A STUDY OF PERSISTENCE ANALYSIS IN CLIMATIC DATA FROM SALVADOR – BRAZIL

Breno M. BRAGA [1]
Amin BASSREI [2]

[1] Bacharel em Geofísica, Geofísico, PETROBRAS S/A.
[2] Engenheiro Eletricista, Doutor em Geofísica Aplicada, Professor Associado, Departamento de Geologia e Geofísica Aplicada, Instituto de Geociências & Centro de Pesquisa em Geofísica e Geologia, Universidade Federal da Bahia. Pesquisador, Instituto Nacional de Ciência e Tecnologia em Geofísica de Petróleo. E-mail: bassrei@ufba.br

RESUMO. Estudo sobre análise de persistência de dados climáticos de Salvador, Brasil. A geometria fractal tem sido usada frequentemente para caracterizar e descrever fenômenos naturais. Suas aplicações variam desde dimensões microscópicas até a compreensão de fenômenos macroscópicos. Baseado nesse princípio nós estudamos um conjunto de dados climáticos, medidos em Salvador, Bahia, através dos métodos da análise espectral e da análise re-escalonada. As séries temporais são os parâmetros físicos da precipitação, pressão atmosférica, evaporação, umidade, radiação solar, temperaturas máxima, mínima e média coletadas mensalmente durante um período de 30 anos, de 1961 a 1990. Para cada parâmetro, nós calculamos o expoente de Hurst *H* e o expoente espectral *b*, atestando a variabilidade dos dados disponíveis. Os resultados para *H* variaram de 0.52 a 0.91, para ambos os métodos, de modo que todos os parâmetros se comportaram como persistentes. A importância do estudo mostrou que diferentes métodos propiciaram os mesmo resultados, o que é um fato significativo para sistemas complexos. O conhecimento de *H* permite o cálculo da dimensão fractal, o que quantifica a complexidade do fenômeno climático.
Palavras-chave: dimensão fractal, expoente de Hurst, climatologia.

ABSTRACT. *Fractal geometry has been used frequently to characterize and describe natural models. Its applications range from microscopic dimensions to the understanding of macroscopic processes. Following this principle, we studied a set of climatic data, collected in Salvador, Bahia, through the methods of rescaled analysis and spectral analysis. The time series are physical parameters of precipitation, atmospheric pressure, evaporation, humidity, solar radiation, maximum, minimum and average temperatures collected monthly over a span of 30 years, from 1961 to 1990. For each parameter, we computed the Hurst exponent H and the spectral exponent b, attesting the variability of the available data. The results for H varied between 0.52 and 0.91, for both methods, in such a way that all the parameters behave as persistent. The importance of such study is to show that different methods provide the same results, which is a significant fact for complex systems. The knowledge of H also provides the computation of the fractal dimension, thus allowing us to quantify the complexity of the climatic phenomena.*
Key-words: *fractal dimension, Hurst exponent, climatology.*

INTRODUCTION

The mechanisms related to the weather are, in a first glance, completely random. However more precise analyses have revealed some order in those phenomena, which is expressed with the form of scale invariance. This scale invariance is the foundation of fractal geometry. Fractal geometry is the field of Mathematics that studies the properties and behavior of fractals. It describes many situations in sciences and technology that cannot be explained by classical geometry. The conceptual history of fractals comes from the measurement attempts of objects which traditional definitions based on Euclidean geometry fail. A fractal is an object that can be divided into pieces, each one similar to the original object.

Fractals have infinite details, are self-similar and scale independent. In many cases a fractal can be generated by a repetitive pattern, typically a recurrent or iterative process. The theory of fractals has been applied to the description of some physical processes in Geosciences, in particular to analysis of geometrical relations in different observation scales. For example, it is possible to understand in a process how a microscopic behavior can influence a macroscopic behavior.

As above mentioned many physical variables seem to have a scale behavior, which means that their power spectra *P(f)* are proportional to the frequency *f*:

$$P(f) = \frac{1}{f^b},$$

or

$$\ln P(f) = -b \ln(f),$$

where *b* is the spectral exponent (Feder, 1989). This implies that although a sequence of values of a given physical property is non-periodic, it preserves its statistics in an observed range scale. This phenomenon is related to fractal geometry. When the value of the spectral exponent is around 1, limited between 0.5 and 1.5, the considered property is said to behave as *1/f* noise. One method to quantify the correlation or long range persistence is through the re-scaled analysis (*R/S*), which provides the so-called Hurst exponent *H*. A non-correlated signal, that is, without persistence, like the white noise has H=0.5.

The Hurst exponent and the spectral exponent are related by

$$b_{cum} = 2H + 1,$$

where b_{cum} is the cumulative spectral exponent of the sequence with spectral exponent *b*.

Beer (1994) used the *R/S* analysis for the characterization of hydrocarbon reservoirs. The spatial series data showed to be persistent, where the value of *H* varied from 0.76 to 0.94, depending on the physical parameter. He also calculated the fractal dimension of the series.

Leonardi and Kümpel (1998) used both the *R/S* analysis and spectral analysis to study the variability of well log data from two deep boreholes (one 4 *km* and the other 9 *km* deep) in the well known German KTB project. The results for the Hurst exponent ranged from 0.59 to 0.84.

Miranda and Andrade (1999) analyzed pluviometric data collected in some cities of the Brazilian Northeast. Depending on the site, the length of the collected data varied from 38 to 72 years. Using the *R/S* analysis and spectral analysis they calculated an average value of 0.709 for *H*.

