A Letters Journal Exploring the Frontiers of Physics

May 2009

EPL, **86** (2009) 30001 www.epljournal.org
doi: 10.1209/0295-5075/86/30001

The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion

L. Lacasa[1], B. Luque[1], J. Luque[2] and J. C. Nuño[3]

[1] *Departamento de Matemática Aplicada y Estadística, ETSI Aeronáuticos, Universidad Politécnica de Madrid Madrid, Spain, EU*
[2] *Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya - Barcelona, Spain, EU*
[3] *Departamento de Matemática Aplicada a los Recursos Naturales, ETSI Montes, Universidad Politécnica de Madrid Madrid, Spain, EU*

received 7 January 2009; accepted in final form 14 April 2009
published online 12 May 2009

PACS 05.45.Tp – Time series analysis
PACS 05.40.Jc – Brownian motion
PACS 89.75.Hc – Networks and genealogical trees

Abstract – Fractional Brownian motion (fBm) has been used as a theoretical framework to study real-time series appearing in diverse scientific fields. Because of its intrinsic nonstationarity and long-range dependence, its characterization *via* the Hurst parameter, H, requires sophisticated techniques that often yield ambiguous results. In this work we show that fBm series map into a scale-free *visibility graph* whose degree distribution is a function of H. Concretely, it is shown that the exponent of the power law degree distribution depends linearly on H. This also applies to fractional Gaussian noises (fGn) and generic $f^{-\beta}$ noises. Taking advantage of these facts, we propose a brand new methodology to quantify long-range dependence in these series. Its reliability is confirmed with extensive numerical simulations and analytical developments. Finally, we illustrate this method quantifying the persistent behavior of human gait dynamics.

Self-similar processes such as fractional Brownian motion (fBm) [1] are currently used to model fractal phenomena of different nature, ranging from physics or biology to economics or engineering. To cite a few, fBm has been used in models of electronic delocalization [2], as a theoretical framework to analyze turbulence data [3], to describe geologic properties [4], to quantify correlations in DNA base sequences [5], to characterize physiological signals such as human heartbeat [6] or gait dynamics [7], to model economic data [8], or to describe network traffic [9–11]. Fractional Brownian motion $B_H(t)$ is a nonstationary random process with stationary self-similar increments (fractional Gaussian noise, fGn) that can be characterized by the so-called Hurst exponent, $0 < H < 1$. The one-step memory Brownian motion is obtained for $H = \frac{1}{2}$, whereas time series with $H > \frac{1}{2}$ shows persistence and antipersistence if $H < \frac{1}{2}$.

While different fBm generators and estimators have been introduced in the last years, the community lacks consensus on which method is best suited for each case. This drawback comes from the fact that fBm formalism is exact in the infinite limit, *i.e.*, when the whole infinite series of data is considered. However, in practice, real-time series are finite. Accordingly, long-range correlations are partially broken into finite series, and local dynamics corresponding to a particular temporal window are overestimated. The practical simulation and the estimation from real (finite) time series is consequently a major issue that is, hitherto, still open. An overview of different methodologies and comparisons can be found in [11–18] and references therein. Most of the preceding methods operate either on the time domain (*e.g.*, aggregate variance method, Higuchi's method, detrended fluctuation analysis, and range scaled analysis) or in the frequency or wavelet domain (periodogram method, Whittle estimator, wavelet method). In this letter we introduce an alternative and radically different method, the visibility algorithm, based on graph theoretical techniques. In a recent paper this new tool for analyzing time series has been presented [19]. In short, a visibility graph is obtained from the mapping of a time series into a network according to the following visibility criterion: Two arbitrary data (t_a, y_a) and (t_b, y_b) in the time series have visibility, and consequently become two connected

