AUTOREGRESSIVE MODELING OF THE VARIABILITY OF AN ACTIVE GALAXY

V. MORARIU[1,2], C. VAMOS[3], A. POP[4], S. SOLTUZ[3], L. BUIMAGA-IARINCA[1],

[1]National Institute of Research and Development for Isotopic and Molecular Technologies,
Department of Molecular and Biomolecular Physics, 400293, Cluj-Napoca, Romania
E-mail: vvm@itim-cj.ro
[2]Academy of Romanian Scientists, 54, Splaiul Independentei, Sector 5, 050094 Bucharest, Romania
[3]"T. Popoviciu" Institute of Numerical Analysis, Romanian Academy, 400110 Cluj-Napoca, Romania
E-mail: vamoscalin@yahoo.com
[4]Astronomical Institute of the Romanian Academy, Astronomical Observatory, 19 Ciresilor,
400487 Cluj-Napoca, Romania
E-mail: andi_pop@yahoo.com

Many natural phenomena can be described by power-laws of the temporal or spatial correlations. The equivalent in frequency domain is the $1/f$ spectrum. A closer look at various natural data reveals more or less significant deviations from a $1/f$ characteristic. Such deviations are especially evident at low frequencies and less evident at high frequencies where spectra are very noisy. We exemplify such cases with a phenomenon offered by astrophysics. The X ray variability of the NGC 5506 galaxy can be better approximated by AR(1) – a first order autoregressive model, than by a $1/f$ model (long-range memory). The same spectra can be more or less easily confused and/or approximated by power-laws. A key step to detect non-power laws in the spectra, already suggested by Mandelbrot, is to average out the spectra.

Key words: Autoregressive model, short-range correlation, galaxies.

1. INTRODUCTION

Many processes in nature exhibit temporal or spatial correlations that can be described by power-laws. Examples can be found in physical, biological, social and psychological systems [1–5]. In case of a spectral analysis, the power-law is of the type $P = 1/f\text{-}\beta$, where f is frequency and β is the long-range correlation exponent. Its value is $0 < \beta < 2$ for most of natural processes. A $1/f$ spectrum is diagnosed by looking at the double log plot which should be linear over the whole range of frequency scale. The slope of the linear fit is the correlation exponent β. However, a general problem encountered in the spectral analysis of various fluctuating systems is that the spectrum is quite noisy. Mandelbrot's recommendation that apparent $1/f$ spectra should be averaged before further interpretation [4], has often been overlooked including some of our previous work [6–9]. Mandelbrot considers that non-averaged

spectra leads to" unreliable and even meaningless results" [4]. An immediate consequence of the averaging procedure is that deviation from a power-law description of the spectra may be better disclosed. Some-times, such a deviation can be obvious 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. Such an example is illustrated in figure 1a for a heat shock protein (Protein Data Bank code: 1gme). The linear fit of the power spectrum appears to be reasonable good, yet it can be easily noticed the plateau at low frequencies (Figure 1b). A simple averaging procedure can result in a clear non-1/f spectrum (Figure 1b).
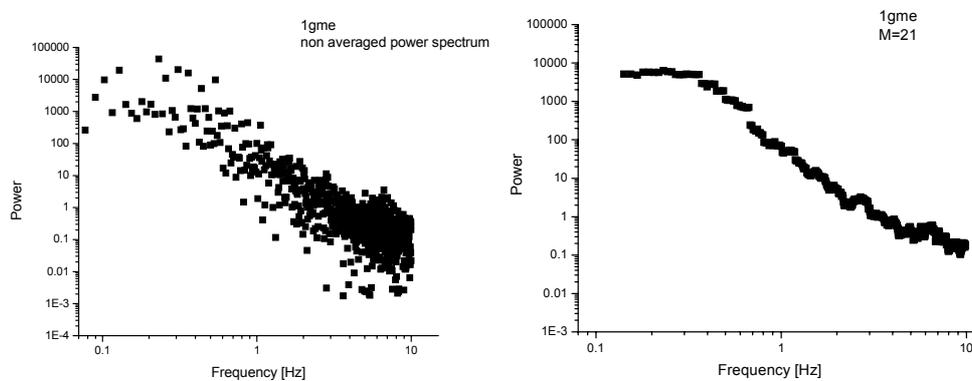

Fig. 1 – a) Non-averaged power spectrum for the atomic mobility in the backbone of a heat shock protein (PDB code: 1gme). The spectrum can be well approximated by a straight line which might suggest a long-range correlation interpretation. b) The averaged power spectrum with M = 21 terms. It clearly indicates the short-range correlation (deviation from linearity). PDB stands for Protein Data Bank, available at http://www.rcsb.org.

