R Ospina²

¹ Departamento de Matemáticas, Universidad Tecnológica de Pereira, Pereira, Colombia
² Escuela de Física, Universidad Industrial de Santander, Bucaramanga, Colombia

E-mail: femesa@utp.edu.co

Abstract. This article presents a methodology for fractal analysis using the Hurst re-staging exponent for the generation prediction of a hydroelectric power plant. Which is an innovative technique to address this problem since literature usually performs it through classical and lately Bayesian statistical processes. The importance of the work implies amalgamation of concepts from fractal theory, chaos, phase space, delay time and Hurst exponent, necessary in the analysis that will determine an estimate of the future behavior of a chaotic series. Thus, the problem is addressed from the qualitative analysis through the theory of chaos.

1. Introduction

It is common to find phenomena that do not have all the necessary information to express the behavior of the system in the form of equations, a clear example of this fact in the mathematical description given of how the population size of prey / predatory species that cohabit through the Lotka-Volterra equations [1,2]. Despite the simplification made to model these phenomena, the equations that determine the system are often very complex.

Thus, under certain hypotheses about the equations, the behavior of these systems varies from converging to a particular state, called equilibrium point, converging to a set of states that are repeated periodically, called the limit cycle or at other times having a behavior very irregular. These sets to which the solutions of the systems converge are called attractors [3]. The problem that often arises when you want to describe in broad strokes the central dynamics of the phenomenon in question, is manifested in the fact that you do not quantitatively dispose of all the necessary basic variables and you only have one or one of them.

Frequently, scientists have only the possibility of making partial observations of the phenomenon or in other cases making measurements of the entire phenomenon is very expensive, so they limit themselves to observing part of it, therefore the need arises to know the states of the system based on the observation of their behavior [4].

One of the most used techniques for the prediction of future values is known as the Hurst exponent which consists of a statistical method that indicates the level of persistence of a time series from the observations, in this way it can be estimated how the series moves from its initial point to determine whether it is a random movement or not [5].
2. Fractal analysis
The fractal is a mathematical object in which the whole is exactly or approximately similar to a part of itself: that is, the whole has the same shape as one or more of its parts; as an illustrative case of this we show the "fractal tree" which is shown in Figure 1.

![Fractal tree](image)

Figure 1. Fractal tree [6].

An object that possesses this quality is statistically similar on different scales, be they spatial or temporal. Hence, we can speak of the fractal dimension as the way in which this object fills space, a definition that helps us to predict the behavior of a system and also postulate that the fractal dimensionality determines the degree of irregularity or variability of a time series.

2.1. Chaos
There are various definitions of chaotic systems as far as mathematics or physics is concerned; however, in ours we will define it as that aperiodic deterministic system, the characteristics that identify these systems being non-linearity, extreme sensitivity to very small changes in their initial conditions and the fact that there is no way to predict the behavior of the system until the process occurs or is calculated. Despite this, the chaotic system is deterministic.

2.2. Hurst re-staging analysis
In fractal analysis of time series, the Hurst exponent is considered the most important or most relevant parameter; since when taking a pair of identical values in a time series, the greater the delay between them, the lower the Hurst coefficient [7]. Hurst's statistical model was based on Albert Einstein's work on Brownian motion, which was basically the model of the particle's random walk. The essence of the theory is that the distance traveled by the particle $R$ is proportionally increased to the square root of time $T$ as showed in Equation (1).

$$R = T^{0.5}. \tag{1}$$

Equation (1) is widely known in finance and is regularly used to annualize volatility or standard deviation. Hurst thought it could be based to check randomness of the overflow behavior on the Nile River and devised the following methodology [8]. Be $x_1, x_2 \ldots x_n$, a time series of $n$ components of the variable to be studied. The average value of this time series is defined as Equation (2):

$$x_m = (x_1 + x_2 + \cdots + x_n)/n. \tag{2}$$

And the standard deviation as Equation (3).
s_n = n^{-1/2} \sqrt{\sum(x_r - x_m)^2}. \quad (3)

The average of the series is given by Equation (4).

\[ z_r = (x_r - x_m); \quad r = 1, ... n. \quad (4) \]

And a series of accumulated sums is generated as the following Equation (5) shows.

\[ Y_r = \sum_{i=1}^{r} z_i; \quad r = 1, ..., n. \quad (5) \]

Note that the last component of the vector \( Y_r \) must be zero, due to the standardization performed in Equation (4). Finally, the adjusted range of \( x \) called \( R_n \) is calculated as shown in Equation (6). It can be shown that the value of \( R_n \) is not negative [9].

\[ R_n = \max(Y_1, ..., Y_n) - \min(Y_1, ..., Y_n). \quad (6) \]

