# RESCALED RANGE ANALYSIS OF THE ASYMMETRY OF SOLAR ACTIVITY

R. OLIVER and J. L. BALLESTER
*Departament de Física, Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain*

(Received 16 July, 1995; in final form 19 July, 1996)

**Abstract.** Previous studies of the north–south asymmetry of solar activity (e.g., Carbonell, Oliver, and Ballester, 1993; Oliver and Ballester, 1994) suggest that the asymmetry time series can be represented by means of a multicomponent model made up of a long-term trend, a weak sinusoidal component (with a period close to 12.1 years) and a dominant random process. Here, we have used the rescaled range analysis to study the variation of the stochastic component of the asymmetry. To avoid the influence of the trend and the sinusoidal component on the result, we have removed both from the original time series. The value obtained for the Hurst exponent ($0.717 \pm 0.002$) suggests that the non-periodic component is a correlated random process.

## 1. Introduction

North–south asymmetry (N–S asymmetry) in solar activity as shown by indices such as sunspot areas (Carbonell, Oliver, and Ballester, 1993; Oliver and Ballester, 1994) or frequency of solar flares (Roy, 1977; Garcia, 1990) is well known.

Carbonell, Oliver, and Ballester (1993) and Oliver and Ballester (1994) reached several interesting results by a statistical study of the north–south asymmetry of sunspot areas between 1874 and 1993. The low-frequency component in the power spectrum of the asymmetry time series suggests the presence of a long-term trend, which was confirmed by a Cox–Stuart test. This long-term trend seems to account for the slow shift from south dominant asymmetry, at the end of the last century and the beginning of the present one, to north dominant during most of this century. Also, the time series contains a deterministic cycle, with a period of 12.1 years, which is clearly revealed in the power spectrum. The remaining component of the time series is practically featureless. A 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 the value of the asymmetry at any time is independent of all previous values of the asymmetry).

On the other hand, the Wold decomposition theorem (e.g., Gottman, 1981) states that any discrete stationary process can be expressed as the sum of two uncorrelated processes, a purely deterministic one and a stochastic one. This means that, in the N–S asymmetry, the trend and the sinusoidal component represent the deterministic process while the stochastic one is yet to be found. Carbonell, Oliver, and Ballester (1993) removed the trend and the sinusoidal component from the N–S asymmetry, and then fitted autoregressive (AR), moving average (MA), and autoregressive-moving average (ARMA) models to the remaining signal, but could

*Solar Physics* **169:** 215–224, 1996.

not find a definite evidence of the stochastic process in it. They concluded that the stochastic process is probably pure white noise that accounts for most of the time series variance.

In order to check their conclusion, we here make use of the rescaled range analysis, which allows us to determine whether the stochastic fluctuations of a signal are caused by a white noise process or not. This test is performed by computing the so-called Hurst exponent (e.g., Mandelbrot and Wallis, 1969a; Feder, 1988), whose value is equal to 0.5 when the analyzed signal contains a non-periodic variation generated by an uncorrelated random process, and different from 0.5 when the signal shows persistence.

When computing the Hurst exponent one has to bear in mind that long-term trends or periodic components can make it difficult to calculate its correct value (Mandelbrot and Wallis, 1969b; Bhattacharya, Gupta, and Waymire, 1983). To avoid this difficulty, we have removed both the trend and the sinusoidal component from the original time series.

About the application of the rescaled range analysis to solar activity, Mandelbrot and Wallis (1969a) calculated the Hurst exponent for sunspot numbers. Recently, Ruzmaikin, Feynman, and Robinson (1994) made a similar study of solar activity using radiocarbon data. For this record they found a constant Hurst exponent $H = 0.84$ between 100 and 3000 years, which indicates persistence of solar activity in such time scales. Also, Komm (1995) analyzed Mt. Wilson rotation measurements and found 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.

