ON A METHOD FOR INVESTIGATING LONG-RANGE CORRELATIONS

Karin Vibe-Rheymer, Jens Timmer†, Jean-Marc Vesin

Signal Processing Laboratory, Swiss Federal Institute of Technology
LTS/DE, EPFL, 1015-Lausanne, Switzerland
e-mail: vibe@ltssg3.epfl.ch
†Zentrum für Datenanalyse und Modellbildung, Albert-Ludwig-Universität
Albertstr. 26-28, 79104 Freiburg im Breisgau, Germany

ABSTRACT

Fractal signals have attracted a lot of attention in various fields lately, and numerous algorithms have been designed to analyze them. Most of them investigate long-term correlations, hence requiring long data sets (i.e. data sets extending to very large time scales). This requirement is however rarely met in practice, which can cast doubts on the reliability of the results. This work tries to partially fill this void by analyzing a method examining long-term correlations for short time series. It is shown the conclusions obtained for long data sets remain valid, but there are some particular cases that should be taken into account before concluding on the fractality of a signal. A practical example, namely heart rate recordings, is taken to illustrate some possible pitfalls that can be encountered when real-world short data sets are to be studied.

1. INTRODUCTION

Fractal signals have attracted a lot of attention in the recent past, since many real-world signals, in particular biomedical signals, seem to have this property [1, 2, 3]. Different methods have been developed to assess the fractal nature of a signal, such as the Hurst exponent [4] or, more recently, the Allan and Fano factors [5].

These algorithms look for long-range correlations, and thus take data sets long enough to get rid of the short-range correlations. Investigation on the behavior of these algorithms has however not been done extensively for data exhibiting only short-range correlations, and it is of interest to have a better understanding of their behavior in this context.

The test that will be studied here is a simplified version of Hurst's method. It was proposed by Buldyrev *et al.* (see [6] and refs therein) for the analysis of biomedical data, namely DNA sequences, but has proved to be useful in other fields of biomedical engineering, such as heart beat or membrane channel openings recordings. The more classical Hurst analysis could also be used, but the former method basically yields the same results, and is faster. Its behavior is studied in the presence of short data sets, and an application to heart rate signals is presented.

2. ROOT MEAN SQUARE FLUCTUATION FUNCTION

The mean square fluctuation function [7] examines the correlations across scales, by showing the differences between samples separated by a time lag l with respect to this lag. If the signal is fractal, its evolution obeys a power law whose exponent is similar to the Hurst one.

In practice, the procedure consists in first computing the difference function $d(l)$

$$d(l,n) = x(n+l) - x(n) \tag{1}$$

where $x(n)$ is the time series of interest. One then computes the mean-square fluctuation function $F(l)$ itself by

$$F^2(l) = \overline{d^2(l,n)} - \overline{d(l,n)}^2 \tag{2}$$

The horizontal bar stands for the average with respect to n, so that this averaging yields the mean square difference for a given lag.

In the case of a fractal signal, the mean square fluctuation function follows a law of the form

$$F(l) \sim Ca^{\alpha} \tag{3}$$

where a is a given parameter, and α is the exponent of the power law. In general, it can take one of the following forms:

(1) $F(l) \sim l^{\alpha}$. This type of behavior implies that there are correlations on all scales, or self-similarity. $\alpha = 0.5$ is a critical value: It is characteristic of

Brownian motion, a specific scale-invariant process that is independent, contrary to classical fractals (see [8], chapter 9, for a thorough explanation on Brownian motion, fractional Brownian motion and their difference). When $\alpha > 0.5$, the process is persistent: An increasing (decreasing) trend in the past implies an increasing (decreasing) trend in the future, whatever the time scale. On the contrary, if $\alpha < 0.5$, the process is anti-persistent, i.e. an increasing trend in the past implies a decreasing trend in the future.

(2) $F(l) \sim 1 - e^{-l/R}$. In this case, there are correlations across scales, but extending only up to a range R. The asymptotic behavior is thus unchanged from the purely random case (i.e. white noise).

(3) When neither of the preceding forms is encountered, further study is necessary. We will focus on this case for the remainder of this work, since the two former cases are already well-known.

