Solar Phys
DOI 10.1007/s11207-016-0892-3

Detailed Analysis of Solar Data Related to Historical Extreme Geomagnetic Storms: 1868 – 2010

Laure Lefèvre[1] · Susanne Vennerstrøm[2] · Mateja Dumbović[3] · Bojan Vršnak[3] · Davor Sudar[3] · Rainer Arlt[4] · Frédéric Clette[1] · Norma Crosby[5]

Received: 4 September 2015 / Accepted: 2 April 2016

Abstract An analysis of historical Sun–Earth connection events in the context of the most extreme space weather events of the last ∼ 150 years is presented. To identify the key factors leading to these extreme events, a sample of the most important geomagnetic storms was selected based mainly on the well-known aa index and on geomagnetic parameters described in the accompanying paper (Vennerstrøm *et al.*, Solar Phys. in this issue, 2016, hereafter Paper I). This part of the analysis focuses on associating and characterizing the active regions (sunspot groups) that are most likely linked to these major geomagnetic storms.

For this purpose, we used detailed sunspot catalogs as well as solar images and drawings from 1868 to 2010. We have systematically collected the most pertinent sunspot parameters back to 1868, gathering and digitizing solar drawings from different sources such as the Greenwich archives, and extracting the missing sunspot parameters. We present a detailed statistical analysis of the active region parameters (sunspots, flares) relative to the geomagnetic parameters developed in Paper I.

In accordance with previous studies, but focusing on a much larger statistical sample, we find that the level of the geomagnetic storm is highly correlated to the size of the active regions at the time of the flare and correlated with the size of the flare itself. We also show that the origin at the Sun is most often a complex active region that is also most of the time close to the central meridian when the event is identified at the Sun. Because we are dealing with *extremely severe storms*, and not the usual *severe storm* sample, there is also a strong correlation between the size of the linked active region, the estimated transit speed, and the level of the geomagnetic event. In addition, we confirm that the geomagnetic events studied here and the associated events at the Sun present a low probability of occurring at

✉ L. Lefèvre
laure.lefevre@oma.be

1 Royal Observatory of Belgium, Brussels, Belgium

2 National Space Institute, DTU Space, Copenhagen, Denmark

3 Hvar Observatory, Faculty of Geodesy, University of Zagreb, Zagreb, Croatia

4 Institute for Astrophysics Potsdam, AIP, Potsdam, Germany

5 Royal Belgian Institute for Space Aeronomy, Brussels, Belgium

Published online: 29 April 2016

low sunspot number value and are associated mainly with the maximum and descending part of the solar cycle.

Keywords Historical data · Extreme events · Solar storms · Geomagnetic storms · Flares · Active regions · Sunspots · Statistics

1. Introduction

In a world where we rely more and more on electronic equipment in our day-to-day life, severe geomagnetic storms and their consequences on Earth are very important phenomena. In an effort to increase the statistics on these infrequent occurrences, we have gathered solar parameters for the most intense storms over the past ~ 150 years.

Geomagnetic storms are generally caused by coronal mass ejections (CMEs), which are often associated with solar flares and solar energetic particles (SEPs) (Gosling *et al.*, 1990; Reames, 1999; Koskinen and Huttunen, 2006; Zhang *et al.*, 2007). Ground-based observations of solar flares are usually performed in the Hα wavelength, but there are records of particularly intense white-light flares (Neidig and Cliver, 1983, Catalog of White Light Flares). Intense flares are often associated with the presence of unusually large and complex active regions on the Sun (Sakurai, 1970; McIntosh, 1990; Qahwaji and Colak, 2007; Qahwaji *et al.*, 2008; Colak and Qahwaji, 2009). CMEs are large bubbles of gas interweaved with magnetic field lines that are ejected from the Sun over the course of several hours. Flares and CMEs are closely related and appear to be different manifestations of a single physical process. Although CMEs are not always associated with large flares, if a large flare does occur, it has a very high probability of being associated with a CME (Priest and Forbes, 2002 and Yashiro *et al.*, 2006).

This article is the second of two companion articles that present a detailed analysis of solar terrestrial connections in the context of extreme space weather events. Based on the well-known aa index availability (Mayaud, 1980; Menvielle and Marchaudon, 2007), the most important geomagnetic storms since 1868 have been selected. Here we focus on what occurs on the Sun, *i.e.* the sunspots and solar flares, while Paper I focuses on what occurs in the interplanetary space and especially at Earth, *i.e.* SEPs, galactic cosmic ray variations, solar wind, and the geomagnetic events themselves.

Since our study encompasses historical events that date back to 1868, it is based on the *most probable associations* between flares and CMEs because of the limited availability of solar data back in time. Its focus is more particularly on solar active regions (hereafter ARs) that are *most likely* related to the above-mentioned extreme events. Of course, we should mention here that great storms do not necessarily originate in eruptive flares and active regions. They can sometimes arise from the disappearance of filaments outside of active regions (Joselyn and McIntosh, 1981) or have a corotating interaction region (CIR) source with a possible embedded CME (Crooker *et al.*, 1993). The main difference between a geomagnetic storm linked to a solar flare and a storm linked to a filament or a CIR is the speed of the associated interplanetary disturbance: it is lower for disappearing filaments than for flares (Cane, 1985; Cane, Kahler, and Sheeley, 1986). For a better view, we also considered the alternative possibilities (filament, CIR) on a case-by-case basis. We also note here that the use of an index that characterizes mid-latitude range geomagnetic variations (the aa index) instead of the Dst ring current index will place the emphasis on flare-related events (*i.e.* explosive) rather than on gradual eruptions, which are characteristic of disappearing filaments (Feynman, 1980). In short, the aa index favors high-speed (energetic) events because it more strongly depends on solar wind velocity than the Dst.

We show the specificity of the considered sample of events in terms of sunspot information because it is indeed the only source of reliable solar information available at the end of the nineteenth century. In this context, this study primarily relies on detailed sunspot catalogs, from which a survey was made in Lefèvre and Clette (2014), and on solar images and drawings. For events after 1996, detailed solar and interplanetary data are readily available, and solar–terrestrial connections are therefore easier to access through various detailed studies. Events that occurred after approximately 1934 during the epoch of the Hα patrol are still relatively easy to study. However, going back to 1868, solar data from catalogs become scarce, as is true for sunspot drawings or images. In this context, the most challenging and rewarding step of this analysis consisted of gathering the information, especially for events before the 1880s.

This work presents the ARs that are most likely associated with the extreme geomagnetic events listed in Paper I and their detailed statistical analysis. Sections 2 and 3 present the data and methods used in this study. Section 4 presents the inferred data associated with each listed geomagnetic event with emphasis on specific events, and Section 5 uses the data presented in the previous section to try and understand the connections between the geomagnetic impact and the presented solar parameters.

2. Data

In this section, we describe the different types of data that are used in this analysis, and the evolution of their availability throughout time. The first subsection briefly describes the data used to assess the level of geomagnetic impact. The next three subsections describe solar data, in the form of parameters of ARs as well as flare parameters. Section 2.5 describes the CME/Interplanetary CME (ICME) data, and the last subsection summarizes the distribution of this information through time. We recall that the availability becomes more sporadic with distance in time for all of the parameters presented here. The description of geomagnetic and CME/ICME data is kept to a minimum; they are described in more detail in Vennerstrøm *et al.* (2016) or Dumbović *et al.* (2015). These data are only summarized here to provide context because they are interconnected and cannot be taken as completely separate problems.

2.1. Geomagnetic Data

The aa index is a simple global geomagnetic activity index, with units of 1 nT. It provides a measure of global geomagnetic activity, which is very valuable because it extends continuously back to 1868 (Menvielle and Marchaudon, 2007). Based on the aa index, we selected the 105 largest storms, all with a peak in aa larger than or equal to 300 nT. In addition, we collected geomagnetic data from a number of individual observatories where long time-series existed. Together with the aa peak values and the peak values of 24-hour running means of aa, these data were used to rank the storms. Figure 1 displays the relation between the peak of the 24-hour running mean of aa (aa24 in the following) and the rank. More details on this selection and ranking process can be found in Vennerstrøm *et al.* (2016). In this article we focus on these 105 storms.

2.2. Visual Data: Images and Drawings of the Sun

We surveyed existing databases for images and drawings of the surface of the Sun in a window of ± 10 days around each geomagnetic storm because in addition to characterizing the active region before the storm, we are also interested in the overall evolution of

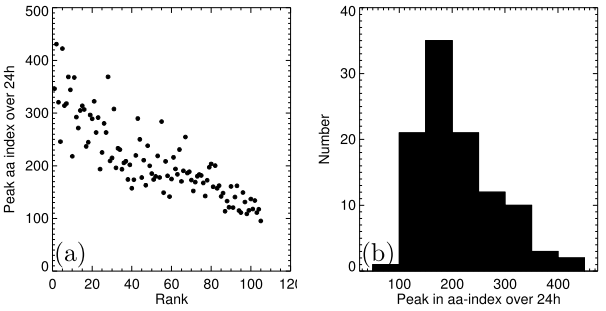

Figure 1 (a) Rank versus peak value of the aa index averaged over 24 hours for each of the 105 storms in our sample. (b) Distribution of the peak value of aa over 24 hours for our sample.

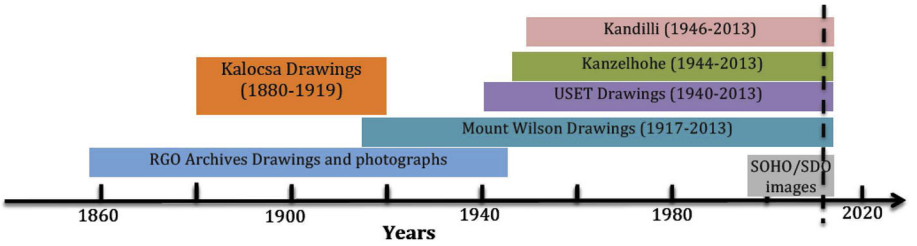

Figure 2 Description of the different sources for the images and/or drawings used in this analysis for the different time periods between 1868 to 2010.

the AR. Figure 2 presents the most important sources of images and drawings that were used in this analysis. Most of them are available online: the Kalocsa drawings can be found at ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-imagery/photosphere/sunspot-drawings/kalocsa, on the NGDC website. The Kandilli and Kanzelhöhe drawings go back to 1946 and 1944, respectively, and can be found at (http://www.koeri.boun.edu.tr/astronomy/) and (http://cesar.kso.ac.at/). The Uccle Solar Equatorial Table (USET) data is available at http://sidc.oma.be/uset/searchusetDrawing.php back to 1955. Drawings before 1955 are kept locally at the Royal Observatory of Belgium (ROB) and are only available upon request for outside queries.

The Mount Wilson drawings, scanned by R.K. Ulrich are available at ftp://howard.astro.ucla.edu/pub/obs/drawings/. For the oldest events, from 1868 to 1880, the only consistent source of images and drawings is the Royal Greenwich Observatory (RGO). These events are only accessible through the Cambridge University Library (UK: http://janus.lib.cam.ac.uk), and we took pictures of the drawings of each event, except for archive MS.RGO.51, which was scanned at the Mullard Space Science Laboratory and made available by S. Matthews. Images and drawings found in the RGO archives cover the period between 1858 and 1979.

2.3. Detailed Sunspot Data: Catalogs and Data Extracted from the Images and Drawings

For the most recent storms, we searched for associated AR parameters in the merged catalog from Lefèvre and Clette (2014) (and references therein). It was built from the very detailed Debrecen Photoheliographic Data (DPD, Győri, 1998; Győri *et al.*, 2005; Győri, Baranyi, and Ludmány, 2011) and data from the Solar Optical Observing Network (SOON) within the

US Air Force bases (USAF). This catalog now covers the period 1982 – 2011 and presents information on ARs: mainly sizes (area and number of individual sunspots), positions, and morphological information, *i.e.* McIntosh types (McIntosh, 1990). It also contains information on the evolution of every individual sunspot inside all the ARs. The USAF data covers group information from 1981 to 2014. For earlier data, we only have access to sunspot group data: we used the well-known RGO catalog of Erwin *et al.* (2013); Willis *et al.* (2013a, 2013b) combined with the USET local catalog created from USET drawings through a software called DIGISUN that was developed at the ROB. This combined catalog RGO/USET extends our time coverage backward from 1982 to 1940 (for the sunspot group information) because the first images from USET were taken in March 1940. Before this, events are covered partly by the RGO catalog (until 1874), partly by analyses of existing images and drawings extracted from the pictures of the RGO archives in the Cambridge University Library. From 1874 to the present, the time sampling is approximately one observation per day, which indicates that our time-resolution for size (or number of spots) variations is of about one day. This might seem a very low resolution, but as we show in this study, this is a sample of extremely large groups of sunspots whose average growth or decay rate is much lower than that for small groups or spots (Howard, 1992, 1993).

To complement these data, we used the work of Maunder (1904), who tried to link geomagnetic events to sunspot regions with the knowledge that was available at that time. This article provides detailed information on a few ARs, particularly the complete lifetime (recurring over a few solar rotations) of the largest regions present on the surface of the Sun at the time of important geomagnetic storms from the end of the nineteenth century. Available images and drawings from the seven events between 1868 and 1874 were analyzed by colleagues from the ROB and from the Astrophysical Institute Potsdam (AIP). In the latter case, a rotational matching of adjacent days was used to eliminate uncertainties in the orientation of the solar disk. The method is nearly identical to the one by Arlt and Fröhlich (2012), except that the differential rotation parameter was not a free parameter but was taken from Balthasar *et al.* (1986). These two independent analyses, based on two different methods, enable us to reach a better accuracy on the extracted sunspot data. Works from J. Vaquero and his team (Vaquero *et al.*, 2008, 2012; Carrasco *et al.*, 2013) also helped us collect important information for events where images were unavailable through other sources.

2.4. Flare Data

For this analysis, we used the flare data available in a time window around the geomagnetic events considered. This flare information, in the form of Hα patrol data, can be found on the NGDC website ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/h-alpha/reports. Flares were first called "eruptions" or "solar disturbances" and finally named flares by observers from the Mount Wilson observatory in the 1940s (Richardson, 1944; Cliver, 1995). In this catalog, flare data go back to April 1938. It starts with a simple flare index, then from 1957 contains the area of the flares (in millionths of the solar disk, μsd), and finally from 1982, it includes the X-ray flare information. The historical flare index is an observational index that combines both intensity and size of the flare, which makes it difficult to compare it to its later counterparts. In terms of flare importance, we therefore rely first on the historical index, second on the size of the flare itself, and last on the X-ray intensity measured. In addition, when the visual information is unavailable, the number of stations that have seen a particular flare can help assess its importance. For a few events, information on flares was found on Mount Wilson scanned images of

Table 1 Indices describing flare importance and the evolution of their nomenclature before (row 1) and after 1966 (row 2). Row 3 lists the common numerical scale used in this work.

Before 1966		1−	1	1+	2−	2	2+	3−	3	3+	
From 1966	S	1F	1N	1B	2F	2N	2B	3F	3N	3B	4
Numerical scale	1	1	1	1.5	2	2	2.5	3	3	3.5	3.5

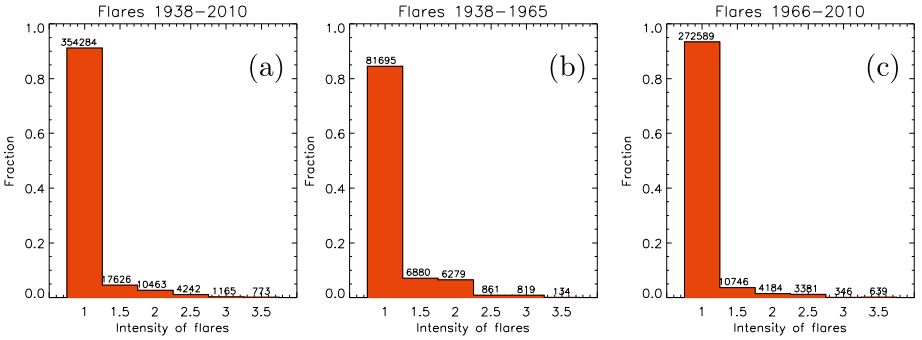

Figure 3 Distribution of the intensity of flares, described using the uniform scale from Table 1, (a) from 1938 to 2010; (b) from 1938 to 1965; (c) from 1966 to 2010 to identify possible differences linked to the evolution of the time coverage.

the Sun. Complementary information about older storms can be found in Newton (1943, 1944) and Hale (1929, 1931) in the IAU's Quarterly Bulletin of Solar Activity (QBSA, http://solarwww.mtk.nao.ac.jp/en/wdc/qbsa.html), in the Catalog of White Light Flares by Neidig and Cliver (1983), or in Cliver, Feynman, and Garrett (1990). Before the official start of the Hα patrol, George Ellery Hale observed flares at different locations and in particular at the Mount Wilson Observatory from 1917.

A second item that should appear here is the fact that from the beginning of the Hα patrol to the present time, the index associated with flare importance has evolved. On 1 January 1966, the scale was changed from 1−, 1, 1+, 2−, 2, 2+, 3−, 3, 3+ to S, 1, 2, 3, 4, with indications of brilliance F (faint), N (normal), and B (bright). The connection between these scales is described in Figure 3 of Švestka (1969). To create a single numerical scale spanning the whole set of Hα flare data used in this study, we adopted a scheme similar to that of Krivský (1973) and Ruždjak *et al.* (1989) (their Table I). Table 1 describes the scale before 1966 and after 1966 compared to the common numerical scale we used in our statistical analysis.

Figure 3 shows the distribution of these flares for the whole sample (a) as well as the variation in this distribution before and after 1966 (b, c). This intensity/surface scale, which can be roughly assimilated to an energy scale, shows a distribution close to the expected power-law described in Hudson (1991) and Schrijver *et al.* (2012) $N(> E_0) \sim E^{-\alpha}$, with $\alpha = 1.8$). Figure 3(a) also gives us a basis on which to compare the distribution of flares in our sample of ARs to the distribution of flares in all ARs during the period of our study (see Section 5).

2.5. Coronal Mass Ejection Data

CMEs constitute large-scale ejections of mass and magnetic flux from the lower corona into the interplanetary medium. The first CME was detected on 14 December 1971 by the white-

light coronagraph onboard NASA's seventh *Orbiting Solar Observatory* (Tousey *et al.*, 1973). CMEs are not preceded by flares, as had first been assumed (less than 20 % of all CMEs are associated with large flares, as explained in Gosling, 1993; CMEs that are associated with flares often appear to start before or at the same time as the onset of the flare: Wagner *et al.*, 1981; Simnett and Harrison, 1985; Zhang *et al.*, 2001), on the other hand, they are clearly related to a common origin. For a historical overview on CMEs see Howard (2006) and references therein. Statistical studies show that CME parameters, *e.g.* velocity or kinetic energy, are correlated with some characteristics of the associated flare, *e.g.*, the soft X-ray peak flux of the integrated flux, and moreover that CMEs associated with large flares are on average faster and larger than non-flare CMEs (Moon *et al.*, 2003; Maričić *et al.*, 2007).

