Fractional Brownian motion analysis for epidemic spreading of diseases

Leonardo Lima (✉ lslima7@yahoo.com.br)

Federal Center for Technological Education of Minas Gerais

Research Article

Keywords: SARS-CoV-2 virus, COVID-19, cases of coronavirus, beginning of pandemic, SIR dynamics

Posted Date: August 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-812062/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Fractional Brownian motion analysis for epidemic spreading of diseases

L. S. Lima

Federal Center for Technological Education of Minas Gerais, 30510-000, Belo Horizonte, MG, Brazil.

(Dated: August 18, 2021)

In this paper, we use the nonlinear fractional stochastic differential equation approach with Hurst parameter H into the range $1/2 < H < 1$ to perform the statistical analysis of public data of number of infected by coronavirus in countries where the number of cases is larger as Brazil and India. We analyze the rises and falls of novel daily cases of coronavirus, where the fluctuation in the official data is treated as a random term in the stochastic differential equation for the fractional Brownian motion. The projection of novel cases in the future is treated as quadratic mean deviation in the official data of novel cases daily since the beginning of the pandemic up the present. Additionally, statistical tests as skewness and kurtosis are performed with aim to estimate the shape of the probability density distribution of novel cases in long time, since the beginning of pandemic, in tentative to estimate the behavior of novel cases in the future.

I. INTRODUCTION

The SARS-CoV-2 virus (COVID-19) emerged at end of 2019 and since then, one whether become a global crisis. The very quick spreading of the disease made the study of its spreading of great importance for its control and prevention.[1–6] On the other hand, the mathematical modeling for spreading of diseases has been performed since the works of Daniel Bernoulli in XVII century for spreading of the small pox. Later, the Bernoulli model has been modified with aim that unbounded growth does not occur and randomness effects are included in the model. At first, was introduced the factor that had the effect of making the variation rate of population growing $\mathcal{N}$, negative, $d\mathcal{N}/dt$. In this case, the model is given by $d\mathcal{N}/dt = -\alpha\mathcal{N}(1 - \nu\mathcal{N})(1 - \eta\mathcal{N})$, where the parameters α and η are at range $\alpha > 0$ and $0 < \eta < \nu$. Furthermore, taking some randomness in these parameters, we obtain a more realistic model to treat real situations. [7–11] In a general way, if the size of the population at time t is $\mathcal{N}(t)$, being the relative rate of growth of the population at time t , $\mathcal{N}(t)$, we have $\mathcal{N}_{eff}(t)$ subjects to some randomness due to the environmental effects, so that $\mathcal{N}_{eff}(t) = \mathcal{N}(t) + \xi(t)$, where the first term $\mathcal{N}(t)$ is deterministic while the second term $\xi(t)$ is random, reflecting the environmental randomness effect. The aim here is to model the novel cases of coronavirus using the nonlinear stochastic differential equation approach with fractional white noise, with Hurst parameter at range $1/2 < H < 1$ to analyze the rises and falls of novel cases daily of COVID-19 reported by the public healthy agencies of countries as Brazil, where the low number tests performed in the population leads to a great uncertainty in the official data, due to possibility that the novel cases dynamics might exhibit long-range correlations and thus may to be described in terms of long-memory processes, such as the fractional Brownian motion.[12] Furthermore, the use of standard stochastic differential equations have already been used in the literature for the study of the behavior of time series for price dynamics in Refs. [13–16]. In addition, the use of standard Brownian motion that corresponds to the case of Hurst index $H = 1/2$ has

already been used recently for the study of spreading of COVID-19 at Ref. [3]. Furthermore, the SIR dynamics expressed using ordinary Itô-stochastic differential equation has been proposed in Refs. [17–23]. Here, we aim to use the equation of the fractional Brownian motion with parameter H at range $H \in (1/2, 1)$ to study the behavior of the curve of novel cases daily. A statistical analysis described by the master equation and transition rates for the infection process was made in Ref.[24] Additionally, the statistical inference in the epidemic model for the Ebola virus was performed in Ref. [25]. Moreover, the discrete-time epidemic model with binomial distribution has been used to study the transmission rate of COVID-19 in Ref. [26]. A statistical model for the health care impact of COVID-19 in India has been studied in [27].