Chierice (2003) analyzed a pluviometric time series from the town of Araras, in the State of São Paulo, in the period between 1955 and 2000. In order to calculate the Hurst exponent, that author used the wavelet method, besides the *R/S* analysis. The resulted for *H* was 0.743, and the fractal dimension was 1.254.

In the present work we use the *R/S* analysis and the spectral analysis to characterize several time series that represent climatic data from the town of Salvador, Bahia, Brazil, from 1961 to 1990. The time series that we used were precipitation, atmospheric pressure, evaporation, humidity, solar radiation, maximum, minimum and average temperatures.

In the next sections of this work we discuss the fundamentals of the *R/S* analysis, the properties of the Hurst exponent, and a review of spectral analysis. We applied both methodologies for the data sets above listed and in all cases the data series showed to be persistent, with the values for *H* varying between 0.602 and 0.848 for the *R/S* analysis, and varying between 0.525 and 0.908 for spectral analysis.

THE *R/S* METHOD

The re-scaled method (*R/S*) was developed by Hurst (1957), and now it is a classical method for the analysis of time series. Originally Hurst worked on problems related to dam constructions, and during a long time he studied the Nile river behavior for the construction of the Assuan dam. His goal was to determine the ideal volume for water storage in a reservoir, based on the water discharge with time.

The *R/S* was then proposed as an attempt to solve the finite reservoir problem submitted to a random input flow. An ideal reservoir must have the following properties: (i) the water volume must be constant; (ii) the water level must be constant after a given period; (iii) the lake can not be overloaded in order to avoid leakage; (iv) the storage capacity must be as small as possible. The conditions (i) and (ii) determine that the water volume released per year must be equal to the reservoir average input volume in the period of τ years.

If the input flow is represented by $\xi(t)$, the average input flow in the period of τ years is defined as:

$$<\xi>_\tau = \frac{1}{\tau}\sum_{t=1}^{\tau}\xi(t).$$

Since the input flow is variable, for each year there will be a difference between the average and the input flow. These differences can be summed and denoted by $X(t, \tau)$ which expression is

$$X(t,\tau) = \sum_{t=1}^{\tau}[\xi(t) - <\xi>_\tau].$$

The maximum and minimum values of $X(t, \tau)$ represent, respectively, the maximum and minimum values which flow through the reservoir during the considered period. The difference between the maximum and minimum values of $X(t, \tau)$ is called range *R*, which satisfies the four conditions above, and is expressed as

$$R(t, \tau) = \text{Max } X(t,\tau) - \text{Min } X(t,\tau).$$

Notice that $R(T)$ depends on the input flow $\xi(t)$, which by its turn depends on the considered period *T*. After many experiments Hurst realized that *R* depended on *T* according to a power law:

$$\frac{R}{S} = C\tau^H,$$

where *C* represents a constant and *H* is now called Hurst exponent. In order to compare phenomena with different origins and characteristics, Hurst concluded that was necessary to divide *R* by the standard deviation *S*, thus making *R/S* a non dimensional variable, where for each scale value *T* there is an associate value for *R/S*. The standard deviation *S* is expressed by

$$S(t,\tau)=\sqrt{\frac{\sum_{t=1}^{\tau}[\xi(t)-<\xi>_{\tau}]^2}{\tau}}.$$

This statistical technique is called Rescaled-Range analysis or *R/S* analysis. If we apply logarithm to the exponential equation we get

$$\log(R/S) = H\log(\tau) + \log(C).$$

If we make a graphic of log*(R/S)* X log*(T)*, we have *H*, which is the straight line angular inclination, and also brings information about the series under study: the Hurst exponent *H* informs about the persistence and the correlation. Correlation *C* is expressed by

$$C = 2^{(2H-1)} - 1.$$

There are three categories for the Hurst exponent:
(1) For $0.5 < H < 1$ the process is persistent, that is, a positive increment in the past increases the possibility of a positive increment in the future, or, a negative increment in the past increases the possibility of a negative increment in the future. The system is said to be positively correlated.
(2) For $0 < H < 0.5$ the process is anti-persistent, that is, a positive increment in the past increases the possibility of a negative increment in the future, or, a negative increment in the past increases the possibility of a positive increment in the future. The system is said to be negatively correlated.
(3) For $H = 0.5$ the process is non-correlated and the behavior is similar to the random movement (Brownian movement). The values of *H* for the cases (1) and (2) indicate that the system refers to a long memory process with a random component, where an event in the past has effect on an event in the future.

In order to understand better these cases we make use of the famous drunk example. If his steps are persistent, one step forward will be followed by another forward, in such a way that he will reach distances further from the origin. On the other hand, if his steps are anti-persistent, one step forward will be followed probably by a step backward, in such away that he will always be near the origin.

The exponent *H* brings information about the geometry of the profile under analysis. By using the expression

$$D = 2 - H,$$

one can compute the fractal dimension *D*. If the value of *H* is smaller the value of *D* will be greater.

SPECTRAL ANALYSIS

The time series are analyzed in the frequency domain through the Fourier transform. Consider that a given time series is represented by a generic time function *x(t)*. Then its Fourier transform is given by

$$\Im\{x(t)\} = X(f) = \int_{-\infty}^{+\infty} x(t)\exp(-i2\pi f)\ dt.$$

The available time series are not continuous functions but rather a set of discrete values. Thus we have to make use of the Discrete Fourier Transform, or from the computational point of view we can use the Fast Fourier Transform (Bracewell, 1986; Hsu, 1970).