The CME–flare relationship enables determining the exact source location of the CME and therefore gives important information on the CME properties, regarding the CME–ICME–geomagnetic storm association. We here rely on the spatial and temporal relations between flares and CMEs established by Vršnak, Sudar, and Ruždjak (2005) and on a more extensive study of CME–flare–geomagnetic storms by Dumbović *et al.* (2015). In addition, we used the ICME catalog by Richardson and Cane (2010), where a number of CME–flare–ICME–geomagnetic storm associations can be found. However, these detailed studies only extend back to 1996. Therefore, older CME information has to be found individually when available.

For the pre-CME detection era we relied on the CME–flare association and assumed that large geomagnetic storms in the pre-CME detection era can be associated with large flares, given the conveniently chosen time criteria. In addition, we used galactic cosmic ray (GCR) data as a means to identify ICMEs, since they are known to cause short-term GCR depressions, so-called Forbush decreases (FD, see *e.g.* Cane, Richardson, and St. Cyr, 2000 and references therein). They often have a specific asymmetric two-step structure and are closely time-related to the passage of the ICME (*e.g.* Cane, Richardson, and Wibberenz, 1996; Dumbović *et al.*, 2011). Therefore they can be used as a substitute for *in situ* measurements to identify ICMEs (in the pre-satellite era). Sometimes, geomagnetic storms are preceded by a sudden increase in the geomagnetic field, which is referred to as a sudden storm commencement (SSC) and is most often followed by an FD. To identify these FDs, we used hourly averaged count rates from nine neutron monitor stations (depending on the availability), corrected for atmospheric pressure taken from the SPIDR website http://spidr.ngdc.noaa.gov/spidr/ in the time period since 1957 to present (Vennerstrøm *et al.*, 2016). More details on the links between ICMEs and geomagnetic storms from this study can be found in Paper I.

2.6. Availability of Data Versus Time

Figure 4 summarizes the different types of data described in the previous subsections and their availability over time. We call the most recent period the *SOHO era*, from 1996 to 2010. The second period corresponds mostly to the availability of continuous flare data, and we therefore call it Hα *patrol era*. This is the time period from 1926 to 1996. It starts before the actual start date of the NGDC catalog covering the Hα patrol (1938) because consistent observations of flares started in 1917 (although observations of flares started as early as 1870), and some data could be recovered from Newton (1943, 1944) and the QBSA. In addition, the spectrohelioscope from the Mount Wilson was perfected by Hale in 1926. The *RGO era* extends from 1874 to 1926, and the earliest data period called *pre-RGO era* extends from 1868 to 1874. We added GCR information to this table as we compared the mean transit speed computed from FDs to the mean transit speed computed from the time of the maximum of the storm (see Section 4).

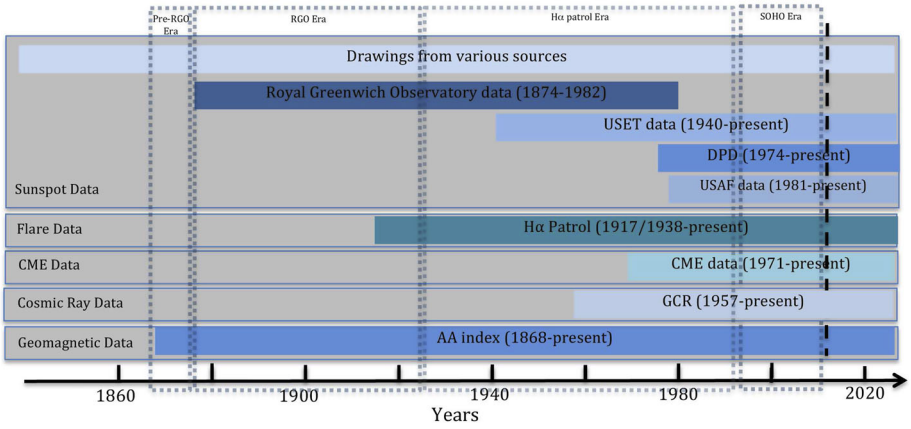

Figure 4 Available parameters and their time coverage. USET is the *Uccle Solar Equatorial Table* data from the ROB. USAF stands for data from the NGDC website (SOON stations). DPD stands for *Debrecen Photoheliographic Data*. RGO is the *Royal Greenwich Observatory Catalog* (NGDC website). Overlaid are the four time periods we chose to use in our study.

3. Methods

As our sample of geomagnetic storms spans almost 150 years, the data and their subsequent analysis present a strong dependency on the time period: the type of data available and the time coverage varies significantly, therefore we divided the methods used in this study into four sections, corresponding to the four different time periods described in the previous section and especially in Figure 4. We start from the most recent period, for which a large choice of solar data is available, and go back to the oldest events for which little information exists and time coverage is partial at best.

3.1. SOHO Era: 1996 – 2010

Following studies by Zhang *et al.* (2003) on a much smaller sample of less intense storms than in the current sample, we chose a four-day window before the geomagnetic storm as our most probable period for corresponding event(s) on the Sun: this gives us a window of approximately 100 hours. In this four-day window, we *chose the most extensive/intense flare*, in terms of flare index, flare area or associated X-ray flux, because it has been shown that geoeffective CMEs are statistically associated with stronger flares (Zhang *et al.*, 2007; Srivastava and Venkatakrishnan, 2004). Independent studies on CME–flare associations by Richardson and Cane (2010) and Dumbović *et al.* (2015) confirm the date and time of the flare that is chosen with this method.

In this context, we also checked the 13 events from this time period against the results from Zhang *et al.* (2007): all of our events can be found in Table 1 from Zhang *et al.* (2007). All of these events are characterized as ICMEs with shocks or current sheets (some are multiple or interacting ICMEs), and none of the events are associated with CIRs. In addition, Figure 5 shows that within our sample of events all of the associated flares occurred inside the selected window of four days. For this period, they even fit in a window of approximately two days.

We used the visual information in the form of drawings from the Observatories of Kanzelhöhe, Kandilli, or Uccle to assess the evolution of the ARs and their complexity on the solar

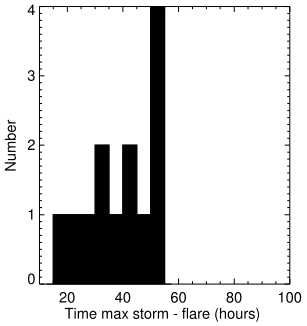

Figure 5 Time between the maximum of a solar flare and the maximum of a geomagnetic storm for associated events. The duration varies between 18 and 54 hours for the 13 events in the SOHO era (1996 – 2010).

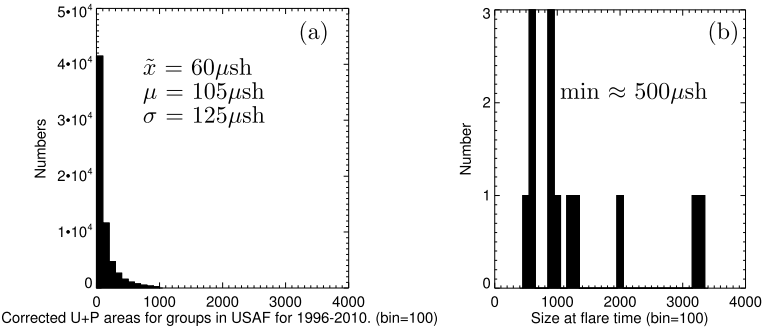

Figure 6 (a) Sizes of all ARs in 1996 – 2010 extracted from the USAF catalog in millionths of the solar hemisphere (μsh). (b) Size of the ARs considered in this study at the time of the flare for 1996 – 2010 (13 candidates).

surface during the solar rotation that encompasses the defined time window. Of course, we also took advantage of the available SOHO-MDI images (Scherrer *et al.*, 1995), but as this historical analysis rests mainly on data from the distant past, we focus on drawings first. We observe that all ARs linked to our sample of 13 geomagnetic events are extremely large and complex regions: sizes are at least 3 σ above the average size of sunspot groups on the same period (Figure 6), and associated ARs from this period present the most complex modified Zurich or McIntosh (Z from Zpc) types D, E, and F (McIntosh, 1990).

Figure 7 shows that 70 % of ARs are within 20 degrees of the central meridian at the time of the flares linked to major geomagnetic events, while approximately 85 % are within 40 degrees.

Then, we used the merged catalog from Lefèvre and Clette (2011) to derive the dynamical parameters of the ARs that were identified. We followed the variations of the AR size (area in μsh) and the number of spots, as well as the morphological type. Figure 8 shows that the size of the ARs seems to be affected by phenomena linked to the associated flare(s): the area shows *sudden changes* (strong variations in growth or decay rate, large slope, local minima or maxima in the size of the region) in a window of ±1 day around flare time. This is shown in Figure 8: identifiable large changes in growth or decay rate appear in 11 cases. Because we target changes in the complexity of the AR, the number of spots can be used as an indicator of change when the variations in size are not as evident. These *sudden changes* are also identified visually as a change of sunspot group configuration (*i.e.* magnetic configuration) in the daily drawings and images. These images and drawings enable a more complete view of the evolution of the complexity of the different ARs on the Sun. We note

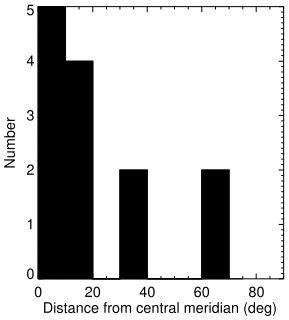

Figure 7 Absolute distance in degrees from the central meridian.

that this *suddenness* is defined with a time-sampling of one day for this historical sample, but because we examine a sample of extremely large groups of sunspots that evolve slowly, it does not have too much influence. Figure 9 shows the variations in size of the studied ARs in this time period with the much better SOHO time coverage (approximately one image per hour): this better sampling does not significantly improve the resolution of the curves from Figure 8.

We note that the goal here is to identify the solar counterparts of geomagnetic events listed in Paper I. These *sudden changes* appear to be almost necessary or favorable conditions for a major eruption or flare but, since no control study was made, they are not necessarily sufficient. However, for this specific study, considering the undeniable presence of a geomagnetic storm of important proportion, we assumed that an ICME was responsible and linked it to the most probable flare and associated AR. As Vennerstrøm *et al.* (2016) concluded that all, or almost all, of the extreme storms investigated here were associated with the passage of an interplanetary shock, this is a reasonable hypothesis. In this case, we can safely assume that the assumed flare, even if not recorded by other means, was associated with a *sudden change* in AR complexity. This is important for time periods when flare information is scarce or nonexistent.

In conclusion: (1) geomagnetic storms are not necessarily caused by ICMEs, but may also be triggered by CIRs (even the large ones), especially when a CME is embedded in the CIR. Based on the analysis of the events in this time period, however, none of the storms are associated with CIRs, they are all caused by ICMEs with shocks (some of them were multiple interacting CMEs). (2) CMEs are not necessarily associated with flares and might not originate from ARs. They can also be associated with erupting filaments and might not be associated with any low-coronal signatures (stealth CMEs). All of the 13 events from the SOHO era are associated with flares and originated in ARs, however.

Thus, the analysis of the events for this time period, where we have *all* the available data, gives us a good argument that the strongest storms are caused by CMEs associated with flares originating from large ARs. Of course, *events that do not follow this rule* carry important implications, but for events before 1996 and older events before the space era, a thorough examination is impossible because we lack observations in the Sun–interplanetary space–Earth chain. This means that a reliable identification of these *events that do not follow the rule* is very difficult, nigh on impossible. We have to assume, based on the analysis in the time period where we have *all* the available data, that most of the events follow the rule.

3.2. Extended Hα Patrol Era: 1926 – 1996

In this time period, we lack information linking flares to CMEs, thus to geomagnetic storms. Taking the above-mentioned hypothesis of a geomagnetic storm linked to an ICME, a flare,

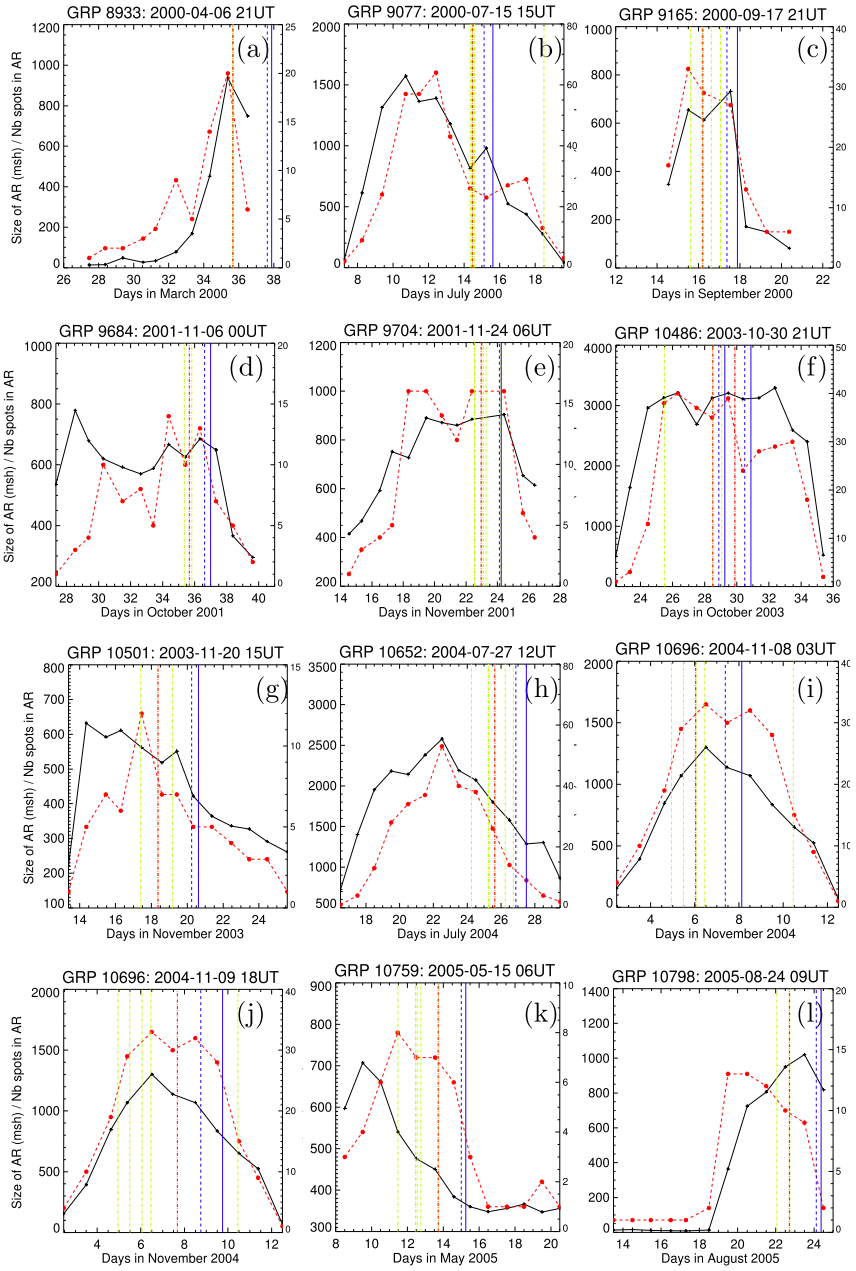

Figure 8 Variations of area (in μsh, in black, left-side scale) and number of spots (in dashed red, right-side scale) of active regions with time (in days) in our sample for the visible part of the solar rotation including each event from the SOHO period. Day numbering starts the first day of the month it appears in. Titles correspond to the NOAA AR number and the time of the associated geomagnetic storm. The time of the flare associated with the storm is plotted as a red vertical line, while other flares of index ≥ 2 from the same region are represented in green. In dashed blue we show the beginning of the storm, and in solid blue the time of the maximum of the storm on Earth. There are only 12 panels (for 13 storms) because case (f) in October 2003 (Halloween storms) shows two flares that caused two successive storms.

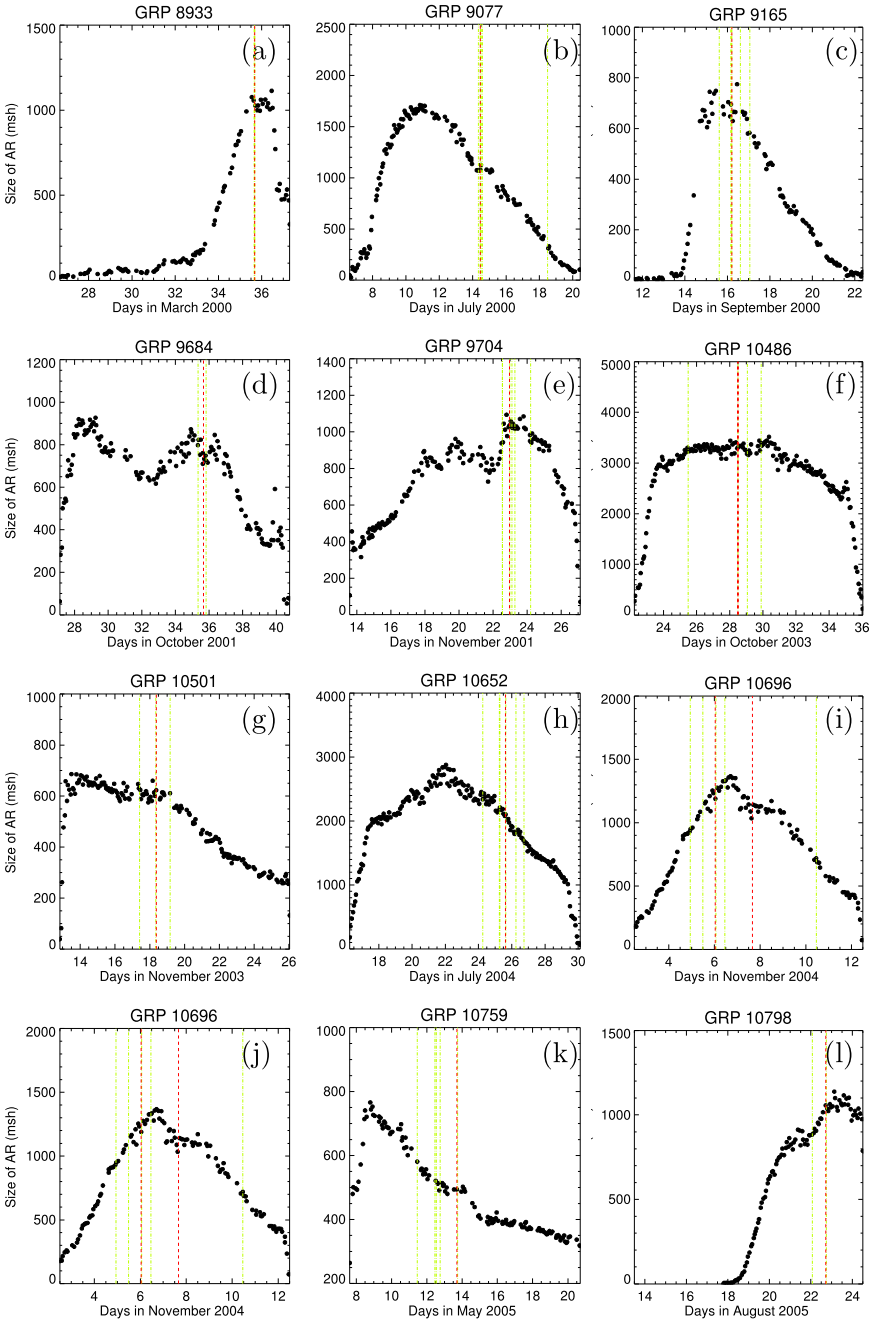

Figure 9 Variations of area (in μsh) of the source active regions with time (in days) in our sample for the visible part of solar rotation including all events during the SOHO Era. This figure presents hourly measurements of the area for the ARs presented in Figure 8. The measurements are extracted from http://fenyi.solarobs.unideb.hu/SDD/SDD.html. Titles correspond to the NOAA group number. The time of the flare associated with the storm is plotted as a red vertical line, while other flares of index ≥ 2 from the same region are represented in green.