Consequently, two confusions may occur when dealing with non-averaged spectra: either a hidden non-power-law remains undisclosed, and/or non-power-laws are easily misinterpreted as power-laws. This work discusses the deviation from power-law behavior, which is characterized by two features: leveling of the spectrum at low frequencies, and a tendency for leveling at high frequencies, where the shape of the spectrum is blurred by the high level of noise ("the Spanish moss" as coined by [4]).

The idea of this work is twofold: first, to disclose cases of non-power-laws by a simple averaging of spectra (other examples from widely different areas of science are described in a preprint [12], and second, to offer an interpretation of such spectra by using an autoregressive model AR(1). This is basically different from a power-law description. While the meaning of a power-law is associated with long-range correlation or memory, the autoregressive model describes a system with short-range memory. The long-range memory is described in terms of a long-range correlation exponent β of the power law P = 1/f^{β}, while the short-

range memory is characterized by the strength interaction φ among consecutive terms as described by an AR(1) model (see further). The literature already reported cases where astrophysical and psychological phenomena are described by autoregressive models rather than by power-laws [10–12].

The spectral approach in the present work was done for two reasons: first, many results in the literature are presented in spectral form and the present work was born out from the observation that $1/f$-like spectra show clear deviations from a power-low, and second, the spectral description can be easily done in an analytical manner.

This paper is organized as follows: first, the main spectral features of an autoregressive model AR(1) are described for various interaction strengths among the terms of the series. Then, is analyzed an astrophysical phenomenon (X-ray emission of a galaxy) where a more complex calculation based on an AR(1) model has already been published [10]. It will be shown that an identical result is obtained with our more simple procedure.

2. THE SPECTRAL CHARACTERISTICS OF THE AUTOREGRESSIVE MODEL

A discrete stochastic process $\{X_n, n = 0, \pm1, \pm2, \ldots\}$ is called autoregressive process of order p, denoted AR(p), if $\{X_n\}$ is stationary and for any n:

$$X_n - \varphi_1 X_{n-1} - \ldots - \varphi_p X_{n-p} = Z_n, \tag{1}$$

where $\{Z_n\}$ is a Gaussian white noise with zero mean and variance σ^2. The real parameters φ_i, $i = 1, \ldots, p$, can be interpreted as a measure of the influence of a stochastic process term on the next term after i time steps. The properties of AR(p) processes have been studied in detail and they are the basis of the linear stochastic theory of time series [13] and [14]. Equation 1 has a unique solution if the polynomial $\Phi(z) = 1 - \varphi_1 z - \ldots - \varphi_p z^p$ has no roots z with $|z| = 1$. If in addition $\Phi(z) \neq 1$ for all $|z| > 1$, then the process is causal, *i.e.* the random variable X_n can be expressed only in terms of noise values at previous moments and at the same moment.
The spectral density of an AR(p) process is:

$$f(\nu) = \frac{\sigma^2}{2\pi} \frac{1}{\left|\Phi(e^{-2\pi i\nu})\right|^2}, \quad -0.5 < \nu \leq 0.5, \tag{2}$$

where ν is the frequency. For an AR(1) process, the spectral density in equation 2 becomes:

$$f(\nu) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \varphi^2 - 2\varphi\cos 2\pi\nu}, \quad -0.5 < \nu \leq 0.5, \tag{3}$$

where φ is the only parameter φ_i in this case. The above mentioned formulas are valid for ideal stochastic processes of finite length.

