A&A 503, 197–201 (2009)
DOI: 10.1051/0004-6361/200811430

Astronomy & Astrophysics

Long-term asymmetry in the wings of the butterfly diagram

N. V. Zolotova[1,2], D. I. Ponyavin[1], N. Marwan[2,3], and J. Kurths[3]

[1] Institute of Physics, Saint-Petersburg State University, Russia
e-mail: [ned;ponyavin]@geo.phys.spbu.ru
[2] Interdisciplinary Center for Dynamics of Complex Systems, University of Potsdam, Germany
[3] Potsdam Institute for Climate Impact Research, Germany
e-mail: [marwan;juergen.kurths]@pik-potsdam.de

Received 27 November 2008 / Accepted 12 May 2009

ABSTRACT

Aims. Sunspot distribution in the northern and southern solar hemispheres exibit striking synchronous behaviour on the scale of a Schwabe cycle. However, sometimes the bilateral symmetry of the Butterfly diagram relative to the solar equatorial plane breaks down. The investigation of this phenomenon is important to explaining the almost-periodic behaviour of solar cycles.
Methods. We use cross-recurrence plots for the study of the time-varying phase asymmetry of the northern and southern hemisphere and compare our results with the latitudinal distribution of the sunspots.
Results. We observe a long-term persistence of phase leading in one of the hemispheres, which lasts almost 4 solar cycles and probably corresponds to the Gleissberg cycle. Long-term variations in the hemispheric-leading do not demonstrate clear periodicity but are strongly anti-correlated with the long-term variations in the magnetic equator.

Key words. Sun: activity – Sun: sunspots – Sun: magnetic fields

1. Introduction

Carrington (1858) discovered that the sunspot mean latitude gradually decreases from the onset to the end of a solar cycle. This effect was later confirmed by Spörer (1894) using other solar cycles, and the established rule was subsequently referred to by his name. The latitude-time diagram of sunspot occurrence forms butterfly-like patterns during the course of the solar cycle (Maunder 1904). Maunder (1922) stated: "Since the origin of the solar spots lies within the Sun, and the northern and southern spots show difference in their behaviour, we must conclude that the Sun is not symmetrical in the constitution of its interior. If then we assume, as the basis of any investigation, that the Sun is symmetrical in its internal constitution, we are making an assumption contrary to the evidence supplied by the behaviour of its surface".

Since 1955, the north-south asymmetry has been investigated by introducing the normalised index *NA* (e.g., Newton & Milsom 1955; Knaack et al. 2004; Ballester et al. 2005; Carbonell et al. 2007; Li et al. 2009). The *NA* measure "loses all information about the latitude variations of sunspots" (Pulkkinen et al. 1999) and hides phase hemispheric relationships (Zolotova & Ponyavin 2006), because of the definition of *NA* as instantaneous amplitude dominance of the northern over southern hemisphere and vice versa.

Waldmeier (1957) suggested a 80-year period for the hemispheric asymmetry of sunspot activity, phase shift, and height variations at solar maxima. However, his scheme cannot explain the strong asymmetry detected between the cycles 19 and 20 (Waldmeier 1971). Outstanding asymmetry observed during the Maunder minimum also disagrees with Waldmeier suggestions.

Variations in sunspot latitudes from 1853 to 1996 were investigated by Pulkkinen et al. (1999). They defined the so-called *magnetic equator* to be the sum of the mean latitudes $\langle\lambda\rangle$ of sunspots in the northern and southern hemispheres $\langle\lambda(N)\rangle_n + \langle\lambda(S)\rangle_n$ (the southern component is negative, n is the time epoch). A systematic variation with a period of about 90 years and an amplitude of 1.3 degrees was detected.

Cross-recurrence plots (CRPs) can be used to study run-time differences (such as locally varying phase shifts) between time series (Marwan et al. 2002). Applying CRPs to the raw monthly sunspot area time series, we have shown that run-time differences between northern and southern sunspot appearances exhibit non-random quasi-regular behaviour (Zolotova & Ponyavin 2006). This approach allows us to determine the asymmetry in the northern and southern sunspot activities as an effect of a synchronisation between hemispheres. It was shown that this synchronisation can be detected on the solar cycle time scale (Zolotova & Ponyavin 2007b). Significant continuous long-term variability in the interhemispheric phase shift was confirmed using wavelet decomposition (Donner & Thiel 2007). They suggested that phenomenon of the north-south asymmetry should not be explained as a phenomenon of phase synchronisation, because hemispheric sunspot activities probably represent just two observables of the same complex system.

