A&A 396, 235–242 (2002)
DOI: 10.1051/0004-6361:20021340

Astronomy & Astrophysics

Was one sunspot cycle in the 18th century really lost?

N. A. Krivova[1,2], S. K. Solanki[1], and J. Beer[3]

[1] Max-Planck-Institut für Aeronomie, 37191 Katlenburg-Lindau, Germany
e-mail: natalie@linmpi.mpg.de; solanki@linmpi.mpg.de
[2] On leave from Astronomical Institute, St. Petersburg University, 198504 St. Petersburg, Russia
[3] Swiss Federal Institute of Environmental Science and Technology (EAWAG), 8600 Duebendorf, Switzerland
e-mail: beer@eawag.ch

Received 17 July 2002 / Accepted 12 September 2002

Abstract. The unusually long 4th solar cycle has recently been proposed by Usoskin et al. (2001) to be composed of two cycles. They argue that a weak and short cycle might have been lost in sparse sunspot data at the end of the 18th century. Here we check this hypothesis in different ways. First, we consider the sunspot number record in greater detail and compare in a statistical sense the sunspot observations of the period in question with those at other times. In a statistical sense the sunspot numbers recorded at the time of the proposed new cycle minimum are extremely untypical for other minima in the solar cycle record, but quite usual for the declining phase of the solar cycle. We also analyse other available proxies of solar activity, such as variations of the cosmogenic nuclides ^{10}Be and ^{14}C as well as auroral activity. These historical records are sufficiently long and provide an independent testimony of the cyclic behaviour of solar activity at the end of the 18th century. We found no evidence for a lost cycle in any of these data sets. Finally, we compare the proposed new cycle with the other cycles in the sunspot record. This reveals that the proposed "missing" cycle has very unusual properties, much more so than the original, standard cycle 4. Taken together, the evidence from these various tests strongly suggests that no cycle was missed and that the official sunspot cycle numbering and parameters are correct.

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

1. Introduction

The cyclic behaviour of the Sun's magnetic field has been known since the beginning of the 19th century. First noticed by Schwabe as an approximately 10-year variation in sunspot numbers, the cyclicity was later found in most other indicators of solar activity. The most complete historical record providing a direct measure of solar activity remains the sunspot number time series, however. For a long time, the Wolf or Zürich sunspot number, R_z (see Waldmeier 1961), has been the prime representation of the number of sunspots. There were, however, no regular observations prior to the 19th century and all earlier data were collected from different inhomogeneous sources, so that the R_z record at that time was incomplete. Recently another time series, the group sunspot number, R_g, has been compiled by Hoyt & Schatten (1998). It is based on the number of sunspot groups only rather than on a combination of groups and individual sunspots. It is expected to be less dependent on whether the smallest spots were detected by a given observer or not and, therefore, less noisy. Furthermore, some additional data from early sources have been incorporated in R_g, although it still contains many gaps in the 18th century.

Send offprint requests to: N. A. Krivova,
e-mail: natalie@linmpi.mpg.de

A particularly critical period lies at the end of the 18th – beginning of the 19th century. This period is poorly covered by observations, probably due to the political turmoil within Europe at that time. During this period the relatively strong sunspot cycles 3 and 4 were followed by an interval of low solar activity known as the Dalton minimum (cycles 5–7). At the same time cycles 4–6 were unusually long, with lengths of about 12 to 14 years.

Recently, Usoskin et al. (2001) have suggested that one solar cycle might have been lost in sparse data between cycles 4 and 5. Their main arguments are: (i) there were very few observations in the years 1792 and 1793 so that this might have been the time of an unnoticed activity minimum; (ii) cycle 4 was unusually long and could have comprised two shorter cycles; (iii) an extra cycle would help to explain the change of phase which happened at that time – the intensity of odd cycles was found to be larger than that of the preceding even cycles for cycles 6 and later, whereas the phase is reversed for earlier cycles.

Now, an absence of observations does not necessarily imply an absence of sunspots. Since the directly observed sunspot record is not as reliable as one would wish during the period in question (although it indicates that sunspots were, in fact, seen in 1792–1793; see Sect. 2), it is necessary to consider all possible indicators of solar activity. Hence, in addition to a

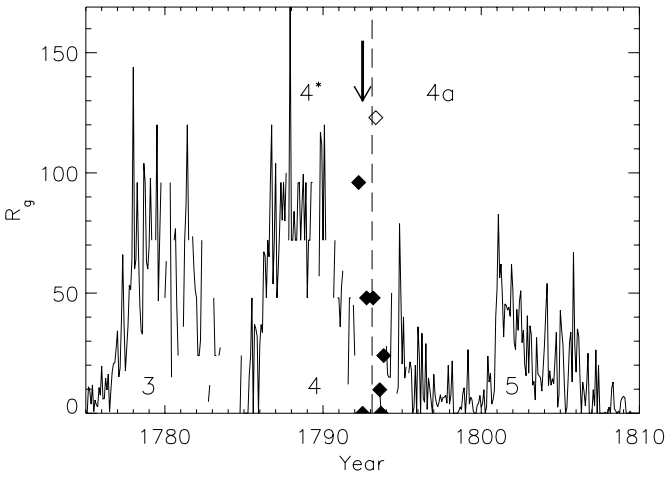

Fig. 1. Monthly group sunspot number vs. time. Isolated monthly data during 1792–1793 are represented by diamonds. The questionable observation by Huber in May, 1793 is denoted by an open symbol. The vertical dashed line marks the time of the suggested minimum. The arrow indicates the time when a naked-eye sunspot was observed in China (Yau & Stephenson 1988). Standard cycle numbers are given at the bottom of the figure.

re-analysis of the sunspot number record (Sect. 2), we consider here three other relevant records, ^{10}Be and ^{14}C cosmogenic isotopes (Sect. 3) as well as auroral activity (Sect. 4), in order to check, if they show any evidence of a lost solar cycle.