The time series found in practice have a finite length and usually they are considered realizations of a finite sample of an AR(1) process of infinite length. Therefore, the changes of the equations 2 and 3 have to be analyzed for a sample with finite length $\{X_n, n=1, 2, ..., N\}$ extracted from an infinite process $\{X_n, n = 0, \pm 1, \pm 2, ...\}$. A detailed analysis of the power spectrum of the AR(1) process and the influence of the finite length is contained in [15]. In this paper some of the main conclusions are discussed.

The sample estimator of the spectral density is the periodogram:

$$I_N(v) = \left| A_N(v) \right|^2 , \tag{4}$$

where $A_N(v)$ is the discrete Fourier transform of the sample:

$$A_N(v) = \frac{1}{\sqrt{N}} \sum_{n-1}^{N} X_n e^{2\pi i n v} . \tag{5}$$

Since the sample contains a finite number of components, there are only N independent values of $A_N(v)$ and $I_N(v)$. Usually, these values are computed for the Fourier frequencies $v_j = j/N$, where j is a integer satisfying the condition $-0.5 < v_j \leq 0.5$. The periodogram of an AR(p) process is an unbiased estimator of the spectral density:

$$\lim_{N \to \infty} \left\langle I_N(v_j) \right\rangle = 2\pi f(v) , \tag{6}$$

where $v_j -0.5/N < v \leq v_j+0.5/N$ [13]. Hence, increasing the sample length N, while the time step is kept constant, the average periodogram becomes a better approximation of the spectral density (equation 6). However, a single periodogram is not a consistent estimator, because it does not converge in probability to the spectral density, *i.e.* the standard deviation of $I_N(v_j)$ does not converge to zero, and two distinct values of the periodogram are uncorrelated, no matter how close the frequencies are when they are computed.

Usually, the spectral density and the periodogram are plotted on a log-log scale. The logarithmic coordinates strongly distort the shape of the graphic because by applying the logarithm, any neighborhood of the origin is transformed into an infinite length interval and the value of $f(0)$ cannot be plotted. For a sample with N terms, the first value of the spectral density is obtained for the minimum frequency $v_{min} = 1/N$. Figure 2a shows the spectral density in equation 3 for $N = 1024$, $\sigma = 1$ and different values of the parameter φ. For $\varphi = 0.90$ and especially for $\varphi = 0.99$, a significant part of the power spectrum is almost linear with a slope equal to -2, which corresponds to $\beta = 2$. A significant part of the spectrum could be regarded as linear for smaller value of φ (for example $\varphi = 0.5$ in figure 2a).

In order to verify this behavior of the AR(1) spectrum, figure 2b includes the derivative of the spectral density from equation 3 in log-log coordinates:
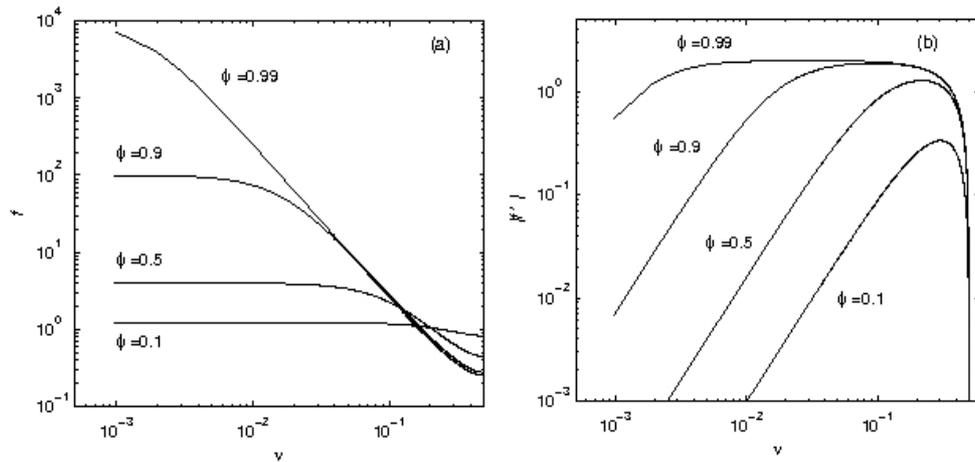
$$f'(v) = -v \frac{d}{dv}(\ln f(v)) \tag{7}$$

Fig. 2 – The spectral density (a) and the absolute value of its derivative (b) of an AR(1) process for $N = 1024$, $\sigma = 1$ and different values of the interaction factor among successive terms φ. It can be noticed that spectra with with larger values of φ can be easily approximated by a straight line (power law) and therefore misinterpreted as a fractal like case.