These properties can be seen in a straightforward way on a log-log plot of $F(l)$. If the signal is uncorrelated, the curve is a straight horizontal line. If the signal has some sort of self-similarity, one has a straight line with a slope different from $\alpha = 0$. If this straight line tends to saturate, it usually means that there are correlations only up to a given range.

3. ANALYSIS OF THE ALGORITHM

The problem of the above analysis is that it is valid only for infinite time series, on which all scales can be observed. In practice, the results are the same for finite data sets, provided their length is sufficient for a large number of scales to be observable, i.e. the difference function can be computed up to large lags. Here again, the problem is that such long time series are often not available, and it is by no means assured that the theoretical results will still be valid in practical situations.

Performing an analytical study of the results for short data sets is not possible in general. It is thus necessary to resort to simulations. The algorithm was thus applied to a wide variety of signals: Synthetic time series, such as fractional Brownian motion, simple Brownian motion, white noise, autoregressive (AR) processes, chaotic signals (among which the Rössler system or the logistic map); noise was superimposed on some of these signals, in order to study the reactions of the algorithm in what is the most usual real-world case. All time series were limited to 1000 samples.

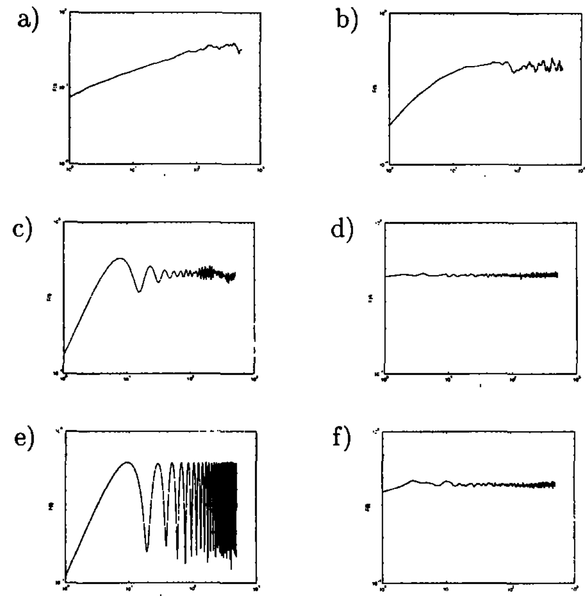

Figure 1: Examples of the behavior of the mean-square fluctuation function for different signals. a) fractional Brownian motion, b) AR(1) process, c) AR(2) process, d) random white noise, e) the Rössler system and f) the Hénon map. Theses plots cover all the observed behaviors.

Figure 1 shows the reaction of the mean-square fluctuation function to several kinds of signals. The behavior of Fig. 1.a) is well predicted by theory: a straight line remains indicative of a fractal, whatever the length of the data. The deviation of the curve close to its end is not surprising either: for larger lags, the number of differences on which the computation of $F(l)$ can be done is reduced, which lowers its accuracy.

Figure 1.d) also corresponds to what was expected. Any signal having a flat spectrum indeed exhibits a flat mean-square fluctuation function whatever the lag and the number of points. The increases in the deviations widths around the trend in Fig. 1.b) and 1.d) also result from the decreasing number of points available to compute the average fluctuation at larger scales.

Figures 1.b) and 1.c) are less trivial to interpret. Figure 1.b) is rather similar to Fig. 1.a), except that the overall shape of the figure is convex. This behavior is characteristic of stochastic signals exhibiting some sort of structure, such as ARMA processes, with the exclusion of fractals. Fig. 1.c) is essentially similar to the preceding one, except for the ripples. Experiments have shown that these ripples are characteristic of periodicities in the signal. In particular, they can be seen in strictly periodic signals, but also in all AR(n) processes (provided $n > 1$), as well as in nonlinear systems.

Figures 1.e) and 1.f) show the results for two chaotic systems. As expected, none of them exhibit a straight line. Moreover, the shape of $F(l)$ depends on the particular process at hand. For instance, the important ripples that are seen on Fig. 1.e) come from the periodicities of the Rössler system.