Although the sunspot cycle is not periodic, with considerable variation from one cycle to the next, the properties of the different cycles do tend to cluster together. One test of whether cycle 4 is a single cycle or is composed of two cycles is to compare the properties of the disputed cycles with those of the remaining sunspot cycles. We describe such a comparison in Sect. 5. Finally, conclusions are presented in Sect. 6.

2. Sunspot observations

In Fig. 1, we have plotted monthly group sunspot number, R_g, for cycles 3–5. In the following we call the cycle from roughly 1785 to 1798 cycle 4, in keeping with standard terminology. To distinguish this from the shorter cycle 4 proposed by Usoskin et al. (2001), we refer to the latter as cycle 4* and to the new cycle proposed by them between 1793 and 1798 as 4a (see Fig. 1). As pointed out by Usoskin et al. (2001), sunspot data are very sparse in the years 1792–1793. There were only 20 observations altogether (Hoyt & Schatten 1998), mostly isolated in the sense that there were almost no months observed successively. Since most of these records came from single observers and were not confirmed by an independent observation, they are considered to be unreliable. Usoskin et al. (2001) have, in particular, cast doubt on the record by Huber, who reported 4 sunspots in May, 1793, and have argued that the beginning of 1793 might have been the time of the minimum between cycles 4 and a new cycle they termed 4a (dashed line in Fig. 1). The questionable point corresponding to Huber's observation is represented by an open diamond, other observations by filled symbols.

It is seen, however, that neglecting this data point does not change the situation significantly. There was, indeed, a week in August–September 1793 when the absence of sunspots was confirmed by three independent observers. This is not exceptional for the declining phase of a cycle and in itself does not constitute strong support for sunspot minimum at that time. Consider, for example, cycle 5: there were no sunspots for 2 and more weeks, e.g., in April–May and June 1803, in July and September 1804, in May and June 1805; whole months were spotless from 1807 onward. Other examples of 6 and more successive spotless days registered 2 to 4 years after the maximum (and about 3 or more years before the next minimum) can be found for cycles 11 (May, 1875), 12 (January, November, 1886), 13 (May, October – November, 1897), 14 (February, April, July, 1910) or 21 (November, 1983). For the last 2.5 years of a cycle such spotless periods are quite common.

Of 9 other observations during 1792–1793 8 were non-zero. In particular, all 3 observations made around the time of the supposed minimum (October 1792–May 1793) confirmed that there were sunspots on the solar disc. We now test how likely it is to obtain the sunspot numbers counted in 1792–1793 at different phases within a sunspot cycle. To this end we consider cycles 9, 10 and 13 which are well sampled and are otherwise similar to cycle 4. Thus 9, 10 and 13 have cycle averaged R_g of 42.7, 41.4 and 39.9, respectively, compared with 42.6 for cycle 4. The cycle lengths are 12.5, 11.2 and 12.1 yr compared with 13.6 years for cycle 4. Taking the dates on which sunspots were recorded in 1792–1793 we created a filter which is then applied to the R_g record. This filter is composed of unit values for a set of days separated in time by exactly the same intervals as the observations in 1792–1793. Everywhere else it is zero. The R_g numbers returned by applying this filter to the R_g record in cycles 4, 9, 10 and 13 are plotted in Fig. 2. For the 3 more recent cycles we illustrate the application of the filter to 2 different times within the cycle, once during the declining phase (the same number of years after sunspot maximum as with cycle 4) and once during sunspot minimum.

Obviously, the pattern seen for cycle 4 is matched much better during the declining phase of a cycle than during minimum. Note that we have left out Huber's observation when constructing the filter, since this had been criticized by Usoskin et al. (2001).

In a next step we applied the filter to the whole declining phase and subsequent minimum of cycles 9, 10 and 13. Thus the filter was placed such that the first date of observation was 1 year after the official maximum of the cycle. Then the average of the R_g values returned by the filter was formed and it was moved by 1 day to a later time. This was repeated until the filter had moved to the following minimum (so that the last temporal window starts at this minimum and ends 2 years after it). The resulting series of average "filtered" R_g values obtained from this running window approach is plotted in Fig. 3 for the 3 cycles.

Also plotted in Fig. 3 is the average R_g level measured in 1792–1793 including all observations (horizontal dotted line), excluding the observation of Huber (dashed) and additionally excluding also the date contributing to the second