One can notice that for $\varphi \geq 0.9$ there is a region where $f' \cong -2$. There is only a maximum value of f' for $\varphi < 0.9$ which corresponds to the center of the "linear" (or "fractal") region of the power spectrum.

For small frequencies, the AR(1) spectral density is strongly stretched in log-log coordinates such that a plateau appears (figure 2a) with a value given by:

$$f(0) = \frac{\sigma^2}{2\pi(1-\varphi)^2} \tag{8}$$

From equation 3 it follows that the plateau corresponds to the small values of v, when the variable term at the denominator can be neglected in comparison with the constant term. Using the quadratic approximation of the cosine function, the condition that the graph of the AR(1) power spectrum has obtained a plateau $v < (1-\varphi)/2\pi\sqrt{\varphi}$. If φ tends to 1, the plateau appears at smaller values of the frequency. Therefore, if N is large enough, the periodogram of an AR(1) sample has a plateau at small frequencies (if N is large, then $v_{min} \to 0$).

A time series $\{x_n, n = 1, 2, ..., N\}$ as a realization of a sample $\{X_n, n = 1, 2, ..., N\}$ from an AR(1) process is considered. Applying the discrete Fourier transform (equation 4) to the time series $\{x_n\}$, and then computing the periodogram (equation 5), values randomly distributed around the spectral density (equation 3) of the AR(1) are obtained. Since the periodogram is not a consistent estimator, by increasing the length N of the sample, the periodogram fluctuations around the theoretical spectral density are not reduced. Consistent estimation of the spectral density may be obtained using averaging of the periodogram on intervals with length of magnitude

order of $\sqrt{N}$ [13]. Choosing the optimum weight function is a difficult task, because, if the periodogram is smoothed too much, then the bias with respect to the theoretical spectrum can become large. From various weight functions [16] the simplest one is used, *i.e.* the averaging with equal weights on symmetric intervals containing M Fourier frequencies, with $M = 1, 3, 5, ..., 21$. Then, the averaged periodogram contains $N - M + 1$ values, because for the first and last $(M - 1)/2$ values of the periodogram the symmetric averaging can not be performed.

Let us consider that an AR(1) model for an averaged periodogram is to be found, *i.e.* to find the values of the parameters φ and σ. The minimum of the quadratic norm of the difference between the averaged periodogram and the theoretical spectral density of the AR(1) model has to be determined. The sample standard deviation of the time series and $\varphi = 0$ are used as initial values for the optimization algorithm.

3. DATA AND METHODS

Different type of astrophysical objects display variability phenomena featured by power-law power spectra, e.g. the light curve of the 3C273 quasar. Also, the amplitude spectrum (in log-log plots) of the intermediate polar AE Aquarii shows a 1 slope [18], while the power spectrum of its radio emission time variability revealed the presence of a red noise described by a power-law. The X-ray variability of Cygbus X-1 system [19] and of Be star γ Cassiopeiae [20] is also featured by power spectra displaying $1/f$ segments. The X-ray variability of active galactic nuclei is also known to show red noise spectra which could be quite well fitted by power-laws [21]. Deviations from the simple power-law behavior are emphasized by several authors. Thus, there are mentioned cases [20, 22, 23, 24] in which different frequencies domains are featured by different slopes, or the power spectra gradually flatten toward low frequencies. Autoregressive analysis has been reported on the variability of X-ray light curves of the active Galaxy NGC5506 [10]. This analysis is compared with the present more simple approach, by using the same data extracted form Hearc Exosat ME archive for the Seyfert galaxy NGC5506.