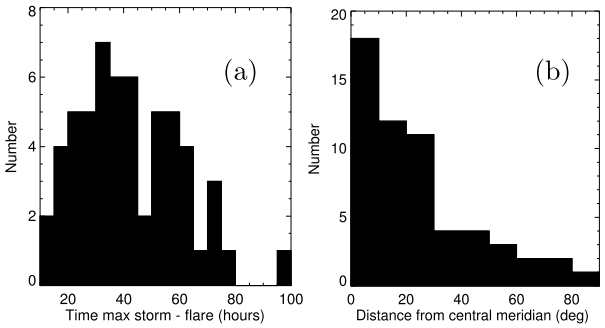

Figure 10 (a) Time (hours) between the maximum of the flare and the maximum of the geomagnetic storm during the Hα patrol era. (b) Absolute value of the distance of the flare from the central meridian in degrees.

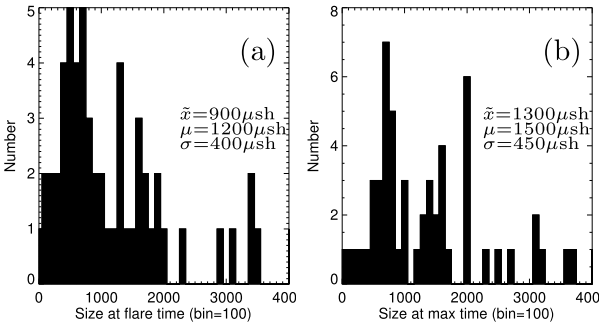

Figure 11 (a) Size (μsh) of the active regions at the time of the flare. (b) Size (μsh) of the active regions at the time of the maximum area.

and associated AR into account, we used the results from the previous time period as a basis and *chose the most extensive or intense flare* (according to the flare intensity described in Table 1, or the flare size or duration when available) in the defined four-day window before the geomagnetic storm. We note that although Figure 5 shows that all 13 events from the SOHO period fit within a window of 72 hours, there is a specific case that compels us to keep the window at four days: the Quebec Storm (13 March 1989), which is ranked number 2 in our list, with a ΔT at 74 hours.

The time delay between the maximum of the flare and the maximum of the storm is shown in Figure 10(a). Because we lack sufficient flare information, a few events could not be diagnosed using these criteria and were *not* taken into account in the statistics of this time period for which flare information is assumed to be available (six events out of 63, six out of 76 if the 13 events from the SOHO period are added). Figure 10(b) shows that the assumption of a short distance from the central meridian at the time of the flare still holds in this time period. We then derived the sunspot parameters for these candidate ARs. Figure 11 shows the sizes of ARs at the time of the flare for these events and at the time when the same AR reaches its maximum size. As for the previous time period, they are significantly larger than average (*cf.* Figure 6(a)).

Finally, we consider the *complexity evolution* of each AR during one solar rotation in this time period: with the variations of the area and number of spots and based on successive drawings in the four-day window. Figure 12 shows a sample of these candidate ARs between 1940 and 1946. For the complete time period from 1926 to 2010 (76 events total), about 80 % of the ARs show clear *sudden variations* in the area close to the associated flare time identified either in the area curve or based on the available drawings. Thirty-one flares occur within one day of maximum size (reaching the maximum size obviously creates a clear rapid inflection in the area), 23 flares occur within a day of the AR area showing clear

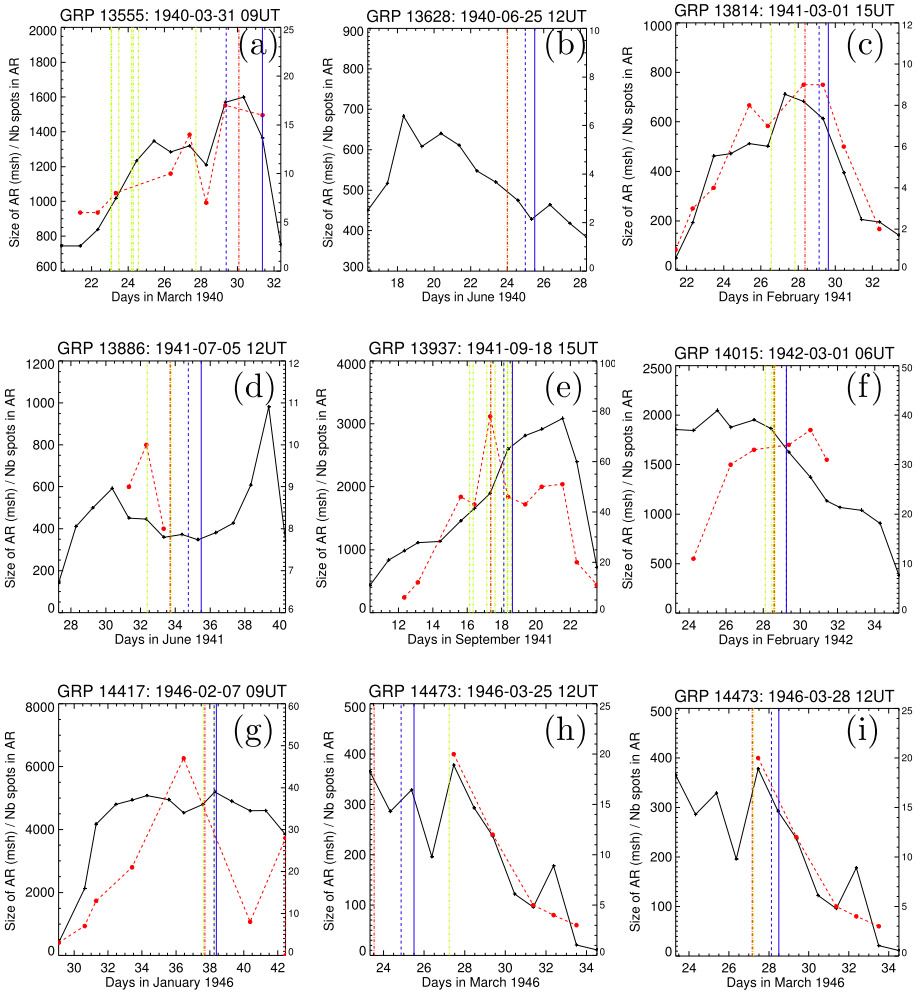

Figure 12 Variations of area (in μsh, black, left-side scale) and number of spots (dashed red, right-side scale) of active regions with time (in days) in our sample for one solar rotation for the extended Hα patrol period. Titles correspond to the NOAA AR number and the time of the associated geomagnetic storm. The time of the flare associated with the storm is plotted as a red vertical line, while other flares of index ≥ 2 from the same region within a window of four days before the storm are represented in green. In dashed blue we show the beginning of the storm and in solid blue the time of the maximum of the storm on Earth. These examples show clear cases of *sudden variations* in terms of solar area.

sudden variations, and six flares occur when the areas show variations, but less clearly. This means that 60 events out of 70 during this time period fit the criteria, which is 86 % (or 76 % if the six cases previously excluded from the statistics are excluded here as well).

3.3. RGO Era: 1874 – 1926

In this period, we lack consistent flare information (except for storms studied in Newton, 1943 or Newton, 1944 and the Catalog of White Light Flares, Neidig and Cliver, 1983),

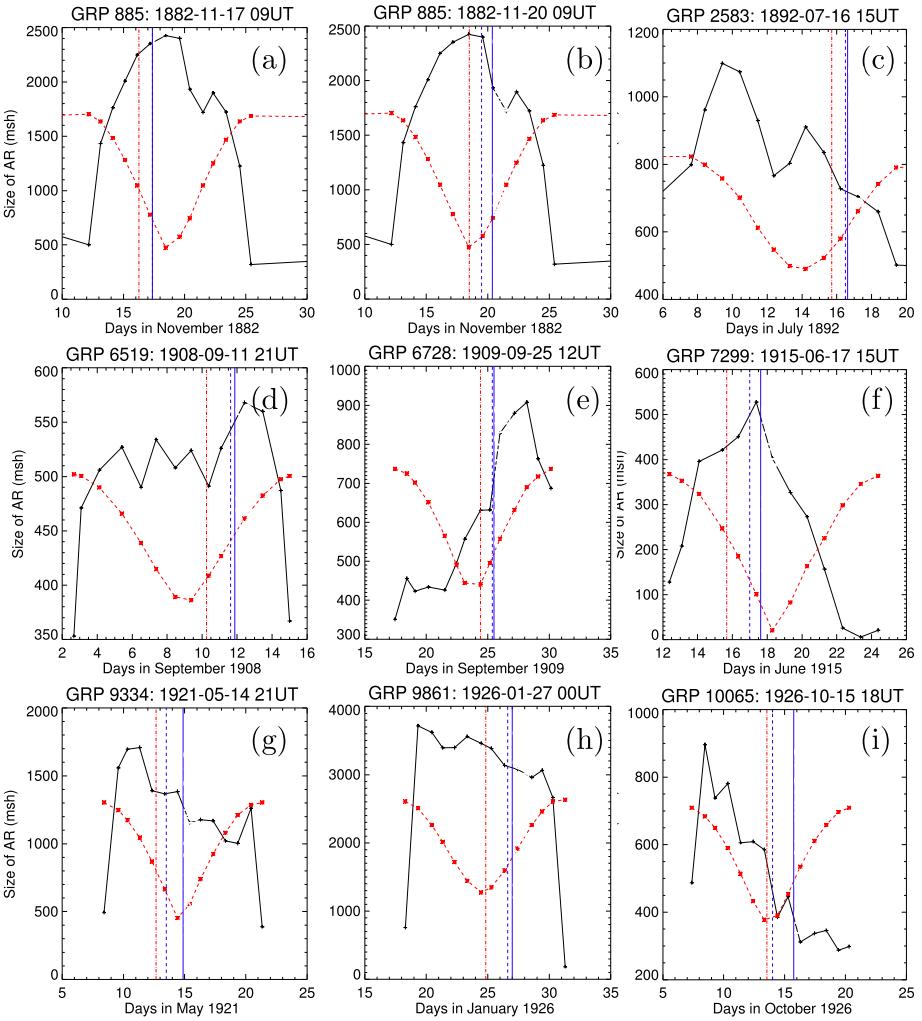

Figure 13 Variations of area (solid line, in μsh) and distance from the disk center (dashed red line, no units, minimum value is closest to the central meridian) of active regions with time (in days) for one solar rotation in the RGO era for cases where a flare was clearly identified in the literature. Titles correspond to group number and time of the associated geomagnetic storm. The flare time is plotted as a red vertical line. In dashed blue we plot the beginning of the storm and in solid blue the time of the maximum of the storm at Earth.

therefore we cannot base our studies on the time of a known flare. Hence, we used assumptions verified from 1926 to 2010 on more than 80 % of the events to link active regions to geomagnetic storms: (a) ARs associated with storms are much larger than average and complex, (b) they are located close to the central meridian in our four-day window, and (c) their area (or more generally, their complexity or configuration) shows *sudden variations* within the time window related to the possible flare. We compiled a list of the largest and most complex regions appearing close to the center of the Sun in our defined time window. These criteria can be combined (if a region is much larger than average, it can also be considered as a likely candidate even if it is located close to the limb).

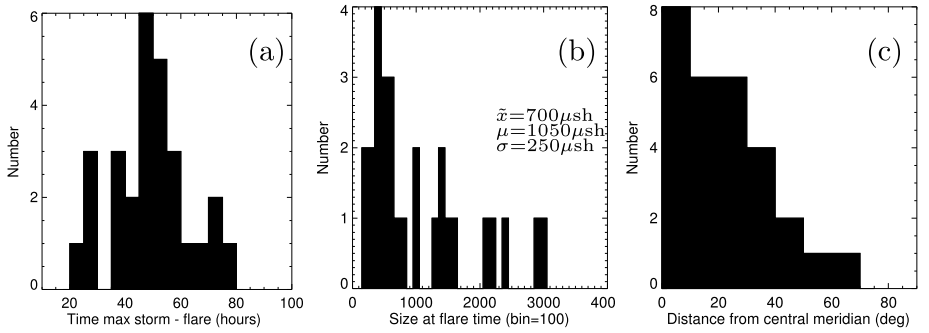

Figure 14 (a) Time (hours) between the maximum of the flare and of the geomagnetic storm during the RGO era. (b) Size (μsh) of the AR at the time of the flare. (c) Distribution of the longitude from central meridian in this sample (degrees) at the time of the flare.

The variations of the area and the number of spots were studied for each candidate AR as shown in Figure 13 for a subsample of ARs with clearly identified flares up to 1926. For the part of the sample where no flare information is available, we chose the day closest to observed *sudden variations* on the day during which a flare occurred. We typically assigned the time of the corresponding drawing or catalog observation in which the effects or causes of the flare were identified. From the time resolution available during that period, we were unable obtain a value more accurate than one day. For the event on 14 February 1892, the date is coherent with a drawing by D.E. Hadden that has been presented in Carrasco *et al.* (2013). The event from March 1898 can also be analyzed using the description and drawing of a large sunspot group from the collection of the Astronomical Observatory of Lisbon by Vaquero *et al.* (2012).

The study of the event on 14 May 1921 (lower left panel in Figure 13), which is the fiercest storm in our list of events, confirms that it is indeed a good choice. The drawing from the Mount Wilson Observatory for 12 May bears the clear inscription "Hα bright", and the evolution of the area of this AR in Figure 13 shows a typical staircase structure that could correspond to large releases of energy, as the downward variation on 12 May attests.

Figure 14 shows the distribution of the time between the flare and the storm, the size of the AR at the time of the flare, and the distance from the central meridian of the same AR at the alleged time of the flare. It can be compared to Figures 6, 7, 10, and 11 from previous periods.

3.4. Pre-RGO Era: 1868 – 1874

In this period we also rely on assumptions verified from 1926 to 2010: systematically compiled catalog information does not exist before 1874. We used the available images and drawings to extract the necessary solar parameters. We collected visual data from pictures taken by hand in the "Old Manuscripts" room in the library of the Cambridge University Library as well as information from various historical sources.

The quality of the visual information from this period induces a lower level of accuracy on measurements. In addition, we do not possess one drawing or picture per day in the window of ±10 days around each event: this implies more gaps in time coverage. We have the same kind of information as in the previous period, but with a lower accuracy in position, size, and time. We note that for three events between 1870 and 1871 we lack a few days of data in the study window.

For these seven events, historical sources and deduced conclusions are as follows,

- 13 May 1869 (rank 45): A synoptic chart can be found in Spoerer (1876) for May 1869. Groups 78 and 80 can easily be identified on the chart. We used the images and drawings from the RGO archives in Cambridge and applied the AIP method (Arlt and Fröhlich, 2012) and manual measurements from ROB to extract sunspot data. The largest group on the surface at this period is in the southern hemisphere (group 78), and it drastically changes size (and configuration, as can be seen from the images between 11 and 12 May). Taking all conclusions from the previous sections into account, we therefore assumed that an associated flare occurred on 11 May at 15:00 UT (time of the image we have in our database).
- 24 September 1870 (rank 61): Images and drawings from September 1870 exist in Proctor (1871), Secchi (1879), plate I, Arcimis (1901), and Spoerer (1876). The synoptic chart from (Spoerer, 1876) would indicate that AR273 is the likely candidate for an association with the storm of 24 September 1870. We also used the retrieved RGO archives and applied the AIP method and manual measurements at ROB to retrieve detailed sunspot information for a few days before and after the geomagnetic storm. As we have full daily coverage (from RGO archives) during its evolution on the solar disk from 19 September to 1 October, we were able to estimate its size variation (with large error bars). Between 22 and 23 September, there is a drop in area (corrected for foreshortening) of more than 40 %, which is why we chose this date as being associated to a chain of events probably linked to a flare.
- 24 and 25 October 1870 (ranks 69 and 50): For this event, we unfortunately lack any exploitable drawings or images for 20, 22, and 23 October. The synoptic chart from Spoerer (1876) and Vaquero *et al.* (2008) indicates that the storm originated from AR 299 (as numbered in the source). Therefore we relied on the less detailed information from the *Astronomische Nachrichten* published in 1871 for the Observations of the Athens Observatory for all of 1870 (Athens, 1871). It describes the number of groups and spots on the Sun around both storms. Inside our defined time window, large changes in these numbers are easily spotted between the 21 and 22 as well as between the 22 and 23, which is why we chose these dates. Hours are very poorly defined for this period: we chose the average time between the last observed drawing (before the change the 14 groups are still there around noon on the drawing) and the first observation by the Athens Observatory the next morning (after the change). We note that the errors on the extracted parameters are very large during this period.
- 12 February 1871 (rank 83): For this event, the drawings retrieved from the RGO archives are very crude. We therefore determined an approximate date of the flare from the variation in the Wolf numbers from day to day (no group and spot numbers) and the responsible region from Spoerer (1876). The number of groups on the Sun during the study period can also be found in Hoyt and Schatten (1998). Then we used a crude drawing from 10 February to determine the number of spots and the morphological type, but the size is impossible to determine because of the quality of the symbols used. With this undeniable lack of precision, we set noon of this day as the time of the flare.
- 9 April 1871 (rank 91): Here we used Spoerer (1876) to determine the group that is most probably linked to the event: group 116. Then we searched in the four-day window for a drastic change in the configuration of this group. In the Wolf numbers, there is a drastic change on 5 April, but the group most probably associated with this event only appears on the east limb on 6 April. The other local peak in the number of spots or groups occurs between 8 and 9 April, and the images from the RGO archives confirm a change

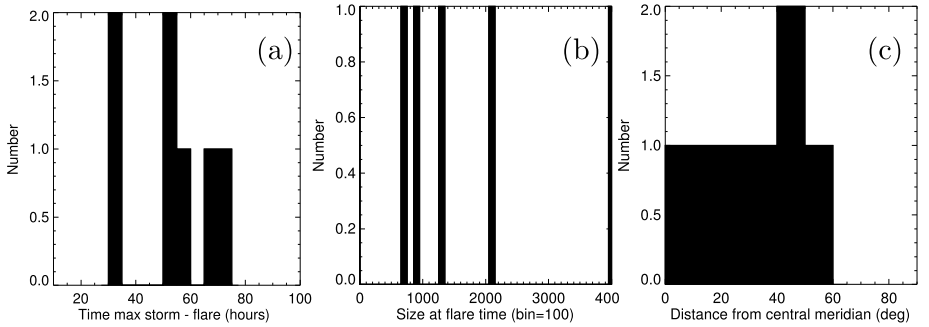

Figure 15 (a) Time (hours) between the maximum of the flare and of the geomagnetic storm during the pre-RGO era. (b) Size (μsh) of the ARs considered in this study at the time of the flare. (c) Distribution of the longitude from the central meridian in this sample (degrees) at the time of the flare.

of configuration during this period. We used the time of the image on 8 April (13:00 UT) as the time of the associated event at the Sun.