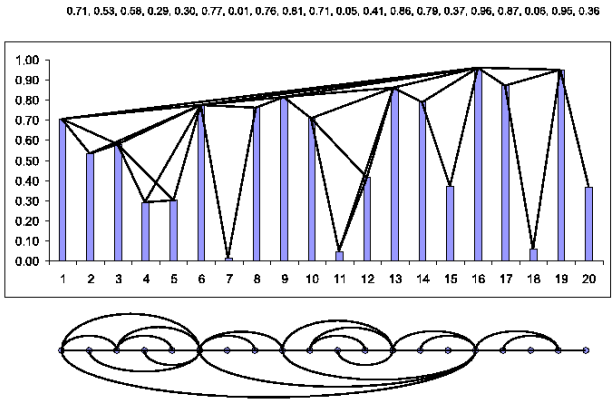

Fig. 1: (Color online) Example of a time series (20 data, depicted in the upper part) and the associated graph derived from the visibility algorithm. In the graph, every node corresponds, in the same order, to a series data. The visibility rays between the data define the links connecting nodes in the graph.

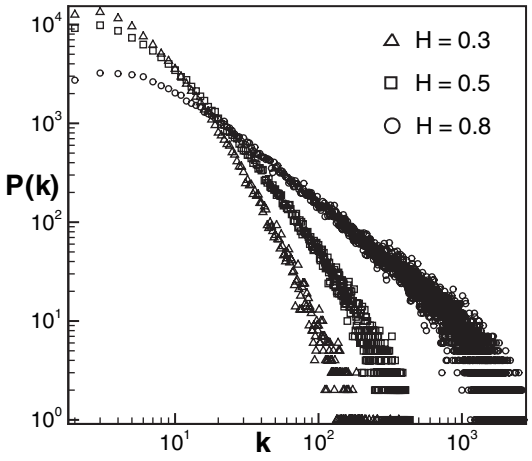

Fig. 2: Degree distribution of three visibility graphs, namely i) triangles: extracted from an fBm series of 10^5 data with $H = 0.3$, ii) squares: extracted from an fBm series of 10^5 data with $H = 0.5$, and iii) circles: extracted from an fBm series of 10^5 data with $H = 0.8$. Note that distributions are not normalized. The three visibility graphs are scale-free since their degree distributions follow a power law $P(k) \sim k^{-\gamma}$ with decreasing exponents $\gamma_{0.3} > \gamma_{0.5} > \gamma_{0.8}$.

nodes in the associated graph, if any other data (t_c, y_c) such that $t_a < t_c < t_b$ fulfills

$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a}. \qquad (1)$$

In fig. 1, we have represented for illustrative purposes an example of how a given time series maps into a visibility graph by means of the visibility algorithm. A preliminary analysis has shown that series structure is inherited in the visibility graph [19]. Accordingly, periodic series map into regular graphs, random series into random graphs, and fractal series into scale-free graphs [20]. In particular, it was shown that the visibility graph obtained from the well-known Brownian motion has got both the scale-free and the small world properties [19]. Here we show that the visibility graphs derived from generic fBm series are also scale-free, that is to say, their degree distribution follows a power law $P(k) \sim k^{-\gamma}$, where k stands for the degree of a given node. This robustness goes further, and we prove that a linear relation between the exponent γ of the power law degree distribution in the visibility graph and the Hurst exponent H of the associated fBm series exists. Therefore, the visibility algorithm provides an alternative method to compute the Hurst exponent and then, to characterize fBm processes. This also applies to fractional Gaussian noise (fGn) [1] which is nothing but the increments of an fBm, and generic $f^{-\beta}$ noises, enhancing the visibility graph as a method to detect long-range dependence in time series.

In fig. 2 we have depicted in log-log the degree distribution of the visibility graph associated with three artificial fBm series of 10^5 data, namely an antipersistent series with $H = 0.3$ (triangles), a memoryless Brownian motion with $H = 0.5$ (squares), and a persistent fBm with $H = 0.8$ (circles). As can be seen, these distributions follow a power law $P(k) \sim k^{-\gamma}$ with decreasing exponents $\gamma_{0.3} > \gamma_{0.5} > \gamma_{0.8}$.

