



# Hurst exponent estimation of self-affine time series using quantile graphs



Andriana S.L.O. Campanharo<sup>a</sup>, Fernando M. Ramos<sup>b,\*</sup>

<sup>a</sup> Departamento de Bioestatística, Instituto de Biociências, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil

<sup>b</sup> Laboratório de Computação e Matemática Aplicada, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil

## HIGHLIGHTS

- The quantile graph (QG) method for the estimation of the Hurst exponent of self-affine time series is presented.
- The QG method is applied to the characterization of different fractional Brownian motions.
- Comparison between  $H$  estimates using the QG method and the exact values used to generate the motions shows an excellent agreement.
- For a given time series length,  $H$  estimation error depends basically on the statistical framework used for determining the exponent of a power-law model.
- The QG method is numerically simple and has only one free parameter, the number of quantile/nodes; with a simple modification, it can be extended to the analysis of fractional Gaussian noises.

## ARTICLE INFO

### Article history:

Received 26 May 2015

Received in revised form 23 September 2015

Available online 22 October 2015

### Keywords:

Self-affine time series

Hurst exponent

Complex networks

Quantile graphs

## ABSTRACT

In the context of dynamical systems, time series analysis is frequently used to identify the underlying nature of a phenomenon of interest from a sequence of observations. For signals with a self-affine structure, like fractional Brownian motions (fBm), the Hurst exponent  $H$  is one of the key parameters. Here, the use of quantile graphs (QGs) for the estimation of  $H$  is proposed. A QG is generated by mapping the quantiles of a time series into nodes of a graph.  $H$  is then computed directly as the power-law scaling exponent of the mean jump length performed by a random walker on the QG, for different time differences between the time series data points. The QG method for estimating the Hurst exponent was applied to fBm with different  $H$  values. Comparison with the exact  $H$  values used to generate the motions showed an excellent agreement. For a given time series length, estimation error depends basically on the statistical framework used for determining the exponent of the power-law model. The QG method is numerically simple and has only one free parameter,  $Q$ , the number of quantiles/nodes. With a simple modification, it can be extended to the analysis of fractional Gaussian noises.

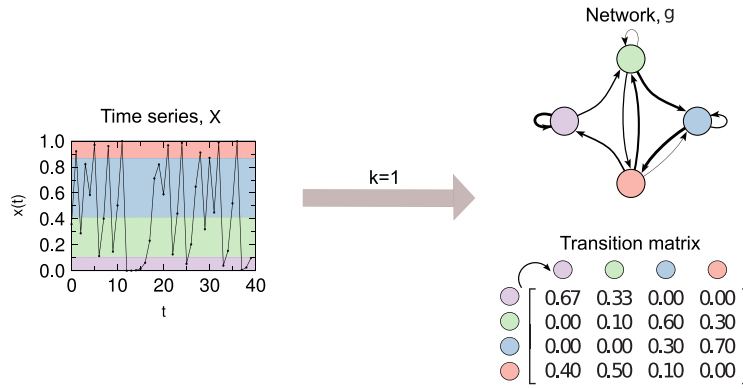
© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

In the last two decades, research on complex networks became the focus of widespread attention, with developments and applications spanning different scientific areas, from sociology and biology to physics. One of the reasons behind the growing

\* Corresponding author.

E-mail addresses: [andriana@ibb.unesp.br](mailto:andriana@ibb.unesp.br) (A.S.L.O. Campanharo), [fernando.amos@inpe.br](mailto:fernando.amos@inpe.br) (F.M. Ramos).



**Fig. 1.** Illustration of the QG method for  $k = 1$ . A time series  $X$  is split into  $Q = 4$  quantiles (colored shading) and each quantile  $q_i$  is assigned to a node  $n_i \in \mathcal{N}$  in the corresponding network  $g$ . Two nodes  $n_i$  and  $n_j$  are then connected in the network with a weighted arc  $(n_i, n_j, w_{ij}^k) \in \mathcal{A}$  where the weight  $w_{ij}^k$  of the arc is given by the probability that a point in quantile  $q_i$  is followed by a point in quantile  $q_j$ . Repeated transitions between quantiles results in arcs in the network with larger weights (represented by thicker lines) and therefore higher values in the corresponding transition matrix. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

popularity of complex networks is that almost any discrete structure can be suitably represented as a graph, whose features may be then characterized, analyzed and, eventually, related to its respective dynamics [1]. Recently several approaches have been proposed for mapping a time series into a complex network representation, based on concepts such as correlations [2,3], visibility [4,5], recurrence analysis [6], transition probabilities [7–10] and phase-space reconstructions [11–13]. These studies have shown that distinct features of a time series can be mapped onto networks with distinct topological properties, opening the door to the analysis of discrete, time-ordered data sets with mathematical tools usually used in the study of geometric shapes and topological spaces.

