STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LII, 3, 2007

SHORT AND LONG RANGE CORRELATION IN FLUCTUATING PHENOMENA

VASILE V. MORARIU*

ABSTRACT. Many natural phenomena from molecular to astrophysical processes may show a fluctuating behavior. Some of them are long-range correlated. However this paper reveals that just as many may show short-range behavior although they can be easily confounded with long-range correlation. The short-range memory of any complexity can be modeled by autoregressive processes.

INTRODUCTION

There are many natural phenomena which present fluctuating characteristics with apparent random properties [1]. The fluctuating characteristic may occur either in time or in space. For example the heart beating or human gating are not constant in time. The length of DNA coding sequences expressed as number of bases represents a fluctuating series along the genome etc. As a general rule we always search for order in such series of data which is not apparent to the naked eye. Order can be measured by investigating the correlation property of the data in the series. A correlation between two variables is a measure of the association or relationship between them. For example height and weight are related - taller people tend to be heavier than shorter people. An important point is that correlation does not always imply a causal relationship. However if correlation exists then research can be directed along specific lines such as to uncover causal relationships.

A fluctuating series of data may be attributed to two different kinds of processes: a) dynamic processes and b) stochastic processes. In the first case the source of fluctuation is a deterministic rule, described very often by a simple equation, which generates chaos. Stochastic processes means of, relating to, or characterized by conjecture and randomness. A stochastic process is one whose behavior is non-deterministic in that a state does not fully determine its next state. Here we distinguish uncorrelated (or random) fluctuation and correlated fluctuation respectively. They represent the subject of the present paper.

A classical example of correlated stochastic process is the so called 1/f noise where the amplitude of the noise is proportional to the inverse of

* National Institute of R&D for Isotopic and Molecular Technology, POBox 700, 400293 CLUJ-NAPOCA 5, ROMANIA

frequency [1]. This is also known as “pink” noise or flickering. Such a kind of noise is quite common in nature: from the flickering light of galaxies to the flickering of the red blood cells and many others. The present work started from the author’s earlier observation that a great deal of quite different fluctuating phenomena appeared to be fitted by the 1/f noise rule. This is generally associated to long range correlation or fractal like characteristics. However it became evident that quite often the 1/f characteristic revealed deviations from the rule of a power law. Evidence has accumulated in time until it was realized that deviations pointed to a significant qualitative difference from a 1/f behavior. These deviations were found to be due to short range correlation phenomena as opposed to the long range correlation of 1/f phenomena. The aim of this paper is to show how these distinctions can be made and to offer some examples from the field of molecular biology.

RESULTS AND DISCUSSION

We illustrate the case of the atomic mobility in the main chain of a protein (fig.1). The series of data were subjected to Fast Fourier Transform (FFT). Due to high dispersion of the data the spectrum can be easily approximated by a straight line and therefore by a 1/f spectrum. However, in this particular case it can be easily noticed that at low frequencies a plateau appears to be present and also at higher frequencies of the data. We found many similar cases which have been overlooked in the literature. Therefore this example shows how a non-1/f spectrum can be easily confused by a 1/f spectrum.

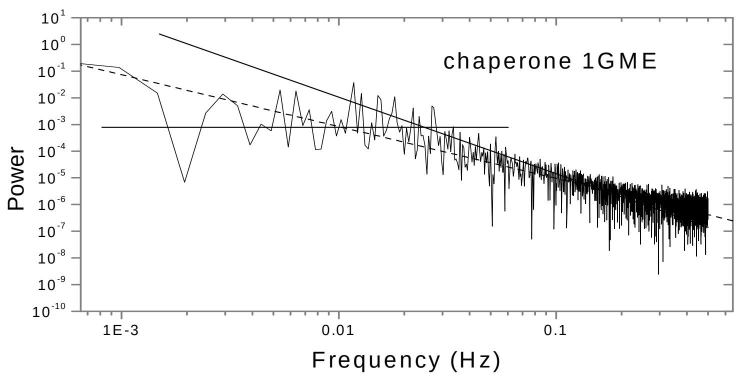

Fig.1. Non-averaged power spectrum for the atomic mobility in the backbone of a heat shock protein (Protein Data Bank code: 1gme). The linear fit (dashed line) of the spectrum may suggest a long-range correlation interpretation while the spectrum is non-linear with an evident plateau at low frequencies (continuous line) and a higher slope line at higher frequencies.