The plan of this paper is the following. In section II, we describe the model. In section III, we present the numerical results by numerical solution of the stochastic differential equation with fractional white noise for different values of H . In section IV, we present our conclusions and final remarks.

II. FRACTIONAL WHITE NOISE ANALYSIS FOR SPREADING OF DISEASE

The fractional Brownian motion (BMF) is defined as a family of Gaussian processes being introduced by Kolmogorov in Ref. [28] and indexed by the Hurst parameter H into the range $H \in (0, 1)$. Being the first application made by Hurst in 1951 to model the long term storage capacity of reservoirs along the Nile River.[29, 30] The fractional Brownian motion was defined in Refs. [31–33] where for each $H \in (0, 1)$, a real-valued Gaussian process $B_H(t)$, $t \leq 0$ is defined such that the expectation $E[B_H(t)] = 0$ and $E[B_H(t)B_H(s)] = \frac{1}{2} [t^{2H} + s^{2H} - (t-s)^{2H}]$ for all $s, t \in \mathbb{R}_+$. If $H = 1/2$, the fractional Brownian motion is the standard Brownian motion or Wiener process. We consider H restricted to range $1/2 < H < 1$ and apply this equation as well as its modifier for the infected numbers by coronavirus $N(t)$ in each day. Hence, we use the nonlinear fractional

stochastic equation with drift term in the form of a logistic model with a randomly term to perform the analysis of spreading of novel cases of coronavirus $N(t)$, given by

$$dN(t) = \mathfrak{h}_0 + \mathfrak{g}(t)dt - \alpha N(t)(1 - \nu N(t))(1 - \eta N(t))dt + \beta(t) \diamond dB_H(t). \quad (1)$$

Furthermore, if $t \rightarrow B_H(t)$ is differentiable into the Schwartz space $S(\mathbb{R})$ of rapidly decreasing smooth functions, being $(S_H)^*_H$ is the dual of $(S)_H$

$$\frac{dB_H(t)}{dt} = \mathcal{W}_H(t), \quad \text{in } (S_H)^*, \quad (2)$$

where $\mathcal{W}(t)_H$ is a fractional white noise defined by

$$\mathcal{W}_H = \sum_{i=1}^{\infty} \left[\int_{\mathbb{R}} e_i(v) \phi(t, v) dv \right] \mathcal{H}_{\varepsilon(i)}(\omega) \quad (3)$$

where $\{e_n\}_{n=1}^{\infty}$ is the orthonormal basis of $L^2_{\phi}(\mathbb{R})$ and

$$\mathcal{H}_{\varepsilon(i)}(\omega) = \int_{\mathbb{R}} e_i(t) dB_H(t). \quad (4)$$

The B_H -integration theory of Eq. (1) is based on using ordinary products path-wise and leads to an integral that for integrands $g(t, \omega)$ is defined by

$$\int_a^b g(t, \omega) \delta B_H(t) = \lim_{|\Delta| \rightarrow 0} \sum_{k=0}^{n-1} g(t_k, \omega) [B_H(t_{k+1}) - B_H(t_k)], \quad (5)$$

where $\Delta: a = t_0 < t_1 < \dots < t_n = b$ is a partition of the interval $[a, b]$ and $|\Delta| = \max_{0 \leq k \leq n-1} (t_{k+1} - t_k)$. These integrals do not have an expectation zero and are called fractional path-wise integrals. Here, we considerate the B_H -integral considered in Ref. [34], defined by

$$\int_a^b g(t, \omega) dB_H(t) = \lim_{|\Delta| \rightarrow 0} \sum_{k=0}^{n-1} g(t_k, \omega) \diamond [B_H(t_{k+1}) - B_H(t_k)], \quad (6)$$

where $\diamond$ denotes the Wick product.[31–34] These integrals use the Wick products and are known as the fractional Itô integrals with reference to corresponding situation for standard Brownian motion.[32, 34] We consider the white noise calculus based on $B_H(t)$, $H \in (1/2, 1)$. Due to the behavior of novel cases displayed in Fig. 1, we consider the deterministic part adjusting to the daily novel cases using a polynomial of n degree (suave curve). The aim is simply provide a smoothed version of the time series using standard methods as the lowest regression. The aim here is to calculate the ensemble of possible trajectories using the fractional Brownian motion as is made in Fig. 3. Moreover, the assumption of a constant rate $\mathfrak{h}_0$ is reasonable when $N(t)$ is large but becomes less reasonable when $N(t)$ is small. In Fig. 1, we display different least square fits to the set of official data. We obtain a better adjust using a higher order polynomial like $N(t) = 15460 -$