- 14 October 1872 (rank 42): Neidig and Cliver (1983) and Secchi (1872) only mentioned a white-light eruption the month after. Based on the drawing of Secchi (1872), it could be the same group as the one that flared during the next Carrington rotation (but we do not have the exact coordinates because only part of the Sun is drawn). Denning (1873) described a very large AR at the center of the disk on 14 October. This is obviously the one we have identified in this analysis: Wolf numbers show a drastic change on 13 October. In addition, the available drawings and images (6, 7, 10, 11, 12, 13, 14, 15, 17 October 1872) also show a large change in configuration on 13 October. We chose the time of the drawing on 13 October as the time of the probable flare associated with the geomagnetic event.

Figure 15 shows the distribution of the time between the flare and the storm, the size of the AR at the time of the flare and the distance from the central meridian of the same AR at the alleged time of the flare. It can be compared to Figures 6, 7, 10, 11 and 14 from previous periods.

4. Results

The parameters that were derived through this analysis are presented in Tables 2 and 3. Table 2 contains a subset of the information on the geomagnetic storms listed in Vennerstrøm *et al.* (2016), the associated flares and quantities representing the transit time between the flare and the geomagnetic storm, which are a good proxy for the mean transit speed or time of CMEs.

First, we introduce the *parameters of the geomagnetic storms* listed in Table 2: (a) the ranking shown earlier in the first column (*rk*, column 1), (b) the time of the maximum and beginning of the storm (in that order) as year, and then month, day, and hour (*yr, mt, mt, mt, bt, bt, bt*, columns 2 to 8), and (c) the peak in the aa index (*pk* col. 9, averaged over 24 hours). The columns (b) and (c) are white, (a) is gray-shaded in Table 2.

The next gray-shaded columns present the *flare parameters*: (d) the time of the flare on the Sun in month, day, and hour (rounded, *ft, ft, ft*, col. 10 to 12), (e) the position of the flare (or the position of the AR assumed to be responsible for it when this flare position is not

Table 2 Historical data: geomagnetic, flare, and transit data.

1 rk	2 yr	3 mt	4 mt	5 mt	6 bt	7 bt	8 bt	9 pk	10 ft	11 ft	12 ft	13 pos	14 fli	15 xray	16 fls (μsd)	17 dth (h)	18 sp1 ($km\,s^{-1}$)	19 sp2 ($km\,s^{-1}$)	20 comments
45	1869	5	13	18	5	13	12	177.6	5	11	15	S25E32	−99		−999	51	926	817	DR
61	1870	9	24	15	9	24	3	216	9	22	12	N16E41	−99		−999	51	1068	817	DR
69	1870	10	24	18	10	24	9	188.7	10	21	20	N11W03	−99		−999	70	683	595	DR, WO
50	1870	10	25	15	10	25	9	185.2	10	22	20	N11W15	−99		−999	67	683	621	DR, WO
83	1871	2	12	21	2	11	12	157.3	2	10	12	N20E55	−99		−999	59	1736	1262	DR, WO
91	1871	4	9	21	4	9	15	120.7	4	8	13	S20E49	−99		−999	32	1602	1302	SP1876
42	1872	10	14	21	10	14	21	219.6	10	13	11	N22E23	−99		−999	34	1225	1225	WL1983, DR, WO
85	1880	8	12	12	8	12	9	141.8	8	9	12	N24E57	−99		−999	72	603	578	MA1904, CAT
95	1881	1	31	18	1	31	6	111	1	30	5	S13W03	−99		−999	37	1666	1126	MA1904, CAT
20	1882	4	17	12	4	16	21	289.2	4	15	12	S20E20	−99		−999	48	1262	868	MA1904, CAT
11	1882	11	17	9	11	17	9	367.6	11	16	6	N19E35	3.5		500	27	1543	1543	MA1904, NE1943
19	1882	11	20	9	11	19	12	296.4	11	18	12	N19E05	3.5		500	45	1736	925	MA1904, NE1943
73	1886	3	30	9	3	30	6	180.1	3	28	12	S17E32	−99		−999	45	992	925	MA1904, CAT
22	1892	2	14	0	2	13	3	263.2	2	11	7	S28E08	−99		−999	65	946	641	MA1904, CAT
82	1892	7	16	15	7	16	12	200.2	7	15	17	S32W16	3.5		1500	22	2192	1893	MA1904, NE1943
81	1892	8	12	18	8	12	12	160.1	8	9	12	S31E28	−99		−999	78	578	534	MA1904, CAT
96	1894	2	25	12	2	25	6	149.2	2	22	12	S32W01	−99		−999	72	631	578	MA1904, CAT
25	1894	7	20	12	7	20	6	225.2	7	18	12	N08E02	−99		−999	48	992	868	MA1904, CAT
24	1894	8	20	3	8	20	0	193.7	8	18	9	N07W18	−99		−999	42	1068	992	MA1904, CAT
84	1898	3	15	21	3	15	0	162.3	3	13	12	S14W24	−99		−999	57	1157	730	MA1904, CAT
78	1898	9	9	18	9	9	12	172.5	9	7	12	S12E21	−99		−999	54	868	771	HAD1899, CAT
6	1903	10	31	12	10	31	3	314.2	10	29	10	S20E32	−99		−999	50	1016	833	MA1904, CAT
72	1907	2	9	18	2	9	6	169	2	7	10	S14E01	−99		−999	56	946	744	DR, CAT
38	1908	9	11	21	9	11	15	174	9	10	6	S06W18	3.5		2800	39	1262	1068	NE1943
3	1909	9	25	12	9	25	9	320.6	9	24	10	S05W08	3.5		1500	26	1811	1602	NE1943

Table 2 (*Continued.*)

1 rk	2 yr	3 mt	4 mt	5 mt	6 bt	7 bt	8 bt	9 pk	10 ft	11 ft	12 ft	13 pos	14 fli	15 xray	16 fls (μsd)	17 dth (h)	18 sp1 (km s^{-1})	19 sp2 (km s^{-1})	20 comments
62	1915	6	17	15	6	17	0	193.9	6	15	16	N02E42	−99		−999	53	1302	886	HA1931
10	1919	8	11	15	8	11	6	218	8	10	14	S07W17	−99		−999	25	2604	1666	DR, CAT
17	1920	3	23	0	3	22	9	236.8	3	21	10	S05E02	−99		−999	38	1811	1096	DR, CAT
1	1921	5	14	21	5	13	12	346.4	5	12	16	N00E32	−99		−999	53	2083	786	MWIL *
86	1926	1	27	0	1	26	15	148.1	1	24	20	N21W04	3.5		11000	52	968	801	NE1943
60	1926	4	15	6	4	14	12	174.8	4	13	12	S19E25	−99		−999	42	1736	992	DR, CAT
53	1926	10	15	18	10	14	0	218.8	10	13	13	N17E05	3.75		4700	57	3787	786	NE1943, WL1983
15	1928	7	8	0	7	7	21	313.8	7	5	12	S28E28	−99		−999	60	730	694	DR, CAT
48	1938	1	22	9	1	21	21	238	1	20	18	N17W27	3.5		−999	39	1543	1068	NE1943 *
33	1938	1	25	18	1	25	9	233.2	1	23	11	N17W65	3.5		−999	55	905	757	NE1950, CAT *
41	1938	4	16	6	4	16	3	173.5	4	15	8.5	N23W11	3.5		2000	21	2314	1984	NE1943, CL1990 *
47	1938	5	11	18	5	11	15	163.1	5	9	12	S08E11	3		−999	54	816	771	NE1944
77	1939	4	24	18	4	24	15	142.6	4	21	16	N28E59	3.5		4500	74	586	563	NE1943
8	1940	3	24	15	3	24	6	368.7	3	23	12	N12E37	3		−999	27	2314	1558	NE1944 *
37	1940	3	31	9	3	29	9	208.6	3	30	2	N10W55	3		−999	31	−2450	1343	FlPat *
102	1940	6	25	12	6	25	0	134.1	6	24	0	S13W14	3		−999	36	1736	1146	FlPat
4	1941	3	1	15	3	1	3	245.8	2	27	20	N16W06	2		−999	43	1344	974	FlPat
12	1941	7	5	12	7	4	18	292.3	7	3	17	N13E07	3		−999	43	1666	974	NE1944
5	1941	9	18	15	9	18	3	422.4	9	17	9	N12W06	3		−999	30	2314	1343	FlPat *
93	1942	3	1	6	3	1	6	161.9	2	28	14	N07E04	3		−999	16	2597	2597	NE1944, CL1990
89	1944	12	16	15	12	15	21	121.2	12	13	16	S22E13	2		−999	71	786	586	FlPat
67	1946	2	7	9	2	7	6	254.7	2	6	17	N27W19	3.5		−999	16	3205	2597	CL1990
66	1946	3	25	12	3	24	21	190.4	3	23	13	N23E44	1		−999	47	1302	886	FlPat *
7	1946	3	28	12	3	28	3	318.2	3	27	4	N20E05	3		−999	32	1811	1343	FlPat
35	1946	7	26	18	7	26	18	193.4	7	25	18	N20E15	3.5		−999	24	1736	1771	WL1983

Table 2 (*Continued.*)

1 rk	2 yr	3 mt	4 mt	5 mt	6 bt	7 bt	8 bt	9 pk	10 ft	11 ft	12 ft	13 pos	14 fli	15 xray	16 fls (μsd)	17 dth (h)	18 sp1 ($km\,s^{-1}$)	19 sp2 ($km\,s^{-1}$)	20 comments
43	1946	9	22	9	9	22	3	289.7	9	21	12	N16W23	2.5		−999	21	2777	2050	FlPat
98	1947	8	22	9	8	22	6	108.7	8	18	7	N10W61	2		−999	98	438	428	FlPat
103	1948	10	17	21	10	17	21	111	10	15	22	N12E18	2		−999	47	886	906	FlPat
32	1949	1	25	18	1	24	18	196.1	1	23	2	N25E00	3		−999	64	1041	649	CL1990
49	1949	5	12	15	5	12	6	199.9	5	10	20	S20E12	3.5		−999	43	1225	999	FlPat
87	1951	10	28	18	10	28	9	113.7	10	28	0	S09E57	1		−999	18	4629	2314	MWIL *
68	1957	1	21	21	1	21	6	186.1	1	20	11	S27W24	3		−999	34	2192	1217	CL1990
71	1957	3	2	3	3	1	12	152.2	2	28	0	N18W35	3		−999	51	1157	829	FlPat
36	1957	9	4	15	9	4	12	205.5	9	3	14	N22W29	3		1400	25	1893	1694	WL1983, CL1990 *
40	1957	9	13	9	9	13	0	157.4	9	11	3	N11W03	3		2100	54	925	779	FlPat *
59	1957	9	29	15	9	29	3	141.5	9	26	21	N24E17	3		2200	66	771	638	FlPat *
23	1958	2	11	0	2	10	15	291.5	2	9	22	S11W14	2.5		1290	26	2450	1623	CL1990
14	1958	7	8	15	7	8	6	304.9	7	7	1	N26W08	3.5		5570	38	1436	1096	CL1990
52	1958	9	4	15	9	4	12	179.9	9	2	21	S07E85	3		244	42	1068	1025	FlPat
70	1959	3	27	15	3	26	6	172.8	3	24	11	N29W73	3		1630	76	968	548	FlPat
104	1959	5	12	15	5	11	21	117.4	5	10	21	N19E46	3.5		8000	42	1736	1025	CL1990
9	1959	7	15	15	7	15	6	344.2	7	14	8	N18E08	3.5		4720	31	1893	1343	CL1990 *
74	1959	7	17	18	7	17	15	183.4	7	16	22	N15W21	3.5		2220	20	2450	2050	CL1990 *
16	1960	4	1	15	3	31	3	306.7	3	30	16	N12E09	3.5		1500	47	1157	885	FlPat*
29	1960	4	30	15	4	29	0	209.1	4	29	5	N11W22	3		2902	34	−8333	1256	FlPat*
44	1960	10	6	21	10	5	21	250.2	10	5	10	S18E41	1.5		456	35	3787	1180	FlPat
28	1960	11	13	6	11	12	12	368.8	11	12	14	N28W01	3.5		3000	16	−20833	2435	FlPat *
90	1960	11	16	0	11	15	12	160.9	11	15	2	N26W33	3.5		1179	22	4166	1855	WL1983 *
97	1961	7	27	9	7	27	3	131.2	7	24	19	N07E10	3		1794	62	744	671	FlPat
65	1961	10	28	18	10	28	6	170.3	10	27	9	N11W09	1		567	33	1984	1256	FlPat *
75	1963	9	22	21	9	21	12	181.7	9	21	0	N11W09	3		1770	45	3472	927	WL1983

Table 2 (*Continued.*)

1 rk	2 yr	3 mt	4 mt	5 mt	6 bt	7 bt	8 bt	9 pk	10 ft	11 ft	12 ft	13 pos	14 fli	15 xray	16 fls (μsd)	17 dth (h)	18 sp1 (km s^{-1})	19 sp2 (km s^{-1})	20 comments
105	1967	2	16	9	2	16	6	95.2	2	13	18	N22W10	3.5		3732	63	694	671	FlPat
13	1967	5	25	21	5	25	9	271.7	5	23	19	N28E24	3.5		1135	50	1096	847	WL1983
101	1969	2	2	18	2	2	15	117.8	2	2	5	N09E74	3.5		640	13	4166	3247	FlPat (CIR?) *
54	1970	3	8	18	3	7	12	177.7	3	7	2	S11E09	3.5		1341	40	4166	1041	FlPat
55	1972	8	4	21	8	4	12	283.8	8	4	7	N14E08	3.5		2589	14	8333	2997	WL1983, CL1990 *
51	1981	7	25	12	7	25	3	173.9	7	23	10	N15W18	2		478	50	1016	833	WL1983
27	1982	7	13	21	7	13	3	263.2	7	12	9	N12E42	3.5	X7.1	1527	36	2314	1157	CL1990 *
39	1982	9	6	12	9	5	21	201.9	9	4	1	N12E38	3	M6.4	2150	59	946	706	(+54 d) *
26	1986	2	8	18	2	7	9	280.5	2	6	7	S07W06	3.5	X1.7	2260	59	1602	706	FlPat *
2	1989	3	13	21	3	13	0	430.9	3	10	19	N31E22	3	X4.5	989	74	786	563	CL1990 *
63	1989	10	20	18	10	20	6	183.7	10	19	13	S27E10	3.5	X13.	800	29	2450	1436	CL1990
79	1989	11	17	18	11	17	6	197.2	11	15	7	N11W26	3.5	X3.2	648	59	886	706	CL1990
58	1991	3	24	3	3	24	3	181	3	22	23	S26E28	3.5	X9.4	314	28	1488	1488	FlPat
57	1991	6	5	15	6	4	12	208.3	6	4	4	N30E65	3.5	X12.	721	35	5208	1190	FlPat
34	1991	11	8	21	11	8	6	230.7	11	6	5	S14W20	1	M4.7	100	64	850	651	FlPat, CL2009 *
99	1992	9	9	5	9	9	0	115.4	9	6	19	S11W41	2	X1.7	530	58	786	718	FlPat
76	2000	4	6	21	4	6	15	167.4	4	4	16	N16W66	2	C9.7H	380	53	860	786	RC 2010, ZH2007
30	2000	7	15	15	7	15	3	214.9	7	14	11	N22W07	3.5	X5.7H	656	28	1500	1488	RC 2010, ZH2007
88	2000	9	17	21	9	17	9	133	9	16	5	N14W07	2.5	M5.9H	723	40	> 750	1041	RC 2010, ZH2007
56	2001	11	6	0	11	5	15	149	11	4	16	N06W18	3.5	X1.0H	643	32	1250	1302	RC 2010, ZH2007
92	2001	11	24	6	11	24	3	140.6	11	22	23	S10W39	3.5	M9.9H	24 733	31	1320	1344	RC 2010, ZH2007 *
21	2003	10	29	6	10	28	21	322.4	10	28	12	S16E08	3.5	X17.2	−999	18	2185	2314	RC 2010, ZH2007
31	2003	10	30	21	10	30	12	307.7	10	29	21	S15W02	2.5	X10.0	542	24	2138	1736	RC 2010, ZH2007
18	2003	11	20	15	11	20	6	244.6	11	18	9	N00E18	2	M3.9	270	54	886	771	RC 2010, ZH2007
80	2004	7	27	12	7	26	21	203.4	7	25	15	N08W33	1	M1.1	156	45	1302	925	RC 2010, ZH2007

Table 2 (*Continued.*)

1 rk	2 yr	3 mt	4 mt	5 mt	6 bt	7 bt	8 bt	9 pk	10 ft	11 ft	12 ft	13 pos	14 fli	15 xray	16 fls (μsd)	17 dth (h)	18 sp1 ($km\,s^{-1}$)	19 sp2 ($km\,s^{-1}$)	20 comments
46	2004	11	8	3	11	7	9	210.6	11	6	1	N10E08	2	M9.3	252	50	> 720	833	RC 2010, ZH2007
64	2004	11	9	18	11	8	18	230.8	11	7	16	N09W17	−99	X2.0H	−999	50	830	833	RC 2010, ZH2007
94	2005	5	15	6	5	15	0	115	5	13	17	N12E11	2.5	M8.0H	311	37	1270	1126	RC 2010, ZH2007
100	2005	8	24	9	8	24	3	136.8	8	22	17	S13W65	1	M5.6H	141	40	790	1041	RC 2010, ZH2007

rk: rank, yr: year, mt: max time (month, day, hour), begin time (month, day, hour), pk: peak value (nT)
ft: flare time (month, day, hour), pos: position of the source AR, fli: flare index, xray: X-ray flare
fls: flare size, dth: time bt-flare time (hours), sp1: speed dth ($km\,s^{-1}$), sp2: speed from mt-flare time
Sources are Cliver (1990): CL1990; Richardson and Cane (2010): RC2010; Catalog of White Light Flares: WL1983; Flare Patrol data: FlPat; Mount Wilson Data: MWIL; Newton (1943, 1944, 1950): NE1943, NE1944, NE1950; Spoerer (1876): SP1876; Maunder (1904): MA1904; catalogs of sunspots: CAT; information found in drawings of various sources: DR; information found in Wolf number variation: WO; Hale (1931): HA1931; Hadden (1899): HAD1899; Cliver (2009): CL2009; Zhang (2007): ZH2007

Table 3 Historical sunspot data linked to geomagnetic storms and flare parameters from Table 2.