## 2. Computation of the Hurst Exponent of a Time Series

The 'rescaled range' analysis (or R/S analysis) was developed to study the problem of water storage and was described in detail by Hurst, Black, and Simaika (1965). This statistical method was used by Mandelbrot and Wallis (1969a) to study the long-run properties of various geophysical records, including sunspot numbers, and has been also reviewed by Feder (1988). Here, we follow them in our application of the method and refer to their 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 a $s$-year period starting at year $t_0 + 1$, that is, let us consider the data set $x_i$, $i = t_0 + 1, t_0 + 2, \ldots, t_0 + s$, where $0 \leq t_0 < N - s$. We denote the average of this subset (i.e., the average water inflow into the reservoir over the $s$-year period) as $\overline{x}(t_0, s)$,

$$\bar{x}(t_0, s) = \frac{1}{s} \sum_{i=t_0+1}^{t_0+s} x_i \,. \tag{1}$$

In an ideal reservoir, designed so as to never overflow nor empty, $\bar{x}(t_0, s)$ also represents the optimum annual water release. In Equation (1) and in what follows, $t_0$ and $s$ 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

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

This definition of the standard deviation (with the factor $s - 1$ in the denominator instead of $s$) is usually considered so as to make it an unbiased estimator of the actual standard deviation of the time series (Bendat and Piersol, 1986; Press *et al.*, 1988).

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

$$y_t(t_0, s) = \sum_{i=t_0+1}^{t_0+t} [x_i - \bar{x}(t_0, s)] \,. \tag{3}$$

In this equation the difference $x_i - \bar{x}(t_0, s)$ is the *departure from the mean* of the influx in the $i$th 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 $s$-year average.) The summation in Equation (3) 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 $s$-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 $s$ years. The range is defined by the formula

$$R(t_0, s) = \max_{1 \le t < s} y_t(t_0, s) - \min_{1 < t \le s} y_t(t_0, s) \,. \tag{4}$$

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,

$$(R/S)(t_0, s) = \frac{R(t_0, s)}{S(t_0, s)} \,. \tag{5}$$

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

$$(R/S)(s) = \frac{1}{n_{t_0}} \sum_{t_0} (R/S)(t_0, s) \,. \tag{6}$$

where $n_{t_0}$ is the integer part of $N/s$ and is the number of values for $t_0$ used. From now on the subscript $s$ in $(R/S)(s)$ will be dropped and the rescaled range will be represented by $R/S$.

If the ratio $R/S$ is proportional to $s^H$, we can obtain $H$, called the Hurst exponent. The Hurst exponent is 0.5 for random noise. For $0 < H < 0.5$, the time series is less correlated than random noise. For $0.5 < H < 1.0$, the time series is more correlated than random noise and is called persistent.

To determine the value of $H$ for a time series, the rescaled range $R/S$ 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 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 (cf., Mandelbrot and Wallis, 1969a). When the lag $s$ is small compared to the length of the time series, a large number of independent estimations of $R/S$ 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 $s$ close to the total amount of data. The average has little statistical significance because only one or a few estimations of $R/S$ are available. Then, very small or very large values of the lag $s$ must not be considered in the determination of the Hurst exponent.

Mesa and Poveda (1993) made a strong case against the use of the pox diagram alone in the determination of $H$ and suggested using other estimation tools. One of these tools relies on the examination of the so-called 'GEOS' diagrams, with $s$ on the abscissa and $(R/S)/s^H$, $0.5 \leq H \leq 1$, on the ordinate. Mesa and Poveda were concerned about the goodness of previous estimations of the Hurst exponent for geophysical time series, which they claimed were erroneously believed to have $H \neq 0.5$. If the Hurst exponent is larger than 0.5, then $(R/S)/s^{0.5}$ will diverge to infinity for large values of $s$. On the contrary, if $H = 0.5$ the points in the GEOS diagram will converge towards a finite value for large $s$. Moreover, it is possible to gain some information on the actual value of the Hurst exponent by plotting a GEOS diagram with $H > 0.5$. A time series with Hurst exponent 0.5 will converge to zero for large $s$, while a time series with Hurst exponent $H$ will converge to a non-zero limit.