Fractional Gaussian noises (fGn) and fractional Brownian motions (fBm) have been used as a theoretical framework to investigate time series emerging in different scientific areas. A fBm is a non-stationary self-affine process with stationary increments given by a fGn [14]. Self-affine random processes are characterized by a power-spectrum with a power-law dependency on frequency as  $P(\omega) \sim \omega^{-\beta}$ , with  $-1 \leq \beta \leq 1$  for noises and  $1 < \beta \leq 3$  for motions. For  $\beta = 0$ , we have a white uncorrelated noise [15]. Noises and motions are also characterized by their Hurst exponents ( $H$ ) [16]. The classical Brownian motion is a fBm with  $H = 1/2$ . Correlation between fBm's increments is negative for  $0 < H < 1/2$  and positive for  $1/2 < H < 1$  [17]. Because of the intrinsic non-stationarity and long range dependence of fBm, the estimation of  $H$  often requires more robust methods than those provided by standard Fourier analysis [18].

In spite of the large number of applications of complex networks methods in the study time series, usually involving the classification of dynamical systems or the identification of dynamical transitions (see Ref. [13] and references therein), establishing a link between a network measure and  $H$  remains an open question. Recently, a linear relationship between the exponent of the power law degree distribution of visibility graphs and  $H$  has been established for noises and motions [19, 20]. Here, we propose an alternative approach for the computation of the Hurst exponent. This new approach is based on a generalization of the method introduced in Ref. [10], in which time series quantiles are mapped into nodes of a graph (here called quantile graph or QG) and vice-versa. To this end  $H$  is computed directly as power-law scaling exponent of the mean jump length performed by a random walker on a QG, for different time differences between the time series data points.  $H$  estimates computed with this method are robust, with a standard error that depends basically on the statistical framework used for fitting a power-law model to the random walk data (for a maximum likelihood estimator, see, for example, Ref. [21]).

This paper is organized as follows. After this Introduction, we describe in Section 2 the QG method for computing  $H$ . Results are presented and discussed in Section 3 while an overall conclusion is given in Section 4.

## 2. Methods

Let the range of values in a time series be coarse-grained into  $Q$  quantiles  $q_1, \dots, q_Q$ , and let  $\mathcal{M}$  be a map from a time series  $X \in \mathcal{T}$  to a network  $g \in \mathcal{G}$ , with  $X = \{x(t) | t \in \mathbb{N}, x(t) \in \mathbb{R}\}$  and  $g = \{\mathcal{N}, \mathcal{A}\}$  being a set of nodes  $\mathcal{N}$  and arcs  $\mathcal{A}$ . Specifically,  $\mathcal{M}$  assigns each quantile  $q_i$  to a node  $n_i \in \mathcal{N}$  in the corresponding network. Two nodes  $n_i$  and  $n_j$  are connected with a weighted arc  $n_i, n_j, w_{ij}^k \in \mathcal{A}$  whenever two values  $x(t)$  and  $x(t+k)$  belong respectively to quantiles  $q_i$  and  $q_j$ , with  $t = 1, 2, \dots, T$  and the time differences  $k = 1, \dots, k_{\max} < T$ . For an illustration of the QG method for  $k = 1$ , see Fig. 1.

Weights  $w_{ij}^k$  are simply given by the number of times a value in quantile  $q_i$  at time  $t$  is followed by a point in quantile  $q_j$  at time  $t+k$ , normalized by the total number of transitions. Repeated transitions through the same arc increase the value of the corresponding weight. With proper normalization, the weighted adjacency matrix becomes a Markov transition matrix  $\mathbf{W}_k$ , with  $\sum_j^Q w_{ij}^k = 1$ . The resulting network is weighted, directed and connected, with  $Q$  being typically much smaller than  $T$ . Note that by randomly moving from one node to the other with probability given by  $\mathbf{W}_k$ , and by assigning the corresponding quantile values to  $x(t)$ , it is possible to reconstruct  $X$  from  $g$ .