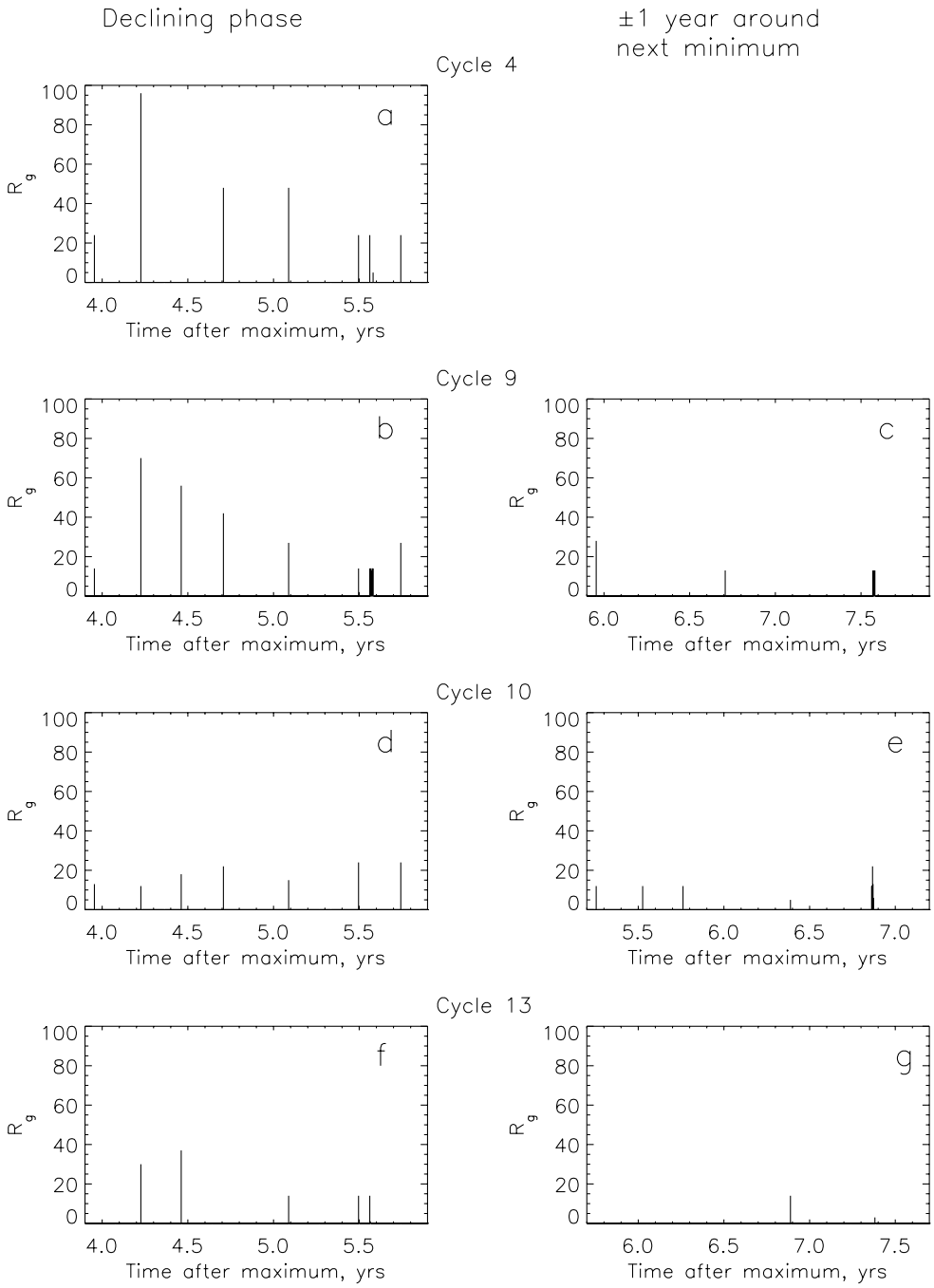

Fig. 2. a) Group sunspot numbers, R_g, observed between 1792 and 1793. **b)–g)** The number of sunspot groups returned by applying the same filter (observing sequence) to cycles 9, 10 and 13 during their declining phases (Frames **b)**, **d)** and **f)**) and the following minima (**c)**, **e)**, **g)**).

highest measured R_g value in this record (the record by Staudacher on 28.04.1792; dot-dashed line). This last level must be considered extremely conservative, even biased, since we have assumed that the observers only overestimated the number of sunspots (in reality we expect small sunspots to be missed more often than false detections) so that the 2 largest R_g values are wrong. Even in this extreme case, however, the average R_g observed in 1792–1793 is far more typical of the declining phase of a cycle, 3–5 years after maximum, than for the minimum phase. This strongly argues against the presence of a sunspot cycle minimum in 1792–1793.

Finally, according to Yau & Stephenson (1988), a naked-eye sunspot was seen in China in 1792. The probability of seeing such a big spot is higher at times not corresponding to sunspot minimum. We consider all naked-eye sunspots observed in East Asia during 1848–1917 (Yau & Stephenson 1988) and compare with the R_g values at the dates when naked-eye spots were reported. We find that the probability of a naked-eye sunspot being reported was approximately 4 times higher during periods when R_g was above the cycle average, as compared to the times when the sunspot number was below the cycle average. Hence, the fact that a naked-eye sunspot was

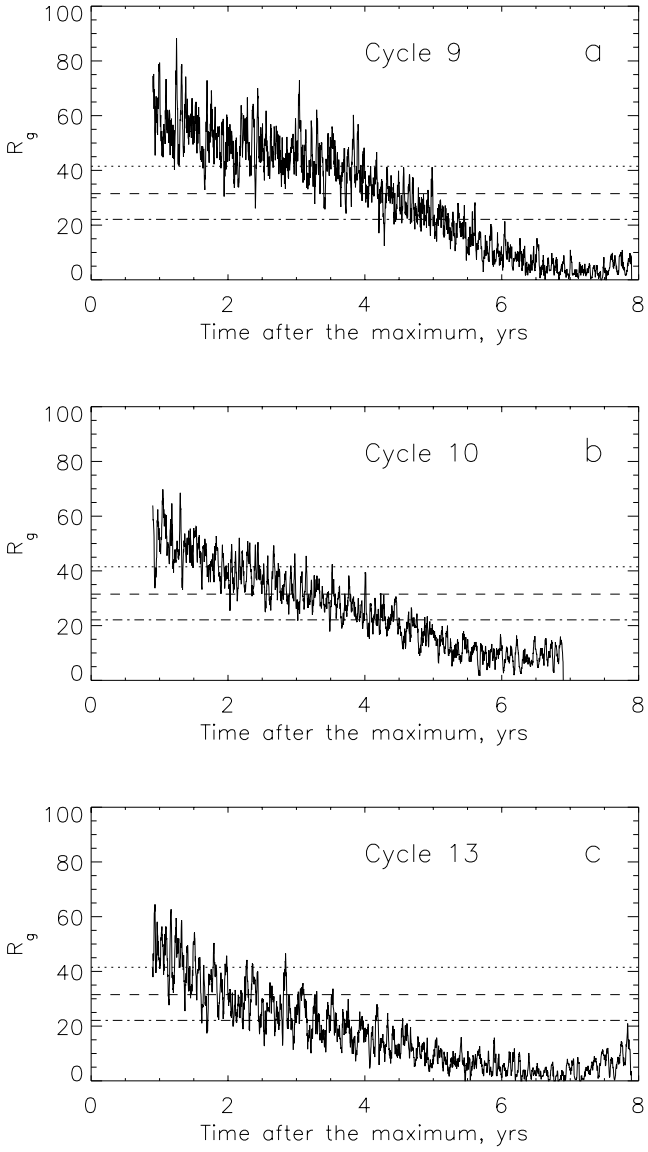

Fig. 3. Average R_g value obtained by applying the filter described in the text. Horizontal lines mark the corresponding values measured in 1792–1793. Dotted line: including Huber's observation; dashed: excluding Huber's observation; dot-dashed: excluding Huber's and the second strongest observation (Staudacher; 28.04.1792).

seen in 1792 further supports the argument that this time did not correspond to sunspot minimum.