## 3. Rescaled Range Analysis of the N–S Asymmetry of Sunspot Areas

The World Data Center of the National Oceanic and Atmospheric Administration provided us with the daily sunspot area in the north and south solar hemispheres, taken from the Greenwich Photoheliographic Results (for the period: January 1874–December 1982) and the United States Air Force (for the period: January 1983–July 1993). These databases total 43558 daily values from which the asymmetry has been computed by means of

$$AS = \frac{N - S}{N + S} , \tag{7}$$

where $N$ and $S$ stand for the total sunspot area in the north and south hemispheres.

After computing the sunspot area asymmetry with the above formula, the long-term trend and the main periodic component in the time series have been removed by subtracting a cubic polynomial and a sinusoidal component with a period of 12.1 years (see Carbonell, Oliver, and Ballester, 1993, for more details). The cubic trend is of the form

$$a_0 - a_1 t + a_2 t^2 + a_3 t^3 , \tag{8}$$

with $t = 1, 2, \ldots, 43558$ the time index and

$$\begin{aligned} a_0 &= 6.90 \times 10^{-3} , \quad a_1 = -2.85 \times 10^{-5} , \\ a_2 &= 2.31 \times 10^{-9} , \quad a_3 = -4.00 \times 10^{-14} . \end{aligned} \tag{9}$$

Moreover, the equation of the sinusoidal component is

$$b \cos \frac{2\pi}{T}(t - 1) + c \sin \frac{2\pi}{T}(t - 1) , \tag{10}$$

where $T = 4434$ days is the period and

$$b = -1.24 \times 10^{-1} , \qquad c = -1.50 \times 10^{-3} . \tag{11}$$

Once the subtractions have been performed, the power spectrum shows that the power at very low frequencies ($< 0.25 \times 10^{-4}$ day$^{-1}$) has decreased by a factor $10^3$ and that the peak at 12.1 years has been eliminated (compare Figure 1(b) with Figure 1(a)). The remaining power corresponds to the stochastic component, which is the one used in our study. To estimate the importance of the three components, the power of each of them in the periodogram has been computed. The cubic trend, the deterministic cycle, and the stochastic component account for 17%, 16% and 67% of the total power in the asymmetry time series.

In the analysis we have used lags between 21 and 10 889 days and have then plotted the values of $R/S$ in a pox diagram (Figure 2). The Hurst exponent has

been computed by discarding small and large values of $s$. Therefore, a straight line has been fitted to the points in Figure 2 with $65 \leq s \leq 3111$, for which the value $H = 0.717 \pm 0.002$ is obtained. The uncertainty in $H$ has been estimated following Press *et al.* (1988) for the case in which the errors associated with the individual points are not known.

The Hurst exponent has been calculated for several different ranges of time lags. Owing to the fine alignment of the points in Figure 2 along a straight line, in all cases the value of H differs from $H = 0.717$ in the last decimal place. Moreover, $H$ has also been computed for weekly and monthly values and a slight increase of the Hurst exponent with the bin size has been found. This can be understood by taking into account that the power in the high-frequency range is reduced as the bin size is increased. According to Feder (1991), $H = (\beta + 1)/2$, with $S \sim f^{-\beta}$ the power spectrum and $f$ the frequency, so the slight increase of $H$ is a consequence of the increase in the bin size.

Figures 3(a) and 3(b) show the GEOS diagrams for $H = 0.5$ and $H = 0.717$. In Figure 3(a) an increasing trend of $(R/S)/s^{0.5}$ for large values of $s$ should be interpreted as an indication that the Hurst exponent of the series is larger than 0.5. On the other hand, Figure 3(b) shows convergence of $(R/S)/s^{0.717}$ towards a finite value, which confirms the value of $H$ found before.