A main reason for this confusion is that the spectrum is quite noisy. Mandelbrot's recommendation that apparent 1/*f* spectra should be averaged before further interpretation has often been overlooked including in our own previous work [2-4]. Mandelbrot considers that non-averaged spectra lead to "unreliable and even meaningless results" [5]. An immediate consequence of the averaging procedure is that deviation from a power law description of the spectra can be easily disclosed otherwise remaining undetected. Sometimes such a deviation can be visible even if the spectrum is not averaged. On the other hand the non-averaged spectrum can be often reasonably fitted by a straight line, i.e. described by an apparent power law.

A further complication arises when the series has non-stationary character. The spectral analysis in fig. 1 is performed on the original series which is non-stationary as shown in fig. 2. The consequence is that the result is affected by the correlation introduced by the trend in the series (non-stationary character). Therefore the correlation described by the spectrum is in error. A simple procedure to avoid the non-stationary character is to define the trend by fitting the series with a higher degree polynomial. For example, in fig. 2, the trend can be described by a 10 degree polynomial. This can be subtracted from the series and the result is a detrended series (fig.3).

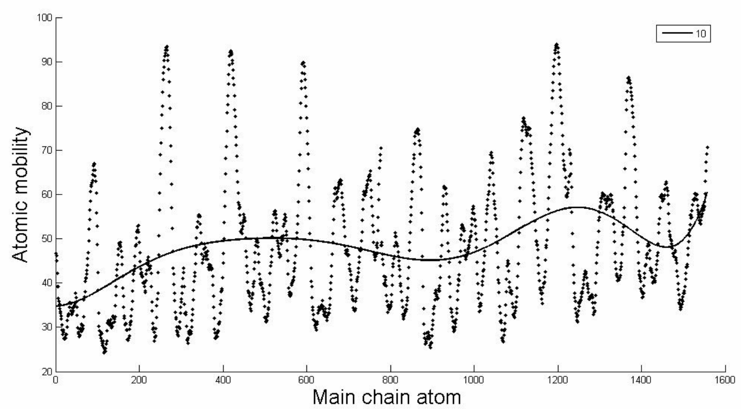

Fig.2. The series of the atomic mobility (crystallographic temperature factor) versus the main chain atom in the heat shock protein (PDB code: 1gme). The series is fitted by a 10 degree polynomial which describes the trend in the series.

It is on this detrended series that we performed Fourier Transform followed by averaging of the spectra. The true shape of the spectrum can be now revealed (fig.4).

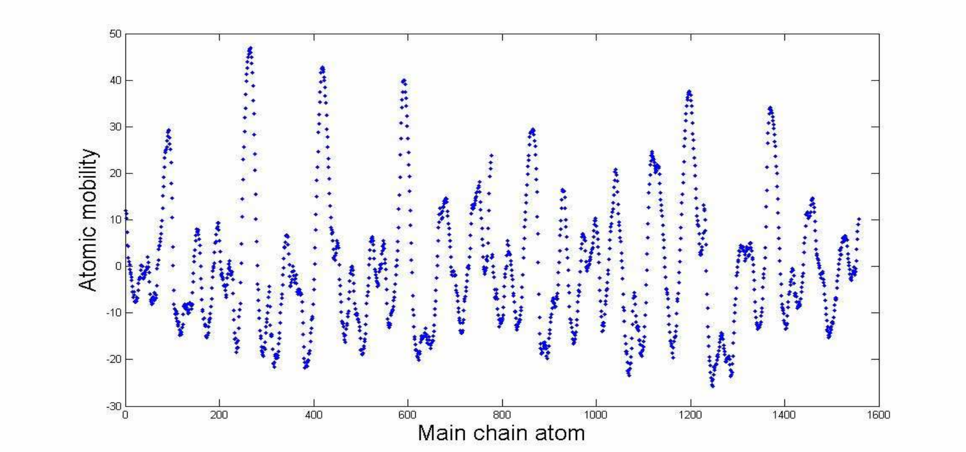

Fig.3. Detrended series shown in fig.2. This is done by subtracting the 10 degree polynomial from the original series.