3. Cosmogenic radionuclides

Cosmogenic radioisotopes, such as ^{10}Be and ^{14}C, are tracers of solar variability. Their production rate is modulated by the interaction of the solar wind with galactic cosmic ray particles and thus provides information on the Sun's magnetic activity. Beer et al. (1990, 1994), analysing ^{10}Be concentrations in Greenland ice, found a clear anti-correlation between sunspot numbers and the ^{10}Be production rate. They showed that ^{10}Be can be used to study variations in solar magnetic activity, in particular, the 11-year cycle. Thus, the dating of the ^{10}Be record is accurate to within one year according to Beer et al. (1990).

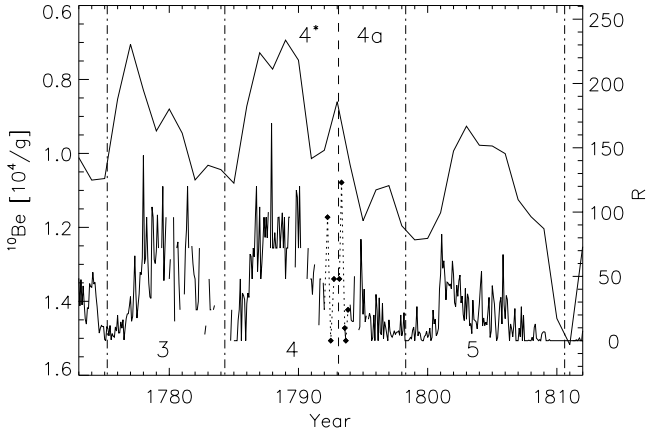

Fig. 4. Left-hand axis and upper curve: ^{10}Be concentration in 10^4 atoms per gram of ice for sunspot cycles 3–5 after applying a low-pass filter with a cut-off of 3 years (note the inverted scale). Right hand axis and lower curve: monthly group sunspot number. Isolated data in 1792–1793 (diamonds) are connected by the dotted line. The vertical dot-dashed lines indicate official sunspot minima, the vertical dashed line the time of the proposed new minimum between cycles 4* and 4a.

All sunspot maxima were found to be related to corresponding ^{10}Be minima, with three ^{10}Be minima (sunspot cycles 1, 6 and 9) being doubled. Smoothing removes short-term fluctuations probably introduced by transport and deposition effects and increases the correspondence between ^{10}Be and sunspot records. Furthermore, Usoskin et al. (2002b) have combined the model of Solanki et al. (2000, 2002) for the modulation of the Sun's magnetic flux with the heliospheric model of Usoskin et al. (2002a) to reproduce the ^{10}Be ice-core concentration starting from the R_g record. They basically confirm the conclusion of Beer et al. (1990, 1994), although on the basis of a more consistent physical description. We now check if there is any indication of an extra solar cycle at the end of 18th century in the ^{10}Be historical record.

The ^{10}Be data for cycles 3–5 are plotted in Fig. 4 together with the monthly R_g record. To remove possible seasonal variations we apply to the ^{10}Be time series a low-pass filter with a cut-off of 3 years, which is sufficiently short not to significantly influence the solar cycle variations. As Fig. 4 illustrates, the ^{10}Be record follows the group sunspot number quite well within the quoted period.

The ^{10}Be ice concentrations show three minima during the years 1784–1798 (which appear as maxima in Fig. 4 due to the inverted scale): (i) The slightly doubled main minimum at 1786–1790, in agreement with the sunspot maximum of cycle 4. (ii) The secondary minimum at 1792.5 (dates are given for the original unfiltered data), when the isolated sunspot data suggest a secondary maximum if Huber's observation is retained, and for when Usoskin et al. (2001) propose the new minimum in R_g. (iii) Finally, a weak minimum at 1796.5, around the time of the suggested maximum of cycle 4a. Now, as Cliver & Ling (2001) and Usoskin et al. (2002b) have shown, the ^{10}Be production can run behind the sunspot number by approximately 1 year for odd cycles, but is almost in phase for even cycles. Hence the accuracy of dating for cycle 4 remains

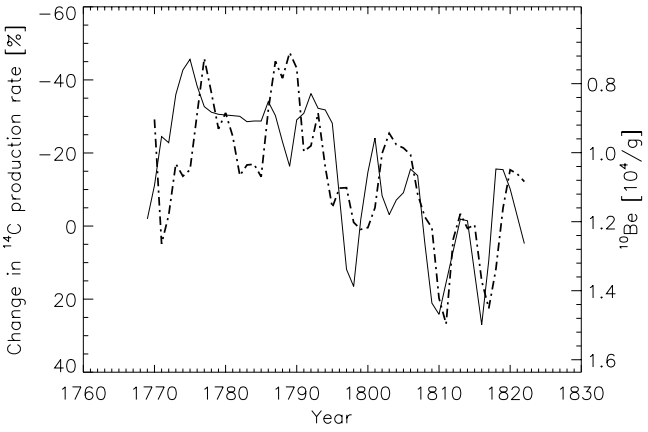

Fig. 5. Left-hand axis and solid curve: Changes in the ^{14}C production rate for the period 1770–1820 (note the inverted scale) following Stuiver & Quay (1980); Stuiver & Braziunas (1993) and Stuiver et al. (1998). Right hand axis and dot-dashed curve: ^{10}Be concentration in Greenland ice (same as in Fig. 4).

within about 1 year and the ^{10}Be data provide no firm evidence for the possibly lost cycle 4a. On the whole, they actually tend to support the isolated "unreliable" sunspot observations and indicate that there might have been a local secondary maximum in solar magnetic activity around the years 1792–1793.