## 4. Conclusions

In this paper, the rescaled range analysis has been applied to the asymmetry of sunspot areas to ascertain whether the variations of its continuum component are due to the presence of a white noise process or not. To this end, the long-term trend and the 12.1-year periodic component have been removed from the asymmetry time series and the Hurst exponent of the remaining signal has been computed. The value $H = 0.717 \pm 0.002$ indicates that the stochastic component of the asymmetry shows persistence and that it must be due to a random but correlated process yet to be found.

From the previous results (e.g., Carbonell, Oliver, and Ballester, 1993; Oliver and Ballester, 1994) and the present one, we can conclude that the north–south asymmetry of sunspot areas can be represented by means of a multicomponent model made up of a long-term trend, a sinusoidal component with a period close to 12.1 years, and a dominant non-white noise component.

The notion of *persistence* in a signal has to be understood as the presence of fairly long time lapses with values mostly above or mostly below the average. In contrast, a white-noise process presents disordered fluctuations around the average. If solar activity is a manifestation of the emergence of magnetic flux through the photosphere, as it is usually assumed, the existence of persistent behaviour in the asymmetry leads to the conclusion that in the past there have been time lapses in which the transport of magnetic flux in the convection zone towards the surface has

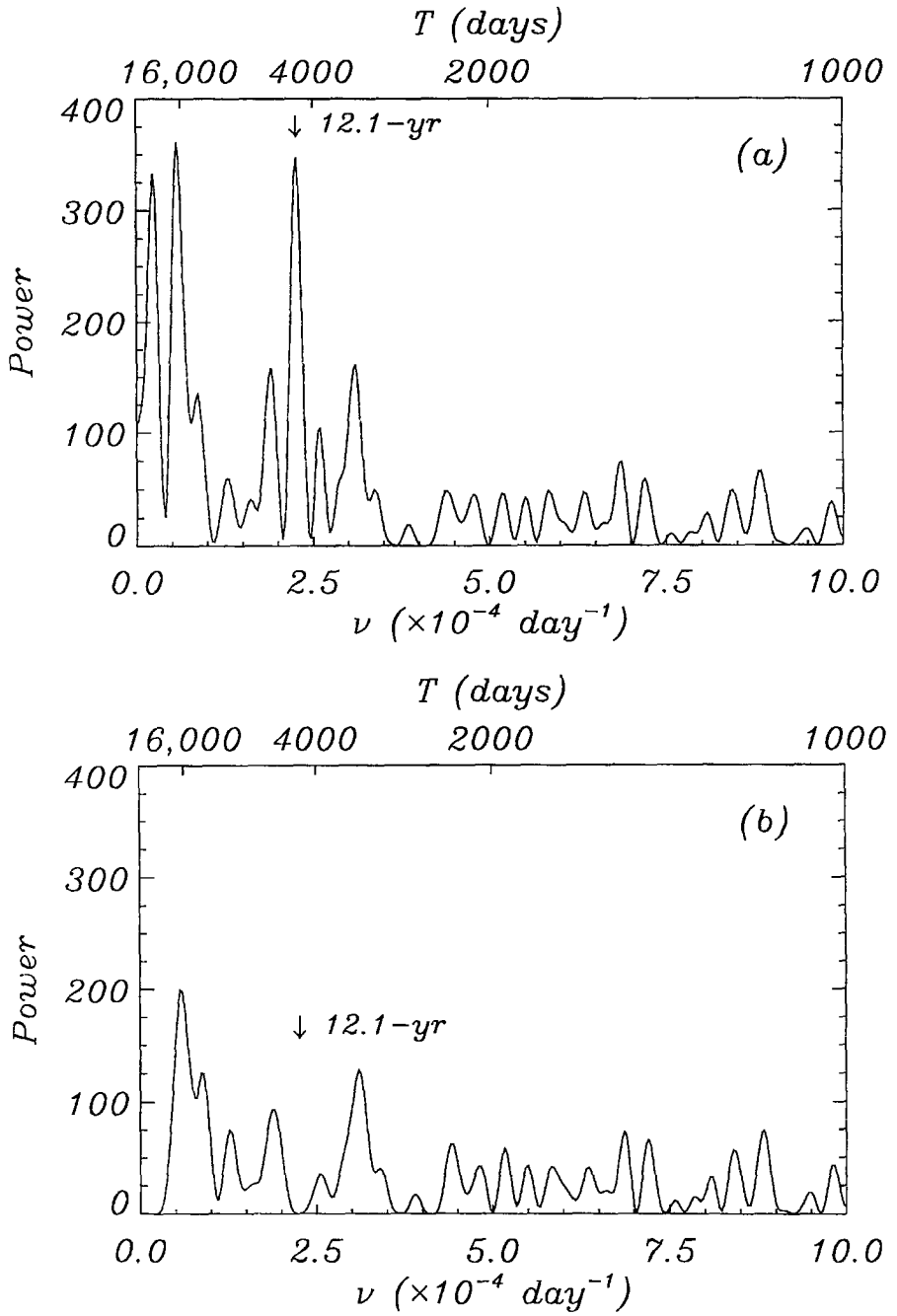


*Figure 1.* (a) Power spectrum of the sunspot area asymmetry between January 1874 and July 1993. (b) Power spectrum of the sunspot area asymmetry without the cubic trend and the 12.1-year period. The effect of the long-term trend removal is to reduce the power at low frequencies, while the subtraction of the periodic signal reveals itself in the disappearance of the peak at 4434 days (12.1 years).

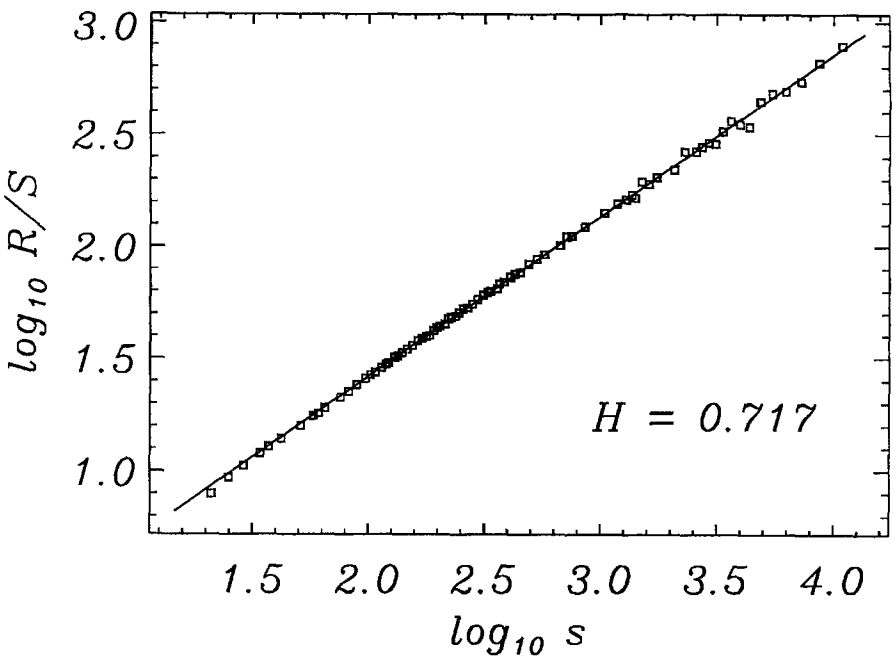


*Figure 2.* (a) Pox diagram of the N – S asymmetry of sunspot areas after detrending and subtracting the deterministic cycle from the original series. The solid line represents a fit to points with lag $s$ between 65 and 3111 days, which suggests a value of the Hurst exponent $H = 0.717 \pm 0.002$.

been favoured in one hemisphere. This has been corroborated by Howard (1974) for the period 1967 – 1973 and by Rabin *et al.* (1991) for the time interval 1976 – 1981. Thus, this imposes an important restriction on models to explain the mechanism by which magnetic flux tubes are carried to the surface.