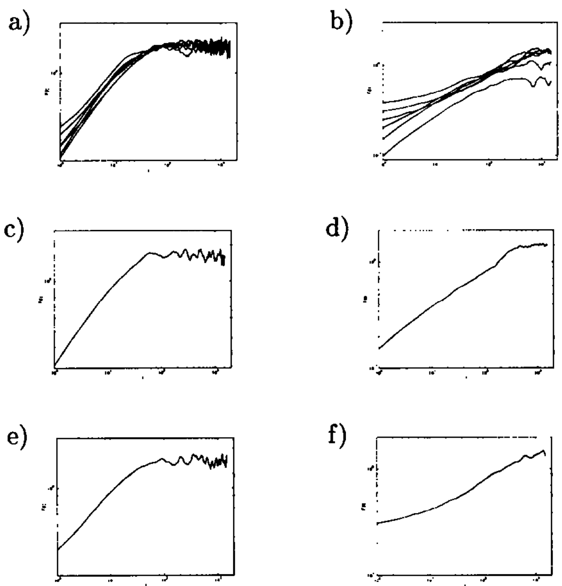

Figure 2: Effect of noise on AR(1) process (a, c, e) and fractional Brownian motion (b, d, f). a) and b) $F(l)$ for different amounts of noise (1%, 5%, 10%, 15%, 20%, 25%), c) and d) 5% of noise, e) and f) 20% of noise (the percentages are the percentage of noise in the overall signal)

When noise is added to a signal, the mean-square fluctuation function is significantly altered. Figure 2 shows the alteration for AR(1) and fractional Brownian motion results, and sums up all possible behaviors that can be encountered in practice. It can be inferred from the theoretical behavior of $F(l)$ that the presence of noise will tend to flatten the curve toward a flat line, which can indeed be observed by comparing Fig. 2.a) and 1.b).

Figure 2.b) also reveals an unexpected effect: the noise renders $F(l)$ concave. This can however be easily explained: Because of the stationarity of the signal, the signal-to-noise ratio tends to be lower for small lags. This results in a flattening of the curve for the smaller lags l, while for larger lags the behavior is distinctly a power law. Since the transition between both must be smooth by definition of $F(l)$, these characteristics can only result in a concave behavior (which is put in evidence on Fig. 2.f).

A similar reasoning can be applied to the AR process and its corresponding $F(l)$ curve. We have seen above that the higher SNR for small lags results in a flattening of the $F(l)$ curve. Thus, when noise is applied to the AR(2) process, the superposition of the noise-free convex curve and the flattening effect of the noise will first result in a straight line, while, when the overall level of noise is high enough, $F(l)$ will take a somewhat concave appearance (Fig. 2.e).

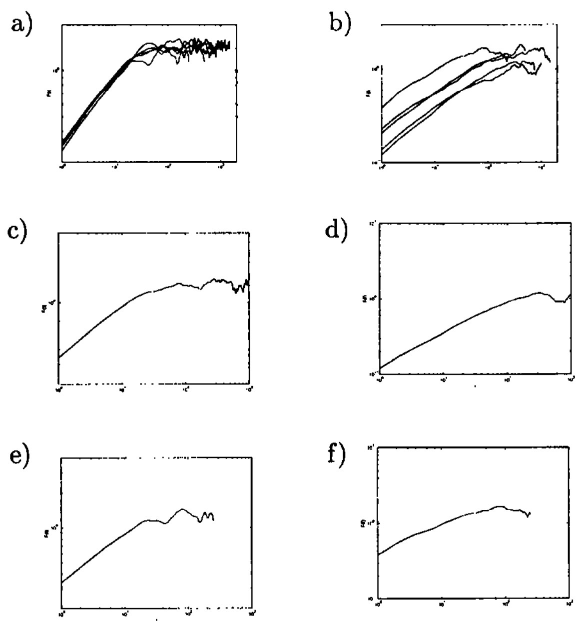

Figure 3: Effect of the length of the data on AR(1) process (a, b, c) and fractional Brownian motion (d, e, f). a) and d) $F(l)$ for (bottom to top) time series of 3000, 2000, 1500, 1000 and 500 samples. b) and e) show $F(l)$ for 1500 samples, c) and f) for 500 samples.