However, we suggest that sunspots, as tracers of the toroidal magnetic field, evolve independently in both hemispheres, while the large-scale poloidal magnetic field is the common "synchroniser" for the entire Sun. This autonomous evolution in strong small-scale magnetic fields and weak hemispheric coupling between them do not contradict the dynamo theory. Large asymmetry during the Maunder minimum, when sunspots appeared commonly only in the southern hemisphere (Ribes & Nesme-Ribes 1993; Sokoloff & Nesme-Ribes 1994), is also indirectly indicative of an independent increase in toroidal flux tubes in both hemispheres. If the north and south represents two observables of the same complex system then the butterfly diagram should always have two wings (not only one).

A hemispheric coupling can be produced by a poloidal large-scale magnetic field formed in the convective zone and/or transequatorial loops in the solar corona as a visible signature of the dynamo process (Jiang et al. 2007). Hemispheric coupling can be continuous or intermittent in time. The last case can result in a strong phase asymmetry.

In this paper, we apply cross-recurrence plots to analyse long-term asymmetries between the northern and southern sunspot distributions. We study the variation in the north-south temporal asymmetry and compare it with the long-term variation in mean sunspot latitudes. Furthermore, we discuss the strong north-south asymmetry at the time of Grand minima and just at the beginning of the 24th solar cycle.

2. Methodology

The recurrence plot was introduced as a tool for visualisation of recurrent states of dynamical systems (Eckmann et al. 1987; Marwan et al. 2007a). Further developments in this method promise new approaches to the analysis of spatio-temporal systems (Vasconcelos et al. 2006; Marwan et al. 2007b). Its bivariate extension, the cross-recurrence plot (CRP), was later used to study run-time differences between two similar systems (Marwan et al. 2002). A CRP analyses the parallel occurrence of states as described by the time series x_i and y_i and is defined by the cross-recurrence matrix $\mathbf{CR}$

$$CR_{i,j} = \Theta(\varepsilon_i - \|x_i - y_j\|), \qquad i,\ j = 1, \ldots, N, \tag{1}$$

where N is the length of the time series x and y, $\|\cdot\|$ is a norm, and Θ is the Heaviside function (Marwan et al. 2007a). The recurrence threshold ε_i can be either fixed or can vary for each index i in such a way that the number of recurrence points (neighbours) for each state x_i is the same. This latter condition causes obviously a CRP with a predefined number of recurrence points (recurrence point density), given by the number of neighbours of each state.

In the special case of $x = y$, the CRP corresponds to the common recurrence plot $\mathbf{R}$, which contains the *line of identity* $R_{i,i} = 1$, expressing the trivial recurrence of a state to itself at the same time. Applying a transformation of the timescale (e.g., a phase shift) to the second time series y, the line of identity becomes bowed or shifted (Marwan et al. 2002; Marwan & Kurths 2005). We call this line a *line of synchronisation (LOS)*, because we can use it to rescale the timescale of the second time series and determine the closest match (synchronisation) of the two time series. We, therefore, note that this synchronisation has a different meaning than the phase synchronisation as a physical phenomenon. It is also important to point out that CRPs require time series of the same observable and similar amplitude variations in both x and y. However, a CRP can be useful for detecting differences in timescales between two different data series or dynamical systems, as shown for geophysical measurements of lake sediments (Marwan et al. 2002; Marwan & Kurths 2005) or phase differences in the occurence of sunspots (Zolotova & Ponyavin 2006). CRPs were initially introduced to study dynamical systems of higher dimensions, i.e., by considering phase space vectors $\boldsymbol{x} \in \mathbb{R}^m$, but can be analogously applied to the analysis of spatio-temporal data (Vasconcelos et al. 2006; Marwan et al. 2007a,b), as in the case of sunspot activity data.