As was stated in the Introduction, Mandelbrot and Wallis (1969a) computed a value $H = 0.96$ for the monthly sunspot numbers from 1749 to 1948. In addition, Ruzmaikin, Feynman, and Robinson (1994) determined a value $H = 0.84$ for time lags between 100 and 3000 years for a tree-ring $^{14}C$ data set. These large values of the Hurst exponent must be interpreted as an indication of quite strong long-run memory in solar activity and proxy records, which reflect the same underlying process, i.e., the amount of emerging magnetic flux. On the other hand, the asymmetry time series reflects an imbalance between the emerging flux in the north and south solar hemispheres, so there is no reason to expect the same value of $H$. The Hurst exponent $H = 0.717$ of N – S asymmetry implies that its continuum component is closer to white noise than that of the whole-disk solar activity.

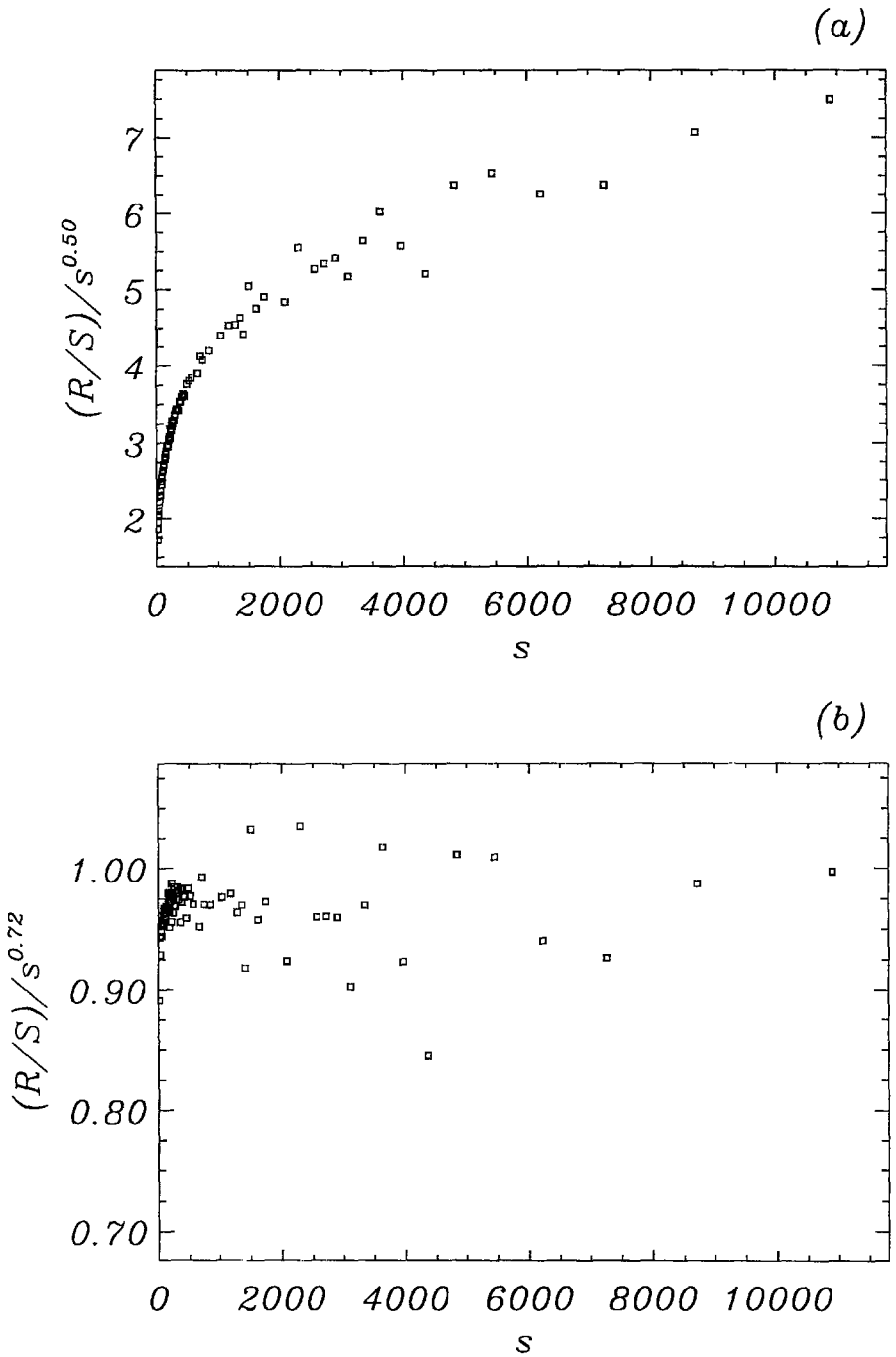


*Figure 3.* GEOS diagrams of the N–S asymmetry of sunspot areas after detrending and subtracting the deterministic cycle from the original series. (a) $H = 0.5$; there is divergence of $(R/S)/s^{0.5}$ for large values of $s$. (b) $H = 0.717$; there is convergence of the points in the diagram towards a finite limit.

## Acknowledgement

We want to acknowledge an anonymous referee for some useful comments that have helped improve the paper.

## References

Bendat, J. S. and Piersol, A. G.: 1986, *Random Data. Analysis and Measurement Procedures*, 2nd ed., John Wiley and Sons, New York, p. 76.