Another cosmogenic isotope that can be used to trace solar magnetic activity back in time is ^{14}C. As for ^{10}Be, its production rate changes due to modulation of the galactic cosmic-ray flux by the varying solar wind and shows a clear anti-correlation with other indicators of solar activity, such as sunspot numbers, geomagnetic *Aa* indices and aurorae (Stuiver & Quay 1980). These changes in the atmospheric ^{14}C level are recorded in tree rings. Analysing tree rings, Stuiver & Quay (1980) have compiled a 860 year long ^{14}C record, which was later extended in time and refined by Stuiver & Braziunas (1993) and Stuiver et al. (1998). For the period of interest, the time resolution of the record is approximately 1 year.

The long residence time of ^{14}C in the carbon system leads to a damping of short-term variations in its production rate. Weak 11-year modulations are superimposed on a much stronger global trend and are sometimes difficult to identify. In Fig. 5 we show therefore not the measured change in ^{14}C concentration, but the corresponding change in production rate which has been derived from the measurements using a carbon cycle model (Stuiver & Braziunas 1993). The record does not present any indications of an additional separate minimum (corresponding to a maximum in solar activity) around 1795, i.e. year of the maximum of cycle 4a according to Usoskin et al. (2001). Nor does the ^{14}C production rate exhibit a maximum around 1792/93, as would be expected if that corresponded to an activity minimum. The very low ^{14}C production rate, about 40% below average during the relatively strong cycle 4, increases steeply up to 20% above the average at the solar activity minimum in 1798.

Of course, due to a strong damping of short-term variations, the ^{14}C record alone cannot be considered as a strong argument against the existence of cycle 4a. What might be more important, however, is that ^{14}C and ^{10}Be records show a very similar behaviour for the period in question. In Fig. 5 the ^{10}Be record discussed above is superimposed on ^{14}C (dot-dashed). The two records agree relatively well, with a correlation coefficient of 0.76. The production processes are very similar for both nuclides, whereas their following geochemical processing is quite different. Therefore, the fact that ^{10}Be and ^{14}C show a very similar behaviour for this period strongly supports the reality and the solar origin of the signal variations.

4. Aurorae

Aurorae are another phenomenon caused by the magnetospheric disturbance from the solar wind particles and can also be used as a proxy of the variability of the solar magnetic field (e.g., Legrand & Simon 1987; Schröder 1992). Historical records of auroral observations are, of course, far from complete and homogeneous. Fortunately, around 1780 weather service stations appeared in Europe that also jotted down northern lights. On the other hand, at the end of the 18th and the beginning of the 19th century Europe was ravaged by war. Not only were observations irregular, but many reports were lost or destroyed. To make matters even worse, the glow of large, distant fires connected with the war might have been mistaken for auroral phenomena. Therefore, great care is needed when analysing the auroral record for the time of interest.

As pointed out by Legrand & Simon (1987), the proper choice of the highest allowed geomagnetic latitude of the reporting stations is also important. Obviously, the highest frequency of aurorae occurs in the auroral zone, at high latitudes. However, the frequent appearance of bright quiet arcs not related to magnetospheric activity, the low population density and the varying size of the auroral oval make such data only weakly related to the solar activity cycle and thus highly unreliable for our purpose. Legrand & Simon (1987) even cite an example of a high latitude station at which the auroral activity reaches its maximum level during the minimum of solar activity and vice versa. From a correlation between the auroral occurrence and the number of geomagnetically disturbed days, they find that only observations from stations at geomagnetic latitudes below 62° should be employed. Auroral observations at high geographic latitudes are also disturbed during the summer time, so that, e.g., scandinavian data should not be used alone but must be complemented by records from other locations.

Taking these and other effects (weather conditions, longitudinal distribution of stations etc.) into account, Legrand & Simon (1987) have carefully analysed available auroral records to compile, to our knowledge, the most complete and homogeneous record of auroral activity for the period 1780–1980 that may be used as a proxy of solar activity. Note that the authors restrict themselves to this period since they consider these data to be relatively reliable. The frequency of auroral occurrence was found to follow very well the records of sunspot numbers and of the geomagnetic *aa* indices. Minima and maxima in auroral activity coincide with those in sunspot numbers, although secondary peaks, sometimes higher than the main ones, often occur during the last 4 years of a cycle, which is well known as recurrent geomagnetic activity.

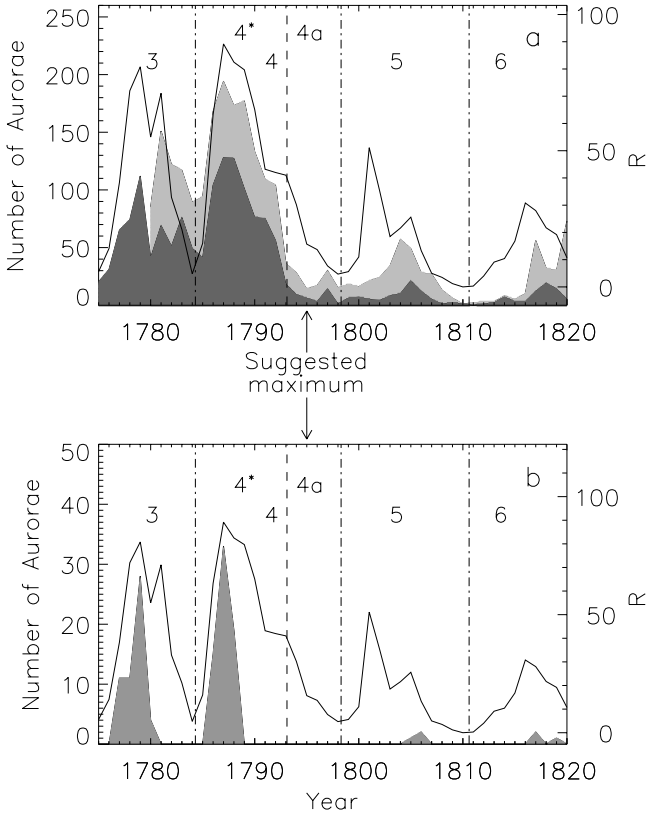

Fig. 6. Auroral frequency in cycles 3–6 (shaded areas) and the yearly R_g record. **a)** Aurorae recorded world-wide at geomagnetic latitudes below 62 N (lightly shaded; Legrand & Simon 1987) and 55 N (heavily shaded; Krivsky & Pejml 1988). **b)** Only reliable observations of aurorae in central Europe (southwards of 53 N, all registered by, at least, two different stations), following Schröder (2002).

