Is there memory in solar activity?

R. Oliver and J. L. Ballester
Departament de Física, Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain
(Received 18 May 1998)

The Hurst effect is a presumed and unexpected behavior of geoastrophysical time series by which these time series have persistence or ''memory.'' The application of Hurst analysis to monthly sunspot numbers [B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 321 (1969)] yielded a Hurst exponent $H=0.86 \pm 0.05$, suggesting that solar activity shows persistence and that the underlying responsible mechanism can guarantee a positive correlation of solar activity during long time lapses, raising, at the same time, the possibility of the existence of long-term memory in solar activity. More recently, radiocarbon data have been used for a similar study [A. Ruzmaikin, J. Feynmann, and P. Robinson, Sol. Phys. **149**, 395 (1994)] resulting in a constant value $H=0.84$ between 100 and 3000 years, which indicates persistence of solar activity in such time scales. Furthermore, Mount Wilson rotation measurements have also been analyzed in the same way [R. W. Komm, Sol. Phys. **156**, 17 (1996)] and the results indicate that temporal variations of solar rotation on time scales shorter than the 11-year cycle are caused by a stochastic process which is characterized by persistence. Here, we have followed the scale of fluctuation approach to show that there is no incontrovertible evidence for the presence of the Hurst effect in sunspot areas and, therefore, that there is no proof of the existence of long-term memory in solar activity. [S1063-651X(98)02311-3]

PACS number(s): 05.45.+b

I. INTRODUCTION

A time series can be characterized by means of the Hurst exponent (H) which reveals whether it shows persistence ($1>H>0.5$), i.e., positive correlation between the present values and those in the distant past; antipersistence ($0.5>H>0$), i.e., negative correlation; or comes from a random process ($H=0.5$) for which the correlation is zero. The Hurst effect can be defined as an anomalous behavior of the rescaled adjusted range, R_n^*, in a time series of record length n. In natural phenomena time series, Hurst [1] found the power relation

$$R_n^* = a n^H, \tag{1}$$

with $a=0.61$ and a mean value $H=0.72$, and thus claimed that natural time series show persistence. However, for independent identically distributed processes (fully random processes) the asymptotic value of H is 0.5. This discrepancy between the values of H for fully random processes and those obtained in geoastrophysical time series is known as the Hurst effect.

This result has been very much in debate since it is very difficult to understand what sort of physical mechanism can assure infinite memory, for instance, that the level of solar activity nowadays will be transmitted over decades and centuries. Mandelbrot and Van Ness [2] pointed out that $H \neq 0.5$ arises in a class of processes with infinite memory that they termed fractional Brownian noises (FBN's). A white noise arises from a Bachelier-Wiener process (Brownian motion) and, in a similar way, a FBN arises from a process (the fractional Brownian motion) in which each increment is a weighted average of all the past increments of a Bachelier-Wiener process. Although FBN's processes are operational, they are not physically founded models and did not arise as the result of the analysis of physical or dynamic properties of processes under study. Several hypotheses have been put forward in order to explain the Hurst effect: (a) The Hurst phenomenon is not a real effect, but a transitory behavior (preasymptotic behavior) produced by the slow convergence to a 0.5 exponent. This means that finite time series without persistence may give Hurst exponents larger than 0.5 [3]; (b) the Hurst phenomenon is due to nonstationarities in the underlying mean of the process [4]; (c) the Hurst phenomenon is a real one due to stationary processes with very large memory, i.e., there are processes in nature having infinite memory.

Solar activity is produced by the emergence of magnetic flux through the photosphere, forming active regions which include sunspots. However, although the behavior of the most characteristic feature of solar activity, namely the 11-year sunspot cycle, is sufficiently well known, the behavior of the nonperiodic component of solar activity is not so well understood. The question arises as to whether it can be characterized as a correlated process, in which persistence or memory is present, or as an uncorrelated random process, in which solar activity at any time is independent of previous history.

Mandelbrot and Wallis [5] used the Hurst analysis to study the behavior of monthly sunspot numbers between 1749 and 1948. Estimation of the Hurst exponent by means of a pox diagram yielded the value $H=0.86\pm 0.05$, which suggests that solar activity presents long-term persistence. Recently, radiocarbon data have been used for a similar study [6] resulting in a constant value $H=0.84$ between 100 and 3000 years, which indicates persistence of solar activity in such time scales. Also, Hurst analysis of Mount Wilson rotation measurements [7] seems to indicate that temporal variations of solar rotation on time scales shorter than the 11-year cycle are caused by a stochastic process which is characterized by persistence.

