Are the sunspots really vanishing?

Anomalies in solar cycle 23 and implications for long-term models and proxies

Frédéric Clette* and Laure Lefèvre

SIDC – Royal Observatory of Belgium, 3 avenue Circulaire, 1180 Bruxelles, Belgium

*corresponding author: e-mail: frederic.clette@oma.be

Received 29 February 2012 / Accepted 9 June 2012

ABSTRACT

Context: The elapsed solar cycle (23) ended with an exceptionally long period of low activity and with unprecedented low levels for various series of solar irradiance and particle flux measurements. This unpredicted evolution of solar activity raised multiple questions about a future decline of the solar cycles and launched a quest for precursor signs of this possible deep solar transition over the last decade.

Aim: We present here a review and overall interpretation of most current diagnostics of solar cycle 23, including the recent disagreements that appeared among solar reference indices and standard solar-based geo-indices, the indication of a changed pattern of internal torsional waves (helioseismology) or the announced fading and magnetic weakening of sunspots.

Methods: Based on a statistical analysis of detailed sunspot properties over the last 24 years, we complete the picture with new evidence of a strong global deficit of the smallest sunspots starting around 2000, in order to answer the question: are all sunspots about to disappear?

Results: This global scale-dependent change in sunspot properties is confirmed to be real and not due to uncontrolled biases in some of the indices. It can also explain the recent discrepancies between solar indices by their different sensitivities to small and weak magnetic elements (small spots). The International Sunspot Index \( R_i \), based on unweighted sunspot counts, proved to be particularly sensitive to this particular small-scale solar evolution.

Conclusions: Our results and interpretation show the necessity to look backwards in time, more than 80 years ago. Indeed, the Sun seems to be actually returning to a past and hardly explored activity regime ending before the 1955–1995 Grand Maximum, which probably biased our current space-age view of solar activity.

Key words. Sun – solar cycle – sunspots – solar activity – space climate

1. Introduction

The last solar activity cycle (23) was marked by an abnormally long duration of 12.6 years, due to the exceptionally long duration of the cycle 23–24 minimum. Almost none of the 50 solar cycle predictions collected by the Solar Cycle 24 Prediction Panel predicted an abnormally late maximum compatible with cycle predictions collected by the Solar Cycle 24 Prediction Group (Pesnell 2008), which suggests panel predicted an abnormally late maximum compatible with cycle predictions collected by the Solar Cycle 24 Prediction Group (Pesnell 2008), which suggests that a lot of progress is still needed in our understanding of the long-term evolution of solar activity.

The last minimum was not only remarkable by its long duration, but also by the very low values reached by different measured solar fluxes and activity indices. Many series spanning only the last decades (1–6 solar cycles) even reached unprecedented values. Exceptionally low levels of the UV irradiance were measured by several instruments. Relative to the solar cycle 22–23 minimum, the SEM UV fluxes were 10% lower and the MgII core-to-wing ratio, a key proxy for the chromosphere and low corona, was 10% lower, an unprecedented level (Didkovsky et al. 2010; Solomon et al. 2010, 2011; Woods 2010b; Lean et al. 2011a). Tapping & Valdés (2011) also show a drop of the minimum \( F_{10.7} \) radio flux below the average of the four previous minima, but it is comparatively small (3%).

In response to those low fluxes, the density of the Earth thermosphere dropped 20% lower than earlier minima (Emmert et al. 2010; Solomon et al. 2010, 2011), while in the ionosphere, the \( f_{F2} \) index was depressed by 10% (Coley et al. 2010; Chen et al. 2011; Lean et al. 2011b; Liu et al. 2011b).