Data for the years 1780–1820 are shown as a lightly shaded area in Fig. 6a. The correlation coefficient between the yearly auroral and R_g records for this period is 0.79. A weak peak is seen in 1797. Is it an observational artifact, a sign of the recurrent auroral activity or an indicator of the lost cycle? The overall profile of the auroral activity during the years 1785–1800 actually follows quite well those of the sunspot number and of the reversed ^{10}Be concentration, showing traces of all three maxima in the inverted ^{10}Be record. Thus the peak at 1797 may, indeed, be real and related to the weak peak in sunspots at 1795 (although the recurrent activity, being very common, looks like an even more probable cause), but does it support an additional cycle? Firstly, the number of registered events in 1797, 30 altogether, is less than the number of aurorae around the time of the suggested minimum in 1792 and 1793 (104 and 36 events, respectively). Secondly, the time of the suggested maximum of cycle 4a corresponds exactly to a local minimum in the auroral occurrence (around 1795–1796; see the light grey shaded area), which never happened during other cycles. Finally, the 1797 peak is very local in time: most aurorae were seen in February–March and at the end of the year, with very few events in the years 1796 and 1798. So even if the peak in auroral frequency at 1797 is real, it does not support the existence of a separate cycle 4a in any significant way.

Table 1. Length of sunspot cycles 1–7 as given by the NGDC. The values in brackets are estimates according to Usoskin et al. (2001).

Cycle Number	Year of Min	Cycle Length (yrs)
1	1755.2	11.3
2	1766.5	9.0
3	1775.5	9.2
4	1784.7	13.6 (8.4)
(4a)	(1793.1)	(5.2)
5	1798.3	12.3
6	1810.6	12.7
7	1823.3	10.6

In Fig. 6a, we also plot (heavily shaded area) another data set compiled by Krivsky & Pejml (1988), which includes all aurorae registered world-wide at geomagnetic latitudes below 55 N. Although it covers a smaller area and therefore includes less individual observations, this record is nevertheless very similar to the first one, and the conclusions reached above apply to it as well.

Finally, in Fig. 6b we plot only the most reliable observations of aurorae in central Europe, southwards of 53 N (Schröder 2002). All events were registered independently by, at least, two observers. Note, that the events associated with the recurrent geomagnetic storms are mostly confined to higher latitudes (Legrand & Simon 1987 and references therein), so that secondary maxima should basically be eliminated now too. The curve agrees well with the data shown in Fig. 6a, except that it now looks less noisy, although the statistics are poorer, and the peaks are narrower. All maxima of solar activity have left a clear mark in the auroral record (the maximum in auroral frequency in cycle 5 seems to be delayed with respect to the R_g maximum, but agrees very well with the official date given by the NGDC and with the maximum in the Zürich sunspot number). No trace is seen of the hypothetical cycle 4a: there are no reliably known auroral events for the period between 1790 and 1800. Consequently, this record also does not support the existence of cycle 4a. Although, on its own it cannot be used to argue strongly against such a cycle either, when combined with the other auroral and radionuclide records the lack of evidence for an additional cycle becomes striking.

5. Sunspot cycle statistics

One argument made by Usoskin et al. (2001) in favour of a new cycle 4a is the unusual length of cycle 4. According to the National Geophysical Data Center (Boulder, USA; NGDC), its length, defined as the time between two activity minima, was 13.6 years (see Table 1). Dates of cycle minima given by NGDC are based on an average of the times when minima are reached in the monthly mean sunspot number, in the smoothed monthly mean sunspot number, and in the monthly mean number of spot groups alone. In addition, the number of spotless days and the frequency of occurrence of "old" and "new" cycle spot groups were taken into account, to the extent possible.

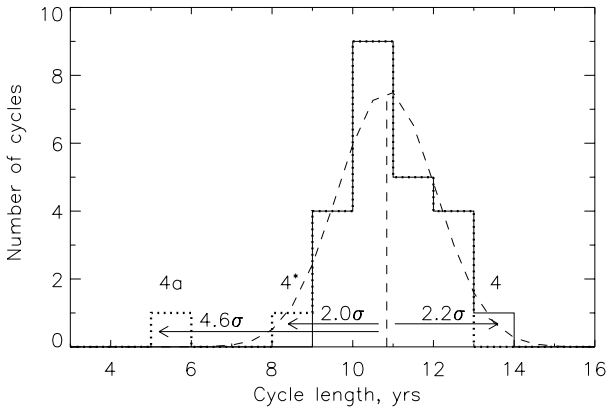

Fig. 7. Histogram of the solar cycle length. Solid line: cycle length following NGDC; dotted line: cycle 4 replaced by cycles 4* and 4a following Usoskin et al. (2001); dashed line: a Gaussian fit.