FIG. 1. Adjusts to the real data through suave curve (polynomial of n degree) to the set of novel cases $N(t)$ registered in Brazil in the period from 14 Mar (2020) up to 15 Jul (2021). The data were supplied by Brazilian ministry of health. We use different polynomials of n degree ($n = 3, 4, 5, 6$) for the adjustment to the data.

$1359t + 34t^2 - 0.3t^3 + 0.001t^4 - 2.0 \times 10^{-6}t^5 + 1.0 \times 10^{-9}t^6$. The variability in the time series for e.g. Brazil is mostly due to case reports not being filed on weekends, leading to a large-amplitude oscillation with a period of 7 days. A rolling average, or simply a bin window of greater or less than 1 day may give very different results. Hence, the aim here is to simply provide a smoothed version of the time series where standard methods can be used (e.g., lowest regression).

The dynamics of novel cases daily $N(t)$, of infected number by coronavirus registered in the Brazil as function of time (days) registered in the period from 14 Mar (2020) up to 11 Jul (2021) is displayed in Fig. 3. The zigzag behavior into the range of large t values reflects in an increase of the uncertainty in the data and low-number of test performed in the population. For modeling of this behavior, we add a random noise term in the logistic model for spreading of infectious number as given by Eq. (1) in order to simulate the effect of rises and falls of novel cases. However, we find that for values of Hurst index far from $H = 1/2$ the zigzag behavior exhibited in the model becomes distant from viewed for the real cases, showing that the Hurst index of the time series must stay around $H = 1/2$. The α , ν and η parameters comes from the logistic model with threshold for spreading of the smallpox with the value found given by $\alpha = \nu = 0.13$. The range of η values is into $0 < \eta < \nu$. $\mathfrak{g}(t)$ comes from the adjustment of least squares to the data set within the period considered. The analysis performed using Brazil data can be applied to other coun-

FIG. 2. Dynamics of novel cases $N(t)$ in India. We use different least squares fits for the time series (polynomials of $n = 7, 8, 9, 10$ degree are plotted as a guide to the eyes). The range of data considered is from 14 Mar (2020) up to 21 Jul (2021).

tries as well, since $\mathbf{g}(t)$ is obtained by least squares fit to the data of each country with an uncertain that is small in the beginning and becomes large for long times. In Fig. 2, we display the curve of novel cases $N(t)$ in India. The period considered is up to 21 Jul (2021). The data in India present a stronger stochastic behavior into middle range displayed in the figure, where in the beginning of the pandemic the oscillation was small. The least square fit gives a suave curve to the data with a polynomial of t given by $N(t) = -28442 + 1545t + 16t^2 - 2t^3 + 0.04t^4 - 0.0004t^5 + 2.0 \times 10^{-6}t^6 - 7.0 \times 10^{-9}t^7 + 1.0 \times 10^{-11}t^8 - 2.0 \times 10^{-14}t^9 + 7.0 \times 10^{-18}t^{10}$. Thus, the effect of the randomness term is to span the fluctuations of the data as displayed in Fig. 3. These daily fluctuations with a weekly cycle of reported cases have been observed in many countries and are due to the diagnostic and data reporting practices [37]. We obtain the time series of the model Eq. (1) using the values found by Bernoulli given by $\alpha = \nu = 0.13$, $\eta = 0.063$ and fractional white noise amplitude $\Gamma = 1.0$. Furthermore, we use a low intensity of fractional white noise β as $\beta = 1.0 \times 10^{-6}$.