The construction of the LOS with the CRP can be performed in different ways. However, an important requirement is that for each point on the x-axis of the CRP a corresponding point on the y-axis must be found. Additional criteria to obtain a good LOS are that the amount of targeted recurrence points by the LOS should converge to the maximum and the amount of gaps in the LOS should converge to the minimum. To achieve this aim, we used a simple two-step algorithm proposed in Marwan et al. (2002), available in the CRP Toolbox for Matlab®. This algorithm provided a trade-off between efficiency and correctness (there may be more optimised algorithms available from other sources).

The algorithm for constructing the LOS $T(i)$ consists of the following steps:

(1) *Initial step*: Find a recurrence point next to the axes origin $(0, 0)$, lying either on the x-axis or on the y-axis. The preferred direction is the x-axis (if in x and y-direction, the distance to the next recurrence point is equal). Thus, we have the two cases
   (a) *Start point on x-axis*: The starting point is at $(i_0, 0)$. The first points of the LOS are $T(0, \ldots, i_0) = 0$.
   (b) *Start point on y-axis*: The starting points is at $(0, j_0)$. The first point of the LOS is therefore $T(0) = j_0$.

(2) *Next recurrence point*: Find the next recurrence point at (i, j) after a previously determined LOS point $(\tilde{i}, \tilde{j})$ by looking for recurrence points in a squared window of size $w = 2$, located with its origin at $(\tilde{i}, \tilde{j})$. If the edge of the window meets a recurrence point $(i, j) = (\tilde{i} + \Delta i, \tilde{j} + \Delta j)$, we follow with step (3), else we iteratively increase the size of the window $w = w + 1$ until a recurrence point is found or the end of the CRP is reached (i.e. $\tilde{i} + w = N$ or $\tilde{j} + w = N$). Note that it is not necessary that $\Delta i \equiv \Delta j$.

(3) *Next LOS point*:
   (a) We consider a window of size $\mathrm{d}x \times \mathrm{d}y$ with its origin at the identified recurrence point (i, j). The parameters $\mathrm{d}x$ and $\mathrm{d}y$ are *rubber parameters* useful for correcting the LOS if extended clusters of recurrence points occur deflecting the LOS to their edge (for $\mathrm{d}x = \mathrm{d}y$ the preferential direction of the LOS will be 45°). Determine the local centre of mass of this window with coordinates s_x and s_y ($s_x = 0$ corresponds to i and $s_y = 0$ to j, respectively). Note that we have to limit $i + \mathrm{d}x \le N$ and $j + \mathrm{d}y \le N$.
   (b) Find the indices k_x and k_y of the first non-recurrence points $CR_{i+k_x,j} = 0$ and $CR_{i,j+k_y} = 0$, where $k_x \in [0, s_x]$ and $k_y \in [0, s_y]$ (i.e., within a column and a row starting from (i, j)). If a non-recurrence point is not found, we set for this index the centre of mass ($k_x = s_x$ or $k_y = s_y$). Note that we have to limit $i + k_x \le N$ and $j + k_y \le N$.
   (c) We consider the point $(i + k_x/2, j + k_y/2)$ as a point on the LOS. We model the LOS between this point and the former point $(\tilde{i}, \tilde{j})$ using linear interpolation (fractional coordinates are floor-rounded).
   (d) Set the last point $(i + k_x/2, j + k_y/2)$ as the new starting point $(\tilde{i}, \tilde{j})$ and repeat with step (3), until $i = N$ or $j = N$.

In the case of sparse CRPs, a manual positioning/correction of the starting point may be necessary. Because we are interested only in the relative deviation of the LOS, we subtracted the monotonically increasing line of identity (i.e., a series running from 1874 to 2008).

3. Data and results

We used the Royal Greenwich Observatory USAF/NOAA data of sunspot area and their latitudes separately for the northern and southern hemispheres. In contrast to our previous analysis (Zolotova & Ponyavin 2006, 2007b) in which raw (unsmoothed)