In order to compare γ and H appropriately, we have calculated the exponent of different scale-free visibility graphs associated with fBm artificial series of 10^4 data with $0 < H < 1$ generated by a wavelet-based algorithm [21]. Note at this point that some bias is inevitably present since artificial series generators are obviously not exact, and consequently the nominal Hurst exponents have an associated error [22]. For each value of the Hurst parameter we have thus averaged the results over 10 realizations of the fBm process. It is well known that to estimate the exponent γ of a (finite size) power law distribution is a tricky task, and that a least-squares fit in a log-log plot sometimes provides misleading estimation [23]. Accordingly, we have employed for this task a maximum-likelihood estimation (MLE) [23] such that

$$\gamma = 1 + n\left[\sum_{i=1}^{n} \log \frac{k_i}{k_{min}}\right]^{-1}, \qquad (2)$$

where n is the total number of values taken into account, $k_i, i = 1, \ldots, n$, are the measured values and k_{min} corresponds to the smallest value of k (the smallest degree in our case) for which the power law behavior holds (typically, $k_{min} = 10$). In fig. 3, we have represented the relation between γ and H (black circles). As can be seen, a roughly linear relation holds (the dotted line represents the best linear fitting $\gamma = 3.1 - 2H$).

That fBm yields scale-free visibility graphs is not that surprising. The most highly connected nodes (hubs) are responsible for the heavy-tailed degree distributions. Within fBm series, hubs are related to extreme values

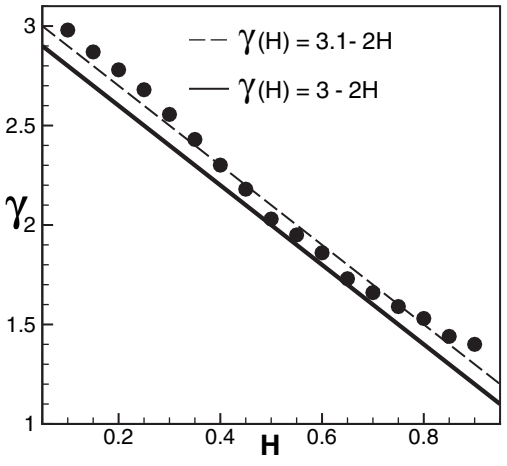

Fig. 3: (Black dots) Numerical estimation of exponent γ of the visibility graph associated with an fBm series with exponent H. In each case, γ is averaged over 10 realizations of an fBm series of 10^4 data, in order to avoid nonstationary biases (the error bars are included in the dot size). The dotted line corresponds to the best linear fitting $\gamma(H) = a - bH$, where $a = 3.1 \pm 0.1$ and $b = 2.0 \pm 0.1$, and the solid line corresponds to the theoretical prediction $\gamma(H) = 3 - 2H$. Both results are consistent. Note that deviations from the theoretical law take place for values of $H > 0.5$ and $H < 0.5$ (strongly correlated or anticorrelated series), where fBm generators evidence finite-size accuracy problems [22], these being more acute the more we move away from the noncorrelated case $H = 0.5$.

in the series, since a data with a very large value has typically a large connectivity, according to eq. (1). In order to calculate the tail of the distribution, we consequently need to focus on the hubs, and thus calculate the probability that an extreme value has a degree k. Suppose that at time t the series reaches an extreme value (a hub) $B_H(t) = h$. The probability of this hub to have degree T is

$$P(T) \sim P_{fr}(T) r(T), \tag{3}$$

where $P_{fr}(T)$ provides the probability that after T time steps the series returns to the same extreme value, *i.e.*, $B(t+T) = h$ (and consequently the visibility in t gets truncated in $t+T$), in other words, the probability that an arbitrary node sees *at least* k other nodes (T time steps in the series correspond to T nodes in the graph). $P_{fr}(T)$ is nothing but the first return time distribution, which has been proved to scale as $P_{fr}(T) \sim T^{H-2}$ for fBm series [24]. Now, $r(T)$ is the percentage of nodes between t and $t+T$ that t may see. The percentage of visible nodes between two extreme values is related to the roughness of the series in that basin, that is, to the way that a series of T time steps folds: this percentage is larger when the series roughness is low, and conversely, it decreases when the series roughness increases. As a matter of fact, the series roughness is encoded in its standard deviation [1], which goes like t^H for fBm. Normalizing, we have $r(T) \sim T^H/T = T^{H-1}$ (this fact has been confirmed numerically). Finally, notice that in this context $T \equiv k$, so eq. (3) converts into