III. RESULTS

We write the BMF model that is a Gaussian process $W_H(t)$, $t > 0$ with zero mean and stationary increments and variance given by $E[W_H^2(t)] = t^{2H}$. The sample path of the BMF model is a fractal curve with fractal dimension $d = 1/H$, where $H = 1/2$ corresponds to the standard Brownian motion whose increments $\Delta W_H(t) = W_H(t+1) - W_H(t)$ are statistically independent. For $H \neq 1/2$ the increments $\Delta W_H(t)$ are

FIG. 3. Dynamics of novel cases $N(t)$ in Brazil. The zigzag behavior of the results is reflected by the stochastic term in Eq. (1). We plot the time series of the model Eq. (1) for different values of Hurst parameter $H = 0.6$, $H = 0.8$. The values for the parameters α, ν used are $\alpha = \nu = 0.13$, $\eta = 0.063$ and the noise intensity is $\beta(t) = 1.0 \times 10^{-6}t^3$.

known as fractional white noise, showing long-range correlation as $E[\Delta W_H(t+h)\Delta W_H(t)] \simeq 2H(2H-1)h^{2H-2}$, for $h \rightarrow \infty$. For $H \in (1/2, 1)$ the increments are positively correlated exhibiting thus a persistent behavior. For $H \in (0, 1/2)$, the increments are negatively correlated, showing anti-persistence. For $H = 1/2$, we write the Wiener increment as $dW_H(t) \sim (\sqrt{dt}R_G)^{2H}$, where R_G is an aleatory generator number with Gaussian distribution of mean zero and variance $\sigma_{W_H}^2 = 1$. In Fig. 3, we plot the time series of the model Eq. (1) for values of noise intensity $\beta(t) = 1.0 \times 10^{-6}t^3$. We have the time series of the variation of novel cases oscillating quickly as displayed in the figure. We plot the time series of the model for values of Hurst parameter $H > 1/2$: $H = 0.6$ and $H = 0.8$. The results for $H = 1/2$ of the Ref. [3] are overlapped with the results for $H > 1/2$ and hence are not displayed here. The low intensity of the value of fractional white noise used in the calculations is such as $\beta(t) = 1.0 \times 10^{-6}t^3$.

In Fig. 4, we plot the half-width of the probability density, $P(N, t)$ as function of time t , $\sigma(t)$. We calculate the variance of the distribution where the standard deviation gives an estimating of novel cases in each day t . The difference to the real data is due to the approach used. We calculate the n^{th} order moments $\mu_n = \langle (x - m_1)^n \rangle$ until the fourth order about the mean or central moments, where we have the following relations: $c_1 = \mu_1$, $c_2 = \mu_2$, $c_3 = \mu_3$, $c_4 = \mu_4 - 3\mu_2^2$. Normalized measures commonly used indicating a deviation from the Gaussian are the kurtosis λ_4 , defined as $\lambda_4 = \mu_4/\sigma^4 - 3$ and the skewness, λ_3 . In Fig. 5, we display the behavior of $\lambda_4(t)$. $\lambda_4(t)$ is numerically calculated by solving Eq. (1)). Moreover, the

FIG. 4. Behavior of half-width of the distribution as a function of t , $\sigma(t)$. The half-width gives an expectation of novel cases in each day t .

kurtosis relates to the deviation of the tail of the distribution as compared to Gaussian $P(N) = \frac{1}{\sqrt{4\pi\beta t}} e^{-N^2/4\beta t}$, whose solution would correspond to Eq. (1) for $\nu = \eta = 0$ and $g(t) \equiv 0$. The range of negative values obtained for the kurtosis indicates that the shape of the distribution is nearest to a Wigner's distribution[10]. Furthermore, at range of small t considered, where the kurtosis is near to zero, we have that the distribution is nearest of a Gaussian ($\lambda_4 = 0$).