1 rk rk	2 grp	3 ne	4 lf days	5 gr μsh/d	6 dr μsh/d	7 long (o)	8 lat (o)	9 n1	10 n2	11 n3	12 szm μsh	13 szf μsh	14 dtd days	15 dist	16 dcm (o)	17 mrp1 McI	18 mrp2 McI	19 comments
45	78	1	13	−99	−99	187.27	−25	5	8	8	2100	2100	0	0.6	32	Fki	Fki	DR
61	273	1	14	1000	−99	87.81	15.8	15	13	13	4000	4000	0	0.75	41	Fki	Fki	DR
69	299	2	14	−99	−99	75.25	11	7	5	10	1100	700	2	0.2	3	Dkc	Dsi	DR, WO
50	299	2	14	−99	−99	87.25	11	7	5	−99	1100	−99	3	0.2	15	Dkc	XXX	DR, WO
83	36	1	14	−99	−99	17.24	20	−99	−99	8	−99	−99	−99	0.8	55	XXX	Fki	DR, WO
91	116	1	−99	137.96	−99	267.74	−20	6	7	6	1100	900	−99	0.8	49	XXX	DKC	SP1876
42	7	1	14	419	−99	224.28	24	9	20	4	4000	1300	3.9	0.25	23	Fkc	Dki	WL1983, DR, WO
85	343	1	14.05	67.9	−43.6	234.49	24.56	9	9	3	444	250	2.99	0.83	57	Eso	Dso	MA1904, CAT
95	412	1	7.08	190.68	−66.89	165.27	−13.04	9	9	−99	581	515	1.25	0.13	3	Dai	XXX	MA1904, CAT
20	726	1	14.36	192.64	−554.45	92.88	−19.31	13	14	9	2258	219	5.9	0.7	6	Eki	Dao	MA1904, CAT
11	885	2	14.22	305.58	−303.96	120.68	19.22	16	20	14	2425	2249	2.23	0.61	35	Fki	Fki	MA1904, NE1943
19	885	2	14.22	305.58	−303.96	120.68	19.22	16	20	20	2425	2425	0	0.28	4	Fki	Fki	MA1904, NE1943
73	1860	1	11.01	118.66	−130.34	115.88	−16.79	19	33	22	849	303	4.91	0.6	32	Fki	Dko	MA1904, CAT
22	2421	1	14.09	313.74	−199.51	255.76	−28.33	40	73	66	3038	2921	−1.07	0.39	8	Fkc	Fkc	MA1904, CAT
82	2583	2	12.75	172.2	−59.65	33.19	−31.32	23	20	15	1099	835	−6.3	0.62	16	Dkc	Dki	MA1904, NE1943
81	2615	2	11.9	−99	−38.44	16.57	−30.37	9	7	23	462	354	−4.15	0.71	28	Dao	Dai	MA1904, CAT
96	3412	1	14	175	−126.49	186.92	−32.04	20	22	28	1742	1541	−1.07	0.41	1	Fko	Fko	MA1904, CAT
25	3629	1	12.8	50.81	−28.2	56.5	7.52	13	32	32	422	422	0.05	0.06	2	Dko	Dko	MA1904, CAT
24	3668	1	12.93	211.16	−252.59	26.5	6.35	34	54	35	1910	1554	−1.92	0.29	18	Fkc	Fkc	MA1904, CAT
84	4702	1	11.86	153.25	−240.12	119.55	−13.17	17	20	18	1552	1293	0.72	0.4	24	Fki	Fki	MA1904, CAT
78	4781	1	13.06	172.97	−327.36	239	−12.07	9	17	13	2235	1449	2.95	0.46	21	Fki	Fki	HAD1899, CAT
6	5098	1	13.29	205.35	−74.98	298.09	−18.84	7	8	10	807	735	−0.8	0.64	32	Dki	Eki	MA1904, CAT
72	6104	1	10.16	99.44	−216.78	12.44	−14.1	7	6	9	766	639	4.16	0.13	1	Fki	Fki	DR, CAT
38	6519	1	13.29	22.17	−77.59	279.74	−5.51	5	8	7	568	491	2.15	0.37	18	Dki	Dko	NE1943
3	6728	1	13.69	51.52	−117.78	306.5	−4.79	9	12	5	908	631	3.84	0.24	8	Dki	Dki	NE1943

Table 3 (*Continued.*)

1 rk rk	2 grp	3 ne	4 lf days	5 gr μsh/d	6 dr μsh/d	7 long (o)	8 lat (o)	9 n1	10 n2	11 n3	12 szm μsh	13 szf μsh	14 dtd days	15 dist	16 dcm (o)	17 mrp1 McI	18 mrp2 McI	19 comments
62	7299	1	12.97	79.93	−72.74	34.44	1.85	15	18	14	528	421	1.74	0.66	42	Dai	Dai	HA1931
10	8994	1	12.8	−99	−16.28	226.94	−6.76	6	1	9	627	531	−7.01	0.36	17	Hhx	Hhx	DR, CAT
17	9143	1	12.93	284.34	−261.72	134.63	−4.97	34	50	35	2690	2086	1.97	0.06	2	Fkc	Fkc	DR, CAT
1	9334	1	13.91	418.74	−132	28.57	0.54	19	36	25	1709	1391	−1.34	0.52	32	Ekc	Ekc	MWIL *
86	9861	1	14	3007.4	−293.8	34.77	21.04	18	25	12	3716	3457	−5.54	0.44	4	Fki	Fki	NE1943
60	9920	1	11.95	54.38	−27.38	50.5	−19.24	5	5	6	279	175	−3.13	0.46	25	Dao	Dko	DR, CAT
53	10 065	1	13.91	397.92	−50.32	171.89	16.68	9	7	18	896	585	−5.09	0.21	5	Dko	Fki	NE1943, WL1983
15	10 658	1	12.97	85.57	−73.67	102.64	−26.55	10	9	14	861	575	−2.14	0.65	28	Dki	Eki	DR, CAT
48	12 673	2	12.93	292.21	−402.21	225.37	17.05	45	30	45	3627	3379	0.54	0.56	27	Fki	Fki	NE1943 *
33	12 673	2	12.93	292.21	−402.21	225.37	17.05	45	30	37	3627	2983	−2.14	0.93	65	Fki	Fki	NE1950, CAT *
41	12 777	1	13.11	14.7	−157.22	174.04	26.89	7	15	4	1512	1045	−4.05	0.56	11	Fhi	Cho	NE1943, CL1990*
47	12 808	1	13.02	163.22	−82.08	189.36	−7.31	10	11	11	1019	1019	0	0.21	11	Ekc	Ekc	NE1944
77	13 245	1	13.15	−99	−39.58	236.84	29.06	6	2	4	815	798	−1.34	0.93	59	Cho	Dho	NE1943
8	13 555	2	12.97	85.6	−412.62	136.52	12.15	10	16	8	1599	1017	6.79	0.6	37	Fko	Eko	NE1944*
37	13 555	2	12.97	85.6	−412.62	136.52	12.15	10	16	16	1599	1599	0.22	0.84	55	Fko	Fko	FlPat*
102	13 628	1	13.02	116.5	−27.18	47.31	−10.19	4	11	4	683	520	−5.67	0.27	14	Hkx	Hkx	FlPat
4	13 814	1	12.93	110.33	−81.43	354.42	14.42	5	10	9	712	683	−0.54	0.43	6	Hkx	Dkc	FlPat
12	13 886	1	14.05	64.62	−638	122.45	13.1	11	9	8	981	359	5.67	0.21	7	Dkc	Dko	NE1944
5	13 937	1	14.22	221.08	−790.67	210.09	11.67	36	51	78	3088	1896	4.2	0.12	6	Fkc	Fkc	FlPat*
93	14 015	1	13.02	64.33	−166.5	197.22	7.13	27	19	17	2048	1865	−3.04	0.27	4	Dkc	Eki	NE1944, CL1990
89	14 238	1	13.82	115.7	−109.2	55.93	−21.9	14	25	26	1010	893	−1.21	0.41	13	Dro	Ekc	FlPat
67	14 417	1	15.16	479.8	−568	299.06	25.93	16	27	25	5202	4799	0.63	0.56	19	Fki	Ekc	CL1990
66	14 473	2	12.17	2.6	−45.88	351.78	23.58	8	20	15	378	365	3.26	0.91	48	Fko	Fki	FlPat*
7	14 473	2	12.17	2.6	−45.88	351.78	23.58	8	20	20	378	378	0.27	0.54	5	Fko	Fko	FlPat
35	14 585	1	15.79	452.5	−927	197.06	22.38	72	117	94	4720	4279	3.62	0.46	15	Fkc	Fkc	WL1983

Table 3 (*Continued.*)

1 rk rk	2 grp	3 ne	4 lf days	5 gr µsh/d	6 dr µsh/d	7 long (o)	8 lat (o)	9 n1	10 n2	11 n3	12 szm µsh	13 szf µsh	14 dtd days	15 dist	16 dcm (o)	17 mrp1 McI	18 mrp2 McI	19 comments
43	14643	1	11.72	−99	−117.36	191.27	20.24	27	4	58	1376	672	−7.91	0.4	23	Dso	Eai	FlPat
98	15106	1	12.93	133.1	−447.67	195	10.8	38	23	9	1359	936	−1.79	0.96	61	Fki	Eko	FlPat
103	15743	1	13.11	172.5	−80.18	260.37	12.66	23	40	29	1282	964	−2.64	0.22	18	Hkx	Dkc	FlPat
32	15871	1	14	158	−382.67	53.52	23	29	37	45	2471	2319	1.21	0.49	0	Fko	Fko	CL1990
49	16042	1	12.93	−99	−36.17	61.31	−16.41	8	10	14	626	393	−5.5	0.35	12	Dkc	Chi	FlPat
87	16892	1	12.12	61.71	−57.27	14.49	−8.89	22	34	26	497	479	1.43	0.75	57	Fki	Fhi	Mt Wilson *
68	17829	1	13.91	59	−45	61.11	−27.17	10	13	12	636	557	−3.04	0.49	24	Dkc	Cki	CL1990
71	17884	1	8.73	235.5	−44	288.8	14.09	12	10	23	512	455	−3.71	0.76	35	Fso	Eai	FlPat
36	18182	1	13.82	243	−173	329.52	24.52	39	79	79	1726	1336	−5.27	0.64	29	Ekc	Fki	WL1983, CL1990 *
40	18194	1	12.62	159	−113	194.36	10.55	42	36	77	1365	664	−4.78	0.13	3	Eki	Eki	FlPat *
59	18222	1	11.14	14.78	−58.67	335.53	23.95	14	10	10	192	153	2.68	0.3	16	Cao	Bxo	FlPat *
23	18500	1	14	70.62	−41.17	15.71	−14.28	19	12	12	756	756	−0.45	0.34	14	Cko	Cko	CL1990
14	18773	1	12.97	169.49	−80.28	200.97	27.08	30	49	49	689	686	−0.71	0.4	8	Dko	Dko	CL1990
52	18889	1	12.93	32.67	−83.75	87.48	−14.38	24	84	2	1047	276	8.44	0.99	85	Eki	Hsx	FlPat
70	19234	1	13.02	154.08	1017	94.57	26.02	51	22	5	2274	1257	−0.94	0.99	73	Eai	Dko	FlPat
104	19336	1	11.99	390.33	−149.44	53.13	18.56	87	82	82	1552	1552	0.45	0.79	46	Fkc	Fkc	CL1990
9	19448	2	14	89.7	−383.25	330.47	15.86	38	61	52	1981	1314	3.04	0.21	8	Dkc	Dkc	CL1990 *
74	19448	2	14	89.7	−383.25	330.47	15.86	38	61	61	1981	1981	0.45	0.61	21	Dkc	Dkc	CL1990 *
16	19810	1	11.05	261	−63.2	130.1	11.15	45	61	54	1982	1715	2.73	0.37	9	Dkc	Dkc	FlPat*
29	19836	1	13.02	116	−24	133.71	10.69	17	8	22	741	393	−5.9	0.48	22	Hkx	Esi	FlPat*
44	20039	1	13.15	25.38	−60.2	126.52	−14.93	12	32	5	496	232	4.91	0.68	41	Eki	Cso	FlPat
28	20075	2	13.06	130	−201.8	25.44	26.99	38	43	56	2040	1519	1.79	0.41	1	Fko	Fkc	FlPat *
90	20075	2	13.06	130	−201.8	25.44	26.99	38	43	40	2040	1690	−0.71	0.7	33	Fko	Fko	WL1983 *
97	20268	1	12.97	184	−47.75	262.97	8.28	20	5	31	696	502	−4.47	0.18	10	Hkx	Dkc	FlPat
65	20334	1	4.93	16	−3	111.72	8.82	4	9	3	32	16	0.94	0.21	9	Bxo	Axx	FlPat *

Table 3 (*Continued.*)

1 rk rk	2 grp	3 ne	4 lf days	5 gr μsh/d	6 dr μsh/d	7 long (o)	8 lat (o)	9 n1	10 n2	11 n3	12 szm μsh	13 szf μsh	14 dtd days	15 dist	16 dcm (o)	17 mrp1 McI	18 mrp2 McI	19 comments
75	20610	1	13.02	49.57	−84.17	311.28	12.95	27	42	42	1311	1311	0.33	0.23	9	Dkc	Dkc	WL1983
105	21044	1	12.04	128	−55.62	170.54	18.25	23	27	19	565	452	−6.17	0.8	10	Eai	Eai	FlPat
13	21164	1	14.09	241.67	−165.78	223.27	26.9	53	79	45	1642	1612	0.54	0.66	24	Fkc	Fkc	WL1983
101	21876	1	4.8	−99	−18.75	272.5	9.23	1	1	1	97	91	−0.71	0.93	74	Axx	Axx	CIR ?? *
54	22307	1	10.78	−99	−22.79	122.6	−11.85	7	4	9	266	68	−5.54	0.11	9	Dao	Cri	FlPat
55	23179	1	14.05	219.25	−72.18	10.36	12.97	12	7	16	1334	1107	−2.95	0.19	8	Dkc	Dkc	WL1983, CL1990 *
51	3224	1	10.11	115	−132.5	15.21	15.88	27	35	43	700	600	1.16	0.32	18	EKI	EKI	WL1983
27	3804	1	14	916.9	−259.4	321.86	13.67	39	91	11	3154	2867	−1	0.69	42	FKI	FKC	CL1990 *
39	3886	1	13.02	550	−60	330.22	13.75	16	14	43	750	340	−2.41	0.7	38	DKC	DAI	(+54 d) *
26	4711	1	12.88	93.4	−128.57	62.28	−8.46	10	26	16	783	623	−0.8	0.15	6	DKI	EKC	FlPat *
2	5395	1	13.87	245.04	−795.65	255.53	33.36	21	28	21	4200	3354	3.57	0.75	22	FKC	FKC	CL1990 *
63	5747	1	13.15	241.28	−77.23	208.57	−26.72	12	12	18	1471	1354	−2.95	0.57	10	DKI	EKC	CL1990
79	5786	1	12.26	24.26	−133.68	252.92	12.18	9	9	9	677	677	−0.09	0.46	26	FKI	FKI	CL1990
58	6555	1	14.98	383.48	−455.8	184.67	−23.83	19	30	39	3149	3138	2.32	0.41	28	FKC	FKI	FlPat
57	6659	1	14.98	275.47	−533.41	246.71	31.45	12	23	6	2694	1813	8.22	0.93	65	EKC	DKC	FlPat
34	6906	1	10.7	176.13	−146.71	87.74	−16.12	4	8	8	846	846	0.13	0.47	20	EKI	EKI	FlPat, CL2009 *
99	7270	1	12.3	139.18	−449.65	37.44	−9.19	7	18	18	1237	1237	0.36	0.78	41	EKI	EKI	FlPat
76	8933	1	11.14	105.12	−158.4	268.4	16.3	6	20	20	946	946	−0.31	0.93	66	EKO	EKO	RC 2010
30	9077	1	14.27	441.47	−160.93	309.76	17.47	26	57	26	1591	879	−3.75	0.21	7	FKI	FAI	RC2010, ZH2007
88	9165	1	10.78	227.56	−98.16	188.6	13.97	12	33	29	674	633	−0.67	0.14	7	DAI	DAO	RC2010, ZH2007
56	9684	1	14.05	156.68	−76.48	138.59	5.82	7	7	10	850	642	−6.3	0.37	18	DKO	DKC	RC2010, ZH2007
92	9704	1	12.84	85.63	−51.33	270.17	−18.01	9	14	16	919	901	−2.5	0.68	39	DKI	EKC	RC2010, ZH2007 size? *
21	10486	2	13.91	289.77	−815.61	286.13	−17.85	24	29	35	3388	3193	3.93	0.42	8	FKC	FKC	RC2010, ZH2007
31	10486	2	13.91	289.77	−815.61	286.13	−17.85	24	29	39	3388	3289	2.5	0.4	2	FKC	FKC	RC2010, ZH2007
18	10501	1	13.24	101.69	−37.89	3.38	3.5	5	7	7	647	579	−2.9	0.23	18	DAO	DHI	RC2010, ZH2007

Table 3 (*Continued.*)

1 rk rk	2 grp	3 ne	4 lf days	5 gr μsh/d	6 dr μsh/d	7 long (o)	8 lat (o)	9 n1	10 n2	11 n3	12 szm μsh	13 szf μsh	14 dtd days	15 dist	16 dcm (o)	17 mrp1 McI	18 mrp2 McI	19 comments
80	10652	1	14	340.41	−271.44	349.13	8.01	21	53	26	2763	2046	−3.08	0.5	33	FKC	FKC	RC2010, ZH2007
46	10696	2	11.1	282.46	−208.12	27.56	7.88	19	33	33	1330	1330	0.49	0.09	8	EKC	EKC	RC2010, ZH2007
64	10696	2	11.1	282.46	−208.12	27.56	7.88	19	33	30	1330	1199	−1.12	0.27	17	EKC	EKI	RC2010, ZH2007
94	10759	1	12.97	−99	−45.1	56.45	11.8	3	3	7	911	497	−5.18	0.3	11	DKO	DHI	RC2010, ZH2007
100	10798	1	11.99	100.22	−188.22	222.66	−9.95	5	9	10	1021	967	0.8	0.86	65	DKC	DKC	RC2010, ZH2007

rk: rank, grp: AR number, ne: number of geomagnetic events, lf: lifetime of AR, gr: growth rate of AR
dr: decay rate AR, long: longitude AR, latitude AR, n1: mean number of spots, n2: number spots AR max time
n3: number spots AR flare time, szm: size of AR at mt, szf: size of AR at ft, dtd: time max size AR-ft
dist: distance from disk center (no units), dcm: dist. from CM, mrp1 and mrp2: McIntosh type at max size and ft
Sources are Cliver (1990): CL1990; Richardson and Cane (2010): RC2010; Catalog of White Light
Flares: WL1983; Flare Patrol data: FlPat; Mount Wilson Data: MWIL; Newton (1943, 1944, 1950):
NE1943, NE1944, NE1950; Spoerer (1876): SP1876; Maunder (1904): MA1904; catalogs of sunspots: CAT;
information found in drawings of various sources: DR; information found in Wolf number
variation: WO; Hale (1931): HA1931; Hadden (1899): HAD1899; Cliver (2009): CL2009; Zhang (2007): ZH2007

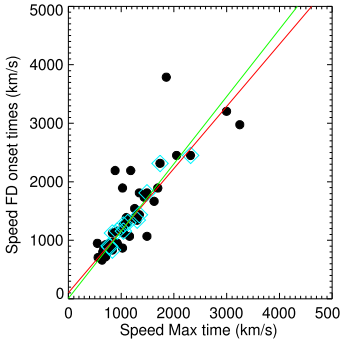

Figure 16 Comparison of a proxy speed computed from the time of flare compared to the time of the Forbush decrease onset associated with our events, SpeedFD (Forbush decreases are computed from 1957 onward, hence the comparison from this date) to a proxy speed computed with the time of the maximum value of the storm, SpeedMax. Light blue squares represent the events from the SOHO era. Overplotted is the linear fit $Y = aX + b$ (red line) to the data ($a = 0.97 \pm 0.07$, $b = 115 \pm 85$). The Pearson correlation coefficient is $r = 0.96$ (w/o four outliers). Considering only the SOHO era, the fit gives $a = 1.1 \pm 0.1$, $b = 10 \pm 140$ with a correlation coefficient of $r = 0.96$ (green line).

available, *i.e.* before the beginning of the Hα-patrol), *position* (col. 13) and (f) the size of the flare in index (*fli*, see Section 2.4), in X-ray flux (*xray*, col. 14 and 15), and its size (area) in millionths of the solar disk or µsd (*fls*, col. 16). For the period between 1926 and 2010, when both the position of the AR on the day of the flare and the position of the flare itself are available, we compared these two parameters. They are not indicated separately in this table because they correspond perfectly to within a few degrees.