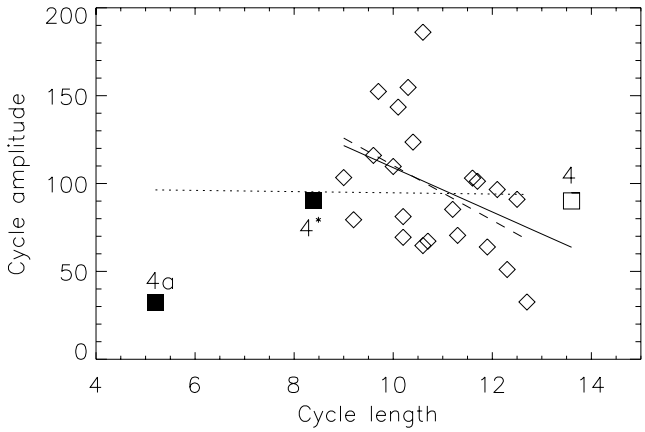

Fig. 8. Cycle amplitude versus cycle length. Each symbol represents a solar cycle. Cycles 1–3 and 5–22 are represented by diamonds. An open square represents the official cycle 4, whereas filled squares denote the new cycles 4* and 4a proposed by Usoskin et al. (2001). Solid, dashed and dotted lines are linear regressions to different subsets of the plotted points (see text for details).

The cited length of the 4th cycle is indeed higher than the mean length (11.2 years) of the solar activity cycle. To test whether this length is unrealistically large we plot in Fig. 7 a histogram of the standard lengths of solar cycles 1–22 (solid). Only the period after 1745 is included, since earlier data are less reliable. Indeed, cycle 4 is the longest cycle since 1750, but its length does not appear to be unrealistically large compared to that of the other cycles. Also plotted is the histogram of lengths after introducing an additional cycle 4a (dotted). It is immediately apparent that cycles 4* and 4a would be the two shortest solar cycles on record. In particular, the new cycle 4a with a length of less than 6 years, is well removed from the distribution. Including the official lengths of earlier cycles (1610–1745) does not change the shape of the cycle-length distribution a lot: the maximum moves to the bin 11–12 years and the histogram becomes somewhat broader (8 to 15 years) and more symmetric. Longer cycles (up to 13–14 years) turn out not to be unusual when these earlier data are included, in particular during and before low activity periods. Evidence, although of lower reliability, for even longer periods before and during the Maunder minimum exists (e.g., NGDC; Beer et al. 1994; Fligge et al. 1999). In contrast, cycles shorter than 8 years have never been registered. If we fit a Gaussian to the distribution of lengths of all cycles plotted in Fig. 7, except 4, 4* and 4a, then 4a lies -4.6σ from the centre of the distribution.

The different parameters of a given sunspot cycle tend not to be completely independent of each other. Thus an inverse correlation with a correlation coefficient, r_c, of -0.3 to -0.4 between cycle amplitude and cycle length is well known (e.g., Dicke 1978; Hoyng 1993): stronger cycles have a tendency to be shorter, as can be seen from Fig. 8. The amplitude is defined here as the maximum in the 12-month running mean of monthly data. In Fig. 8 cycles 1–3 and 5–22 are represented by diamonds. Cycle 4 is given by an open square, cycles 4* and 4a by filled squares. The solid, dashed and dotted lines show linear regressions for the standard case (i.e. fits to cycles 1–22), the standard case with cycle 4 dropped and the case with the split cycle 4 (i.e. fits to 1–3, 4*, 4a, 5–22) respectively. The corresponding values of the correlation coefficient r_c are -0.41, -0.44 and -0.01, which implies that cycle 4 is statistically not untypical, whereas cycle 4a is really exceptional.

Another test is based on the overall profiles of sunspot cycles. These are asymmetric, with rising phases being shorter than the declining phases. Is the shape of the standard cycle 4, with a long declining phase, untypical? Let us compare profiles of different cycles. We take skewness, s (i.e. the 3rd moment as defined below), as a quantitative measure of a cycle's profile:

$$s = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{R_g - \langle R_g\rangle}{\sigma_{R_g}}\right),$$

where N is the number of monthly R_g records within the cycle and σ_{R_g} is the standard deviation of R_g from the mean value $\langle R_g\rangle$. In Fig. 9a we show the correlation of s with R_g averaged over the cycle, in Fig. 9b the correlation with cycle length. The skewness shows a very good anti-correlation ($r_c = -0.71$) with the mean R_g. The standard cycle 4 fits well into this relation, whereas cycle 4a again deviates – it is much more asymmetric than any other observed cycle. Even worse is the situation with the skewness vs. cycle length relation. As Fig. 9b shows, shorter cycles are generally more symmetric. For the standard cycles 1–22, the correlation is not particularly strong but statistically significant ($r_c = 0.41$). Cycle 4a, being the shortest and at the same time the most asymmetric cycle, is a complete outlier in Fig. 9b.

6. Concluding remarks

Based on the scarcity of sunspot observations in the years 1792–1793, the unusual length of cycle 4 and the change in phase of the relative strength of odd and even cycles at around that time, Usoskin et al. (2001) proposed that sunspot cycle 4 is composed of 2 shorter cycles, which we call 4* and 4a. Here we compare the two hypotheses, that of a single cycle 4 vs. two cycles in the light of additional data, statistical studies of the sunspot number record and a comparison with sunspot cycle properties. We have statistically tested how likely it is that the number of sunspots seen during the few observations