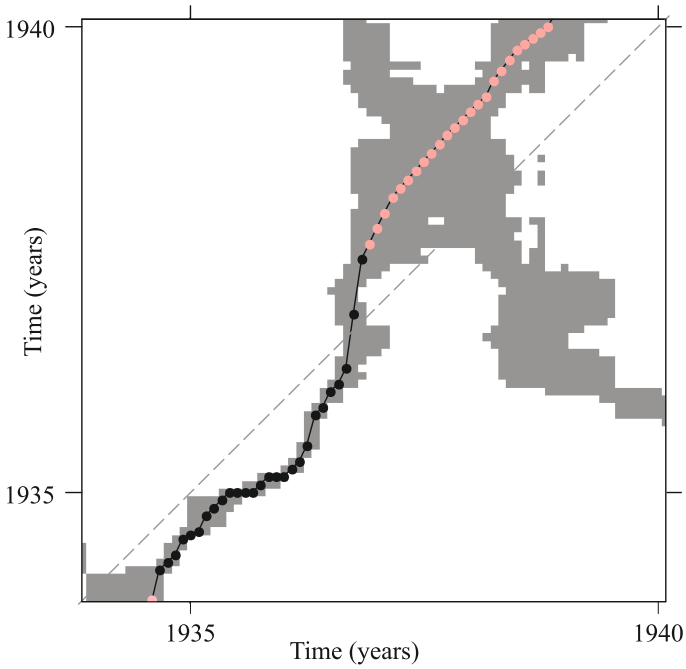

Fig. 1. Fragment of the cross-recurrence plot of smoothed monthly sunspot area for the cycle 17. LOS is presented by black (significant) and pink (insignificant) points. The leading of southern hemisphere changes to the northern leading near 1937.

data were analysed, we used filtered (smoothed) data. Smoothing of the data allows an accurate detection of the LOS in the CRP. To achieve this, the monthly sunspot area data was filtered using an optimal 20-month moving average. The length of the moving average filter was empirically determined as a trade-off by ensuring as much detail as possible and at the same time a smooth LOS without large jumps. To be able to identify the slight wandering of the magnetic equator, the latitude data for each Carrington rotation was filtered using a 10-month moving average.

The CRP of the northern and southern hemispheres sunspot area data was calculated using a varying recurrence threshold ε, preserving a recurrence point density of 10%. For the construction of the LOS, rubber parameters dx = 30 and dy = 30 were used.

In general, a CRP can contain extended clusters of recurrence points for which the states of the considered systems change slowly. Solar cycle extremes (minima and maxima) cause regions of uncertainties (extended clusters of recurrence points) in the description of the timescale mismatch between both systems. In contrast, ascending and descending phases of the solar activity exhibit rather clear segments on the LOS, in which we are able to detect the phase differences between the northern and southern hemisphere. To reduce these uncertainties in the following analysis, we consider only these particular segments of the LOS, whose underlying recurrence structure does not exceed a width of a half year (i.e., 6 points). As an example, we may consider the 17th solar cycle (Fig. 1). Until 1935, the solar minimum caused an extended cluster of recurrence points. The following ascending phase in the both hemispheres is more-or-less free from ambiguities and is found to contain a phase shift. After 1937, the solar maximum again causes an extended cluster of recurrence points. For clearness, we label the LOS as uncertain (pink colour). In the following, we consider only the remaining clearly defined segments of the LOS (black points, Fig. 2b).

The change in sign of the LOS indicates the change of the leading role in the hemispheres (LOS > 0 – north leading, LOS < 0 – south leading). A general change from mainly positive values to mainly negative values appears between 1925 and 1930, whereas a change from negative to mainly positive values appears between 1965 and 1970 (Fig. 2b, vertical gray lines). Thus, from the beginning of the 12th to the maximum of the 16th cycle, the northern hemisphere dominates in leading, after which, until the minimum before the cycle 20, the southern hemisphere leads, and then until the present, the northern hemisphere leads again. These findings coincide with the identified phase difference between the activities in both hemispheres as studied by Donner & Thiel (2007). However, the timescale differences derived by the LOS increased to a maximum of 25 months, whereas, because of averaging effects, the phase difference derived by Donner and Thiel did not exceed 10 months. For instance, the north-south lag in the beginning of cycle 20 is about 13 months (Zolotova & Ponyavin 2007a). The measured long-term variation in the hemispheric leaderships can be confirmed by cross-correlations, local phase relationship from cross-wavelet analysis (Zolotova & Ponyavin 2007b), and by joint-recurrence plots. However, all these methods have the disadvantage that they need to consider sequences of data points, causing an averaging of the property being measured. Using the CRP approach, we are able to compare the timescales of the two data series on a point-by-point basis.