$$P(k) \sim k^{H-2} k^{H-1} = k^{2H-3}, \tag{4}$$

which provides a linear relation between the exponent of the visibility graph degree distribution and the Hurst exponent of the associated fBm series:

$$\gamma(H) = 3 - 2H, \tag{5}$$

in good agreement with our previous numerical results. Note in fig. 3 that numerical results obtained from artificial series deviate from the theoretical prediction for strongly-correlated ones ($H > 0.5$ or $H < 0.5$). This deviation is related to finite-size effects in the generation of finite fBm series [22], and these effects are more acute the more we deviate from the noncorrelated case $H = 0.5$. In any case, a scatter plot of the theoretical (eq. (5)) *vs.* the empirical estimation of $\gamma(H)$ provides statistical conformance with a correlation coefficient $c = 0.99$.

To further check the consistency of the visibility algorithm, an estimation of the power spectra is performed. It is well known that fBm has a power spectrum that behaves as $1/f^\beta$, where the exponent β is related to the Hurst exponent of an fBm process through the well-known relation [25]

$$\beta(H) = 1 + 2H. \tag{6}$$

Now according to eqs. (5) and (6), the degree distribution of the visibility graph corresponding to a time series with $f^{-\beta}$ noise should be again power law $P(k) \sim k^{-\gamma}$, where

$$\gamma(\beta) = 4 - \beta. \tag{7}$$

In fig. 4, we depict (triangles) the empirical values of γ corresponding to $f^{-\beta}$ artificial series of 10^6 data with β ranging from 1.2 to 2.8 in steps of size 0.1 [26]. For each value of β, we have again averaged the results over 10 realizations and estimated β through MLE (eq. (2) with $k_{min} = 10$). The straight line corresponds to the theoretical prediction eq. (7), showing good agreement with the numerics. In this case, a scatter plot confronting theoretical *vs.* empirical estimation of $\gamma(\beta)$ also provides statistical conformance between them, up to $c = 0.99$.

Finally, observe that eq. (6) holds for fBm processes, while for the increments of an fBm process, known as a fractional Gaussian noise (fGn), the relation between β and H turns to be [25]

$$\beta(H) = -1 + 2H, \tag{8}$$

where H is the Hurst exponent of the associated fBm process. We consequently can deduce that the relation between γ and H for an fGn (where fGn is a series composed by the increments of an fBm) is

$$\gamma(H) = 5 - 2H. \tag{9}$$

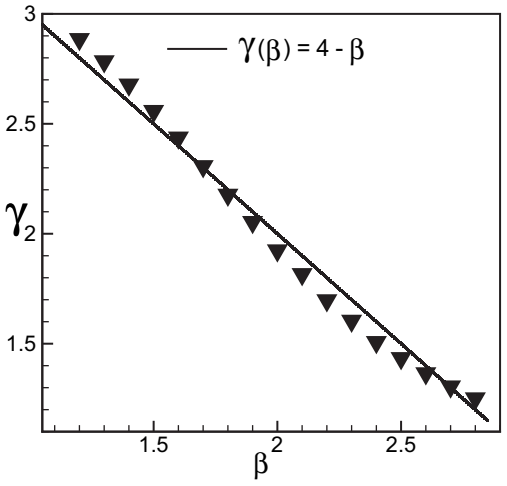

Fig. 4: (Black triangles) Numerical estimation of exponent γ of the visibility graph associated with an $f^{-\beta}$ noise. In each case, γ is averaged over 10 realizations of an $f^{-\beta}$ series of 10^6 data, in order to avoid nonstationary biases (the error bars are included in the triangle size). The straight line corresponds to the theoretical prediction in eq. (7).