Regarding the solar corpuscular output, record low solar wind densities, pressure and interplanetary magnetic fields (IMF) were recorded. The IMF strength was 30% below past minima (McComas et al. 2008; Smith & Balogh 2008; Fisk & Zhao 2009; Heber et al. 2009; Cliver & Ling 2010; de Toma et al. 2010; Jian et al. 2011). This weakening can at least partly be associated with a parallel weakening of the observed polar magnetic field, by 40% (Sheeley 2008; Wang et al. 2009; Janardhan et al. 2010). In response, one could expect a corresponding increase of the galactic cosmic ray (GCR) flux recorded at Earth, in particular at low energy. While an increase by 20% was indeed observed in the low-energy GCRs compared to the previous minimum (Heber et al. 2009; McDonald et al. 2010; Mewaldt et al. 2010; Jian et al. 2011), no significant excess was observed relative to the range of the previous four solar minima. This apparent insensitivity of the GCR flux to the weaker solar wind can be due to a non-dipolar configuration of the corona and heliosphere, which is anomalous in the minimum phase of the solar cycle. This complex heliospheric
configuration could explain the unusual solar wind properties recorded at Earth, i.e., in the ecliptic plane (de Toma 2010).

This exceptional minimum thus led to a completely new and unexplored regime in the Sun-Earth connection. However, solar indices and fluxes did not only stray out of the maximum range recorded thus far but the mutual relations between those quantities seem to have changed as well.

2. Index discrepancies

Since about 2000, back at the maximum of cycle 23, a divergence was observed between the two primary long-term solar indices, namely the international sunspot index $R_i$ and the $F_{10.7}$ radio flux, with the $R_i$-based $F_{10.7}$ proxy falling 15% below the measured $F_{10.7}$ flux (Svalgaard & Hudson 2010; Lefèvre & Clette 2011; Lukianova & Mursula 2011; Fig. 1).

This disagreement first called for a careful verification of any source of possible biases in those indices. Regarding $R_i$, a key element determining the long-term stability of the index is the pilot station of the network used by the World Data Service “Sunspot Index” (Clette et al. 2007): the Specola Solare Ticinese observatory in Locarno. A 50-year long intrinsic trend control of Locarno observations was carried out, based on the continuous record of the sky quality and on cross-comparisons between the counts from the primary observer and from several other observers making observations in alternation. Those tests do not show any significant trend, in particular over the last decade (Bianda 2011, priv. comm.). A systematic comparison between the Locarno sunspot number and a series of independent individual stations indicates a change since about 1998. Primarily, the dispersion between the stations increases strongly after 1998 compared to the previous solar cycle and initial rise of cycle 23. On average, $R_i$ forms the lower envelope of all the other stations (Lefèvre & Clette 2011). Still, with various amplitudes, all stations give values that fall below the $F_{10.7}$ proxy for $R_i$ (Johnson 2010), while they agree with this proxy before that time. As the trend is common to many different stations, this confirms that it is a real solar effect, but intrinsic variations at individual sites prevent a quantitative assessment.

In this respect, the independent index that is the most comparable with $R_i$ is the American sunspot number $R_A$ produced by the AAVSO also from a statistical treatment over an international network of more than 40 stations. A comparison with $R_i$ (Fig. 2) shows large disagreements in the early part of the series before 1990, when the $R_A$ index suffered from now well-identified processing flaws (Schaefer 1997; Coffey et al. 1999; Hossfield 2002). On the other hand, the agreement with $R_i$ is excellent over the last two decades and in particular since 2000. This gives an additional indication that the absolute scale of both independent sunspot indices remained stable and consistent within 5% on average over the entire cycle 23, excluding a bias or trend as high as the 15% discrepancy with $F_{10.7}$.

Next to the $R_i$-$F_{10.7}$ comparison, it turns out that similar disagreements were found in classical proxies based on the $F_{10.7}$ flux. Since 2000–2003, UV flux proxies relevant to Earth aeronomy overestimate the actual fluxes by 10 to 15% (Kane 2003; Solomon et al. 2010; Chen et al. 2011; Lean et al. 2011b; Liu et al. 2011a, 2011b; Wintoft 2011). This breakdown of standard pre-2000 proxies is also reported for $F_{2}$, TEC and thermospheric density, with typical overestimates by >10% (Bergerot et al. 2010; Bruinsma & Forbes 2010; Emmert et al. 2010; Lühr & Xiong 2010; Lean et al. 2011b; Liu et al. 2011b). This indicates that using chromospheric standards like the $F_{10.7}$ or near UV fluxes leads to an underestimation of the actual drop in coronal emissions (EUV and X-ray irradiance, and solar wind).