Here, we have tried to assess the presence of the Hurst effect in solar activity using a different procedure, called the scale of fluctuation approach, which has been applied to a very representative feature of solar activity such as sunspot areas, which provide us with an indication of the amount of magnetic flux emerging through the photosphere.

II. DATA AND METHODS

In this study, we have used daily sunspot areas between 1874 and 1993 grouped in four-week bins, which yields a time series with 1555 data values. Before performing the Hurst analysis, we have carried out a Cox-Stuart test, whose statistic indicates that the mean of the time series is not stationary. Then, to ensure the stationarity of the mean, we have fitted and subtracted from the time series a sinusoidal function with a period equal to that of the solar cycle, to remove the deterministic cycle, and a second-order polynomial, to remove the underlying long-term trend. A new application of the Cox-Stuart test reveals that, after the above process of detrending, the mean is now stationary.

The ''rescaled range'' analysis (or R_n^* analysis) was developed to study the problem of water storage and was described in detail by [8]. This statistical method was also used by [5] to study the long-run properties of various geophysical records, including sunspot numbers, and has been reviewed by [9]. Here, we follow them in our application of the method and refer to these works for a complete description of the analysis procedure.

Let x_i, $i=1,2,\ldots,N$, be an observed data series whose Hurst exponent is to be computed. In the hydrological context the x_i may be the annual water input into a dam or reservoir during N consecutive years. Let us now restrict ourselves to an n-year period starting at year t_0+1, that is, let us consider the data set x_i, $i=t_0+1,t_0+2,\ldots,t_0+n$, where $0 \leqslant t_0 \leqslant N-n$. We denote the average of this subset (i.e., the average water inflow into the reservoir over the n-year period) as $\overline{x(t_0,n)}$,

$$\overline{x(t_0,n)}=\frac{1}{n}\sum_{i=t_0+1}^{t_0+n} x_i . \tag{2}$$

In an ideal reservoir, designed so as to never overflow nor empty, $\overline{x(t_0,n)}$ also represents the optimum annual water release. In Eq. (2) and in what follows, t_0 and n in brackets are used to indicate a dependence on these two parameters.

Furthermore, the standard deviation of the x_i during the same period is estimated with the formula

$$S(t_0,n)=\left\{\frac{1}{n-1}\sum_{i=t_0+1}^{t_0+n}[x_i-\overline{x(t_0,n)}]^2\right\}^{1/2}. \tag{3}$$

This definition of the standard deviation (with the factor $n-1$ in the denominator instead of n) is usually considered so as to make it an *unbiased* estimator of the actual standard deviation of the time series [10].

Next, a new variable y_t, $t=1,2,\ldots,n$, is defined as follows:

$$y_t(t_0,n)=\sum_{i=t_0+1}^{t_0+t}[x_i-\overline{x(t_0,n)}]. \tag{4}$$

In this equation the difference $x_i-\overline{x(t_0,n)}$ is the *departure from the mean* of the influx in the ith year. Hence, a year in which the reservoir receives less water than is released yields a negative value of this quantity (the opposite happens when the water influx lies above the n-year average). The summation in Eq. (4) gives the *accumulated departure* from the mean (i.e., the net gain or loss of stored water) during the first t years of the period considered. The dimensions of the reservoir depend on the fluctuations in the accumulated departure and should be such that the reservoir never empties nor overflows. The storage capacity required to maintain the mean discharge over the n-year period is called the *range* (represented by R) and is equal to the difference between the maximum and the minimum accumulated departure over the n years. The range is defined by the formula

$$R(t_0,n)=\max_{1\leqslant t\leqslant n} y_t(t_0,n)-\min_{1\leqslant t\leqslant n} y_t(t_0,n). \tag{5}$$

The range so defined will take values on very different scales when different phenomena are studied. Therefore, it is convenient to substitute it by the *rescaled range*, equal to the range divided by the sample standard deviation,

$$R^*(t_0,n)=\frac{R(t_0,n)}{S(t_0,n)}. \tag{6}$$

Now one can consider the dependence of the rescaled range on the time lag n. However, there still remains one arbitrary parameter, t_0, which should be eliminated. To this end the values $t_0=0,n,2n,\ldots$ are selected so that the entire data set is divided into as many nonoverlapping n-year periods as can be constructed. For each of these subsets the rescaled range $R^*(t_0,n)$ is computed as outlined above and the rescaled range, R_n^*, for the time lag n is finally defined as the average of those values,

$$R_n^*=\frac{1}{n_{t_0}}\sum_{t_0} R^*(t_0,n), \tag{7}$$

where n_{t_0} is the integer part of N/n and is the number of values for t_0 used.