The output of the Fourier transform applied to the time series is an amplitude distribution in relation to the associated frequencies. From the Fourier transform we can compute the amplitude spectrum for each parameter.

The amplitude spectrum of a given function *x(t)* is calculated from the real and imaginary parts of the function in the frequency domain *X(f)*,

$$A(f) = \sqrt{\mathrm{Re}[X(f)]^2 + \mathrm{Im}[X(f)]^2}.$$

The power spectrum is given by the amplitude spectrum to the square:

$$P(f) = [A(f)]^2.$$

In this work the spectral densities *P(f)* of the available data are computed through periodograms $P^*(f)$, which are commonly used to obtain the spectral densities. The periodogram provides the frequency decomposition in a given interval for the time series, as is defined as,

$$P^*(f) = \frac{1}{2\pi N_{\max}}\left|\sum_{n=1}^{t}[x(t)-\mu]\exp(-ift)\right|^2,$$

where $N_{\max}$ is the number of samples, *x(t)* is the time series, and μ is the series average. From the periodogram we can compute spectral exponent *b*, which is the angular coefficient of the power spectrum. Leonardi and Kümpel (1998), recommended a smoothing procedure in the power spectrum, and in the input, instead to use the time series coefficients, they proposed the deviations, that is, the difference between the coefficients and the mean μ, as can be seen in the above equation.

They also applied the Tukey-Hanning window which is a smoothing procedure in order to reduce the dispersion. One computes the average value using the anterior and posterior values, according to the equation

$$P(f) = \frac{1}{4}P^*(f-1) + \frac{1}{2}P^*(f) + \frac{1}{4}P^*(f+1).$$

The spectrum analysis of a given time series shows how the energy is distributed in a frequency range. For the white noise the energy is distributed in all frequencies, that it, the power spectrum can be approximated by a smooth curve with $b = 0$.

CLIMATIC DATASET

In this work we have used the data colleted by the Instituto Nacional de Metereologia (INMET). The data was measured at the Salvador Station which belong to the INMET's Fourth District. The Station is localized at the coordinates 13° 01' south latitude and 38° 31' west longitude. The height at the station is 51.41 *m*.

The analyzed data set was constituted of the following parameters: precipitation (*mm*), atmospheric pressure (*hPa*), evaporation (*mm*), humidity (%), solar radiation (W/m^2), as well as maximum, minimum and average temperatures (*°C*). All the data was sampled in monthly values, acquired over a period of 30 years, from 1961 to 1990, in a such way that the number of samples for each parameter is 360.

The precipitation, given in *mm*, can be seen in Figure 1. We can notice a periodic behavior, with relatively few peaks above 400 *mm*. The atmospheric pressure, given in *hPa*, which can be seen in Figure 2, presents an abnormal behavior for the first years, but then becomes periodic for the rest of the interval. The evaporation, given in *mm*, which can be seen in Figure 3, presents an abnormal behavior in two regions, as well as a very non-regular pattern for whole series. The humidity, given in percentage, and the solar radiation, given in W/m^2, can be seen, respectively, in Figure 4 and Figure 5. Both present a regular pattern of periodicity. The register of temperature, given in *°C*, can be seen in Figure 6 (maximum), in Figure 7 (minimum) and in Figure 8 (average). Notice a very regular pattern of periodicity, with very few peaks above 32 *°C* for the maximum temperature, and above 28 *°C* for the average temperature.

APPLICATION OF THE R/S METHOD

We used the climatic data above described for the R/S analysis, adapting the algorithm developed by Beer (1994). For each simulation, besides the main input data which is the time series, the following parameters are introduced: the numbers of observations *N*; the number of samples in each sequence *τ*; and the starting time *t*.

The utility of *t* is to break the series into sub-series, avoiding the sample superposition, and thus generating independent values of the *R/S* estimate. *N* was always 360, while we used 15 values for *τ*: 2, 3, 6, 12, 15, 18, 24, 30, 36, 40, 45, 60, 72, 90, 120 e 180. The following values were used for *t*: 1, 10, 20, 30, 40, 100, 200 e 300.

For the determination of the *H* exponent, we plotted the values of log *R/S* as a function of log τ, and then we performed a least squares adjustment for the calculation of the angular coefficient, which is the exponent *H*. Figure 9 shows the *R/S* analysis for precipitation, Figure 10 for atmospheric pressure, Figure 11 for evaporation, Figure 12 for humidity, Figure 13 for solar radiation, Figure 14 for maximum temperature, Figure 15 for minimum temperature, and Figure 16 for average temperature.

We compared our results with those obtained by other researchers, published in four different sources. They also used the *R/S* analysis either on time and spatial series. Chierice (2003) analyzed precipitation data from the city of Araras, for the period from 1955 to 2000, obtaining *H* = 0.743. Beer (1994) used different log data from boreholes and obtained *H* ranging from 0.790 to 0.946.

Miranda and Andrade (1999) used the *R/S* analysis to compute the Hurst exponent in precipitation time series in several cities from the Brazilian Northeast region. One of the cities was Salvador, and for that town they computed three values of *H*: 0.51, 0.48 and 0.53. In the present work we adapted the methodology proposed by Leonardi and Kümpel (1998) for spatial series. They used *R/S* analysis to study the persistence from well log data from to research deep boreholes in Germany. The values for *H* varied from 0.6 to 0.8.

Our values for *H* are presented at the next section (Table I).