The next three columns present *parameters related to the CME–ICME transit time* or a proxy of these quantities: (a) the time between the flare and the maximum of the geomagnetic storm in hours (*dth*, col. 17), (b) a proxy of the mean transit speed computed from the difference between the flare time and the time of the beginning of the storm, SpeedBeg (*sp1*, col. 18) and (c) another proxy of the mean transit speed as the speed corresponding to the time in (a), SpeedMax (*sp2*, col. 19). After 1992, (b) is not speedBeg, it corresponds to the speed extracted from Richardson and Cane (2010): it is completely compatible with SpeedMax on that period. The proxy of the mean transit speed computed from Forbush decreases, SpeedFD was also computed, but is not presented in this table because the proxy SpeedFD is so similar to the proxy SpeedMax that it brings no additional information: Figure 16 shows that the estimated error on the proxy is close to nonexistent because both fits (from 1957 to the present or from 1996 to the present) are compatible with $X = Y$, or equivalent within the quoted error bars. Column 20 presents comments about this association process, where applicable, and references when available.

Table 3 presents *parameters related to the ARs*. The first column repeats the rank (see Table 2). The second column gives the available numbering for the AR linked to the geomagnetic event (*grp*), while the third column tells us how many times the region is associated with a geomagnetic event from the list of Vennerstrøm *et al.* (2016) (*ne*). This enables us to see one aspect of the complexity of the event. The next column presents the lifetime of the considered AR (*lf*, col. 4). However, this parameter does not appear very significant because each AR is renamed when it reappears for another rotation. As a consequence, most lifetimes are very close to half of a solar rotation (13 to 14 days).

On the other hand, ARs showing a lower lifetime (typically much shorter than half a solar rotation) might give us significant information because they are certainly outliers. There are

four regions that show lifetimes between four and nine days in Table 3. In 1881, region 412 appears at the center of the disk on 29 January and grows very rapidly until it passes the limb on 14 February. But it seems to reappear on 20 February: these apparitions in subsequent rotations are difficult to account for considering that the existing catalogs renumber each region as it passes the limb. In March 1957, region 17884 also appears at the center of the disk, thus these two cases are not short-lived active regions. However, regions 20334 and 21876 in 1961 and 1969 are small active regions that are studied as presenting *anomalous parameters* below.

Columns 5 and 6 present the growth and decay rates of the ARs (in millionths of the solar hemisphere or μsh per day, *gr and dr*). Columns 7 and 8 present the mean heliocentric longitude and latitude of the ARs (in degrees, *long and lat*). Columns 9, 10, and 11 present the mean number of spots in these ARs and the number of spots at maximum size time and flare time (*n1, n2, and n3,* respectively). Columns 12 and 13 present the area of the ARs at maximum size and flare time (μsh, *szm and szf*). Column 14 presents the time between the maximum size of the flare and the AR maximum size (in days, *dtd*), while columns 15 and 16 present distances from solar disk center and central meridian (*dist and dcm*). And lastly, columns 17 and 18 present the McIntosh types of the regions at maximum size and flare time (*mrp1 and mrp2*). Values of −99, −999, and XXX indicate that the value is unavailable at the time we were writing this article. Column 19 presents comments about this association process, where applicable, and references when available; it is the same as column 20 of Table 2.

In these two complementary tables, some events present *anomalous parameters* (negative speeds, very high speeds, very simple McIntosh type, low flare index) indicated by a (*) in the *comments* column: we present them here with more detailed comments.

- 14 May 1921 (rank 1): According to Silverman and Cliver (2001), there is no available source in the literature to find a flare that could be linked to this event. However, between 2001 and more recent periods, solar drawings from Mount Wilson have been digitized, and a very interesting inscription is found on the drawing from 12 May 1921 (see Figure 17). We note that while this article was being refereed, Lundstedt, Persson, and Andersson (2015) noted the same feature. It clearly indicates that a flare was observed in Hα on this day: as the only available source for this flare, we used the date of this drawing. The variations in the area of the AR considered responsible (only one AR on the Sun at that time) are displayed in Figure 13. We note that according to Lundstedt, Persson, and Andersson (2015), this flare might be related to a first CME that arrived at Earth on 13 May, thus clearing the way for another CME arriving on 14 May and increasing its impact.
- 22 – 25 January 1938 (ranks 48 and 33): The second event shows a flare of index 3+ from Newton (1950), but we were only able to identify a transit time of 32 hours and not the time of the storm used in their study, therefore we cannot clearly establish the time of the flare from it. We used the drawings and measurements and found a change around 23 January. The associated flare could be closer to 24 January, but this would put it almost on the back side of the Sun, therefore it appears improbable. This point is not taken into account in the statistics because of its uncertainty. Considering the results from Newton (1950), we list a flare of index 3+ between 23 and 24 January 1938.
- 16 April 1938 (rank 41): Within our window, two candidate flares might be associated with this geomagnetic storm: a flare on 13 April, and another one on 15 April. As Cliver, Feynman, and Garrett (1990), we chose the flare on 15 April because it is linked to a large SEP event. However, the flare on this day seems to be shorter than the flare two days before (which was seen by more observers). An additional constraint is the estimate of the

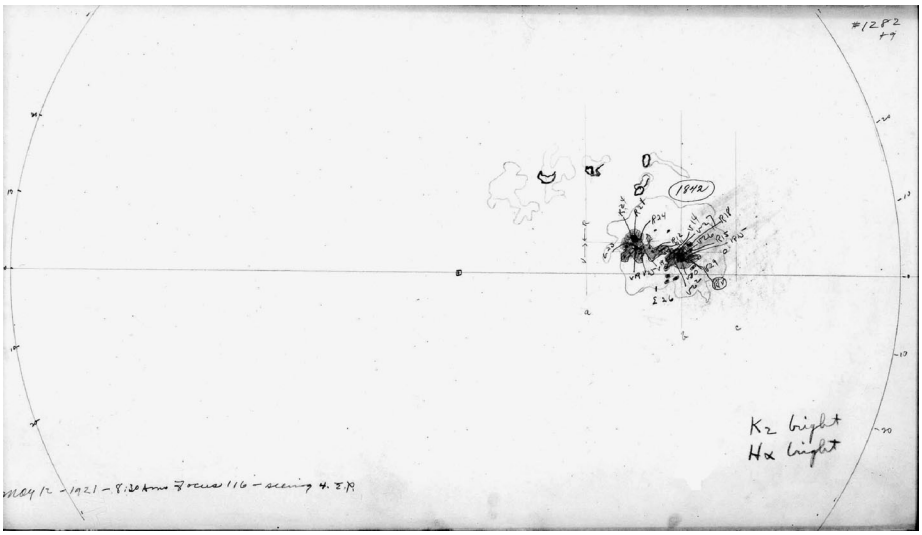

Figure 17 Drawing made of the Sun on 12 May 1921 at the Mount Wilson Observatory. The sunspot drawings are provided by the synoptic program at the *150-Foot Solar Tower of the Mt. Wilson Observatory*.

mean transit speed (≈ 600 km) that would have resulted from a flare on 13 April, which would clearly be an outlier.

- 24 and 31 March and 1940 (ranks 8 and 37): For the storm on 31 March, the Speedbeg proxy speed is negative because this is not a simple geomagnetic or solar event: the storm started before the flare selected here because of another solar event, the geomagnetic conditions were already unsettled. Between 24 and 31 March, four index 3 flares took place that originated from the same AR on 23, 24, 27, and 30 March. There were most certainly complex interactions of the ensuing disturbances in the interplanetary medium. For the storm of 31 March, we chose the flare that seemed to be more developed. On 30 March, the flare lasted about an hour compared to about 30 min for the flare preferred by Cliver, Feynman, and Garrett (1990) on 27 March. Considering the amount of information available for this event, neither the flare on 27 March nor that on 30 March can be completely discarded as sources for the storm on 31 March, but for consistency with the rest of our study, we chose the largest flare based on the available criteria.
- 18 September 1941 (rank 5): Figure 12 shows the variation in area and number of spots for this AR during its passage on the solar disk. The peak in the number of spots is clearly one of the highest of our sample, but it is most interesting that it is the highest on the day when a flare of index 3 was recorded, and not at the time the AR was the largest. The variation in the number of spots is very strong as well, while the variation in area (corrected) is the steepest in the time we observe the flare.
- 25 and 28 March 1946 (ranks 66 and 7): For the first storm, the index indicated for the flare is 1 because there were only two flares of index 1 in the time window for the 25 (starting on the 21). Lacking more information (such as the size of the flare itself), we chose the flare inside the same region as the one from the 28. This event was not taken into account in the statistics because it does not fit the criteria. However, it is very likely that those flares come from the same region, considering it is the only large region on the Sun. Moreover, the time coverage for flares at that point in time (1946) is still not very

high, therefore it is also possible that the flare started before the registered time and was missed at its maximum.

- 28 October 1951 (rank 87): The speed is high but is within the limits established by Cliver, Feynman, and Garrett (1990).
- 4, 13, and 29 September 1957 (ranks 36, 40 and 59): These regions seem to show sympathetic flares in two ARs located on the northern around 22 and 11 degree W. The event from 29 September seems to be linked to the same AR as the one from 4 September, 25 days later. In addition, for the storm on 4 September 1957, we chose the index 3 flare on 3 September and not the index 2 flare of 2 September (Cliver, Feynman, and Garrett, 1990), according to our criteria. We note that in this case, the estimated mean transit speed associated with this event is high, but within an acceptable range for a storm of this rank.
- 30 April 1960 (rank 29): One computed speed is negative because the beginning of the storm (the threshold chosen) is before the flare associated with the event. This simply means that this is a complex event: geomagnetic conditions were already unsettled when the flare occurred. Another flare erupted on the 28 at 02:00 UT, which was much smaller in size and duration (approximately four hours compared to 15 min) and from a different region. With the addition of several smaller flares from the northern region, it might have triggered active conditions before the (assumed to be) faster ICME launched on 29 April caused it to peak on 30 April. We also note that both flares on 1 and 30 April 1960 are linked to the same active region, returning for a new solar rotation, although the group number is different.
- 13 and 16 November 1960 (ranks 28 & 90): For 13 November, the first computed proxy speed is negative (SpeedBeg, sp1) and the second speed (SpeedMax, sp2) appears to be very high. These two storms occurred in close succession on 13 and 16 November. According to our criteria, we chose the largest flare (in size first) in the window before the first storm and then excluded it from the choice of the flare associated with the storm on the 16. The second set of proxy speeds for the 16 is also very high, but, as explained, the only viable candidate was chosen. Cliver and Crooker (1993) list a class 2+ flare on 11 November with a size of 1571 µsd compared to 3000 for the flare on 12 November (3+).
- 28 October 1961 (rank 65): The only active region on the Sun, close to the largest flare in the time window, is a Axx, per convention according to the McIntosh morphological classification (McIntosh, 1990). However, Cliver and Crooker (1993) linked this storm to a filament that disappeared between 25 and 26 October 1961.
- 2 February 1969 (rank 101): This storm is associated with a Axx, per convention (McIntosh, 1990). There is no pronounced Forbush decrease during the storm, but the cosmic ray data show a prolonged ($>$ 10 days) shallow ($\approx$ 3 %) and symmetric depression, which might indicate the passage of a CIR (starting approximately one week before the start of the storm). This is supported by the fact that we find an additional similar depression approximately 27 days later. The fact that we observe the SSC at the beginning of the storm indicates a passage of the interplanetary shock, which could have prolonged the CIR-induced CR depression and caused the geomagnetic storm. However, since there is no indication of the additional ICME-depression that is typically seen in two-step Forbush decreases, it seems that even if there had been an ICME driving the shock, it missed the Earth.
- 4 August 1972 (rank 55): Although a speed of almost 3000 $\mathrm{km\,s^{-1}}$ seems very high, it is derived from simply choosing the largest flare in the time window according to our criteria. This speed agrees with the results of Cliver, Feynman, and Garrett (1990). We note that there is indeed another (white-light) flare in our time window, on 2 August, but

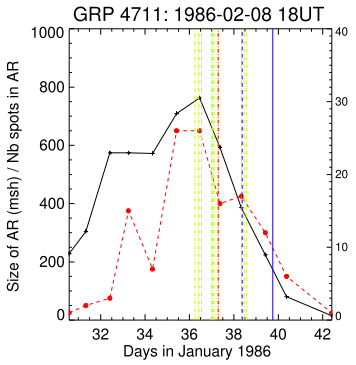

Figure 18 Variations of area (in μsh, black, left-side scale) and number of spots (dashed red, right-side scale) with time (in days) for AR 4711 from 31 January to 11 February 1986. The time of the main flare associated with the storm is plotted as a red vertical line, while other flares of index ≥ 2 from the same region are plotted in green. In dashed blue we show the beginning of the storm and in solid blue the time of the maximum of the storm at Earth.

it is smaller than the Hα flare on 4 August, hence our choice. In addition, there was a very large SEP event on 4 August.

- 13 July and 6 September 1982 (ranks 27 and 39): We chose the flare on 12 July at 09:00 UT, as did Cliver, Feynman, and Garrett (1990). The next flare of 5 September 1982 occurred exactly twice 27 days later, at the same position. This therefore appears to be the same AR that reappears two rotations later: still a very productive AR in terms of flares.
- 8 February 1986 (rank 26): AR 4711 was flare productive as Figure 18 attests. We chose the X1.7 flare from 6 February because the larger X3.0 flare is outside of our four-day window (by almost half a day). The high level of the associated storm is most probably caused by the already unsettled conditions before the ICME created by the solar event on 6 February (associated with the flare) arrived at Earth. Moreover, Garcia and Dryer (1987) indicated that the X3.0 flare on 4 February is linked to a disturbance that passed 1 AU on 6 February, too fast to have been responsible for the storm we are referring to. Cliver, Feynman, and Garrett (1990) gave preference to the flare on 7 February (M5.2) as the most likely candidate for the maximum of the storm, stating that the solar wind was faster than for the previous flare. However, this is a very complex geomagnetic event with five different peaks (Vennerstrøm *et al.*, 2016), which means that there were most certainly complex interactions between ICMEs in the interplanetary environment. Considering the complexity of this event, we chose the strongest flare within our time window. In conclusion, the disturbance associated with the flare on 6 February added to the previous disturbance caused by the flare on 5 February (AR 4711), and the faster disturbance caused by another flare on 7 February might be the link to our maximum on 8 February.
- 13 March 1989 (the famous Québec Storm, rank 2): Cliver, Feynman, and Garrett (1990) and Drake and Gurman (1989) placed the source flare on 10 March. Similarly, the NGDC list of flares for the four days before the event shows that the largest flare occurred in AR 5395 on March 1989 (X4.5). On the other hand, AR 5395 was very productive in terms of flares, as can be seen in Figure 19, and complex interactions in the interplanetary space can explain such a low transit speed.
- 8 November 1991 (rank 34): Cliver *et al.* (2009) attributed this high-ranking storm to the disappearance of a 25-degree-long solar filament in the southern solar hemisphere. However, they stated that the disappearance of this filament can be attributed to the rapid growth of an active region. Consistently, AR 6906 shows an M4.7 flare on 6 November, which we chose as the source of the ensuing geomagnetic storm. The timing is completely consistent with the results from Cliver *et al.* (2009), and the chosen AR, although not alone, seems to be indeed related to this event.

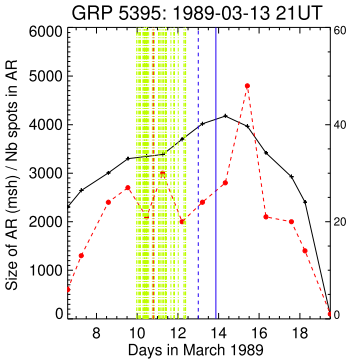

Figure 19 Variations of area (in µsh, black, left-side scale) and number of spots (dashed red, right-side scale) with time (in days) for AR 5395 from 7 – 20 March 1989 around the *Quebec Storm* event. The time of the main flare associated with the storm is plotted as a red vertical line, while other flares of index ≥ 2 from the same region are plotted in green. In dashed blue we show the beginning of the storm and in solid blue the time of the maximum of the storm at Earth.

- 24 November 2001 (rank 92): As we are listing outliers in this list, the event is worth mentioning. The NOAA/NGDC reports places this flare at a size of 24 733 µsd, while the flare events list (http://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/h-alpha/events/) places it at 12 411 µsd: it is clearly an outlier in terms of flare area.

To extend this study some more, we also compared our list of storms with SEP events from different lists: Fritzová-Švestková and Švestka (1966), Švestka (1966), Cliver, Feynman, and Garrett (1990), Dierckxsens *et al.* (2015), and Crosby *et al.* (2015). The list of Cliver, Feynman, and Garrett (1990) contains 23 events and extends from 1938 to 1989; most of our events overlap except for a few events that do not appear in our list, or do appear, but are associated with different sources on the Sun (six events). There are three reasons for these differences: (1) Cliver, Feynman, and Garrett (1990) used a different ranking process of storms (based on $A_p{}^*$), (2) we used a higher selection threshold, and (3) we chose a different source AR/flare (event of September 1957). Of course, during this 1938 – 1989 period some of the storms from our list do not appear in the list from Cliver, Feynman, and Garrett (1990). This does not necessarily mean that there was no SEP event associated with the events from our list, but it might mean (1) that there were several proton events (Cliver, Feynman, and Garrett, 1990 only retained simple events where there was a single major proton flare), or (2) that the SEP events were too weak to be detected. This is confirmed by a simple exercise: if we take polar cap absorption (PCA) events as a proxy for SEP events for 1938 – 1962 (Fritzová-Švestková and Švestka, 1966; Švestka, 1966) and compare their PCA list with our sample of storms during the same period, we count 41 storms, of which 30 have a proxy-SEP association. This means that during this period, ≈ 75 % of the events seem to be indeed associated with an SEP, and most probably a fast CME. By extension, the flare association is most probably correct. If we compare our list of events during the most recent period (SOHO era) to the SSE list from Dierckxsens *et al.* (2015) or Crosby *et al.* (2015), we find that 11 out of 13 events show an associated SEP event. For the two events that do not show an associated SEP event, we can still confirm the presence of a CME–ICME.

In conclusion, although not all the events from our storm list can easily be associated with an SEP event, the most recent storms (SOHO era) are all associated with a CME (*cf.*

Section 3.1) and almost all to a large SEP event. For older events for which we have proxy data, an SEP association is very probable, thus our assumptions about large flares associated with fast CMEs seem to hold.