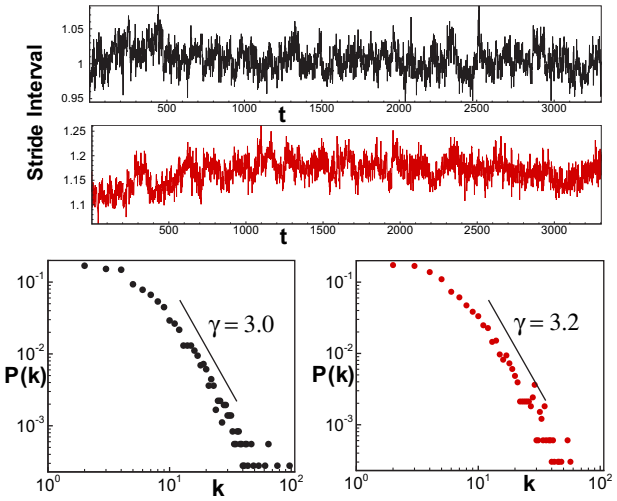

Fig. 5: (Color online) Black signal: Time series of 3595 points from the stride interval of a healthy person in fast pace. Red signal: Time series of 3304 points from the stride interval of a healthy person in slow pace. Bottom: Degree distribution of the associated visibility graphs (the plot is in log-log). These are power laws where $\gamma = 3.03 \pm 0.05$ for the fast movement (black dots) and $\gamma = 3.19 \pm 0.05$ for the slow movement (red dots), which provides $\beta = 1$ and $\beta = 0.8$ for fast and slow pace, respectively, according to eq. (7), in agreement with previous results [27,28].

Notice that eq. (9) can also be deduced applying the same heuristic arguments as for eq. (5) with the change $H \to H - 1$.

In order to illustrate this latter case, we finally address a realistic and striking dynamics where long-range dependence has been recently described. Gait cycle (the stride interval in human walking rhythm) is a physiological signal that has been shown to display fractal dynamics and long-range correlations in healthy young adults [27,28]. In the upper part of fig. 5, we have plotted two series describing the fluctuations of walk rhythm of a young healthy person, for slow pace (bottom series of 3304 points) and fast pace (up series of 3595 points), respectively (data available in `www.physionet.org/physiobank/database/umwdb/` [29]). In the bottom part we have represented the degree distribution of their visibility graphs. These ones are again power laws with exponents $\gamma = 3.03 \pm 0.05$ for fast pace and $\gamma = 3.19 \pm 0.05$ for slow pace, derived through MLE with $k_{min} = 10$ (observe that for rather small series, the specific determination of k_{min} may have a relatively important effect in the maximum-likelihood estimation of γ, with typical variations of 5%, and consequently one has to take specific care when determining it [30]). According to eq. (7), the visibility algorithm predicts that gait dynamics evidence $f^{-\beta}$ behavior with $\beta = 1$ for fast pace, and $\beta = 0.8$ for slow pace, in perfect agreement with previous results based on a Detrended Fluctuation Analysis [27,28]. These series record the fluctuations of walk rhythm (*i.e.*, the increments), so according to eq. (9), the Hurst exponent is $H = 1$ for fast pace and $H = 0.9$ for slow pace, that is to say, dynamics evidences long-range dependence (persistence) [27,28].

As a summary, the visibility graph is an algorithm that maps a time series into a graph. In so doing, classic methods of complex network analysis can be applied to characterize time series from a brand new viewpoint [19]. In this work, we have pointed out how graph theory techniques can provide an alternative method to quantify long-range dependence and fractality in time series. We have reported analytical and numerical evidences showing that the visibility graph associated to a generic fractal series with Hurst exponent H is a scale-free graph, whose degree distribution follows a power law $P(k) \sim k^{-\gamma}$ such that: i) there is a universal relation between γ and the exponent β of its power spectrum that reads $\gamma = 4 - \beta$; ii) for fBm signals (where H is defined such that $\beta(H) = 1 + 2H$), the relation between γ and H reads $H(\gamma) = \frac{3-\gamma}{2}$ while for fGn signals (the increments of an fBm where H is defined as $\beta(H) = -1 + 2H$), we have $H(\gamma) = \frac{5-\gamma}{2}$.