APPLICATION OF SPECTRAL ANALYSIS

The total number of samples used in this work was 360 for a time span of 30 years. Since each sample is related to one month, the time interval is $\Delta t = 1/12$ year. The frequency interval Δf is

$$\Delta f = \frac{1}{N\Delta t},$$

where *N* is the number of samples and Δt is the time interval. Thus, for the present work we have,

$$\Delta f = \frac{1}{360 \cdot (1/12)} = 0.0333 \quad cicle/year,$$

in such a way the Nyquist frequency is,

$$f_N = \frac{1}{2\Delta t} = \frac{1}{2 \cdot (1/12)} = 6 \quad cicles.$$

The Nyquist frequency, which is also called folding frequency is the frequency which value is half of the sampling frequency. The frequencies which are higher than f_N are aliased, that is, are mixed with the lower frequencies, becoming not distinguishable, characterizing the ambiguity situation.

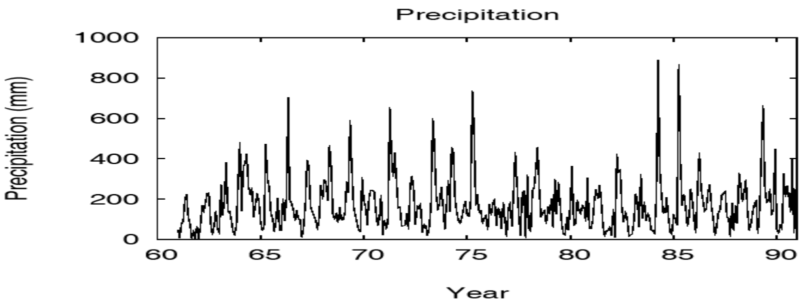

Figure 1. Precipitation (*mm*), measured at the Salvador Station, from 1961 to 1990.

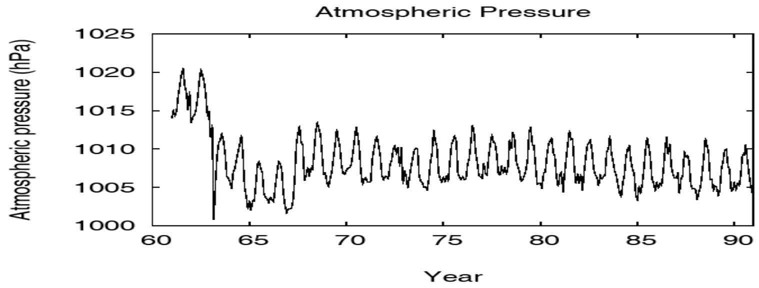

Figure 2. Atmospheric pressure (*hPa*), measured at the Salvador Station, from 1961 to 1990.

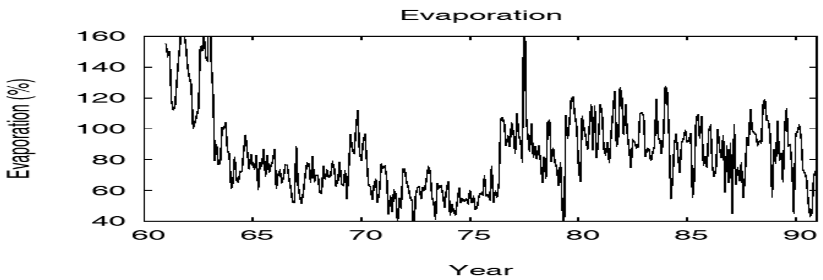

Figure 3. Evaporation (*mm*), measured at the Salvador Station, from 1961 to 1990.

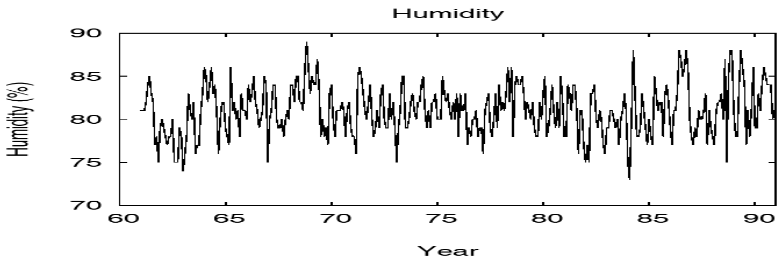

Figure 4. Humidity (%), measured at the Salvador Station, from 1961 to 1990.

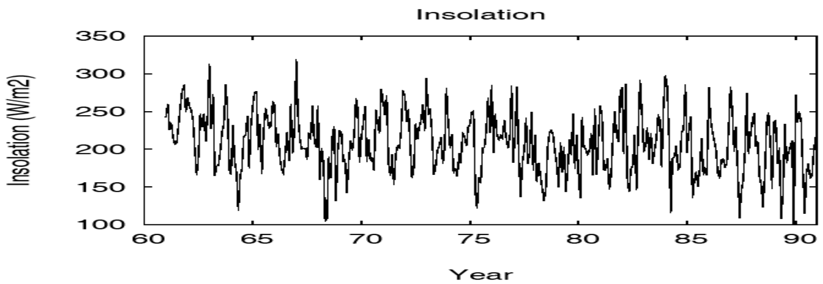

Figure 5. Solar radiation (*W/m*^{2}), measured at the Salvador Station, from 1961 to 1990.