With  $\mathbf{W}_k$  in hand, it is possible to perform a random walk on the quantile graph (QG)  $g$ . After  $l = L$  jumps of length  $\delta_{l,k}(i, j) = |i - j|$ , with  $i, j = 1, \dots, Q$  being the node indices as defined by  $\mathbf{W}_k$ , the mean jump length  $\Delta(k)$  can be computed as follows:

$$\Delta(k) = \frac{1}{L} \sum_{l=1}^L \delta_{l,k}(i, j). \quad (1)$$

For a white noise, since all jump lengths, from  $Q - 1$  to  $0$ , are equally probable,  $\Delta_{\beta=0}(k)$  is trivially equal to  $\frac{Q-1}{2}$  for all values of  $k$ . For motions, the mean jump length shows a power-law scaling with  $k$  as  $\Delta(k) \sim k^H$ . In other words, a log-log plot of  $\Delta(k)$  versus  $k$  can be used to estimate Hurst exponent of a fBm ( $1 < \beta < 3$ ). For noises ( $-1 < \beta < 1$ ), the same approach for estimating  $H$  can be applied, provided the elements of the signal are first cumulatively summed into a fBm (see, for example, Ref. [22]).

The numerical procedure for estimating  $H$  using the QG method is simple and straightforward to be implemented. For a fixed time series length  $T$ , the computational cost grows linearly with the number of quantiles/nodes  $Q$ .

### 3. Results

The motions and noises used to illustrate the applicability of QG method have been generated with the wavelet-based synthesis method [23], for different values of  $H$ , with  $\beta = 2H - 1$  for noises, and  $\beta = 2H + 1$  for motions. Unless when indicated otherwise, the values  $Q = 50$  and  $T = 10\,000$  have been used in the computations. Results have been averaged over 100 realizations.

Fig. 2 displays QG transition matrices  $\mathbf{W}_k$ , for  $k = 1$  and  $\beta$  values ranging from  $-1$  to  $2$ . Observe how well the QG topology mimics the properties of the original time series. White noise, for example, is mapped into a random graph, whose weights are uniformly distributed along the columns and rows of  $\mathbf{W}_k$ . As we move from anti-correlated to correlated time-series,  $\mathbf{W}_k$  changes from an anti-diagonally to a diagonally dominant matrix. For the  $\beta = 2$ , standard Brownian motion is mapped into a network in which neighboring nodes are highly correlated.

A self-affine random process such as a fBm is characterized by its Hurst exponent, with  $0 < H < 1$ . In Fig. 3, log-log plots of  $\Delta(k)$  versus  $k$  are presented, for different values of  $\beta$ . For each  $\beta$ , note the more than two-decades of power-law scaling, which can be used to compute estimates of the Hurst exponent  $\hat{H}_{QG}$ . Note that extending the scaling region requires increasing the value of  $k_{\max}$  and, thus, the length  $T$  of the time series.