The adjusted range \( R_n \) represents the maximum distance that the process moves in a given time \( n \), then, one could think of applying Equation (1) taking \( T = n \); however, this only applies to Brownian movements, that is, time series with zero mean and standard deviation 1. Hurst found that a generalization of Equation (1) can be written as shown in the following expression Equation (7).

\[ \left( \frac{R}{S} \right)_n = cn^H, \quad (7) \]

where the value \( \frac{R}{S} \) is called the scaled range of \( x \) and \( c \) is a constant. This scaled range is expressed in terms of the local standard deviation. In general, the value of the scaled range changes with the increase of \( n \) by means of an exponential law dependent on parameter \( H \), which is called the Hurst exponent. Then calculating the logarithm of Equation (7) you have the Equation (8).

\[ \log \left( \frac{R}{S} \right)_n = \log(c) + H \log(n). \quad (8) \]

Therefore, the Hurst exponent can be calculated from the linear fit on the graph of the logarithm of time \( n \) of Equation (8) this value will be number \( t \) number \( (0; 1) \) depending on the characteristics of the series studied. The following Equation (9) is used to calculate the Hurst exponent.

\[ \ln \left( \frac{R}{S} \right)_n = a + H \ln(n), \quad (9) \]

where \( \frac{R}{S} \) corresponds to the statistic that depends on the size of the series and is defined as the range of variation of the series broken by its standard deviation, \( H \) is the Hurst coefficient, \( n \) is the number of samples is already a constant.

The Hurst coefficient is used for long-term memory arrest in time series [10]. If found, there would be clear evidence of the nonlinear behavior of the series to be analyzed. According on the value of the Hurst exponent, the following classifications could be made for the time series:

- If \( H = 0.5 \): it will indicate that it is an independent process that is to say the past data does not affect the futures and the series would be of time would be totally random.
- If \( 0.5 < H \leq 1 \), it indicates persistent time series, that is, characterized by long-term memory effects.
If $0 \leq H < 0.5$, it means antipersistence in the time series. An anti-persistent system covers less distance than a random one, in the case of an erratic particle. In order for it to happen, it must reverse itself more frequently than a random process.

2.3. Phase space

Phase Space is a way of representing a dynamic system. It consists of the construction of a space that has as many dimensions as the number of variables necessary to specify the state of the original system. Each coordinate axis of this space represents one of the variables that make up the system. In essence, the phase space represents, with a certain vector structure, the set of possible states in which the modeled system may be [11].

The orbit of a particular state will be represented by a curve or trajectory in this space; this representation of the system allows a qualitative description of the temporal evolution of the model we are studying. One of the most used techniques for the reconstruction of phase spaces is the delay time method, if you have a time series of a variable scalar it is possible to construct a vector, Equation (10).

\[
x(t_i), \quad i = 1, \ldots, N.
\]  

(10)

The phase space in time can be described as follows Equation (11):

\[
X(t_i) = [x(t_i), x(t_i + \tau), x(t_i + 2\tau), \ldots, x(t_i + (m - 1)\tau)],
\]  

\[
\quad \text{where each consecutive element } x \text{ represents one of the observations, } i \text{ goes from } 1 \text{ to } N - (m - 1) \tau, \tau \text{ is known as the delay time or lag time, } m \text{ is the reconstructed space dimension or embedding dimension and } M = N - (m - 1) \tau \text{ is the number of points (states) in the phase space.}
\]

According to the embedding theorem, when these parameters are calculated correctly, the reconstructed dynamics using this formulation are equivalent to the dynamics of an attractor at the origin of the phase space, which means that the invariant characteristics of the system are preserved [12].

2.4. Delay time $\tau$

The choice of the delay time for the reconstruction of the attractors is fundamental since in order for the embedding theorem to be fulfilled, an appropriate value of $\tau$ must be chosen which must meet certain criteria such as:

- Be a multiple of the sampling time of the series, since there are only data for these times. An interpolation scheme to "get more data" is as uncertain as an estimate of the series derivative.
- If the lag time is too short, the $x(t_i)$ and $x(t_i + \tau)$ coordinates that will be used in the data reconstruction vector will not be independent enough. That is, not enough time has passed for the system to evolve in the state space.
- If $\tau$ is very large, any relationship between $x(t_i)$ and $x(t_i + \tau)$ is insignificant and is lost numerically.