We next compared the LOS with the magnetic equator derived by Pulkkinen et al. (1999) between 1853 and 1996 (Fig. 2a). To extend the time span to the present day, we used a running mean of daily sunspot latitude data for the entire Sun (equivalent to the magnetic equator definition) from 1980 to July 2008 (200 days moving average; Fig. 2a, blue line). Gaps correspond to the absence of sunspots during solar minima. By fitting a sine to the magnetic equator, Pulkkinen et al. (1999) found three sign changes. Two of them coincide with the already mentioned epoch (1925–1930 and 1965–1970), but with negative values before 1925–1930 and positive values before 1965–1970. They found an additional change in 1875. Obviously, the latitudinal distribution of sunspots (magnetic equator) is in antiphase with the long-term leadership variation in the sunspot occurrences.

Since our time series starts in the middle of 1874, we are unable to detect a change in the hemispheric leadership during the minimum of the 12th cycle. In reattempting a change detection, it would be useful to consider, e.g., Carrington's and Spörer's data.

Another test of the obtained results that we have obtained, is a comparison with the separate mean sunspot latitudes of northern and southern hemisphere. We used the Greenwich data (as used for the Butterfly diagram) and for each Carrington rotation calculated the mean latitude separately for each hemisphere. The 10-point (Carrington rotation) moving average clearly indicates that up to the maximum of the solar cycle 16, the north leads in time, but with a preference for sunspots to emerge at higher latitudes in the south (Fig. 2c). After this epoch and until the minimum of cycle 20, the situation is reversed (south is leading but sunspots belonging to higher latitudes occur in the Northern hemisphere). After cycle 20, the state is restored back to a situation that appeared between cycle 12 and 16.

The differences between these mean sunspot latitudes of northern and southern hemisphere (Fig. 2d) clearly reproduce variations in the magnetic equator. The results also coincide