Overall, this recent disruption in statistical relations that had remained valid since the start of the corresponding measurements suggests a true and deep change in solar activity and solar output, without equivalent over the last six solar cycles. So, this prompted us to look further back in time and check if similar episodes occurred before and to determine how exceptional the past extended cycle is in a multi-century perspective.
3. Cycle 23 versus the past history

Using the sunspot index record of the last 24 cycles for which the data coverage is the most comprehensive, we first superimposed the declining phases of all cycles by aligning them on a tie point corresponding to the downward crossing at \( R_i = 25 \) in the smoothed monthly sunspot number (Fig. 3). This superimposed-epoch plot first indicates that the recent 23–24 minimum was not an extreme case. It also shows that the cycle rise phases after a minimum are clustered in four families: fast, moderate, late and weak rises. The latter group corresponds to cycles 4, 5 and 6, belonging to the Dalton minimum. The 23–24 minimum matches the group of late rises, in contrast with the previous minimum 22–23 which was an example of the fastest rises. Among the slowly rising cycles, the most recent example is the 14–15 minimum dating back from 1913. We note also that, assuming a recurrence of past activity patterns, this comparison of all cycle minima suggests that a low cycle 24 should follow this rather long minimum, with a maximum \( R_i < 90 \).

What also characterized the last minimum was the succession of long periods without any sunspot and the high overall total of spotless days (817 days). In Figure 4, we plotted the total number of spotless days for all activity minima of the last 250 years. Although the peak of spotless days for 23–24 is rather high, it is exceeded by four other minima and comes close to intermediate values found for most past solar cycle, with the exception of recent cycles 19–22. Here, on the contrary, the uninterrupted succession of low spotless days counts of cycles 19–22 forms a unique episode over more than 200 years of systematic sunspot observations.

Therefore, while cycle 23 looks more like a return to a normal moderate activity regime, the recent volley of strong cycles rather marks a rare enhanced-activity anomaly at secular timescales, identified as a Grand Maximum (Usoskin et al. 2007; Abreu et al. 2008). Modern measurements and proxies, as well as their interpretation, may thus suffer from a Grand Maximum bias, as they essentially rest on data spanning only those last 50 years. Simply extrapolating modern observations and models over the distant past may thus prove risky and unreliable.

The above indications of a global and unpredicted solar transition marked by this deep minimum then raise a new question: Can other indicators or precursor signs of a global physical change inside the Sun be identified in current or past data?

4. Odd changes during solar cycle 23

4.1. Changes in zonal and meridional flow patterns

Recent flux-transport dynamo models show the crucial role of superficial and meridional flows in the production and growth of toroidal magnetic fields feeding each solar cycle.
meridional flux transport could explain the weaker dipolar magnetic field (Wang et al. 2009; Janardhan et al. 2010).

8.5 m/s in 2000 to a maximum of 13.0 m/s in 2007–2009, 2011). Following a 11-year modulation, the velocity rose from 2010, 2011; Hathaway & Rightmire 2010, 2011; Komm et al. 2011). It thus took a longer time to reach the upper latitude of 4.5°/year, compared to 5.1°/year for cycle 23 (Howe et al. 2009, 2011; Antia & Basu 2010, 2011; Komm et al. 2011). It thus took a longer time to reach the upper latitude of spot formation (~30°), explaining the late onset of the current activity cycle.

Moreover, the poleward branch corresponding to the migration of cycle 23 magnetic “following” fluxes to the pole was still missing as of late 2010. As in a flux-transport scheme, this migration induces the polar reversal at the next solar maximum and feeds the new dipolar field, this raises a further issue: Will there be no cycle 25, and thus will a Grand Minimum follow cycle 24? As we have just barely reached the phase of the cycle when this branch is expected to form and cycle 24 was also particularly late, this must be confirmed with more data from cycle 24 collected over the next 2 or 3 years.