To determine the value of H for a time series, the rescaled range R_n^* is computed and the results are presented in a ''pox diagram'' (in which the logarithm of the rescaled range is plotted versus the logarithm of the time lag). The Hurst exponent is then given by the slope of a straight line fitted to the points in the pox diagram. However, not all points in the diagram should be given the same weight [5]. When the lag n is small compared to the length of the time series, a large number of independent estimations of R_n^* can be calculated. They have a considerable scatter so that their average could be meaningless. On the other hand, the opposite happens for values of n close to the total number of data: their average has little statistical significance because only one or a few estimations of R_n^* are available. Then, very small or very large values of the lag n must not be considered in the determination of the Hurst exponent.

A more precise definition of the Hurst effect was introduced by [11], which stated that a sequence of random variables exhibits the Hurst effect with $H>0.5$ if $n^{-0.5}R_n^*$ converges in distribution, as n goes to infinity, to a nonzero random variable. This is in contrast with the functional central limit theorem [12,13], which states that, for processes belonging to the Brownian domain of attraction, the expected value and variance are [14]

$$\hat{\mu}=E(n^{-0.5}R_n^*)=(\theta\pi/2)^{0.5}, \tag{8}$$

$$\hat{\sigma}^2=\mathrm{var}(n^{-0.5}R_n^*)=\theta(\pi^2/6-\pi/2), \tag{9}$$

respectively, θ being the scale of fluctuation, or correlation length scale, which was first proposed by [15]. Then, convergence of sample values of $n^{-0.5}R_n^*$ into the asymptotic interval given by $\hat{\mu}\pm 2\hat{\sigma}$ allows us to accept the hypothesis of nonexistence of the Hurst effect in a time series. Such a test can be improved if an independent estimate of the limit of the sequence $n^{-0.5}R_n^*$ is known, so the only extra parameter needed is the scale of fluctuation. The scale of fluctuation of a stationary random series can be defined as

$$\theta=\lim_{n\to\infty} n\gamma(n), \tag{10}$$

n being the time lag and $\gamma(n)$ the variance function. There are several methods to estimate the scale of fluctuation from a record of finite length $x(t)$, with $0\leq t\leq T_0$, which is a representation of a stationary random process $X(t)$. Among them, and for the sake of simplicity, the approach chosen here has been to obtain consistent estimators of θ by using the variance function $\gamma(n)$ [16]. The usual definition of the variance function for the above $X(t)$ is

$$\gamma(n)=\frac{\sigma_n^2}{\sigma^2},$$

which measures the reduction of the point variance σ^2 under local averaging. However, given the correlation function, $\rho(\tau)$, the variance function for a random series can be obtained as follows:

$$\gamma(n)=\frac{1}{n}+\frac{2}{n}\sum_{\tau=1}^{n}\left(1-\frac{\tau}{n}\right)\rho(\tau). \tag{11}$$

Now, to obtain θ it is necessary to adopt an approximate model for the variance function. For wide-band processes such as the present one, there is a family of analytical models for the variance function described as [16]

$$\gamma(n)=\left[1+\left(\frac{n}{\theta}\right)^m\right]^{-1/m}. \tag{12}$$

Vanmarcke [16] discussed the values of m which can be used in Eq. (12), pointing out that any $m>0$ is acceptable. In our case, varying m yields similar results for the behavior of the estimated scale of fluctuation, so $m=1$ has been chosen, i.e.,

$$\gamma(n)=\frac{\theta}{\theta+n}. \tag{13}$$

However, the variance function estimator is biased and we need an unbiased estimator to estimate θ. Then, following Vanmarcke [16], a corrected estimate of the variance function, $\hat{\gamma}_c$, is given by

$$\hat{\gamma}_c(n)=\gamma(T_0)+\hat{\gamma}(n)[1-\gamma(T_0)], \tag{14}$$

T_0 being the time series length and $\hat{\gamma}(n)$ the estimated variance function. Now, we can use Eq. (13) to express both $\gamma(T_0)$ and $\hat{\gamma}_c(n)$ in terms of $\hat{\theta}$, which can then be substituted in Eq. (14) to obtain the following explicit solution for the estimated scale of fluctuation:

$$\hat{\theta}=\frac{\hat{\gamma}(n)T_0 n}{T_0[1-\hat{\gamma}(n)]-n}. \tag{15}$$

In fact, the estimation of $\hat{\theta}$ is by itself a test of the existence of the Hurst effect for stationary processes. Due to the functional central limit theorem, if $\hat{\theta}$ is finite the exponent is $H=0.5$ and there is no Hurst effect. On the contrary, when $\hat{\theta}$ is infinite there is a Hurst effect.