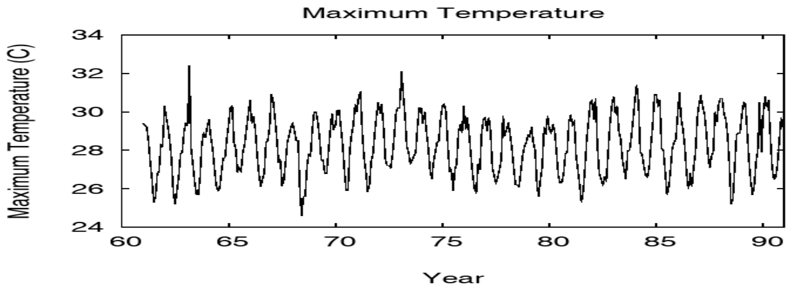

Figure 6. Maximum temperature (^{o}C), measured at the Salvador Station, from 1961 to 1990.

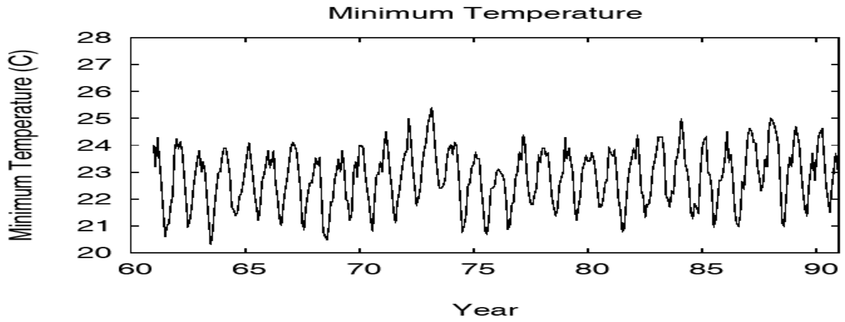

Figure 7. Minimum temperature (^{o}C), measured at the Salvador Station, from 1961 to 1990.

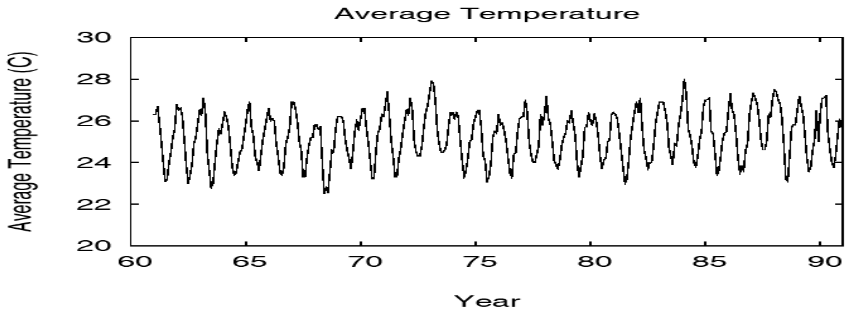

Figure 8. Average temperature (^{o}C), measured at the Salvador Station, from 1961 to 1990.

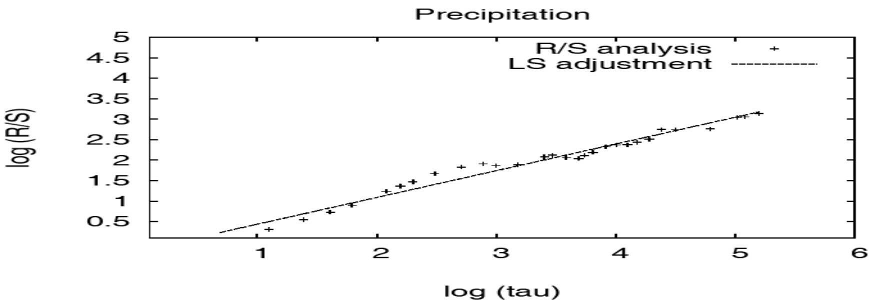

Figure 9. R/S analysis for precipitation.

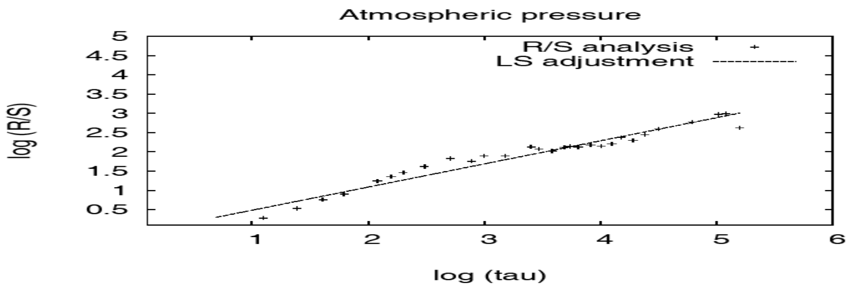

Figure 10. R/S analysis for atmospheric pressure.

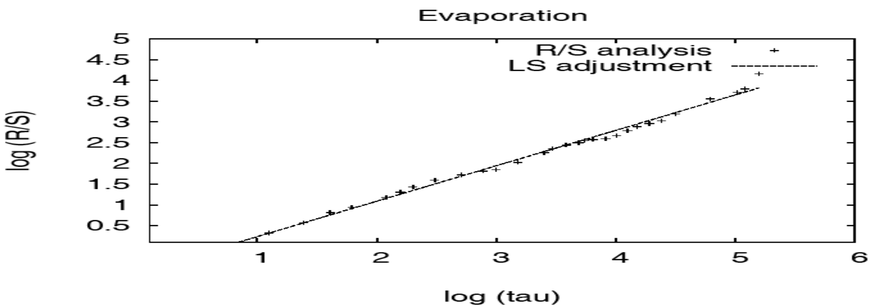

Figure 11. R/S analysis for evaporation.