The effects of the length of the recording on $F(l)$ are by no means surprising. It has already been mentioned that the accuracy of the computation for a given lag depends on the number of points in the time series. It is indeed obvious that, if the lag between samples is of only some points, the averaging performed in eq.(2) will be much closer to the expected value than the average made for a lag close to half the duration of the recording. A decreased number of samples in the recording will thus decrease the amount of lags for which the computation of $F(l)$ will be accurate, hence the size of the usable part of the $log[F(l)]$ vs l curve.

The displacement of the curve origin that can be seen on Fig. 3.b) is purely random, and comes from the fact that a new realization of the fractional Brownian motion process has been used for each simulation.

Computing the exponent of the process is not a difficult task, but it will not be done, since the result is not accurate. It is indeed difficult to automatically select the suitable part of the curve. Besides, given the limited amount of data that is available, the numerical result will be unreliable.

In most practical cases, all that is available from a system is the observation of a time series, for which the amount of noise in unknown. The observation time, hence the amount of data, is limited, and little is known about the system itself. It can then be of help to summarize the results obtained so far according to the qualitative behavior of the $F(l)$ curve. The different results can be:

- **A straight line of non-nil slope**: In this case, the process under study is likely to be a pure fractal, and a fractal analysis can be conducted in order to characterize it. It is however possible that the underlying process is a random process with a non-flat spectrum, with some noise superimposed on it.

- **A straight line of slope** $\alpha = 0$: The process under study is a wide-band process. All that can be said is that it is not a fractal one, and that there are no periodicities in it. Nothing can be said about the amount of noise in the signal, and the available amount of data will have no influence on this fact.

- **A convex curve**: In this case, further studies must be conducted in order to determine the type of process one is dealing with. It can also be the result of the analysis of a fractal signal with a very limited amount of data (typically less than what is necessary to cover at least one period of the signal).

- **A curve with ripples**: The ripples are always indicative of some kind of periodicities in the signal. It is to be noted that they can have various origins: On Fig. 1.c), they come from a stochastic process, while on Fig. 1.e), their origin is to be found in a deterministic process.

- **A concave curve**: This always indicates the presence of noise in the signal. This shape can either be the result of a fractal process with noise superimposed on it, or of a non white process that is severely plagued with noise.

With this in mind, it is now possible to apply the mean square fluctuation function to real world data, namely heart-rate recordings.

4. APPLICATION ON HEART RATE

The mean square fluctuation function was applied to two groups of patients, control subjects, suffering from no known disease, and patients having recently undergone a heart transplant. Typical results are given on Fig. 4.

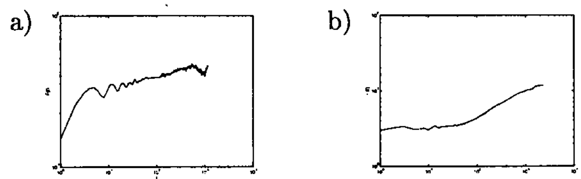

Figure 4: Typical mean square fluctuation function for a) control subjects, b) heart transplant patients

Figure 4.a) indicates that the normal heart rate is not fractal, and that some sort of periodicity is present in the signal. The latter can be readily explained, since respiration modulates the heart rate in a regular fashion. Interesting to note is the fact that this periodic feature even more readily visible in the mean square fluctuation function than it is in the corresponding frequency periodogram:

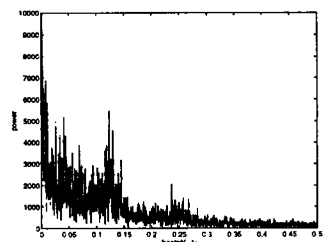

Figure 5: Periodogram of the heart rate recording of a control subject. The respiration peak is located between 0.1 and 0.15 $beats^{-1}$

The assessment of fractality is more complicated to settle. The heart transplant results are indicative of a fractal component plagued with noise (compare with Fig. 2.f). However, Fig. 4.b) does not give any evidence of fractality in normal human heart beat. This seems at first sight in contradiction with the results found in many articles, such as [9, 1]. This may however be explained: In the present work, only short-term correlations are taken into account, while the other papers examine the correlations exclusively on middle to long time scales. It is very likely that there exist short-term correlations that hide the long-range structure of the healthy heart rate.