The reliability of this methodology has been confirmed with extensive simulations of artificial fractal series and real (small) series concerning gait dynamics. To our knowledge, this is the first method for estimation of long-range dependence in time series based on graph theoretical techniques advanced so far, with potentially broad applications. Some open questions for further research include: i) a methodic comparison of the accuracy and computational efficiency of the visibility algorithm with standard methods, ii) the generalization of the visibility algorithm to handle with fractal signals with superposed nonlinear trends (polynomial, sinusoidal, *etc.*),

iii) its robustness when the signal is polluted with noise, and iv) a generalization to handle with multifractal structures.

* * *

The authors thank O. Miramontes and F. Ballesteros for helpful suggestions and two anonymous referees for their interesting comments. This work was partially supported by the Spanish Ministry of Science Grant FIS2006-08607.

REFERENCES

[1] Mandelbrot B. B. and Van Ness J. W., *SIAM Rev.*, **10** (1968) 422.
[2] de Moura F. A. B. F. and Lyra M. L., *Phys. Rev. Lett.*, **81** (1998) 17.
[3] Muzy J. F., Bacry E. and Arneodo A., *Phys. Rev. Lett.*, **67** (1991) 25; Kiyani K. *et al.*, *Phys. Rev. Lett.*, **98** (2007) 2111101.
[4] Golombek M. P. *et al.*, *Nature*, **436** (2005) 44.
[5] Peng C.-K. *et al.*, *Nature*, **356** (1992).
[6] Ivanov P. Ch. *et al.*, *Nature*, **399** (1999) 461.
[7] Haussdorf J., *Hum. Mov. Rev.*, **26** (2007) 555.
[8] Matos J. A. O. *et al.*, *Physica A*, **387** 15 (2008) 3910.
[9] Leland W. E. *et al.*, *IEEE/ACM Trans. Netw.*, **2** (1994) 1.
[10] Mikosch T. *et al.*, *Ann. Appl. Probab.*, **12** (2002) 2368.
[11] Karagiannis T., Molle M. and Faloutsos M., *IEEE Internet Comput.*, **8** (2004) 57.
[12] Weron R., *Physica A*, **312** (2002) 285.
[13] Pilgram B. and Kaplan D. T., *Physica D*, **114** (1998) 108.
[14] Kantelhardt J. W., *Fractal and multifractal time series*, to be published in *Springer Encyclopaedia of Complexity and System Science* (2008) preprint arXiv: 0804.0747.
[15] Podobnik B. and Stanley H. E., *Phys. Rev. Lett.*, **100** (2008) 084102.
[16] Carbone A., *Phys. Rev. E*, **76** (2007) 056703.
[17] Mielniczuk J. and Wojdyllo P., *Comput. Stat. Data Anal.*, **51** (2007) 4510.
[18] Simonsen I., Hansen A. and Nes O. M., *Phys. Rev. E*, **58** (1998) 3.
[19] Lacasa L., Luque B., Ballesteros F., Luque J. and Nuño J. C., *Proc. Natl. Acad. Sci. U.S.A.*, **105** (2008) 4972.
[20] Albert R. and Barabasi A. L., *Rev. Mod. Phys.*, **74** (2002) 47.
[21] Abry P. and Sellan F., *Appl. Comput. Harmon. Anal.*, **3** (1996) 377.
[22] Horn G. A. *et al.*, *Perform. Eval.*, **64** (2007) 162.
[23] Newmann M. E. J., *Contemp. Phys.*, **46** (2005) 323.
[24] Ding M. and Yang W., *Phys. Rev. E*, **52** (1995) 1.
[25] Adison P. S., *Fractal and Chaos: An Illustrative Course* (IOP Publishing Ltd.) 1997.
[26] The series have been generated by a method where each frequency component has a magnitude generated from a Gaussian white process and scaled by the appropriate power of the frequency. The phase is uniformly distributed on $[0, 2\pi]$. See `http://local.wasp.uwa.edu.au/pbourke/fractals/noise/` for a source code.
[27] Goldenberger A. R. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 2466.
[28] Hausdorff J. M. *et al.*, *J. Appl. Physiol.*, **80** (1996) 1448.
[29] Goldberger A. L. *et al.*, *Circulation*, **101** (2000) 215.
[30] Clauset A., Shalizi C. R. and Newman M. E. J., to be published in *SIAM Rev.* (2009) arXiv:0706.1062.