5. Analysis

5.1. Correlation with the Level of the Geomagnetic Event

To understand how the different parameters are linked to the ensuing geomagnetic storm, we wish to assess the level of correlation between the solar flare, CME, ICME, and geomagnetic parameters: we selected a sample of significant quantities and cross-correlated them. For this purpose, we used two different correlation schemes: the Pearson correlation coefficient, which gives us the level of linear correlation between two quantities, and the Spearman correlation coefficient, which is based on rank and gives us the level of nonlinear correlation between two quantities (Pearson, 1895, Spearman, 1904). The Spearman test is nonparametric and uses a common approximation when there are double values. The Pearson correlation significance is shown for a one-tailed test. Figure 20 shows how the significance varies for the different numbers of common values in our dataset. This first part of the analysis has one purpose: to assess the existence of possible correlations and select candidate parameters for a more thorough check. The numbers should be taken as an indicator of the relative level of correlation between two parameters. High confidence levels indicate that the relations are possible and unlikely to have occurred by chance.

Table 4 shows the level of correlation for Pearson and Spearman for pairs of parameters associated either with the Sun or with the level of the geomagnetic storm. The size of the recorded flare (in µsd) or the flare index do not appear in this table because we were only able to consider their variations over 54 events out of the 105 (flare size was not measured consistently before 1957, and flare index before the 1920s). We therefore reduced the set of eleven parameters (ddc, lcm, dcm, nbs, fls, sfl, fli, ΔT, speed, aa, and issn, *cf.* Table 4) to correlate by pairs to the rest. For these nine parameters (columns of Table 4), we can study 102 events out of the total of 105: we excluded only three of the oldest events for which the number of spots inside an AR was impossible to determine because we lack precise drawings at the time of the event.

Principal component analysis or PCA (Chatfield and Collins, 1980) has also been considered to study this large sample of parameters, but it appears not to be relevant in this case because the variance values of the first dimensions were too low regardless of the combination of parameters we considered. Indeed, PCA or SVD (singular value decomposition) is

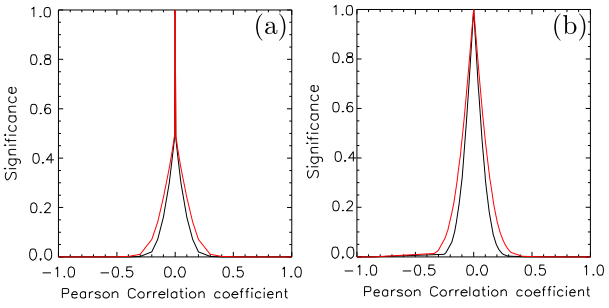

Figure 20 (a) Pearson significance levels for 54 (red) and 102 (black) elements. 90 % significance corresponds to everything below 0.1 on this graph. (b) Spearman significance levels for 54 (red) and 102 (black) elements.

Table 4 Pearson and Spearman correlation coefficients of nine selected parameters from this study: ddc (distance to disk center or dist in Table 3), lcm (longitude from central meridian), dcm (distance from central meridian, lcm without sign), nsflare (number of spots in AR at the time of the flare, n3 in Table 3), szf (size of the AR at flare time, msh), dth (time in hours between flare and max of storm), sp2 (proxy mean transit speed computed from dth), pk (peak of aa averaged over 24 hours) and ISSN (daily international sunspot number). These correlations are for 102 values in common, they are indicated in $100 \times r$. Pearson significance thresholds are as follows: 0.23 corresponds to 99 % (red), 0.16 to 95 % (magenta) and 0.13 to 90 % (pink).

	ddc	lcm	dcm	nsflare	szf	dth	sp2	pk	ISSN
ddc	*	-10/-12	84/79	-17/-17	7/5	14/8	-2/-8	-18/-18	5/2
lcm	-10/-12	*	-13/-14	13/20	8/15	4/-7	-5/7	-8/-7	-11/-6
dcm	84/79	-13/-14	*	-24/-26	-5/-5	11/5	-3/-5	-27/-25	5/-5
nsflare	-17/-17	13/20	-24/-26	*	39/44	-15/-8	12/8	19/19	36/24
szf	7/5	8/15	-5/-5	39/44	*	-15/-15	18/13	28/27	22/28
dth	14/8	4/-7	11/5	-15/-8	-15/-15	*	-88/-100	-23/-23	0/-4
sp2	-2/-8	-5/7	-3/-5	12/8	18/13	-88/-100	*	17/23	1/4
pk	-18/-18	-8/-7	-27/-25	19/19	28/27	-23/-23	17/23	*	1/5
ISSN	5/2	-11/-6	5/-5	36/24	22/28	0/-4	1/4	1/5	*

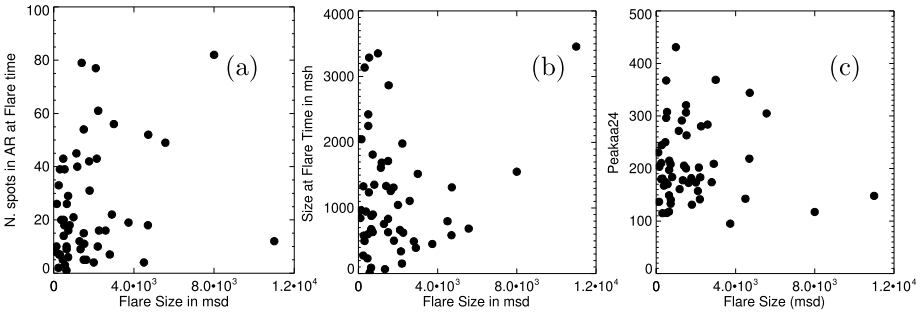

Figure 21 (a) Size of the flare (μsd) versus number of spots in the corresponding AR at the time of the flare. (b) Size of the flare (μsd) versus size of the corresponding AR at the time of the flare (msh). (c) Size of the flare (μsd) versus the value of peak in aa24 index.

typically used to reduce the dimensions of a problem. Usually, most of the variance can be found in the first two dimensions (more than 80 %), which means that most of the similarities and differences will be accessible on these two arbitrary axes. However, this method is sensitive to noise and outliers, and in our case, this creates nine or eleven dimensions that contain similar amounts of variance thus preventing us from simplifying the problem.

Some high correlations, like the one between the transit time dth and our proxy of the mean transit speed, were not taken into account in this analysis because they are obvious and not related to the subject at hand. However, we note that the associated Pearson correlation coefficient is 0.88, while the Spearman correlation coefficient is 1: this corresponds to the fact that the relation between the two is, in fact, nonlinear. The two coefficients for each relation can therefore help us understand the nature of the correlation. In the next paragraphs, we go into more detail for cases where the correlation between the two considered quantities is significant (colored cases in Table 4).

As expected, the available flare sizes appear correlated with the numerical flare index scale described in Section 2.4. A possible correlation was found between the size of the flare and the number of spots within the AR at the time of the flare (see Figure 21(a)). A very weak linear correlation is found between the observed size of the flare at its peak

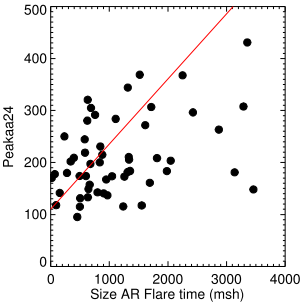

Figure 22 Relation between the size of the AR at the time of the flare and the geomagnetic impact (aa 24 h).

value and the size of the corresponding AR in white light (see Figure 21(b)). We note that the flare size value is recent (from 1957) and is very dependent on the time coverage, *i.e.* if the time coverage is not 100 %, the observed value can be very far from the actual peak value. On the other hand, the size of the region itself at maximum or at flare time suffers less from such a lack in time resolution, as shown in Section 3.1. As mentioned in previous sections, the size of an AR is correlated with the number of spots inside it during most of its lifetime, which links the correlation of the flare size to both the number of spots and size of AR at the time of the flare. This is partly seen in Table 4, and the individual relation was checked for the regions for which we have data in this study. This means that the correlation between these quantities and the flare size is related.

And last, as shown in Figure 21(c), the size of the flare does not appear to be strongly correlated to the peak in the aa24 index at first. However, the lower end of the distribution in flare size vanishes at higher aa indices (figure not shown here). This is in agreement with the conclusions of Howard and Tappin (2005), *i.e.* that statistically speaking, the largest flares are most often associated with the strongest geomagnetic storms and the smaller storms are rarely associated with large flares.

However, if we now look at the correlation of the level of geomagnetic activity with the size of the ARs at flare time in our sample, the picture becomes sharper: the correlation appears in Figure 22. It is reasonably good minus a few outlying points (*i.e.* storms ranked 75 and 93 in Table 2). As mentioned above, this might be because the flare size varies very rapidly (on a timescale of minutes) and is thus very sensitive to the time coverage, while the size of the AR at flare time will not be affected by having only daily measurements. This is consistent with results from Section 5.3.2, which shows that our sample of ARs is much larger and much more complex than the average AR. It also confirms results by Srivastava and Venkatakrishnan (2002) and Aulanier *et al.* (2013), who suggested that the larger the area of the individual sunspot, the larger the total magnetic energy that can be available, in principle for release and consequent driving of the CME.

The connection between the position of the AR on the Sun and its geomagnetic impact appears significant, as stressed by significant correlation coefficients between the distance to central meridian at the time of the flare and the peak in the 24-hour running mean of aa, *aa24*. This is similar to the results of Wang *et al.* (2002): according to them, 83 % of the geoeffective CMEs erupt within 30 degrees of the central meridian. This relation is shown in Figure 23. These two quantities do not appear to be clearly correlated at first glance, but when the distribution are considered in different bins of the *aa24* index, regions located close to the limb are not connected to the strongest geomagnetic events. This is in line with Section 5.3.1 and previous results on the preferred positions of the ARs linked to extreme solar events (Akasofu and Yoshida, 1967, Gonzalez *et al.*, 1996, Srivastava *et al.*, 1997, Srivastava *et al.*, 1998).

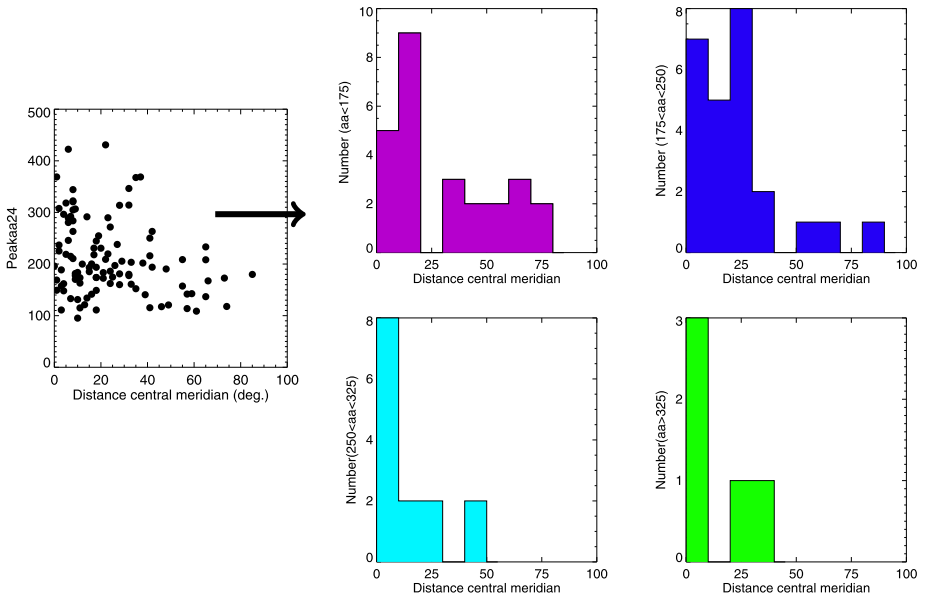

Figure 23 Left: Relation between the distance from the central meridian of our ARs at the time of the flare and the geomagnetic impact (aa over 24 h). Right: The distribution of the distance from the central meridian of our ARs at the time of the flare is shown for four different bins of aa index: aa $<$ 175 nT, 175 $\leq$ aa $<$ 250 nT, 250 $\leq$ aa $<$ 325 nT, and aa $\geq$ 325 nT.

The connection between our proxy of the mean transit speed of the event from the Sun and its geomagnetic impact seems to be non-negligible, as shown in Figure 24, although not so clearly. To clarify this, Figure 24 shows the distribution in *mean transit speed* in different bins of *aa24* level. It demonstrates that when the geomagnetic impact is higher, the CMEs–ICMEs that are assumed to be related transit faster. This study supports results by Cliver, Feynman, and Garrett (1990) as well as Srivastava and Venkatakrishnan (2002) or Howard and Tappin (2005), but on a significantly larger sample.

5.2. Occurrence of Geomagnetic Events or Flares Throughout the Solar Cycle

The high level of correlation between the International Sunspot Number (ISSN V1.0, http://www.sidc.be/silso/versionarchive) and the number of flares is shown in Figure 25. These flare numbers were normalized to account for an obvious level of incompleteness before the end of Cycle 19. There were approximately eight times more facilities observing flares after Cycle 19 than before.

Figure 26 shows in which phase of the solar cycle (minimum, maximum, rising, declining, values from the NGDC: http://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-indices/sunspot-numbers/cycle-data/Table_cycle-dates_maximum-minimum.txt) the flares are preferentially located, while Figure 27(a) shows in which phase the flares associated with our sample of extreme events appear.

Figure 27(b) presents the distribution of the ISSN on the day of these flares, while Figure 27(c) shows the typical distribution of the daily values of the sunspot number since 1870. There seems to be no clear threshold below which no events or flares occur, but considering the whole sample of 105 events, the probability of an event to occur below a sunspot

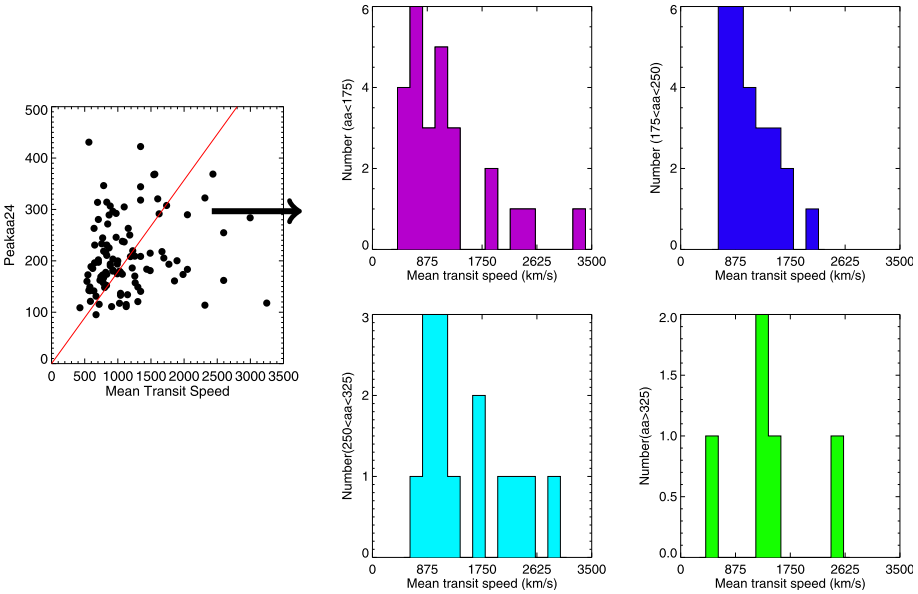

Figure 24 Left: Relation between the proxy of the mean transit speed of travel of the ICME and the geomagnetic impact (aa 24 h). Right: The distribution of this proxy mean transit speed is shown for four different bins of the aa index: aa $<$ 175 nT, 175 $\leq$ aa $<$ 250 nT, 250 $\leq$ aa $<$ 325 nT, and aa $\geq$ 325 nT.

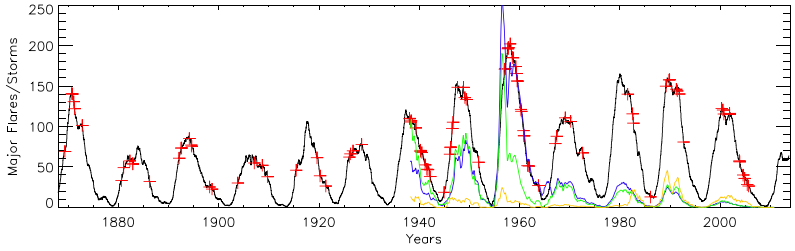

Figure 25 Number of flares of intensity (index) 1, 2, and 3 per month from 1938 to the present (in blue, green, and yellow, respectively). The number of flares has been normalized to account for the evolution of the overall coverage between 1938 and now. Overplotted is the International Sunspot Number to relate solar cycles to these numbers (black). Red crosses represent the dates of the events related to the yearly smoothed sunspot number.

number of 50 is lower than 15 %. Less than 30 % of events occur during the minimum or rising phase. More than 70 % of the extreme geomagnetic events occur in the maximum or decaying phase of the solar cycle and when the sunspot number is greater than 50. However, when the sunspot number is below 50 or in the minimum or rising phase of the solar cycle, the probability of seeing an extreme event drops drastically. The possibilities of predictions remain limited with that information, but it is still clear that this sample of events is not typical.

The comparison of Figures 26 and 27(a) shows that flares, most often associated with geomagnetic events, occur mostly in the same phases of the cycle (although with some

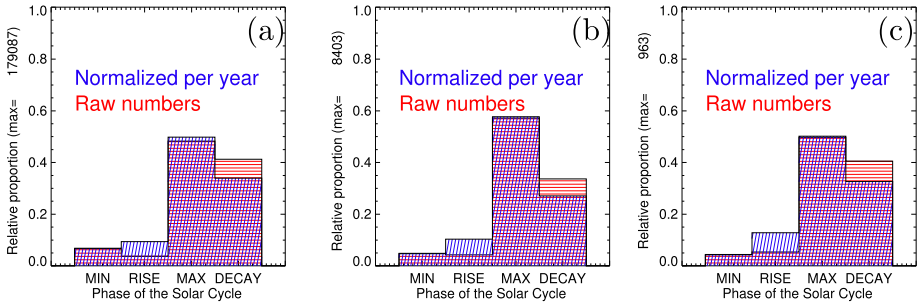

Figure 26 Distribution of the flares in the different phases of the solar cycle. We take as reference the minima and maxima from the NGDC. We consider the minimum to be ±1.5 years around minimum time and the maximum ±1.5 years around maximum time. (a) Flares of index 1, (b) index 2, and (c) index 3.

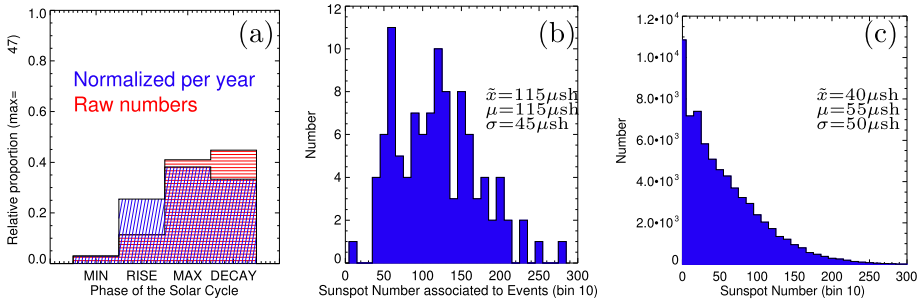

Figure 27 (a) Distribution of the 105 storms or events in the different phases of the solar cycle. We consider that the minimum is ±1.5 years around minimum time and the maximum is ±1.5 years around maximum time. (b) Distribution of the 105 storms or events in values of daily ISSN based on values from the SIDC (not smoothed). (c) Distribution of the daily ISSN based on values from the SIDC (not smoothed, from 1870).

differences in relative frequency), which means that if the extreme events occur in these phases, it is *mainly because* most of the flares occur in the same phases of the solar cycle.