In this way, a lag time is sought that is large enough so that $x(t_i)$ and $x(t_i + \tau)$ are somewhat independent but not so much that all statistical dependence is lost. There are different techniques for calculating the delay time, in this article we propose the use of the method based on the sample auto-correlation function. This is defined for a time series scalar $x_n$ of N samples as Equation (12) and Equation (13):

\[
\rho(\tau) = \frac{\sum_{n=1}^{N}(x_{n+\tau} - \bar{x})(x_n - \bar{x})}{\sum_{n=1}^{N}(x_n - \bar{x})^2}
\]  

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x(t_i)
\]  

(12)  

(13)
Where $\bar{x}$ denotes the sample mean of the series $x_n$. The criterion for the selection of $\tau$ is to choose the lag time where $\rho(\tau)$ reaches its first zero. In this way $x_n$ and $x_{n+\tau}$ will be linearly incorrect. In [10] it has been found that this criterion does not work in all situations, because long memory processes take a long time to reach the value of zero for $\rho(\tau)$, which makes their use impractical because Data samples are finite. Therefore, an alternative criterion is suggested which consists in choosing the lag time $\tau$ where a value of $\frac{1}{e}$ is reached. This method is more robust to choose an adequate lag, however there is no evidence to show that $\frac{1}{e}$ is a universal factor of the autocorrelation function criterion for the choice of adequate lags [11].

3. Analysis and results

In order to estimate the behavior of a chaotic series, the behavior of the Colombian electricity demand of the last 70 months was analyzed as shown in Figure 2.

![Figure 2. Colombian electricity demand.](image)

With the data obtained, the Hurst exponent will first be found. For this, the total amount of data is divided into 7 subgroups, in total these divisions must ensure that the exponent of the global set and that of the subsets remains without many variations. Figure 3 shows the graphical behavior of the data, the $R/S$ method and the Hurst exponent. Where the exponent can be calculated as the intersection between the 3 graphs.

![Figure 3. Hurst exponent behavior.](image)
When performing the calculations, a Hurst exponent of 0.9798 was obtained, which indicates that the system is persistent, therefore it is possible to estimate future values with a low degree of uncertainty. In order to reconstruct the phase space of the system, the delay time was calculated using the methodology proposed in [7] for this purpose the autocorrelation function was plotted as shown in Figure 4.

![Autocorrelation function](image)

**Figure 4.** Autocorrelation function.

As shown in the Figure 4, the value of $\tau$ is between 16 and 17, for this case 16 will be taken as the delay time for the reconstruction of the attractor as shown in Figure 5.

![Reconstruction of the attractor](image)

**Figure 5.** Reconstruction of the attractor with $\tau = 16$.

4. **Conclusions**

The reconstruction of the phase space of a chaotic system is a very important tool for estimating future values for cases in which there are not enough means to know the equations that govern the system. The calculation of the attractor poles of the system does not guarantee a completely accurate prediction for large time values, however, thanks to the calculation of the Hurst exponent it is possible to know the nature of the time series and thus reduce the degree of uncertainty of the prediction. The implementation of the theory of modern physics to classical engineering problems provide better solutions to the problems of chaos and information theory.
References

[1] Ferenti T 2017 Biomedical applications of time series analysis IEEE 30th Neumann Colloquium (Budapest: IEEE)

[2] Zhang Y, Tan X, Xi H, Zhao X 2008 Real-time risk management based on time series analysis 7th World Congress on Intelligent Control and Automation (China: IEEE)

[3] Vélez A, González H, Guerra B 2014 Attractor reconstruction for plethysmographic biosignals International Conference on Electronics, Communications and Computers (México: IEEE)

[4] Lyle J, Charlton P, Bonet E, Chaffey G, Christie M, Nandi M, Aston P 2017 Beyond HRV: Analysis of ECG signals using attractor reconstruction Computing in Cardiology (CinC) (Francia: IEEE)

[5] Kong L, Yang C, Wang Y, Gui W 2009 Generic phase space reconstruction method of multivariate time series Chinese Control and Decision Conference (China: IEEE)

[6] Canals M, Olivares R, Labra F, Caputo L, Rivera A, Novoa F 1998 Caracterización de la geometría fractal del árbol bronquial en mamíferos Revista Chilena de Anatomía 16(2) 237

[7] Alperovich Y, Alperovich M, Spiro A 2017 Trends modeling and its impact on hurst exponent at stock market fractal analysis Tenth International Conference Management of Large-Scale System Development (MLSD) (Moscow: IEEE)

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

[9] Mandelbrot B 1963 The variation of certain speculative prices Journal of Business 36(4) 394

[10] So M K P 2000 Long-term memory in stock market volatility Applied Financial Economics 10(5) 519

[11] Richards G 2000 The fractal structure of exchange rates: measurement and forecasting Journal of International Financial Market, Institutions and Money 10(2) 163

[12] Di Matteo T, Aste T, Dacorogna M 2005 Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development Journal of Banking & Finance 29(4) 827