These indications of fractality or no fractality are however not safe-proof. The last section has shown that several signals could be at the origin of the obtained plots. In the case of heart transplant recordings, two types of signals can match the results: either a fractional Brownian motion of some sort, on which noise is added, or an AR(1) process with a low signal-to-noise ratio. The normal heart rate, on the other hand, could be modeled either by an AR(1) process, a stochastic process of white spectrum, or fractional Brownian motion with some noise added. In all cases, an AR(2) process should be added in order to model the influence of the respiration on heart rate.

All hypotheses were tested, the accuracy of the modeling being assessed by the mean absolute error between the $F(l)$, the time series and the power spectral density of the model and the signal. The best models found for heart transplant patients are fractional Brownian motions of exponent $\beta = 2.0$, together with an AR(2) process taking the respiration into account, and some background noise. The control subjects, on the other hand, are best modeled with an AR(1) process, on which an AR(2) process and some noise are added, for the same reasons as before.

It can thus be concluded that an isolated heart rate (the transplanted heart is not linked with the central nervous system) seems to be fractal, while the healthy heart is not, or at least that the latter exhibit short-range correlations that are strong enough to overshadow a possible long-range structure.

5. CONCLUSION

This paper addressed the behavior of a method investigating long-range correlations for short data sets, including the effect of observation noise. It has been shown that, if some properties are identical to the ones predicted by theory, some strange unpredicted behaviors can appear, such as concave of convex curves. Each case is analyzed, and it is shown that a convex curve is indicative of a non-fractal non-periodic system, while concave curves are indicative of a fractal process contaminated with noise. Ripples may appear in any configuration, and are in every case indicative of the presence of some kind of periodicities in the signal.

This method has also been applied to real-world signals, in this case heart rate data. The results show that some evidence of fractality is visible, but this remains to be confirmed. It has indeed be seen that no final conclusion could be drawn directly from the $F(l)$ plots, since similar plots can have different origins. While being of much use to test if a signal could have fractal properties, it is not sufficient to assess the nature of a signal, at least in the case of short data sets. Further inquiries, such as hypothesis testing or modeling need to be performed in order to reach reliable conclusions.

6. REFERENCES

[1] R.G. Turcott and M.C. Teich. "Fractal character of the electrocardiogram: Distinguishing heart-failure and normal patients". *Annals of Biomedical Engineering*, Vol. 24, pp. 269–293, 1996.

[2] K.M. Stein and P. Kligfield. "Fractal clustering of ventricular ectopy in dilated cardiomyopathy". *The American Journal of Cardiology*, Vol. 65, pp. 1512–1515, June 1990.

[3] D.T. Kaplan, M.I. Furman, and S.M. Pincus. "Techniques for analyzing complexity in heart rate and beat-to-beat blood pressure signals". *IEEE Computers in Cardiology*, Vol. -, pp. 243–266, 1991.

[4] Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freemen and Company, 1977.

[5] S.B. Lowen and M.C. Teich. "Estimation and simulation of fractal stochastic point processes". *Fractals*, Vol. 3, No. 1, pp. 183–210, 1995.

[6] S.V. Buldyrev, A.L. Goldberger, and S. Havlin. "Fractals in biology and medicine: From DNA to heartbeat". In A. Bunde and S. Havlin, editors, *Fractals in Science*. Springer-Verlag, 1994.

[7] C-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, and A.L. Goldberger. "Long-range correlations and non-Gaussian behavior of the heartbeat". *Physical Review Letters*, Vol. 70, No. 9, pp. 1343–1346, 1993.

[8] J. Feder. *Fractals*. Plenum Press, New-York, 1988.

[9] J.P. Saul, P. Albrecht, R.D. Berger, and R.J. Cohen. "Analysis of long-term heart rate variability: methods, 1/f scaling and implications". *Computers in Cardiology 1987*, pp. 419–422, 1988.