Many series of data have non-stationary characteristics, so the application of Fourier transforms to the data results in misleading spectra. A common procedure to avoid this complication is to use detrended fluctuation analysis (DFA) [25]. This results in a correlation exponent free of the correlation introduced by the trend. However, in our case it is essential to obtain the corresponding spectrum, as the shape of the spectrum gives the relevant information (either a power-law or a non-power law is operative). Consequently, an important preliminary step is to remove non-stationary characteristics in the series. We performed detrending by subtracting a polynomial fit from the original series. The problem is to determine the right polynomial fit. We performed 1 to 20 degree polynomial fits and generally found that a polynomial degree around 10 gives the most reliable result

for φ and σ. The values of φ and σ also depended on the averaging procedure of the spectra so that optimizing the values of φ and σ involved optimizing both the detrending and the averaging procedures.

The above mentioned polynomial fitting was chosen as its accuracy is comparable to the moving average method and to an automatic method for the estimation of a monotone trend. The same work also showed that polynomial fitting for a $1/f$ noise proved to have the best performance [26].

The succession of operations can be summarized as following: *i*) Detrend the series of data by subtracting various degrees of polynomial fits; *ii*) Discrete Fourier transform of the series; *iii*) Periodogram averaging using 1–21 terms; *iv*) Fit the spectrum to an AR(1) model. The resulting parameters are the interaction factor φ and the dispersion σ. Their values depend on the degree of the polynomial fit used for the detrending procedure. Finally choose the values of φ and σ, by analyzing the plot of φ and σ against the polynomial degree and the number of averaging terms.

4. AR(1) FITTING PROCEDURE FOR X RAY EMISSION OF THE NGC5506 GALAXY

The variability of X-ray flux from galaxies has been previously described as flickering or $1/f$ fluctuation [10]. A specific problem in astronomical observations is the observational noise, as well as other misleading systematic effects occurring in power spectra. As a result, a specific model was considered in order to generalize AR processes and to estimate the hidden autoregressive process [10]. The model is known as the Linear State Space Model (LSSM) in which the observational noise is explicitly modeled. The results reported by König and Timmer for the first order process AR(1) are the parameter $\varphi = 0.994$ and the standard deviation $\sigma = 0.722$.

König and Timmer analysis was compared with the present more simple approach and proved to be in a very good agreement. The data series are represented in Figure 3a.

The first problem is to remove the deterministic trend that can be found by examining the shape of the signal. Timmer and König did not extract any deterministic trend in their analysis. Three different trends can be obtained by different polynomial fitting degrees (Figure 3b). The degree of these polynomials equals the minimum degree of a class of polynomial trends q, which has very similar shapes. When the degree of the polynomial trend increases, the shape of the trend does not change monotonically. At certain polynomial degrees, the trend has more significant changes, while for the following polynomial degrees the shape remains practically unchanged. The trend reduces to a constant, which equals the average of the temporal series when $q = 0$. There is no significant improvement for a linear trend ($q = 1$), while for $q = 2$ the polynomial trend describes the global shape of the signal. Only for $q = 5$ the polynomial trend describes the different

behavior of the first and last half of the signal. The polynomial trend can follow better the details of the signal when $q = 9$, however, numerical oscillations arise at the end, which cannot be associated with real variations. These numerical oscillations increase with the degree of the polynomial, therefore they were not considered. The choice of these four representative values for the degree of the polynomial trend will be quantitatively confirmed by the variation of parameters characterizing the autoregressive model.
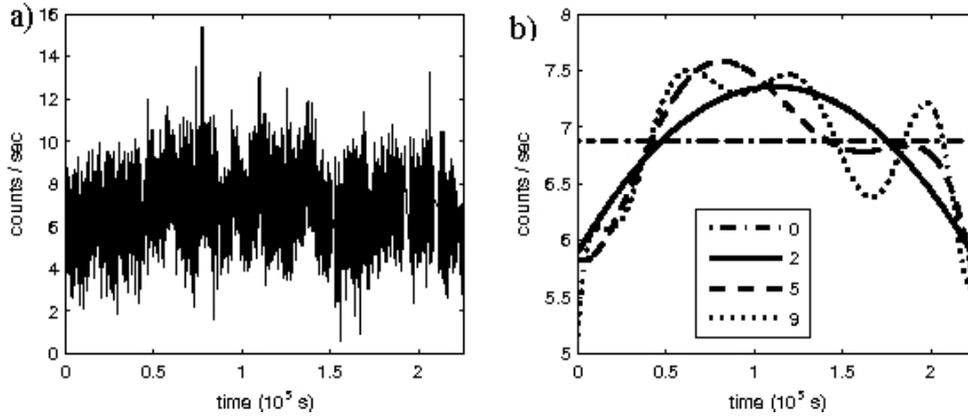