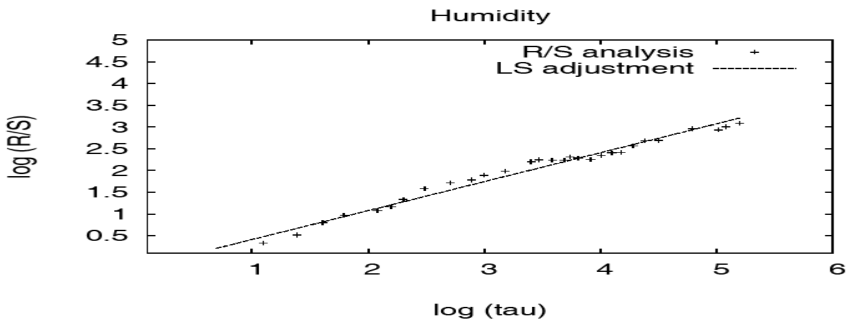

Figure 12. R/S analysis for humidity.

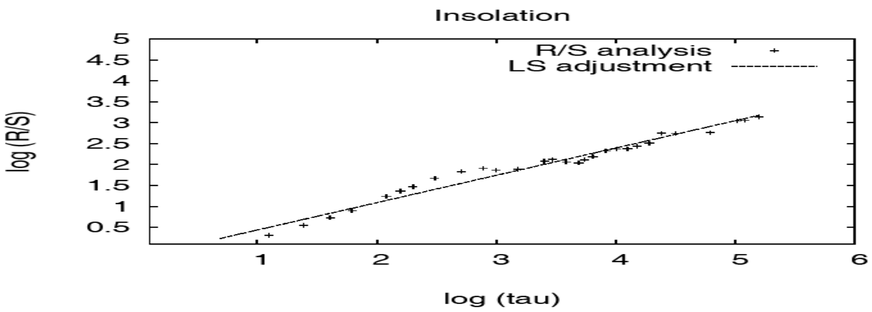

Figure 13. R/S analysis for solar radiation.

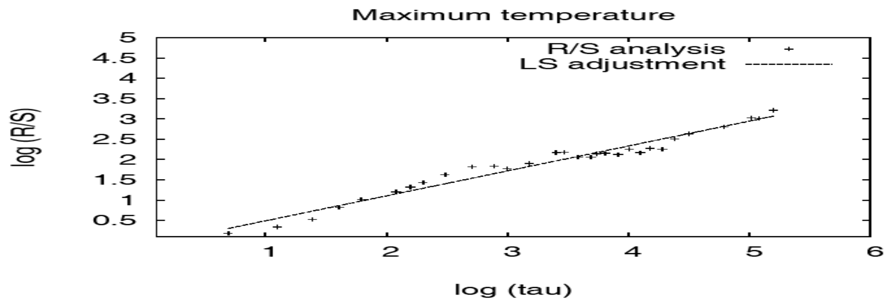

Figure 14. R/S analysis for maximum temperature.

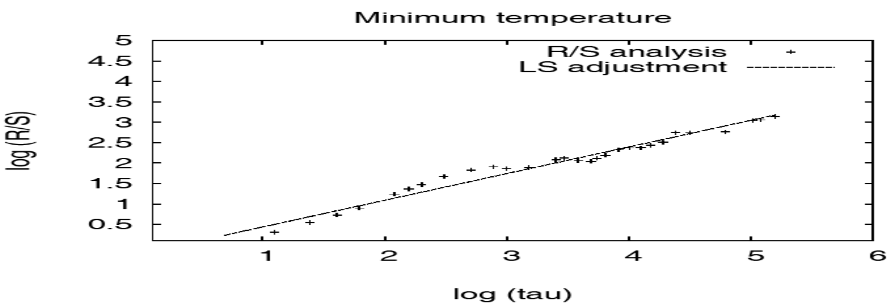

Figure 15. R/S analysis for minimum temperature.

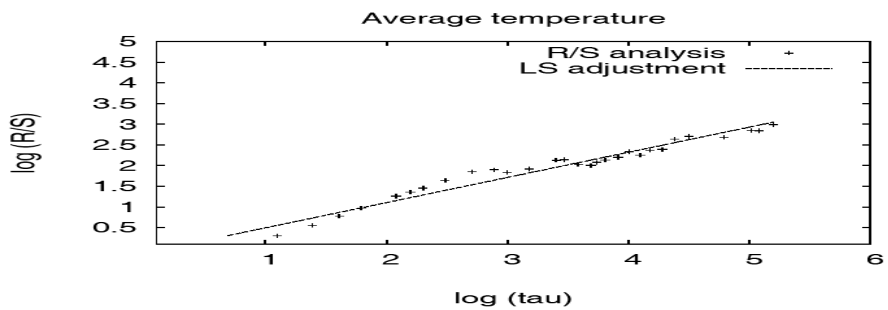

Figure 16. R/S analysis for average temperature.

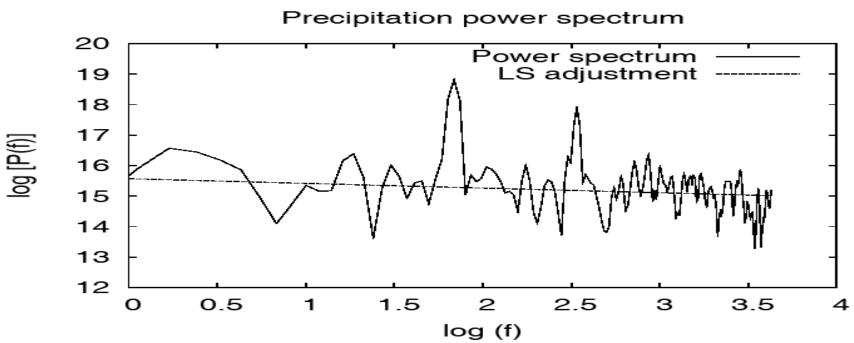

Figure 17. Power spectrum for precipitation.