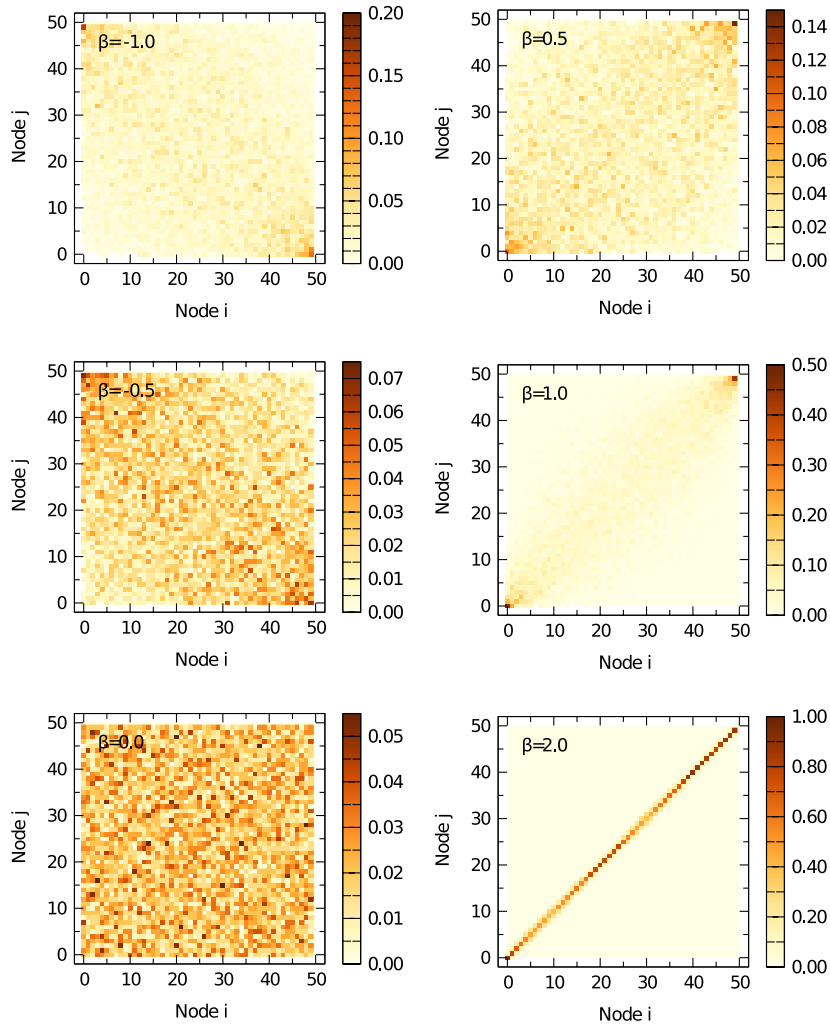
A comparison between  $\hat{H}_{QG}$  and  $H$ , the exact value used to generate the time series, is shown in Fig. 4. The linear scaling regions in Fig. 3 were used to estimate the Hurst exponent  $\hat{H}_{QG}$ , by means of a linear regression [24,25] of log-transformed coordinates; error bars represent 95% confidence intervals. Observe the strong linear correlation between  $\hat{H}_{QG}$  and  $H$ .

The number of quantiles  $Q$  defines the partitioning level of the amplitude range of the original time series. In this sense, it plays a similar role as the alphabet size in symbolic analysis [26]. Its selection involves a trade-off between information loss and computational burden. For the simplest, binary case,  $Q = 2$  and the time series is mapped into a two-node network, which captures only the coarsest patterns of the original signal. Higher values of  $Q$  are able to discriminate more details, but require longer time series for the transition probabilities to properly converge (since more nodes implies less jumps between two given nodes) and, thus, are more demanding in terms of computational effort to compute  $H$ . In the limit, when  $Q$  equals the number of distinct values in the time series, there is no information loss, provided the number of data points is large enough for the statistics to converge. In practice, much smaller values of  $Q$  can be used to estimate  $H$ . To illustrate this point, we plot in Fig. 5,  $|\hat{H}_{QG} - H|$  versus  $\beta$ , for  $Q = 50$  and  $100$ . Note that the absolute difference between QG estimates and exact values of  $H$  is smaller than  $0.02$  for all  $\beta$ , and weakly dependent on the choice of  $Q$ , the only free parameter of the QG method.

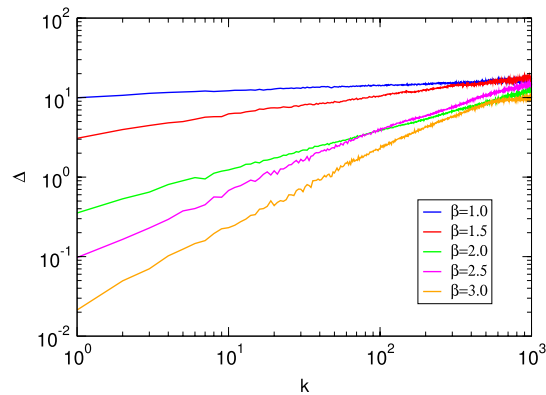
### 4. Conclusions

Many natural signals present a fractal-like structure and are characterized by two parameters,  $\beta$ , the power-spectrum power-law exponent, and  $H$ , the Hurst exponent. Here the QG method for estimating  $H$  was presented and applied to fractional Brownian motions with different  $\beta$  values. Results were compared to the exact  $H$  values used to generate the motions and showed agreement. For a given time series length, estimation error depends basically on the statistical framework used for determining the exponent of a power-law model. The QG method is numerically simple and has only one free parameter,  $Q$ , the number of quantiles/nodes. With a simple modification, it can be extended to the analysis of fractional Gaussian noises.