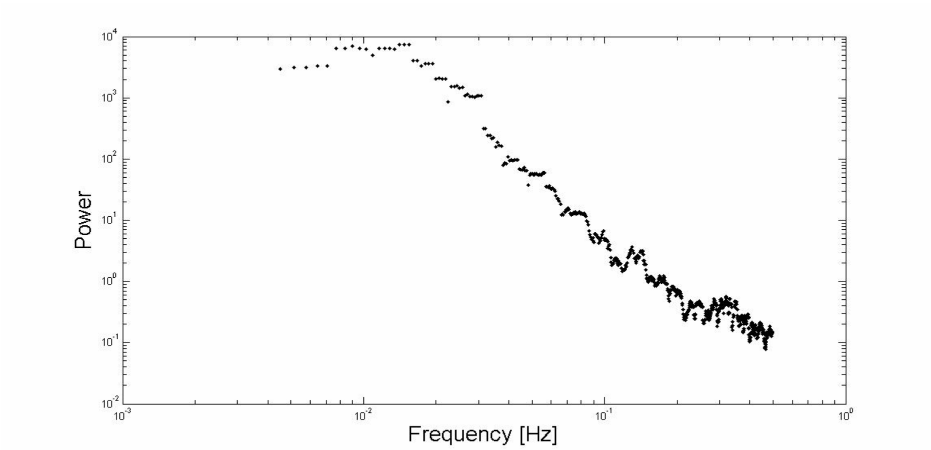

Fig.4. Averaged spectrum for the backbone atomic mobility of a heat shock protein (PDB code: 1gme). The number of averaged terms was M=13. The nonlinearity of the spectrum can be easily noticed. The averaging procedure was performed with a MATLAB program.

It can be noticed that the spectrum is non-linear, in other words it is not long-range correlated. A long–range correlation is visible on such double log plot as a linear dependence. Its slope represents the long-range correlation exponent β. However in this case long-range correlation is not operative therefore the spectrum cannot be described by a β exponent.

What can be done to characterize quantitatively such a spectrum? This is a more complicated question to answer in this limited space. The answer lays in the extreme large possibilities offered by autoregressive models [6-7]. Basically the most simple autoregressive model introduces an interaction between successive terms in a random series. Suppose $x_1, x_2, x_3\ldots$ is a random uncorrelated series. We can introduce a correlation in the series by constructing a new series following the simple rule [3]:

$$x_1,\ (x_1+x_2)0.5,\ [x_3+(x_1+x_2)0.5]0.5,\ \{x_4+[x_3+(x_1+x_2)/0.5]0.5\}0.5\ldots \qquad (1)$$

It can be seen that each term contains some information about the previous term and the further are the following terms, the less memory about the initial terms in the series. The multiplying factor, in the above example is 0.5, is a kind of strength of interaction among the terms which may vary between $0<\varphi<1$. If the factor $\varphi=1$ the series represent the integral series of the initial random series (or random walk). The further the term in the new series the less correlated with the initial terms in the series. Obviously such a series is short-range correlated. The example above is known as autoregressive model of order one or AR(1). A more formal description of an AR(1) model is given by the equation:

$$X_t = \varphi X_{t-1} + \varepsilon_t \qquad (2)$$

where ε_t is a white noise process with zero mean and variance σ^2, while φ is the interaction parameter [8]. The parameter values φ have to be restricted for the process to be stationary which means that $|\varphi|<1$. Examples of spectra for different values of φ for AR(1) processes are shown in fig. 5.