In the case of fully random data, for which $\rho(\tau)=0$ when $\tau>0$, from Eqs. (11) and (12) we obtain

$$\gamma(n)=\frac{1}{n} \tag{16}$$

and

$$\hat{\theta}=\frac{n}{n-1}, \tag{17}$$

which gives $\hat{\theta}=1$ for $n\to\infty$. In addition, the pox diagram yields the value 0.5 for the Hurst exponent.

Then, the procedure to follow in order to obtain the value of $\hat{\theta}$ is to calculate the correlation function $\rho(\tau)$ from the detrended time series, then use Eq. (11) to get $\gamma(n)$, and take this value as $\hat{\gamma}(n)$ in Eq. (15) to compute $\hat{\theta}$. Once $\hat{\theta}$ has been obtained, the interval $\hat{\mu}\pm 2\hat{\sigma}$, to which $n^{-0.5}R_n^*$ should con-

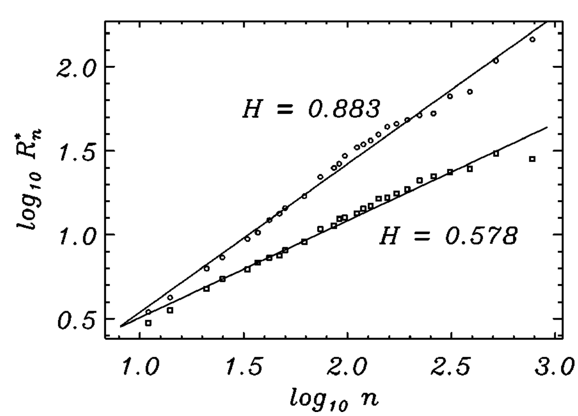

FIG. 1. Pox diagram for sunspot areas (○) and Gaussian random data (□). The Hurst exponent for sunspot areas is 0.883 while for Gaussian random data it is 0.578. R_n^* and n are given in units of 28 days.

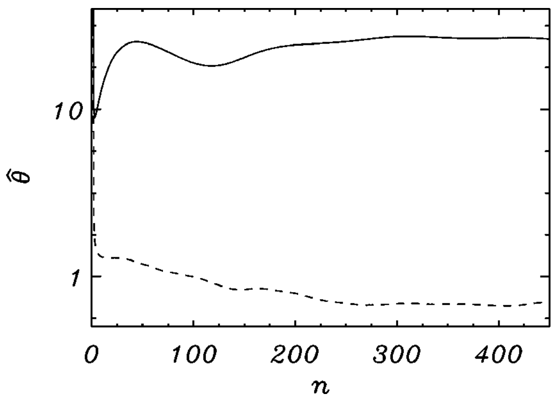

FIG. 2. The scale of fluctuation ($\hat{\theta}$) versus time lag (n) for sunspot areas (solid) and Gaussian random data (dashed). Convergence of $\hat{\theta}$ towards the values 25 (sunspot areas) and 0.8 (Gaussian random data) can be observed. The length of both data series is $T_0 = 1555$. n and $\hat{\theta}$ are given in units of 28 days.

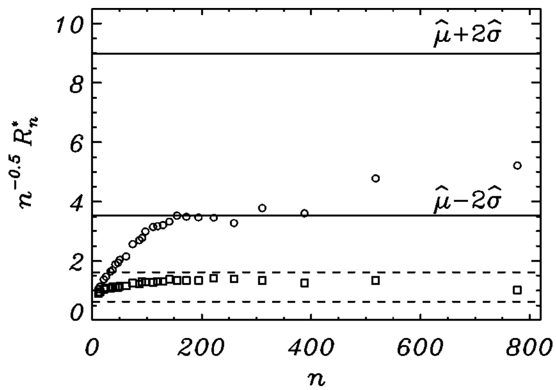

FIG. 3. GEOS diagram for sunspot areas (○) and Gaussian random data (□). Asymptotic convergence into the interval $\hat{\mu} \pm 2\hat{\sigma}$ (indicated by two solid lines for the sunspot data and two dashed lines for the Gaussian random data) points out that the presence of the Hurst effect in sunspot areas is doubtful. R_n^* and n are given in units of 28 days.

verge, must be determined. The reality of this convergence limit can be checked by plotting $n^{-0.5}R_n^*$ versus n (called the GEOS, geophysical record, diagram in [17]).