4.2. Fading sunspots

Next to those global patterns, detailed high-precision measurements of the peak magnetic fields and darkness contrast inside the core of sunspot umbrae were carried out systematically during 10 years by Livingston and Penn in the infrared neutral Fe line at 1564.8 nm (Livingston & Penn 2009; Penn & Livingston 2011). By compiling a set of 2950 sunspot observations made every 3 months (monthly over recent years), they find a steady decrease of the average magnetic field at a rate of 50 Gauss/year (Fig. 5). A simultaneous decrease of the IR contrast of sunspot umbrae measured in SoHO/MDI magnetograms over cycle 23. The smoothed monthly sunspot index is overplotted to show the evolution of the solar cycle (black curve). The overlaid green line corresponds to the linear trend measured by Penn & Livingston (2011). The red curve corresponds to a 11-year periodic modulation, while the blue line connecting the two last minima suggests a slight decrease in the recent minimum, though much lower than the Penn-Livingston trend (Figure adapted from Watson & Fletcher 2011).

However, a parallel study by Watson & Fletcher (2011) was made using SoHO/MDI magnetograms, which have a lower measuring accuracy by a factor of four for individual measurements but includes all sunspots since 1996 (30084 spots). Thanks to the larger sample and to the longer time interval, which also includes the rising phase of cycle 23, their data suggest that the average sunspot magnetic field may follow a solar cycle modulation instead of steady trend (Fig. 5). They also find that the last minimum was slightly below the previous one. Although this difference is barely significant, it implies that the corresponding slope cannot be more than half the Penn & Livingston value (23 ± 4 Gauss/year). Using the same IR line as Livingston & Penn (2009), Rezaei et al. (2012) also conclude on a solar cycle modulation and the absence of a trend. Only measurements collected during the current rising phase of cycle 24 up to the next maximum will tell if the past sunspot field decline will reverse and follow the cycle or if it forms a new persistent anomaly.

5. A size-dependent sunspot deficit

In order to probe such global changes in sunspot properties and diagnose the simultaneous drift of the sunspot index $R_l$ versus other indices, we exploited the combined information of two rich sunspot catalogs spanning the last two solar cycles.
average, such weak features would also exceed the threshold and vanish entirely, thus reducing the count. On above the minimum threshold of 1500 Gauss may fall below slightly lose contrast, the smallest spots with core field just all decline of the sunspot core field. While large spots will only deficit of small spots may be another manifestation of the over-

More detailed information about the catalogs and statistical methods can be found in Lefèvre et al. (2011) and Lefèvre & Clette (2011).

By comparing the distribution of sunspots and sunspot group sizes and morphology (McIntosh classification) during cycles 22 and 23, we find that the number of large spots and of groups containing at least a large spot at maximum development (types C, D, E, F) is almost equal for both cycles.

On the other hand, the number of small spots without penumbra and of groups without large spots (types A and B) dropped by a factor of more than 2 in cycle 23, in particular after 2000, i.e., between the two peaks forming the maximum of that cycle (Fig. 6). This small-spot deficit is present not only in small isolated active regions but is also observed in small secondary spots in large groups (type C, D, E, F) though with a lower ratio of about 1.3 (Fig. 7). This deficit is confirmed by a similar study, which only considers group sizes but covers four solar cycles 19–23 (Kilcik et al. 2011) and indicates that cycle 23 differs from all cycles 19–22.

Considering the sunspot group lifetime, Lefèvre & Clette (2011) find a corresponding deficit of sunspots with the shortest lifetimes. As small spots are also short lived, it confirms the above deficit. In addition, it indicates that the deficit only affects a subpopulation of the A and B types: those with lifetimes below 5 days. This result may be related to an overall trend in growth/decay rates of sunspots found by Javaraiah (2011): on average, sunspot groups have a slower growth and faster decay in cycle 23.