A. Detrended fluctuation analysis

There are several estimators for the Hurst exponent, such as rescaled range (R/S) analysis and detrended fluctuation analysis (DFA) that differ basically by the choice of the fluctuation measure. The DFA method is a technique used to analyse time series that appear as long-memory processes. The exponent is like the Hurst exponent however, the DFA method might be applied to signals whose dynamics is non-stationary. Considering a bounded time series $x(t)$ of length N , where $t \in \mathbb{N}$, the summation in first converts in an unbounded process: $x(t) = \sum_{i=1}^t (x_i - \langle x \rangle)$, $i = 1 \dots, N$. Where x_i compose a series of size N in which $\langle x \rangle$ represents its mean. The cumulative sum, x_t is divided in windows of length n where in each segment ν is calculated the function

$$F(\nu, n) = \left\{ \frac{1}{n} \sum_{t=1}^N [x(t) - y_\nu(t)]^2 \right\}^{\frac{1}{2}}, \quad (7)$$

where t varies from beginning to end of each segment. $y_\nu(t)$ is the correspondent value of $x(t)$ calculated in each polynomial adjusted to each segment ν . Thus we obtain

the fluctuation of second order given by [38]

$$F_2(n) = \left\{ \frac{1}{N} \sum_{\nu=1}^N [F(\nu, n)]^2 \right\}^{\frac{1}{2}} \quad (8)$$

This process of detrending followed by fluctuation is repeated over different n values and a log-log graphic $F(n) \propto n$ is made. The value into the range $H \in (0.5, 1)$ indicates a non-stationary and unbounded behavior of the time series. The rescaled range method (RS) is an older and more standard method. In this case, the Hurst exponent is directly obtained by the RS method, where whether $H \in (0.5, 1.0)$ indicates a time series with long-term switching between high and low values in adjacent pairs. For the calculation of the Hurst exponent by DFA method, we must be aware of the performance of different estimation methods that have been systematically tested. This tell about the accuracy of the method. When analysing real data, we must confront many complicating factors where the behavior of several factors has been analyzed for ideal time series and real-world data, such as strong trends, non-stationarity, nonlinearity, and Hurst exponent being larger than 1. There are many efforts to improve the estimators making them more suitable for real data. However, it is well known that over all estimators, the DFA method is one of best performance being less sensitive than other methods by the choice of the scaling range.[39]

IV. FINAL REMARKS

In summary, we analyzed the spreading of novel cases $N(t)$ of coronavirus using the nonlinear stochastic differential equation with a fractional white noise (nonlinear fractional Brownian motion approach) to model the rises and falls of novel cases of coronavirus in countries as Brazil. The variability in the time series for instance, in Brazil is mostly due to the case reports not being filed on weekends, leading to a large-amplitude oscillation with a period of 7 days. Furthermore, the uncertainty in the results generated by the low number of tests performed in the populations and to underreporting generates an large uncertainty in the official results and consequently, becomes necessary the addition of randomness in the differential equations for growing of infected number. We find a large deviation from the real data for different values of Hurst indexes $H > 1/2$ (persistent case) and thus out of the value $H = 1/2$ (Brownian case), indicating that the approach of Brownian motion ($H = 1/2$) must give a better description to the behavior of the real data.

Acknowledgment This work was partially supported by National Council for Scientific and Technological Development (CNPq), Brazil.

REFERENCES

FIG. 5. Behavior of the kurtosis as a function of t , $\lambda_4(t)$. The range of negative values gives an estimating of the shape of distribution which becomes nearest of a Gaussian for $\lambda_4 = 0$ at range of large t values since the firsts cases reported.