Bhattacharya, R. N., Gupta, V. K., and Waymire, E.: 1983, *J. Appl. Prob.* **20**, 649.
Carbonell, M., Oliver, R., and Ballester, J. L.: 1993, *Astron. Astrophys.* **274**, 497.
Feder, J.: 1988, *Fractals*, Plenum Press, New York, p. 149.
Feder, J.: 1991, in T. Riste and D. Sherrington (eds.), *Spontaneous Formation of Space-Time Structures and Criticality*, Kluwer Academic Publishers, The Netherlands, p. 113.
Garcia, H.: 1990, *Solar Phys.* **127**, 185.
Gottman, J. M.: 1981, *Time-Series Analysis*, Cambridge University Press, Cambridge, p. 102.
Howard, R.: 1974, *Solar Phys.* **38**, 59.
Hurst, H. E., Black, R. P., and Simaika, Y. M.: 1965, *Long-Term Storage: an Experimental Study*, Constable, London.
Komm, R. W.: 1995, *Solar Phys.* **156**, 17.
Mandelbrot, B. B. and Wallis, J. R.: 1969a, *Water Resour. Res.* **5**, 321.
Mandelbrot, B. B. and Wallis, J. R.: 1969b, *Water Resour. Res.* **5**, 967.
Mesa, O. J. and Poveda, G.: 1993, *Water Resour. Res.* **29**, 3995.
Oliver, R. and Ballester, J. L.: 1994, *Solar Phys.* **152**, 481.
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: 1988, *Numerical Recipes. The Art of Scientific Computing*, 3rd ed., Cambridge University Press, Cambrige, pp. 456 and 507.
Rabin, D. M., DeVore, C. R., Sheeley, N. R., Harvey, K. L., and Hoeksema, J. T.: 1991, in A. C. Cox, W. C. Livingston, and M. S. Matthews (eds.), *Solar Interior and Atmosphere*, The University of Arizona Press, Tucson.
Roy, J. R.: 1977, *Solar Phys.* **53**, 61.
Ruzmaikin, A., Feynman, J., and Robinson, P.: 1994, *Solar Phys.* **149**, 395.