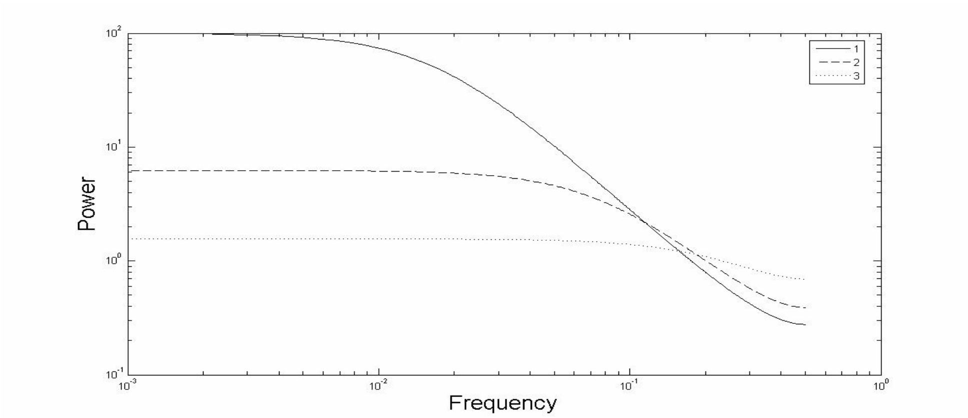

Fig. 5. Spectra of autoregressive models of first order, AR(1) for 1) $\varphi=0.9$; 2) $\varphi=0.6$; 3)$\varphi=0.2$.

AR(1) spectra show features already encountered in the examples above: a plateau at lower frequencies and the beginning of a plateau at higher frequencies. The length of the plateau at lower frequencies is evident for longer series of data. Shorter and real series may miss the plateau and noise has high amplitude at higher frequencies. As a result the spectra can be reasonably fitted by a straight line and the false conclusion that the series is long-range correlated like in fig.1. We explored a large number of series not only for proteins but also for DNA, from cell biology (flickering of red blood cells) and cognitive psychology (generation of apparent random numbers by human subjects). We found that about a quarter of the cases can be described by AR(1) models. We also found that more complicated spectra can be described by higher order autoregressive models AR(p). The relevant parameter in case of AR(1) processes is the interaction constant φ, while for higher order autoregressive processes AR(p), several constants of interaction are operative. These interaction constants are sensitive to various characteristics and influence the system, so that they can be profitably used for comparative studies of various systems.

CONCLUSIONS

This work shows that spectral analysis of fluctuating phenomena should be carefully treated in order to reveal the long-range or short-range correlation property. It is necessary that the series should be first detrended, than subjected to spectral analysis and the spectra should be averaged by using a reasonably high number of terms. Spectra which are not of 1/f type, i.e. long-range correlated, are most likely short-range correlated. They can be modeled by autoregressive processes of first or higher order. While the long range correlated systems can be characterized by the correlation exponent, a short-range correlated systems is characterized by the values and numbers of interaction constant between the terms.

REFERENCES

1. E. Milotti, *arxiv preprint,* **2002** physics/0204033
2. V.V. Morariu, A. Coza, Physica A **320,** 461 (2003)
3. A. Coza and V.V. Morariu, Physica A **320**, 449 (2003)
4. V.V. Morariu, A. Isvoran, O. Zainea, Chaos, Solitons and Fractals **32,** 1305 (2007)
5. B.B. Mandelbrot, “Multifractals and 1/f noise” Springer, New York, **1998**
6. P.J. Brockwell and R.A. Davies, “Time Series: Theory and Methods” 2nd ed. Springer, New York, **1991.**
7. J.D. Hamilton, “Time Series Analysis”, Princeton University Press, **1994**.
8. V.V.Morariu, L.Buimaga-Iarinca, C. Vamoş, Ş.Şoituz, Fluctuation and Noise Letters **7**, L249 (2007)