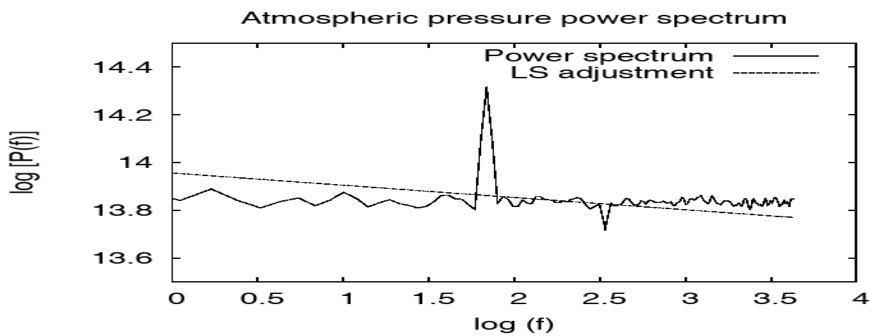

Figure 18. Power spectrum for atmospheric pressure.

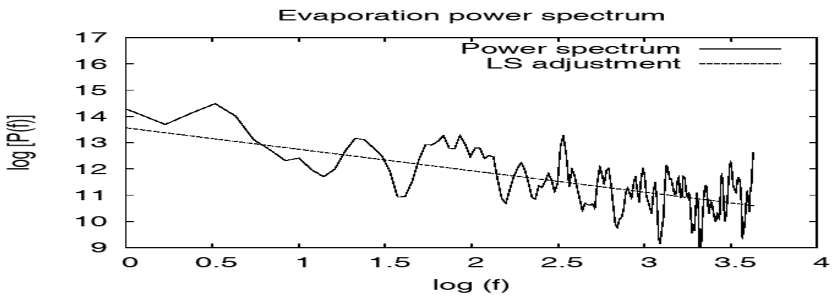

Figure 19. Power spectrum for evaporation.

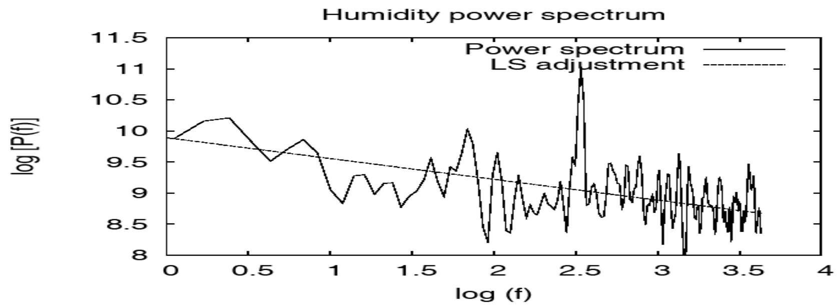

Figure 20. Power spectrum for humidity.

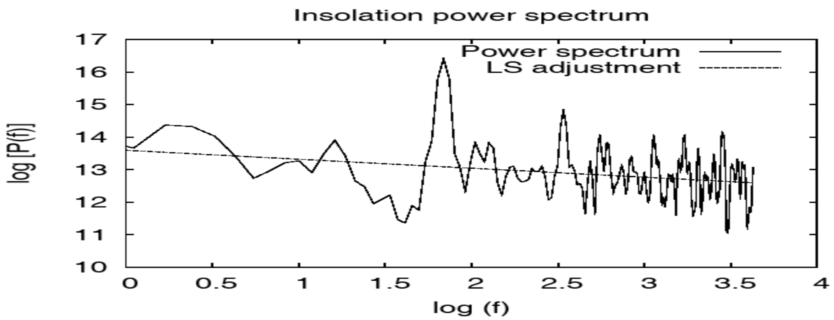

Figure 21. Power spectrum for solar radiation.

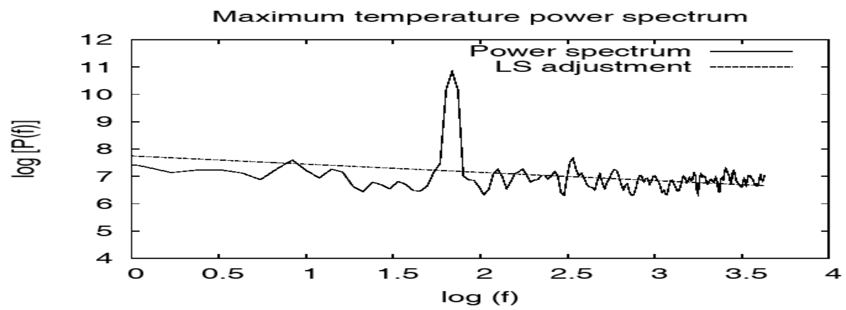

Figure 22. Power spectrum for maximum temperature.