Fig. 3 – a) The X-ray time series of galaxy NGC5506; b) Trends of the time series obtained for three different polynomial fittings (0, 2, 5 and 9). The trend reduces to a constant when $q = 0$. For $q = 2$, the polynomial trend describes the global shape of the signal. For $q = 5$, the polynomial trend describes the different behavior of the first and last half of the signal. The polynomial trend can follow the details of the signal when $q = 9$.

The calculation of the periodogram can not be done using equation 5, since there are missing data in twenty regions. The other values of the series are separated by time interval $dt = 30s$. Therefore, the series can be considered as equidistant and they have $T = 7532$ values, with 584 missing values. A zero value is assigned to these data, so they do not contribute in equation 5 to the value of the periodogram.

The periodogram is presented in Figure 4a for the signal where only the mean value was subtracted, and in Figure 4b for averaged periodogram with $M = 21$ values. It can be seen that the averaging procedure results in loosing data for the lower frequencies which describe the plateau of an autoregressive model.

At frequencies higher than the cutting frequency $v_0 = 0.02$, the shape of the periodogram changes to a white noise spectrum, as it can be seen in Figure 5b. Then, only one part of the spectrum, for $v > v_0$, can be modeled with an AR(1) process. The values $\varphi = 0.991$, $\sigma = 0.757$ and $\varphi = 0.985$, $\sigma = 0.744$ are obtained by fitting the lower frequencies part of the periodogram to an AR(1) process for the non-averaged and the averaged periodogram respectively. These values are very close to those reported by König and Timmer.
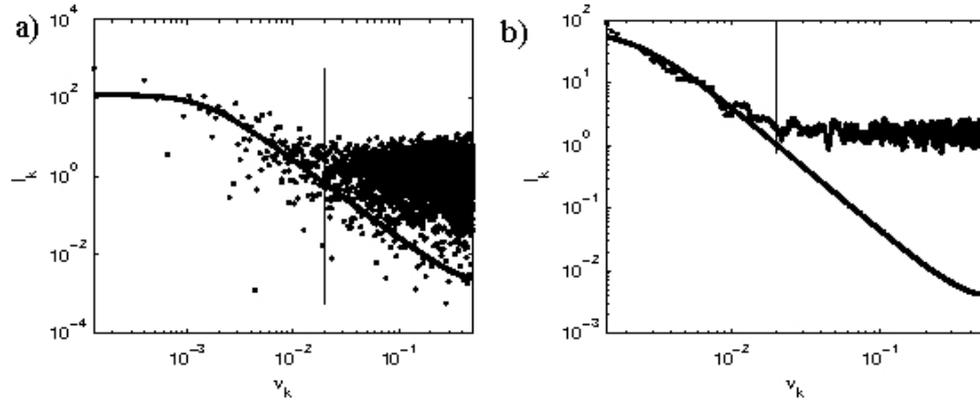
Fig. 4 – a) Periodogram of the time series for galaxy NGC5506. The mean value q = 0 was subtracted; b) The averaged periodogram using a rectangular window with M = 21 values. The averaging procedure results in loosing data for the low frequencies that describe the plateau of an autoregressive model. At frequencies higher than the cutting frequency $\nu_0 = 0.02$, the shape of the periodogram changes to a white noise spectrum. Only one part of the spectrum, for $\nu > \nu_0$, can be modeled with an AR(1) process.