-
- [1] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, *Nature Medicine* 26, 855 (2020).
 - [2] M. Gatto, E. Bertuzzob, L. Maria, S. Miccolid, L. Carraro, R. Casagrandia, A. Rinaldog, *PNAS* 117, 10484 (2020).
 - [3] Leonardo S. Lima, *Scientific Reports* 11, 12177 (2021).
 - [4] D. Fanelli, F. Piazza, *Chaos Solitons Fractals* 134, 109761 (2020).
 - [5] S. Boccaletti, W. Ditto, G. Mindin, A. Antangana, *Chaos Solitons Fractals* 135, 109794 (2020).
 - [6] F. Ndairou, I. Area, J. J. Nieto, D. F. M. Torres, *Chaos Solitons Fractals* 135, 109846 (2020).
 - [7] Steven E. Shreve, *Stochastic Calculus for Finance II Continuous-Time Models*, (Springer, New York, 2004).
 - [8] C. Gardiner, *Stochastic Methods, A Handbook for the Natural and Social Sciences*, (Fourth edition, Springer, New Zealand, (2009)
 - [9] B. Øksendal, *Stochastic Differential Equations An Introduction with Applications*, sixth edition, Springer-Verlag, Oslo (2013).
 - [10] P. Richmond, J. Mimkes, S. Hutzler, *Econophysics and Physical Economics*, Oxford, UK (2013).
 - [11] Kurt Jacobs, *Stochastic Processes for Physicists*, Cambridge University Press, New York (2013).
 - [12] G. L. Vasconcelos, *Braz. J. Phys.* 34, 1039 (2004).
 - [13] L. S. Lima, S. C. Oliveira, *Chaos, Solitons & Fractals* 136, 109792 (2020).
 - [14] L. S. Lima, L. L. B. Miranda, *Physica A* 490, 828 (2018).
 - [15] L. S. Lima, G. K. C. Santos, *Physica A* 512, 222 (2018).
 - [16] L. S. Lima, S. C. Oliveira, A. F. Abeilice, J. H. C. Melgaço, *Physica A* 526, 120932 (2019).
 - [17] A. Simha, R. V. Prasad, S. Narayana, arXiv:2003.11920v3 [q-bio.PE] (2020).
 - [18] M. Mahrouf, A. Boukhouima, H. Zine, E. M. Lotfi, D. F. M. Torres, N. Yousf, *Axioms* 10, 18 (2021).
 - [19] Z. El Kharrazi, S. Saoud, 7th International Conference on Optimization and Applications, ICOA (2021).
 - [20] A. M. Kareem, S. N. Al-Azzawi, *J. Phys. Conf. Ser.* 1818, 012121 (2021).
 - [21] Z. El kharrazi and S. Saoud, 7th International Conference on Optimization and Applications (ICOA) 1, (2021).
 - [22] Zizhen Zhang, Anwar Zeb, Sultan Hussain, Ebraheem Alzahrani, *Adv. Differ. Equ.* 2020(1): 451 (2020).
 - [23] A. Tesfay, T. Saeed, A. Zeb, D. Tesfay, A. Khalaf, J. Brannan, *Adv. Differ. Equ.* 2021(1):228 (2021).
 - [24] T. Tomé, M. J. de Oliveira, *Braz. J. Phys.* 50, 832 (2020).
 - [25] P. Lekone, B. Finkensta, *Biometrics* 62, 1170 (2006).
 - [26] Sha He, Sanyi Tang, Libin Rong, *Math. Biosciences and Engineering* 17, 2792 (2020).
 - [27] K. Chatterjee, K. Chatterjee, Arun Kumar, S. Shankar, *Med. J. Armed Forces India* 76, 147 (2020).
 - [28] A. N. Kolmogorov, *Acad. URSS (N.S.)* 26, 115 (1940).
 - [29] H. E. Hurst, *Trans. Amer. Soc. Civil Eng.* 116, 400 (1951).
 - [30] H. E. Hurst, *Proc. Inst. Civil Engineers, Part I*, Chapter 5, 519 (1956).
 - [31] V. Ness, *SIAM Review* 10, 422 (1968).
 - [32] Y. Hu, B. Øksendal, *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 6, 1 (2003).
 - [33] T. E. Duncan, Y. Z. Hu, B. Pasik-Duncan, *Proceedings of the IEEE Conference on Decision and Control*, (2014).
 - [34] Y. Hu, B. Øksendal and A. Sulem, *Stochastic Analysis and Related Topics*, Vol. 5, eds. H. Körezlioglu, B. Øksendal, A. S. Üstünel, Birkhäuser, 203 (1996).
 - [35] T. E. Duncan, Y. Hu, B. Pasik-Duncan, *J. Control Optim* 38, 582 (2000).
 - [36] J.-P. Bouchaud, R. Cont, *Eur. Phys. J. B* 6, 543 (1998).

- [37] A. Bergman, Y. Sella, P. Agre, A. Casadevall, *Clinical Science and Epidemiology* 5, e00544-20 (2020).
- [38] Gao-Feng Gu, Wei-Xing Zhou, *Phys. Rev. E* 82, 011136 (2010).
- [39] Ying-Hui Shao, Gao-Feng Gu, Zhi-Qiang Jiang, Wei-Xing Zhou and Didier Sornette, *Scientific Reports* 2, 835 (2012).