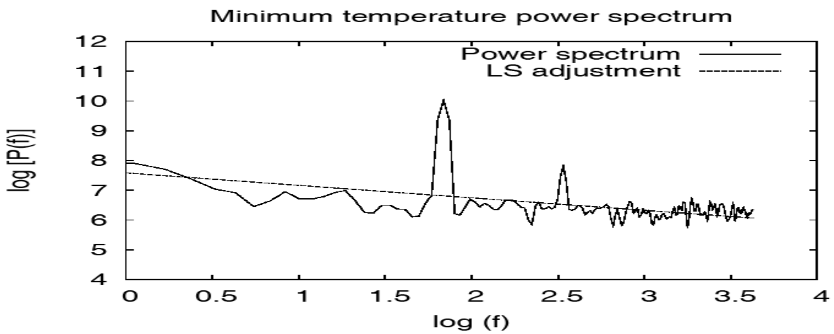

Figure 23. Power spectrum for minimum temperature.

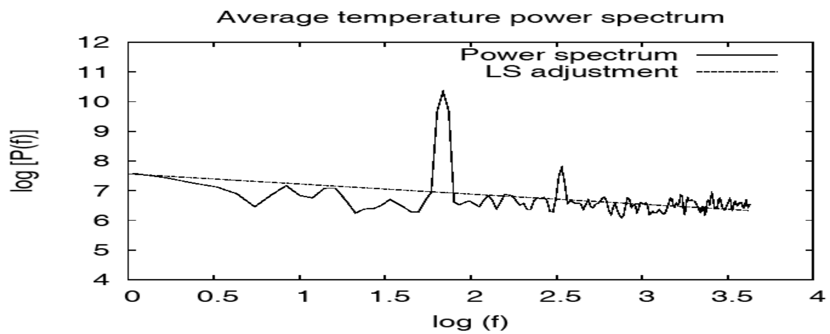

Figure 24. Power spectrum for average temperature.

Figure 17 shows the power spectrum analysis, as well as the adjustment by least squares for precipitation, Figure 18 for pressure, Figure 19 for evaporation, Figure 20 for humidity, Figure 21 for solar radiation, Figure 22 for maximum temperature, Figure 23 for minimum temperature, and Figure 24 for average temperature.
The relation between the spectral exponent *b* and the Hurst exponent is given by,

$$H = 2\, b_{cum} + 1.$$

Since the value of *b* accumulates with the series, we use a recurrence value b_{cum}, where $b = b_{cum} + 2$. The values for H can be seen in Table I.

Table I: *H* computed using R/S analysis and spectral analysis.

Parameter	*H* by R/S analysis	*H* by spectral analysis
Precipitation	0.618	0.576
Atmospheric pressure	0.602	0.525
Evaporation	0.869	0.908
Humidity	0.676	0.667
Solar radiation	0.655	0.637
Maximum temperature	0.617	0.650
Minimum temperature	0.658	0.709
Average temperature	0.625	0.670

CONCLUSIONS

The values of *H* in the studied times series, using the *R/S* analysis, ranged between 0.602 and 0.848. As mentioned before if the value of *H* is between 0.5 and 1.0, the series is said to be persistent. Thus, our data set which describes atmospheric phenomena is persistent, that is, the data are positively correlated, discarding the possibility of being a random distribution. When the Hurst exponent is different from 1, the series is said to be self-affine. Series with such characteristics, which is the case here, can correlate the fractal dimension and the *R/S* analysis as showed above. Fractal dimension is a useful tool to characterize atmospheric phenomena, and in this work the fractal dimension varied between 1.092 and 1.475. In this work the time interval was one month, however a different time interval would not change the fractal dimension, since it has a self-similar attribute. We also computed the Hurst exponent using spectral analysis. Due to the fluctuation in the power spectra we averaged and weighted the data, improving greatly the results. The values for *H* varied between 0.525 and 0.908. Thus we succeed to compare the values of *H* from the two methods. The two results for each time series are close, being the difference less than 10%. The only exception was the atmospheric pressure, where the difference was around 13%. The analysis of persistence of a given signal can be performed in a time series, like in the present work, or in a spatial series, for instance, in a well log. Our results agree with other works using climatic and geophysical data: many series in nature are persistent.

ACKNOWLEDGMENTS

Amin Bassrei would like to thank CNPq for the support to the project 308.875/2007-9 (research fellowship), CNPq and PETROBRAS for the support to the INCT – Geofísica do Petróleo program, and FINEP for the support to the Rede CT-PETRO de Geofísica de Exploração – Rede 01.

REFERENCES

BEER, R. Extensão da análise R/S para caracterização geoestatística local e regional de reservatórios de hidrocarbonetos. Dissertação (Mestrado) – Universidade Estadual de Campinas – UNICAMP, Campinas, 1994.

BRACEWELL, R. N. **The Fourier transform and its applications**. New York: McGraw-Hill, 1986.

CHIERICE, R. A. F. O **Uso de wavelets na determinação do expoente de Hurst de uma série temporal diária de chuvas do município de Araras - SP de 1955-2000.** Dissertação (Mestrado) – Universidade do Estado de São Paulo – UNESP, Rio Claro, 2003.

FEDER, J. **Fractals**. New York: Plenum Press, 1989.

HSU, H. P., **Fourier Analysis**. New York: Simon and Schuster, 1970.

HURST, H. E. A suggested statistical model of some time series which occur in nature. **Nature**, v. 180, p. 494-494, 1957.

LEONARDI, S.; KÜMPEL, H.-J. Variability of geophysical log data and the signature of crustal heterogeneities at the KTB. **Geophysical Journal International**, v. 135, p. 964-974, 1988.

MIRANDA, J. G. V. & ANDRADE, R. F. S. Rescaled range analysis of pluviometric records in northeast Brazil. **Theoretical and Applied Climatology**, v. 63, p. 79-88, 1999.