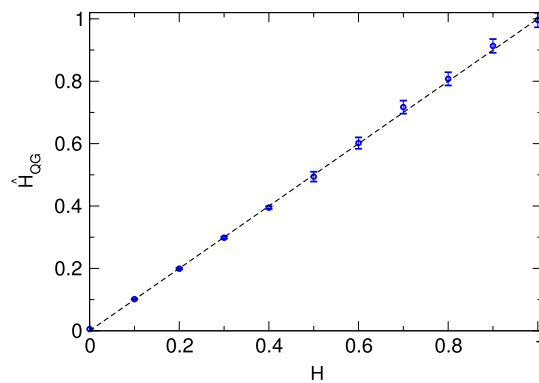
Summarizing, the use of quantile graphs (QGs) for estimating the Hurst exponent of fractional motions and noises is proposed. The QG method permits to quantify features such as long-range correlations or anti-correlations associated with the signal's underlying dynamics, expanding the traditional tools of time series analysis in a new and useful way.



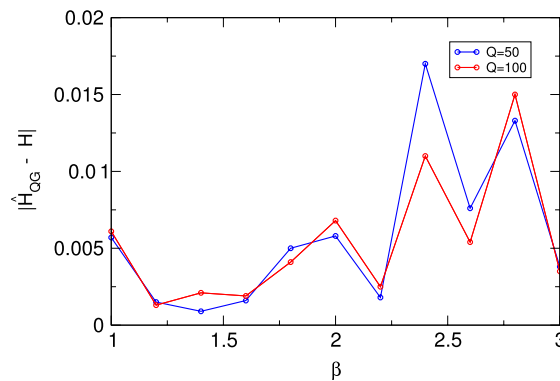
**Fig. 2.** QG transition matrices for different values of  $\beta$  and  $k = 1$ . We use  $T = 10,000$  time points of each self-affine time series and construct QG's using  $Q = 50$  quantiles by applying the map  $\mathcal{M}_{QT}$ . The resulting transition matrices display clear differences in topologies. When  $\beta = 0$ , the corresponding time series is a white noise and the weights in the corresponding weighted adjacency matrix are uniformly distributed. As  $\beta$  increases (decreases), the probability of a given node be connected to its immediate neighbor also increases (decreases) and the weights in the corresponding adjacency matrix are mainly distributed over the secondary (primary) diagonal.



**Fig. 3.** A log–log plot of the mean jump length  $\Delta(k)$  versus  $k$  for different  $\beta$  values, obtained by performing 100 random walks in the corresponding QG's. As  $\beta$  increases the corresponding time series become more correlated, the distance between any given pair of nodes  $n_i$  and  $n_j$  increases and therefore  $\Delta$  also increases. Moreover,  $\Delta(k)$  versus  $k$  are characterized by more than two-decades of power-law scaling for all  $\beta$  values.



**Fig. 4.** Comparison between  $\hat{H}_{QG}$  and corresponding exact values for the Hurst exponent. The linear scaling regions for all  $\beta$  values in Fig. 3 were used to estimate the Hurst exponent  $\hat{H}_{QG}$ , using linear regression [24,25] of log-transformed coordinates; error bars represent 95% confidence intervals. Note the strong linear correlation between  $\hat{H}_{QG}$  and  $H$ . These estimations were performed using linear regression.



**Fig. 5.** Absolute difference between QG estimates and  $H$  exact values as a function of  $\beta$ , for  $Q = 50$  and  $100$ . The small value of  $|\hat{H}_{QG} - H|$  attests the robustness of the results showed here, regardless of the value of  $Q$ .