III. RESULTS AND CONCLUSIONS

Figure 1 shows the pox diagram for both Gaussian random data and solar activity data. While for random data it yields the value $H = 0.578$, close to 0.5 as expected, for sunspot areas the value of H is 0.883, similar to the one obtained by [5], which would erroneously indicate the presence of the Hurst effect. Figure 2 displays a plot of the estimated scale of fluctuation versus time lag for sunspot areas and Gaussian random data, from which the value of $\hat{\theta}$ for both data sets can be estimated. For sunspot areas, $\hat{\theta}$ stabilizes around 25, so that the convergence interval of $n^{-0.5}R_n^*$ is 6.26 ± 2.72, while for Gaussian random data $\hat{\theta}$ tends to 0.8 (fully random data have $\hat{\theta} = 1$) and the convergence interval is 1.12 ± 0.49. The stabilization of $\hat{\theta}$ around a finite value would indicate that there is no memory in these time series.

To check this conclusion, we have plotted the GEOS diagram for the same data sets (Fig. 3), showing an apparent convergence of $n^{-0.5}R_n^*$ within the intervals determined above. However, the limited length of the sunspot data series does not allow us to discard the possibility of the existence of the Hurst effect in the data. A much longer data set could help determine whether the circles in Fig. 3 converge into the asymptotic interval (i.e., there is no Hurst effect) or else diverge as $n \to \infty$ (i.e., the data show the Hurst effect).

In summary, we have used the scale of fluctuation approach to investigate the presence of long-term memory in sunspot areas time series, finding that the estimated scale of fluctuation converges to a finite value which would mean that the time series satisfies the functional central limit theorem (convergence within an interval), in contrast with what should be expected for processes with infinite memory (divergence) for which the scale of fluctuation increases in a continuous way. Since our sample record has a stationary mean, the discrepancy with the pox diagram value $H \approx 0.9$ obtained from the same data could be understood in terms of a slow convergence to $H = 0.5$ (preasymptotic behavior), due to an insufficient length of the record.

We can conclude that there is no incontrovertible evidence for the presence of the Hurst effect in the nonperiodic component of sunspot areas time series. This would mean that, contrary to earlier published research works, it is not possible to make any strong statement about the presence or not of temporal memory in such time series and that, for this reason, one cannot assure whether or not the present solar activity level will influence solar activity in the distant future. The present analysis can be extended to other astrophysical time series, in order to assess the reality of the Hurst effect in astrophysical phenomena.

[1] H. E. Hurst, Trans. Am. Soc. Civ. Eng. **116**, 770 (1951).
[2] B. B. Mandelbrot and J. W. Van Ness, SIAM (Soc. Ind. Appl. Math.) Rev. **4** (10), 422 (1968).
[3] J. D. Salas, D. C. Boes, V. Yevjevich, and G. G. S. Pegram, J. Hydrol. **44**, 1 (1979).
[4] V. Klemes, Water Resour. Res. **10**, 675 (1974).
[5] B. B. Mandelbrot and J. R. Wallis, Water Resour. Res. **5**, 321 (1969).
[6] A. Ruzmaikin, J. Feynmann, and P. Robinson, Sol. Phys. **149**, 395 (1994).
[7] R. W. Komm, Sol. Phys. **156**, 17 (1996).
[8] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-Term Storage: An Experimental Study* (Constable, London, 1965).
[9] J. Feder, *Fractals* (Plenum Press, New York, 1988).
[10] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (J. Wiley and Sons, New York, 1986).
[11] R. N. Bhattacharya, V. K. Gupta, and E. Waymire, J. Appl. Probab. **20**, 649 (1983).

[12] P. Billingsley, *Convergence of Probability Measures* (J. Wiley, New York, 1968).
[13] W. Feller, *An Introduction to Probability Theory and Its Applications* (J. Wiley, New York, 1971).
[14] W. Feller, Ann. Math. Stat. **22**, 427 (1951).
[15] G. I. Taylor, Proc. London Math. Soc. **2**, 196 (1921).
[16] E. Vanmarcke, *Random Fields: Analysis and Synthesis* (The MIT Press, Boston, 1988).
[17] O. J. Mesa and G. Poveda, Water Resour. Res. **29**, 3995 (1993).