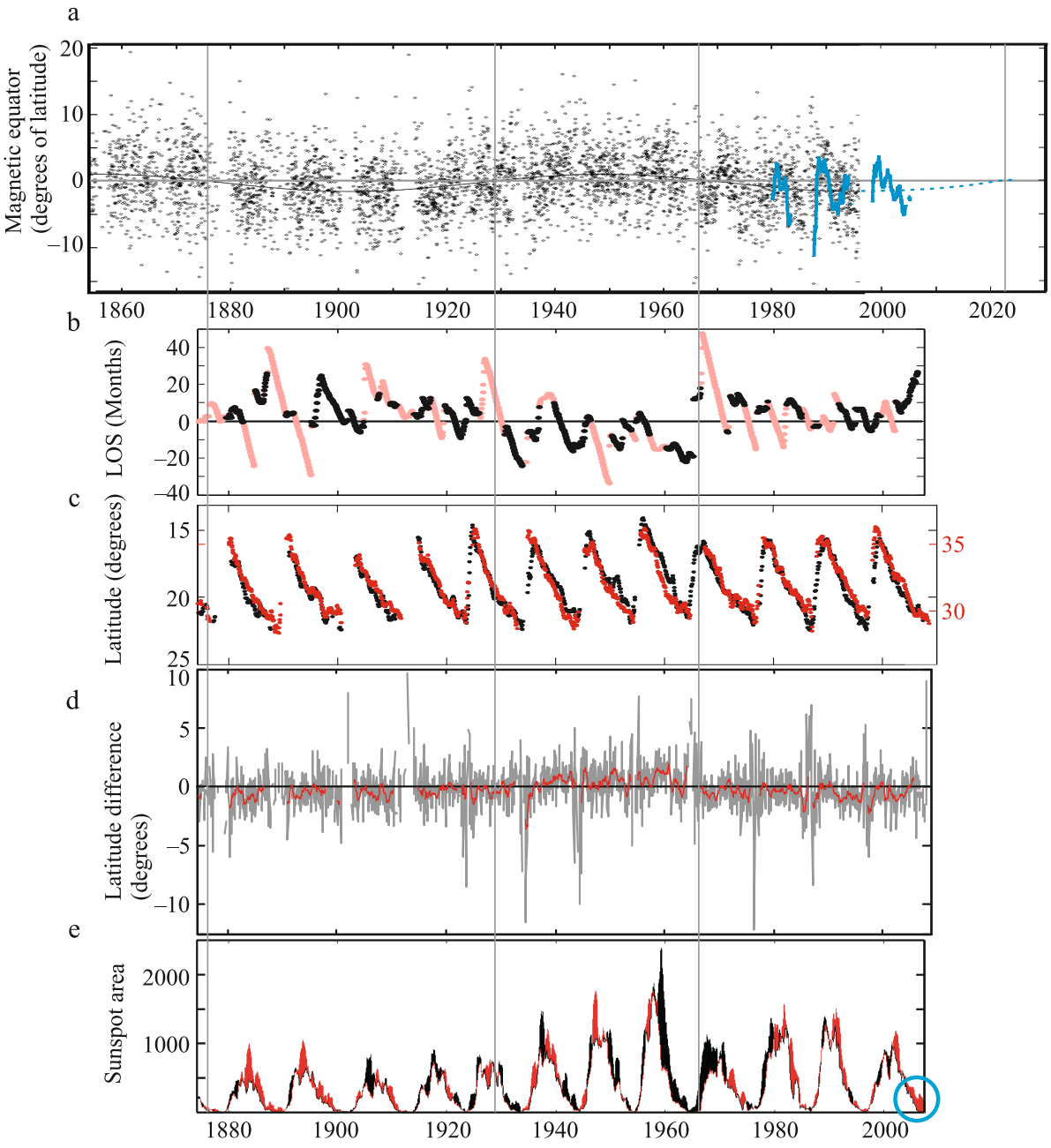

Fig. 2. **a)** Magnetic equator according to Pulkkinen et al. (1999) – black colour, the mean sunspot latitude extended by us – blue colour. Gaps corresponds to days without sunspots. **b)** LOS: rejected deviations are pink. **c)** Smoothed mean sunspot latitudes for the north (black) and south (red). **d)** Their difference is shown by red. All symbols are the same apart from for raw data – grey. **e)** Yearly smoothed sunspot area. The north dominates the amplitude measurement – black, south – red.

with the weighted magnetic equator derived by Pulkkinen et al. (1999) for unsmoothed (grey) and smoothed data (red).

4. Conclusion and discussion

We have performed a cross-recurrence plot analysis of sunspot activity data and have found a long-term persistence in phase leading of one of the hemispheres, which is close to the Gleissberg cycle (Gleissberg 1967). Thus, the wings of the butterfly diagram exhibit a long-term asymmetrical evolution. According to the analysed sunspot area data, two significant changes in the predominant hemispheric leading have been detected since 1874. The first change occurred near 1928 (16th cycle maximum), the second in late 1968 (minimum between the cycles 19 and 20). Phase-leading was found to be in antiphase with the mean latitudes of sunspots in the two hemispheres.

Pulkkinen et al. (1999) found regular magnetic equator variations of periods of about 90 years and an amplitude of 1.31 ± 0.13 degrees. Later, Pelt et al. (2000) defined the solar cycle as a spatiotemporal rather than a purely temporal entity. They presented evidence of a Gleissberg cycle in long-term changes in both length and symmetry of the Schwabe cycle.

We have extrapolated the position of the magnetic equator using data from 1980 until July 2008 (Fig. 2a, blue dashed line). In general, we found that its position has shifted southwards, and we may expect a next zero crossing just after 2020. However, time intervals between zero crossings are not equal to each other. The northern hemisphere leads by about 50–55 years, whereas

the southern hemisphere leads by only 40–45 years. This variability in the period is similar to the occurrence rate of Grand minima and maxima, which is driven not by long-term cyclic variability, but rather by a stochastic or chaotic process (Usoskin et al. 2007). However, historical activity data are still of too short duration to resolve this problem.

The new solar cycle 24 started on January 4th 2008 in the North, since the northern hemisphere leads in time and dominates in power. The unusually long solar cycle 23 exhibits in the last part of the current minimum a significant north-south asymmetry of the sunspot area, reaching 500 units of millions of a hemisphere for unsmoothed and 250 for smoothed data (Fig. 2e, blue circle). For the whole interval of the Greenwich records, the asymmetry during the solar minima does not reach such a stable large amplitude. We expect that some more dramatic change in the solar dynamics may occur in the near future.