As in the former example, the averaging interval of periodogram has a strong influence on the parameters of the AR(1) model. This dependence is shown in figure 5 as a function of *M* for different degrees of the polynomial trend. For small values of *M*, the values of φ and σ for different polynomial trends are quite different. Their variability for $M > 9$ is significantly reduced, therefore the averaging procedure can eliminate the fluctuation of the periodogram. This is why $M = 11$ is used in this investigation.
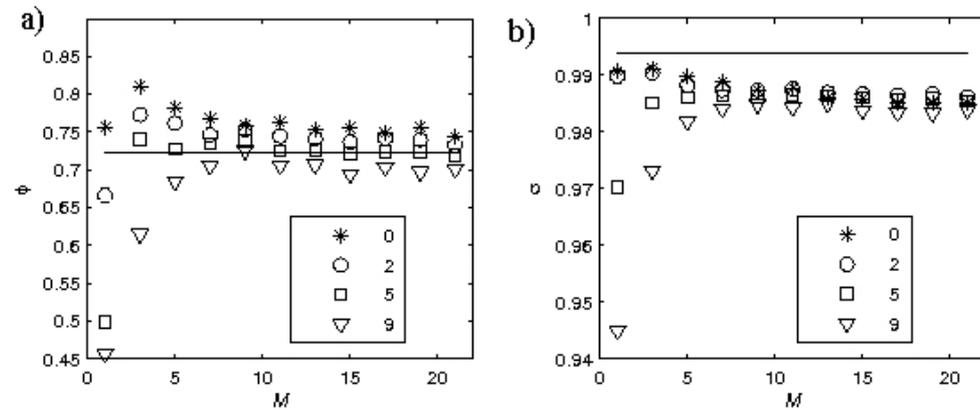

Fig. 5 – a) Dependence of the parameters (a) φ and (b) σ for the AR(1) model on the polynomial trend degree (from 0 to 9) for different *M* values. Continuous lines represent the values reported by [10]. For small values of *M,* the values of φ and σ for different polynomial trends are quite different. For $M > 9$, their variability is significantly reduced, so the averaging procedure can eliminate the periodogram of fluctuation.

Another parameter which has to be analyzed in order to fit an autoregressive model is the cutting frequency v_0. It can be noticed that at frequencies smaller than 0.02 the periodogram presents oscillations, which can be caused by the white noise at higher frequencies. The dependence of the autoregressive parameters on the value of v_0 is presented in Figure 6. It shows that dispersion of the noise does not depend significantly on the periodogram interval used for the fitting procedure.

However, for all the degrees of the polynomial trend eliminated from the initial signal, the value of φ decreases with the increase of v_0. Only the first two values are almost equal. Such a decrease can be explained by the influence of the observational white noise in the modeled part of the periodogram. Consequently, the cutting frequency was established at $v_0 = 0.005$.
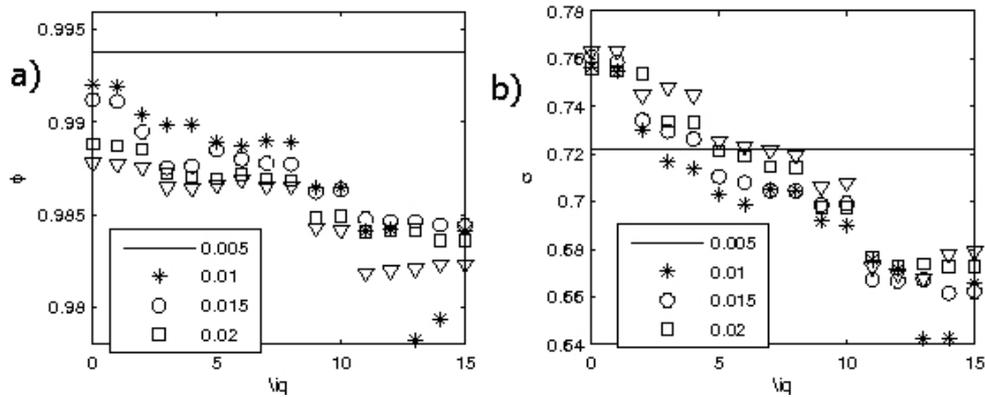

Fig. 6 – The dependence of the autoregressive parameters (a) φ and (b) σ on the degree of polynomial trend q.

First, the stepwise variation at $q = 2$, 5 and 9 can be noticed, which corresponds to the discussion mentioned at the beginning of this section. Another discontinuity at $q = 11$ was no longer used, as the oscillations at the end of the interval become too large. Also, it can be noticed that the autoregressive parameters decrease as the degree of the polynomials increases. This is due to the fact that higher polynomial degrees can better describe the oscillations of the signal, while for lower degrees they are modeled by the autoregressive process. As Köning and Timmer did not remove any deterministic trend from the signal, a comparison with their results should consider the results for $q = 0$. The relative error for φ is 0.2% and 5% for σ, which confirms the spectral autoregressive modeling method of the time series proposed in this work.