Considering the simultaneous magnetic fading of sunspots mentioned in the previous section, it is possible that this global deficit of small spots may be another manifestation of the overall decline of the sunspot core field. While large spots will only slightly lose contrast, the smallest spots with core field just above the minimum threshold of 1500 Gauss may fall below the threshold and vanish entirely, thus reducing the count. On average, such weak features would also exceed the threshold only for a shorter time, which would lead to a reduction of average lifetimes.

6. Diverging indices: an interpretation

6.1. $R_i$ versus other indices

We can expect that the above deficit affecting only the small sunspots will leave a different imprint in different solar indices. Indeed, the international sunspot index $R_i$ gives an equal weight to all sunspots and to all sunspot groups regardless of their actual size. Therefore, $R_i$ will have a good sensitivity to the small-scale deficit, especially as the smallest spots and groups largely outnumber the large ones. By contrast, the value of other indices based either on total areas (sunspot area and CaII plage indices) or on total fluxes ($F_{10.7}$, MgII c/w, SEM, etc.) will be largely dominated by the largest active regions and magnetic features. This weighting in favor of the largest structures is an advantage for some applications (e.g., irradiance proxies) but we expect that it makes those indices only weakly sensitive if not entirely blind to the small scale but global change now observed on the Sun.

Likewise, we speculate that the small-spot deficit may provide a clue to the disagreements between chromospheric and other photospheric and coronal indices. Indeed, assuming that the disappearance of small spots is associated with the weakening of the corresponding magnetic fields below the 1500 G threshold, vanishing spots then become intermediate magnetic flux elements (100–1500 G), which must then bring an additional contribution to the plage and facular components. In the chromosphere, the photospheric sunspot deficit is thus transformed into a chromospheric excess ($UV, F_{10.7}$) which may at least partly compensate and mask the magnetic weakening and lead to smaller index decreases compared to photospheric indices. On the other hand, at shorter coronal wavelengths (EUV, X), there is no equivalent mechanism reversing a deficit into an excess due to the global sunspot fading. An additional contribution from low-latitude coronal holes may have actually amplified the decline of coronal emissions during the last minimum (de Toma 2010; Woods 2010a).
6.2. Implications for dynamo theories

The sunspot weakening and in particular, the fact that only small sunspots show a global change, suggests that a scale-dependent change has occurred and involves specifically the small magnetic elements emerging at the solar surface. This decoupling between small- and large-scale elements seems more difficult to explain with a single source mechanism for all sunspots. So, the change observed over the last cycle should bring up the question: Are there two separate mechanisms for spot formation and decay?
Currently, three possible mechanisms can be invoked for the conversion of toroidal field to dipolar field in dynamo models of the solar cycle (Muñoz-Jaramillo et al. 2010; Nandy 2011; Nandy et al. 2011):

- turbulent diffusion,
- meridional flow,
- downward pumping.

Their relative contribution is currently very uncertain. If more than one of those mechanisms is at work, the relative strength of those mechanisms can also vary independently over a solar cycle and from one cycle to the next. Moreover, in models, those mechanisms show strong radial variations just below the surface. So, different mechanisms may be dominant at different depths and thus will have different effects on small and large magnetic elements.

A few models invoking the existence of a superficial dynamo have been proposed over the past few years. Brandenburg (2005) develops a distributed dynamo involving the shear layer just below the surface, while Schatten (2009), using a cellular-automata model, proposes a mechanism of near-surface aggregation of magnetic flux elements acting at a local scale and leading to the formation of small spots. The recent odd behavior of cycle 23 may provide new support to those ideas and new constraints to models. It thus calls for new modeling efforts to address the question: Is a local near-surface dynamo at work next to the deep-seated global dynamo and how is such a superficial dynamo coupled with the deep dynamo and the 11-year cycle?

7. Conclusion

The odd and unpredicted changes that marked cycle 23 and the deep and long 23–24 transition obviously question various aspects of our current understanding of the long-term evolution of solar activity. With all elements collected here, we should thus clearly reformulate our title “Are sunspots really vanishing?” as: Which sunspots are vanishing? When and why?