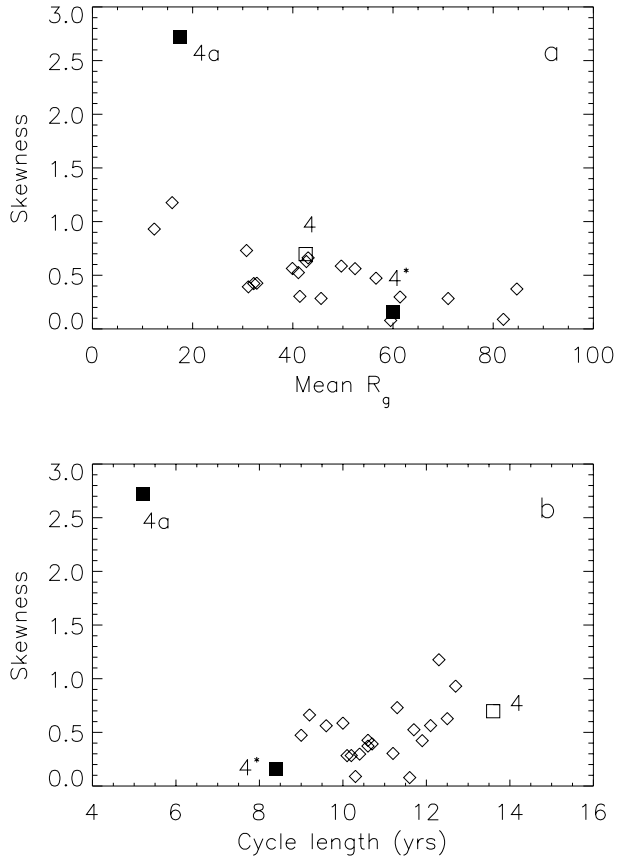

Fig. 9. a) Skewness of sunspot cycles vs. cycle averaged R_g. **b)** Skewness vs. cycle length. Symbols are as in Fig. 8.

carried out in 1792–1793 would be recorded if these observations had been made during a bona fide sunspot minimum. We find that the observed values agree far better with the sunspot cycle records during the declining phases of other, well sampled sunspot cycles than with the minima associated with these cycles.

Usoskin et al. (2001) argue that these observations are unreliable. This is indeed the case for some (we excluded from our analysis the observation by Huber criticised most strongly by them), but we do not see an obvious reason why the observations should systematically overestimate the number of detected sunspots. Nevertheless, the above conclusion remains unchanged even if we additionally exclude the observation giving the highest group sunspot number during 1792–1793 (besides that of Huber). Consequently, there appears to be no basis for the first argument of Usoskin et al. (2001) that the scarcity of observation in 1792–1793 could hide a sunspot cycle minimum.

Furthermore, the presence of a naked-eye sunspot in chinese records presents an additional statistical argument against 1792/1793 being the time of a sunspot minimum.

To complement the sunspot data for this period we have additionally considered ^{10}Be, ^{14}C and auroral records. These also do not provide any evidence for an extra cycle. Although on its own each of these data sets cannot rule out that possibility, taken together these records provide rather strong support for the standard cycle numbering.

Finally we considered which solution – 1 or 2 cycles – provides better agreement with the properties of the other solar cycles. Here the result is quite clear: the standard cycle 4 fits into sunspot cycle statistics just as well as the cycle 4* proposed by Usoskin et al. (2001) to replace it. For example, as Usoskin et al. (2001) correctly point out cycle 4 is the longest cycle on record since 1750. Equally, however, cycle 4* is the shortest. The new and very short cycle 4a proposed by Usoskin et al. (2001) is by contrast a very unusual cycle indeed. Not just does it lie 4.6σ away from the centre of the distribution of cycle lengths, but is also a true outlier when considering other properties of a sunspot cycle, such as the correlation of the length of a cycle with its strength as well as the skewness of the cycles. Hence the second argument of Usoskin et al. (2001) is also found not to hold.

The only argument remaining for two cycles is the phase shift of the Gnevyshev-Ohl rule. It is questionable whether such a phase shift indeed calls for an additional cycle. It may well be that a phase shift is occasionally produced by the solar dynamo. A phase shift alone therefore cannot counter the whole body of evidence that we have presented against the presence of a cycle minimum around 1792–1793. All the above arguments taken together provide very strong support for the proposal that cycle 4 was a single sunspot cycle and that no cycle has been lost.

Acknowledgements. We are grateful to T. Pulkkinen and W. Schröder for providing auroral data and related information. We also acknowledge helpful discussions with R. Muscheler.

References

Beer, J., Baumgartner, S. T., Dittrich-Hannen, B., et al. 1994, in The Sun as a Variable Star: Solar and Stellar Irradiance Variations, ed. J. Pap, C. Fröhlich, H. Hudson, & S. Solanki (Cambridge: Cambridge Univ. Press), IAU Colloq., 143, 291
Beer, J., Blinov, A., Bonani, G., et al. 1990, Nature, 347, 164
Cliver, E. W., & Ling, A. G. 2001, ApJ, 551, L189
Dicke, R. H. 1978, Nature, 276, 676
Fligge, M., Solanki, S. K., & Beer, J. 1999, A&A, 346, 313
Hoyng, P. 1993, A&A, 272, 321
Hoyt, D. V., & Schatten, K. H. 1998, Sol. Phys., 179, 189
Krivsky, L., & Pejml, K. 1988, Publ. Astron. Inst. Chech. Acad. Sci., 75 (data are available from NGDC)
Legrand, J.-P., & Simon, P. A. 1987, Ann. Geophys., 5A, 161
Schröder, W. 1992, J. Geomag. Geoelectr., 44, 119
Schröder, W. 2002, Beitr. Geschich. Geophys. Kosm. Phys., 3, No. 5, in press
Solanki, S. K., Schüssler, M., & Fligge, M. 2000, Nature, 408, 445
Solanki, S. K., Schüssler, M., & Fligge, M. 2002, A&A, 383, 706
Stuiver, M., & Braziunas, T. F. 1993, The Holocene, 3, 289
Stuiver, M., & Quay, P. D. 1980, Science, 207, 11
Stuiver, M., Reimer, P. J., & Braziunas, T. F. 1998, Radiocarbon, 40, 1127
Usoskin, I. G., Mursula, K., & Kovaltsov, G. A. 2001, A&A, 370, L31
Usoskin, I. G., Alanko, K., Mursula, K., & Kovaltsov, G. A. 2002a, Sol. Phys., submitted
Usoskin, I. G., Mursula, K., Solanki, S. K., Schüssler, M., & Kovaltsov, G. A. 2002b, J. Geophys. Res., submitted
Waldmeier, M. 1961, The sunspot-activity in the years 1610–1960 (Zürich: Schulthess)
Yau, K. K. C., & Stephenson, F. R. 1988, QJRAS, 29, 175