During the Maunder minimum (1645–1715), it is interesting that the butterfly symmetry was broken, and a significant amplitude difference between the hemispheres became evident. Sunspots were observed only in the Southern hemisphere (Ribes & Nesme-Ribes 1993; Sokoloff & Nesme-Ribes 1994). We propose that a similar amplitude and/or phase asymmetry occured during the Dalton minimum (1795–1823). Suppressed and delayed activity in the driven hemisphere was possibly responsible for the unusual length of the 4th solar cycle and subsequent Grand minimum (Zolotova & Ponyavin 2007a).

According to the topological kinematic model of the solar dynamo, the differential rotation is the major cause of transforming the poloidal field into toroidal flux tubes that emerge in the form of bipolar sunspots (Babcock 1961; Leighton 1969). The difference in the appearance of sunspots in both hemispheres depends apparently on the north-south asymmetry of rotation. Variations in the north-south asymmetry in the solar rotation seem to be observed at the Sun on a long timescale (see, e.g., Pulkkinnen & Tuominen 1998) and can be associated with a long-term systematic asynchrony and asymmetry of the hemispheres. Further studies are necessary to establish such relationships.

Acknowledgements. This research has been supported by the INTAS Fellowship Grant for Young Scientists Ref. No. 06-1000014-6022. For the calculation of CRP and LOS, we have used the CRP Toolbox for Matlab®, available at tocsy.pik-potsdam.de. The Royal Greenwich Observatory USAF/NOAA data of sunspot area and their latitudes separately for the northern and southern hemispheres is available at http://solarscience.msfc.nasa.gov/greenwch.shtml.

References

Babcock, H. W. 1961, ApJ, 133, 572
Ballester, J. L., Oliver, R., & Carbonell, M. 2005, A&A, 31, L5
Carbonell, M., Terradas, J., Oliver, R., & Ballester, J. L. 2007, A&A, 476, 951
Carrington, R. C. 1858, MNRAS, 19, 1
Donner, R., & Thiel, M. 2007, A&A, 475, L33
Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. 1987, Europhys. Lett., 4, 973
Gleissberg, W. 1967, Sol. Phys., 2, 221
Jiang, J., Choudhuri, A. R., & Wang, J. 2007, Sol. Phys., 245, 19
Knaack, R., Stenflo, J. O., & Berdyugina, S. V. 2004, A&A, 418, L17
Leighton, R. B. 1969, ApJ, 156, 1
Li, K. J., Gao, P. X., Zhan, L. S., & Shi, X. J. 2009, ApJ, 691, 75
Marwan, N., & Kurths, J. 2005, Phys. Lett. A, 336, 349
Marwan, N., Thiel, M., & Nowaczyk, N. R. 2002, Nonlin. Proc. Geophys., 9, 325
Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. 2007a, Phys. Rep., 438, 237
Marwan, N., Kurths, J., & Saparin, P. 2007b, Phys. Lett. A, 360, 545
Maunder, E. W. 1904, MNRAS, 64, 747
Maunder, E. W. 1922, MNRAS, 82, 534
Newton, H. W., & Milsom, A. S. 1955, MNRAS, 115, 398
Pelt, J., Brooke, J., Pulkkinen, P. J., & Tuominen, I. 2000, A&A, 362, 1143
Pulkkinen, P. J., & Tuominen, I. 1998, A&A, 332, 748
Pulkkinen, P. J., Brooke, J., Pelt, J., & Tuominen, I. 1999, A&A, 341, L43
Ribes, J. C., & Nesme-Ribes, E. 1993, A&A, 276, 549
Sokoloff, D., & Nesme-Ribes, E. 1994, A&A, 288, 293
Spörer, G. 1894, Pub. Potsdam Obs., 10, 144
Torrence, C., & Compo, G. P. 1998, Bull. Am. Met. Soc., 79, 61
Usoskin, I. G., Solanki, S. K., & Kovaltsov, G. A. 2007, A&A, 471, 301
Vasconcelos, D. B., Lopes, S. R., Viana, R. L., & Kurths, J. 2006, Phys. Rev. E, 73, 056207
Waldmeier, M. 1957, Z. Astrophys., 43, 149
Waldmeier, M. 1971, Sol. Phys., 20, 332
Zolotova, N. V., & Ponyavin, D. I. 2006, A&A, 449, L1
Zolotova, N. V., & Ponyavin, D. I. 2007a, A&A, 470, L17
Zolotova, N. V., & Ponyavin, D. I. 2007b, Sol. Phys., 243, 193