A key conclusion is the following: the disruption of proxy relations between standard indices and the parallel global scale-dependent change in sunspots are strong indications that long-term irradiance and solar forcing reconstructions require more than a linear rescaling of current proxies. Recent findings presented in this review suggest that the change took place at a local scale in individual sunspots inside active regions. Therefore, this means that further progress will require the inclusion of information at the level of individual active regions (size, morphology, growth/decay). Moreover, as a Grand Maximum bias may affect the current standard indices and proxies, new proxies should be based on information over the actual distribution and parameters of individual active regions at epochs of moderate and low solar cycles, in order to remain valid and accurate over past centuries for solar cycles before cycle 19.

Clearly, issues of the current and future solar cycles can only be addressed by connecting the current abundant solar and space weather data to the past record of solar activity (including Grand Minima, when the solar cycle largely vanishes and solar activity stays at its lowest level). As this record consists primarily or even exclusively of sunspot observations beyond the past 50–100 years, it highlights the renewed importance of recovering sunspot data before the mid-20th century (photographic plates, drawings). This quest was partly included in recent or current FP7-funded European projects (SoTerIA, COST-ES1005-TOSCA, COMESEP). In this context, new detailed sunspot catalogs are in construction and approaching completion (ROB USET catalog, Debrecen Photographic Data catalog) using dedicated digitization tools (e.g., Fig. 8). However, few catalogs include data older than ~1950 and they are sparse and incomplete, which limits their use for global homogeneous studies.

Even before a sunspot catalog or database can be built, the base observations must first be available in digital form. It turns out that many collections (photographic plates, visual drawings) are not digitized and thus largely inaccessible to scientific exploitation. A good recent illustration of the potential of such a data recovery work was given by Artl & Abdolvand (2011), with the reconstruction of the butterfly diagram for cycles in the
19th century, using all original drawings from Schwabe (discoverer of the 11-year cycle). A rich unexploited information going beyond the global daily sunspot number is thus available and can directly address the above mentioned far reaching issues.

Finally, one issue remains open: will the recent changes persist, reverse or amplify in cycle 24? Although we are still only in the early rise phase of cycle 24, the latest index values and ratios suggest a reversal of the trends observed observed in cycle 23 and a recovery, after the extreme values of the past minimum (see Figs. 9 and 10). It is still unclear if all indicators will finally return to their pre-2000 values. If they do, such a return to pre-cycle 23 values would bring another strong confirmation that indices did not get flawed but faithfully tracked an unexpected solar evolution. Continued observations are needed in coming years to establish if cycle 24 indeed proves to be again similar to past cycles or is heralding a true enduring transition.

A hint of what to expect can be found in the 25-year perspective offered by the sunspot index series. For this, we looked for the best fits between the early rise phases of the last 24 cycles and the current rise of cycle 24 over the last 12 and 36-months respectively (Figs. 11 and 12). We can then obtain the observed ranges for cycle amplitude and time of maximum, without including any hypothesis. The 12-month fit indicates that the recent quite steep rise of cycle 24 is only compatible with rather strong solar cycles (maximum $R_i$ of 100–165; Fig. 11). By contrast, the 36-month fit, which also takes into account the past extended minimum, is only compatible with lower past cycles (80–120; Fig. 12). On the other hand, both 12- and 36-month fits lead to a similar range for the time of maximum. The next maximum could occur between +1.4 and +2.5 years after the last monthly smoothed $R_i$ value of July 2011, that is, between 2012.9 (November 2012) and 2014.1 (January 2014). The latter value corresponds to a rather long maximum (cycle 16) with a very flat plateau following an early switching-off of the initial steep rise, which may occur around +1 year, that is, already by mid-2012. Considering the unpredicted anomalies of cycle 23, even such a peculiar scenario should thus be kept in mind.

Acknowledgements. This work partly rests on funding from the European Community’s Seventh Framework Program (FP7/2007–2013) under Grant Agreement No. 218816 (SOTERIA project, www.soteria-space.eu). We also acknowledge the support from the Belgian Solar-Terrestrial Center of Excellence (STCE) funded through the Belgian Science Policy Office (BELSPO).

References