## Acknowledgments

A.S.L.O.C. acknowledges the support of FAPESP: 2014/05145-0 and 2013/19905-3. The support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Brazil) 501221/2012-3 and 303437/2012-0 is acknowledged by both authors. All figures were generated with PyGrace (<http://pygrace.github.io/>) with color schemes from Colorbrewer (<http://colorbrewer.org>).

## References

- [1] L.D.F. Costa, F.A. Rodrigues, G. Travieso, P.R.V. Boas, Characterization of complex networks: A survey of measurements, *Adv. Phys.* 56 (2007).
- [2] J. Zhang, M. Small, Complex network from pseudoperiodic time series: Topology versus dynamics, *Phys. Rev. Lett.* 96 (2006).
- [3] Y. Yang, H.J. Yang, Complex network-based time series analysis, *Physica A* 387 (2008).
- [4] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuno, From time series to complex networks: The visibility graph, *Proc. Natl. Acad. Sci. USA* 105 (2008).
- [5] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: Exact results for random time series, *Phys. Rev. E* 80 (2009).
- [6] N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series, *Phys. Lett. A* 46 (2009).
- [7] G. Nicolis, A.G. Cantú, C. Nicolis, Dynamical aspects of interaction networks, *Internat. J. Bifur. Chaos* 15 (2005).
- [8] P. Li, B.-H. Wang, Extracting hidden fluctuation patterns of Hang Seng stock index from network topologies, *Physica A* 378 (2007).
- [9] A.H. Shirazi, G.R. Jafari, J. Davoudi, J. Peinke, M.R.R. Tabar, M. Sahimi, Mapping stochastic processes onto complex networks, *J. Stat. Mech. Theory Exp.* (2009).
- [10] A.S.L.O. Campanharo, M.I. Sirer, R.D. Malmgren, F.M. Ramos, L.A. Amaral, Duality between time series and networks, *PLoS ONE* 6 (8) (2011).
- [11] X. Xu, J. Zhang, M. Small, Superfamily phenomena and motifs of networks induced from time series, *Proc. Natl. Acad. Sci. USA* 105 (2008).
- [12] Z. Gao, N. Jin, Complex network from time series based on phase space reconstruction, *Chaos* 19 (2009).
- [13] R. Donner, M. Small, J. Donges, N. Marwan, Y. Zou, R. Xiang, J. Kurths, Recurrence-based time series analysis by means of complex network methods, *Internat. J. Bifur. Chaos* 21 (4) (2011).
- [14] J.W. Kantelhardt, Fractal and multifractal time series, in: R.A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, 2009, pp. 3754–3779.
- [15] A. Bunde, *Fractals in science*, in: A. Bunde, S. Havlin (Eds.), *Fractals in Science*, Springer, 1994.
- [16] H.E. Hurst, Long-term storage capacity of reservoirs, *Trans. Amer. Soc. Civ. Eng.* 116 (1951).
- [17] G. Rangarajan, M. Ding, Integrated approach to the assessment of long range correlation in time series data, *Phys. Rev. E* 61 (2000).
- [18] C.-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Long-range anticorrelations and non-Gaussian behavior of the heartbeat, *Phys. Rev. Lett.* 70 (1993).

- [19] X.-H. Ni, Z.-Q. Jiang, W.-X. Zhou, Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks, *Phys. Lett. A* 373 (2009).
- [20] L. Lacasa, B. Luque, J.C. Nuno, The Visibility Graph: a new method for estimating the Hurst exponent of fractional Brownian motion, *EPL, Europhys. Lett.* 86 (2009).
- [21] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, *SIAM Rev.* 51 (4) (2009).
- [22] A. Eke, P. Hermán, J. Basingthwaite, G. Raymond, D. Percival, M. Cannon, I. Ikrényi, Physiological time series: distinguishing fractal noises from motions, *European J. Physiol.* 439 (2000).
- [23] P. Abry, F. Sellan, The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation, *Appl. Comput. Harmon. Anal.* 3 (4) (1996) 377–383.
- [24] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*, 2001. URL: <http://www.scipy.org/> [Online; accessed 21.09.15].
- [25] N. Draper, H. Smith, *Applied Regression Analysis*, A Wiley-Interscience Publication, Wiley, New York, NY, 1998.
- [26] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, *PLoS ONE* 74 (2) (2003).