5. CONCLUSIONS

We propose an AR (1) model for the description of the X ray emission from a galaxy. The two parameters needed for such a description of the spectrum is φ –

the strength of interaction among consecutive terms in the time series and the dispersion σ of the data. Our more simple calculation procedure resulted in similar results with those obtained by a more elaborate calculation. The parameter φ proved to be sensitive and therefore can be profitably exploited to investigate various effects on the fluctuating system. On the other hand, the spectrum of such an emission can easily be confused and/or approximated by power-laws. The most important step in disclosing the nature of fluctuations is to average out their spectra. Apparent 1/*f* spectra should be cautiously treated and averaging should be compulsory. The present method of calculation is being currently tested for other phenomena arising from molecular biology, cell biophysics and cognitive psychology.

REFERENCES

1. Stanley H.E., Ostrowsky N., Dordrecht, *Correlations and connectivity: geometric aspects of physics, chemistry and biology*: Kluwer (1990).
2. Bunde A., Havlin S., Berlin, *Fractals in science*: Springer (1994).
3. Milotti E., *1/f noise: a pedagogical review. Arxiv preprint, Physics* /0204033 (2002).
4. Mandelbrot B.B., *Multifractals and 1/f noise,* New York: Springer (1998).
5. Bassingthwaighte J.B. *et al.*, *Fractal physiology*, New York: Oxford University Press (1994).
6. Morariu V.V., Coza A., Physica A, 320, 461–474 (2003).
7. Morariu V.V., Isvoran A., Zainea O., *Chaos, Solitons and Fractals*, 32, 1305–1315 (2007).
8. Morariu V.V., Coza A., *Fluctuation and Noise Letters*, 1, L111–L116 (2001).
9. Morariu V.V., Coza A., Chis M. A., Isvoran A., Morariu L.C., *Fractals*, 9, 379–391 (2001).
10. König M., Timmer J., Suppl. Ser., 124, 589–596 (1997).
11. Timmer J., Schwarz U., Voss H.U., Kurths, J. Phys. Rev. E, 61, 1342–1352 (2000).
12. Thornton Th.L., Gilden D.L., Provenance Psychonomic Bulletin & Rev., 12, 409–441 (2005).
13. Brockwel P.J., Davies R.A. *Time series: theory and methods*, 2nd edn. New York: Springer (1991).
14. Hamilton J.D., *Time series analysis*, Princeton University Press (1994).
15. Vamos, C., Soltuz, S.M., Craciun, M. arxiv:0709.2963 (2007).
16 Stoica P., Moses R.L, *Introduction to spectral analysis,* New Jersey: Prentice Hall (1997).
17. Morariu V.V., Vamoş C., Pop A., Şoltuz S.M., Buimaga-Iarinca L., Zainea O., arxiv [q-bio] 0808.1021 [2008].
18. Elsworth Y.P., James J.F., Mon. Not. R. Astr. Soc., 198, 889–896 (1982).
19. Lochner J.C., Swank J.H., Szymkowiak A.E Astrophys. J., 376, 295–311 (1991).
20. Smith M.A., Robinson R.D., ASP Conf. Ser., 292, 263-274 (2003).
21. Gaskell C.M., Klimek E.S., Astron. Astrophys. Trans., 22, 661–679 (2003).
22. Uttley P., McHardy I.M., Mon. Not. R. Astr. Soc., 323, L26L30 (2001).
23. Papadakis I.E., Lawrence A., Mon. Not. R. Astr. Soc. 272, 161–183 (1995).
24. McHardy I.M., Papadakis I.E., Uttley P., Page M.J., Mason K.O., Mon. Not. R. Astr. Soc., 348, 783–801 (2004).
25. Peng C.-K., Buldyrev S.V., Havlin S., Simons M., Stanley H.E., Goldberger A.L., Phys. Rev. E, 49, 1685–1689 (1994).
26. Vamos C., Phys. Rev. E, 75, 036705 (2007).