5.3. Distribution of Active Region Parameters

5.3.1. Position of Active Regions

Figure 28(a) shows the positions on the solar disk of all the ARs (and associated flares) that we have associated with our sample of geomagnetic events. It clearly shows that the ARs most probably responsible for the considered geomagnetic storms are located preferentially close to the center of the Sun.

Figures 28(b) to (d) show the distribution of different positions of the considered ARs: (1) the distance from the disk center in solar radii and (2) the distance from the central meridian (CM) of the solar disk, and (3) the latitude of the ARs. Figure 28(b) shows that even when we consider only events in the SOHO period or back to 1938, the associated ARs are close to the central meridian at the time of the flare (for the period from 1938 to the present, 70 % of ARs are within 30 degrees of the CM and 87 % are within 50 degrees). In two cases, ARs located close to the limbs caused large events (85 degrees in the east, and 73 in the west). In the distribution in distance from the disk center this tendency is less

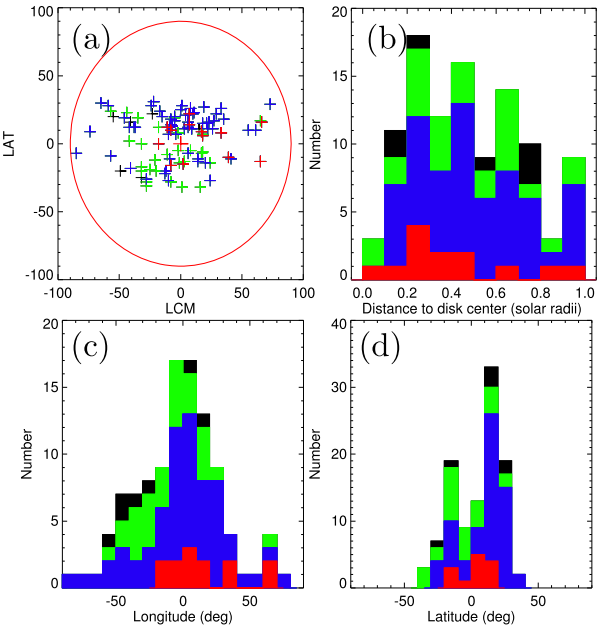

Figure 28 Distribution of the positions of the source ARs. Colors correspond to different time periods. Black covers the whole sample (1868 – 2010), green the RGO era from 1874 to 2010. Blue covers data from the extended flare patrol era from 1926 to 2010, and red covers only the SOHO era (1996 – 2010). (a) Positions of our sample of identified ARs on the solar disk (degrees). (b) Distance from the center of the solar disk (in solar radii) (c) Distance from the central meridian in degrees; east is negative, west is positive. (d) Heliographic latitude of the ARs (degrees).

evident, but shows clearly in a distribution plot of the angular distance from the disk center (arcsin(ddc), not shown here).

Investigations of a possible east-west asymmetry in the distribution in longitude of ARs linked to major geomagnetic events are not conclusive. Figure 28(c) shows an asymmetry biased toward the western hemisphere between roughly 1926 and 2010, and especially for the SOHO era, as developed in Wang *et al.* (2002) and Zhang *et al.* (2003). However, when the sample is extended back in time, this bias seems to disappear: it is probably an effect of the small number of events.

5.3.2. Complexity of Active Regions

On the day of the flare, about 88 % of ARs present McIntosh Z-type (Zpc) D, E, or F. Figure 29 shows the predominance of the most complex regions in our specific sample compared to the distribution in an unbiased sample from the USAF catalog over more than two solar cycles. It clearly shows that flares occur essentially in complex ARs with a complex magnetic configuration (Zirin and Liggett, 1987; Sammis, Tang, and Zirin, 2000; Ternullo *et al.*, 2002). The remaining less complex regions are C-type (5 %), H-type ARs (3 %), and small ARs of type A or B (3 %). C- and H-type regions can still be considered complex: two of the H-types are large ARs in a phase of decay, and another is just appearing on the east limb. This makes this group's Hsx classification unclear at the time of the flare, but it clearly is larger as it reaches Eki a few days later. The peculiar cases of A- or B-type ARs need to be studied in more detail, but this is the subject of another study.

Figure 30 shows the distribution of the maximum sizes of the ARs and their sizes at the time of the flare. For comparison, the left panel shows the distribution of the sizes of all groups in the Royal Greenwich Observatory catalog in the period of 1874 – 1982. The mean sizes of the ARs considered in our sample at the time of the flare or at their maximum development are much larger than the mean size of groups in general.

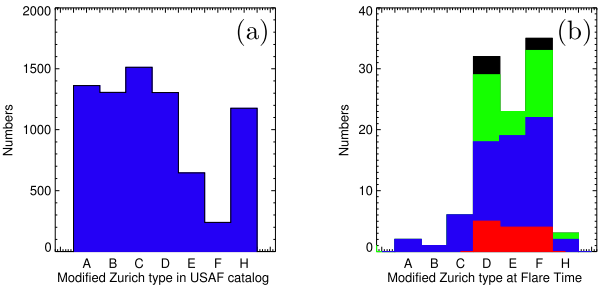

Figure 29 (a) Distribution of the McIntosh morphological type of ARs from 1981 to 2011 in the USAF catalog for the Holloman station (Z-type from McIntosh or modified Zurich type). (b) Distribution of the described morphological type of identified ARs on the solar disk at flare time for the different time periods using the same color scheme as in Figure 28.

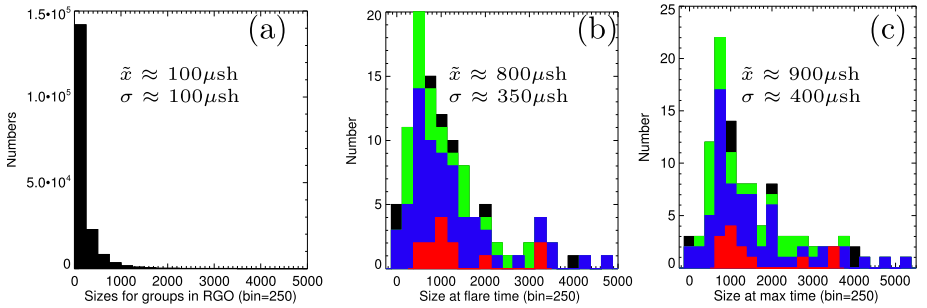

Figure 30 Distribution of the size of ARs in the RGO catalog and our sample. (a) Distribution of AR areas in the RGO catalog (1874 – 1982) in bins of 250 µsh. (b) Distribution of the area of ARs at the time of flare using the same color scheme as Figure 28. (c) Distribution of the maximum area of ARs. Median ($\tilde{x}$) and standard deviation (σ) values are indicated.

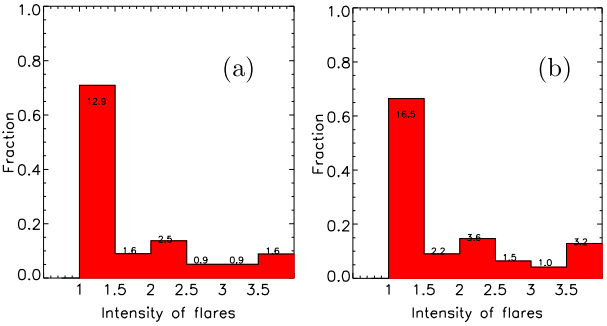

Figure 31 (a) Average number of flares in our sample in each bin of intensity per AR in the period 1938 – 2010. (b) Average number of flares in each bin of intensity per AR in the period with more reliable flare information (1966 – 2010).

5.3.3. Distribution of Flares Within Active Regions

Figure 31 shows the number of flares of index 1, 2, and 3 associated with our sample of ARs (with a distinction before and after Cycle 19, marking the level of completeness). From 1966 to the present, there is no significant incompleteness in the coverage of the flares (Bouwer *et al.*, 1982). The distribution in energy is less steep than for the whole sample of flares (see Figure 3), *i.e.* these ARs are typically in the tail of the normal-sample distribution. The particularity of these events appears in an atypical distribution of energy in the associated ARs.

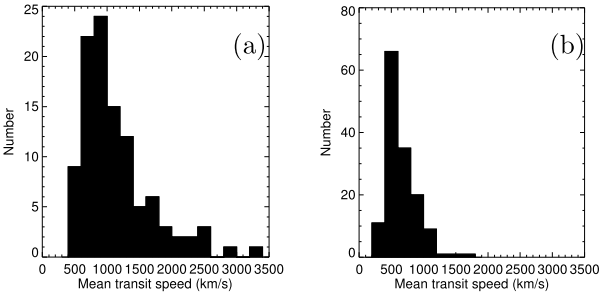

Figure 32 (a) Distribution of our proxy mean transit speed (bin = 200 km s^{-1}) computed from the storm max time – time of flare (km s^{-1}) for this study (column 19). (b) Distribution of the mean transit speed extracted from CME–ICME pairs for the SOHO era.

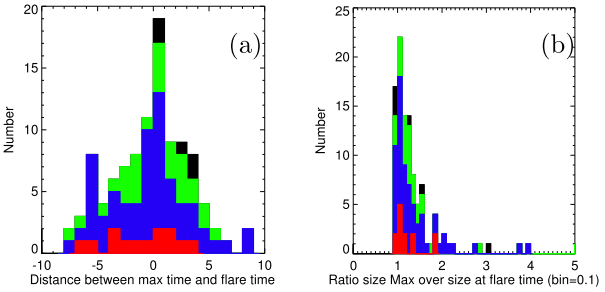

Figure 33 (a) Time between maximum size of AR and flare in days. (b) Ratio between the size of the AR at maximum development and its size at the time of the associated flare. Both panels use the same color scheme as Figure 28.

For the whole sample of flares, there are about 300 times more flares of index 1 than there are of index 3 or greater (see Figure 3(c) after 1966). For our sample of *extreme events*, this ratio reaches ≈ 10 to 1, *i.e.* 10 flares of index 1 for one flare of index 3, as shown in Figure 31. This sample of storms and their associated solar counterparts do not show the energy distribution associated with more typical geomagnetic events: they show a larger proportion of high-intensity flares. This can be linked to the level of complexity of the studied geomagnetic storms and the fact that the aa level is already high for a large part of storms at the time of the SSC (Vennerstrøm *et al.*, 2016).

5.4. Temporal Characteristics

Figure 32 shows the distribution of (1) the proxy we used for the mean transit speed of events from the Sun to Earth in our study compared to (2) the mean transit speed established from CME-ICME pairs during the SOHO era from studies by Zhang *et al.* (2007), Gopalswamy *et al.* (2001), Schwenn *et al.* (2005), Richardson and Cane (2010), and Dumbović *et al.* (2015). The distribution of mean transit speed in our sample is clearly not typical as it peaks closer to 1000 km s^{-1} than the 650 km s^{-1} of the SOHO era study, but it also shows a high-speed tail that is most uncommon. In addition, as Moon *et al.* (2003) stressed in their article, the acceleration and speed of the CMEs are very well correlated with the energy release in the flare. This confirms that this is no typical sample of associated flares.

Figure 33(a) shows that the flare typically occurs within two days before or after the AR has reached its maximum size. In addition, the size of the AR on the day of the associated flare is very close to its size on the day of its maximum development (*cf.* Figure 33(b)). This indicates that the flare occurs when the AR is the most complex, thus when it stores a lot of magnetic energy.

6. Conclusions

We have extended the work from Vennerstrøm *et al.* (2016), referred to as Paper I, back to the surface of the Sun to show which static and dynamic properties of the Sun and mostly sunspots could explain why they are often precursors or sources of the most extreme geomagnetic storms.

First, we pointed out the complexity of the ARs that can be linked to our sample of geomagnetic events: they are significantly larger and more complex (type, extent, area, and number of spots inside an AR) than the average ARs: almost 90 % of ARs present the most complex McIntosh type (D, E, or F), and their areas are much larger than the average area of ARs. This can be seen as an interesting counterpart to the complexity of the storms mentioned in Paper I and is mainly related to the fact that more complex ARs produce more flares.

As expected, in our sample the position of the AR at the time of the flare, *i.e.*, regarding the central meridian, is also a determining parameter for the level of the storm that will follow: approximately 70 % of ARs are within 30 degrees of the central meridian in longitude, and the distribution is centered and peaks at the central meridian.

Following the conclusions from Paper I regarding high solar wind speed and short duration SSC (faster moving shocks) for this sample of 105 storms, we conclude that the transit time (directly linked to the speed of the CME, and thus to the transit speed of the ICME, Wang *et al.*, 2002) deduced from the flare time and the maximum of the storm determined by this study is statistically shorter for larger events.

Moreover, although it is very difficult to link the size of the flare to any other index, there seems to be a significant correlation between the size of the AR at the time of the flare and the level of the geomagnetic storm that is induced at Earth. There are also indirect correlations (but undeniably present) between the flare size itself and AR parameters, such as the size of the AR and the number of spots inside it. The flare index is correlated to our proxy of the mean transit speed of the ICME, which is itself connected to the level of the geomagnetic storm.

Hence we join the conclusions of previous works about the speed of the ICME (we did not determine the initial speed of the CME), the position of the associated source region, and the level of the associated flare (Srivastava and Venkatakrishnan, 2002, 2004; Zhang *et al.*, 2007, and Dumbović *et al.*, 2015) with a different approach based solely on historical data. When we add to these conclusions the huge size and complexity of the associated ARs, all of these characteristics point in the same direction: more energy at the source of the phenomenon results in more energetic (faster) CMEs, larger flares, and associated ARs. Although this looks much like the "big flare syndrome" mentioned by Kahler (1982), this is a very specific sample of very energetic events.

In addition, we reached a very interesting conclusion about the dynamic properties of our sample of associated ARs. The evolution of an AR (area, complexity, number of spots) during the solar rotation, including the solar flare linked to the sample of geomagnetic storms, shows that a large majority of the ARs involved show sudden changes within a day of the supposed flare (these sudden changes are defined in Section 3). This seems to suggest that the release of magnetic energy involved in the process can be detected in the complexity or size of the AR at the source of the chain of events. At this point, the causality is unclear, however: is the rapid change in AR complexity at the source of, or is it a consequence of the release of magnetic energy? In addition, these sudden changes appear to be necessary conditions for a large flare, but they are not necessarily sufficient. A lack of sufficiency is true in general for all of the sunspot active region parameters (active region size, complexity, location) because a favorable region may not erupt near disk center or may present a

northward-leading field in the CME. In addition, although this mechanism is sometimes apparent in our sample of the most extreme events, it is entirely possible that this criterion appears blurred when the level of the associated geomagnetic event is lowered.

The probability of flaring is strongly linked to the number of ARs on the Sun (*i.e.* also to the sunspot number). However, from this study, there is no definitive threshold in sunspot number below which no flares and no associated geomagnetic events can be found. The extreme events and their associated flares of our sample mostly occurred during the maximum and declining phases of the solar cycle. When all the flares observed between 1938 and now are considered, the largest flares also have a tendency to show when the cycle is at its maximum or is declining. This would prove that extreme events are linked mainly to the largest flares. The probability of a large geomagnetic storm is the combined probability of a large (area) and complex (McIntosh classification) sunspot group combined with the probability of flaring of such a group (McIntosh, 1990; Qahwaji and Colak, 2007 and references therein).

And last, we join the conclusions of Paper I concerning the complexity of the storms presented here. Numerous ARs associated with the most extreme storms presented here can be associated with several storms, sometimes when the same AR reappears about 28 days later for another solar rotation. It is not just the storm itself that is multiple, the ARs all show multiple flares of high energy in the time window and outside of it. This frequency is also an important parameter in the subsequent level of the geomagnetic storm.

Problem storms are geomagnetic storms for which no clear association can be made. They appear in studies by Dodson and Hedeman (1964), Dodson, Hedeman, and Mohler (1979), Joselyn and McIntosh (1981), McAllister *et al.* (1996), Cliver *et al.* (2009). In our sample of 105 extreme geomagnetic storms, we count only two cases of *problem storms*: October 1961 and February 1969. This is because most studies of *problem storms* examined samples of geomagnetic storms that, although they may contain a few storms as extreme as those from our sample (the October 1961 storm appears in Dodson and Hedeman, 1964 and the November 1991 storm appears in Cliver *et al.*, 2009, for example), contain mostly severe storms of a lesser degree. From a purely statistical point of view, *problem storms* represent less than 2 % of the cases, which means that although they are worth mentioning and studying (*cf.* Section 4), for the conclusions drawn from our sample, they are negligible.

Using historical data to increase the statistics, we established the most probable solar characteristics associated with extreme geomagnetic events from 1868 to 2010. In addition, we are now able to distinguish which events followed the most probable behavior and which seemed to be outliers. We refer here to the case of well-known problem storms (like those in October 1961 and February 1969), but also to events whose parameters do not follow the established correlations. This is the case of the March 1989 storm, for example, for which the mean transit speed is a clear outlier considering its rank. To reach in-depth conclusions about outlier cases (possibly missing alarms) but also false alarms, case-by-case analyses are needed. However, this is the subject of another study.

Acknowledgements This work has been conducted in the frame of the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 263252 (COMESEP). We also acknowledge the support from the Belgian Solar–Terrestrial Center of Excellence (STCE) funded through the Belgian Science Policy Office (BELSPO). L.L. would like to thank O. Lemaître (Royal Observatory of Belgium) for measuring solar data on drawings for this work, and J.M. Vaquero (Universidad de Extremadura, Spain) for his invaluable help with the historical data. M.D., D.S. and B.V. acknowledge financial support by Croatian Science Foundation under the project 6212 "Solar and Stellar Variability". The *Mt. Wilson 150-Foot Solar Tower* has been operated by UCLA since 1985, with funding from NASA, ONR and NSF, under agreement with the Mt. Wilson Institute. Prior to 1985 the program was operated by the Hale Observatories with support by the Carnegie Institute of Washington.

References

Akasofu, S.-I., Yoshida, S.: 1967, The structure of the solar plasma flow generated by solar flares. *Planet. Space Sci.* **15**, 39. DOI. ADS.

Arcimis, A.T.: 1901, *Astronomia Popular – Tome I*, Montaner y Simón Editores, Barcelona.

Arlt, R., Fröhlich, H.-E.: 2012, The solar differential rotation in the 18th century. *Astron. Astrophys.* **543**, A7. DOI. ADS.

Athens, O.: 1871, Beobachtungen der Sonne im Jahre 1870 auf der Sternwarte zu Athen. *Astron. Nachr.* **77**, 145. ADS.

Aulanier, G., Démoulin, P., Schrijver, C.J., Janvier, M., Pariat, E., Schmieder, B.: 2013, The standard flare model in three dimensions, II: upper limit on solar flare energy. *Astron. Astrophys.* **549**, A66. DOI. ADS.

Balthasar, H., Lustig, G., Woehl, H., Stark, D.: 1986, The solar rotation elements i and omega derived from sunspot groups. *Astron. Astrophys.* **160**, 277. ADS.

Bouwer, S.D., Donnelly, J., Falcon, J., Quintana, A., Caldwel, G.: 1982, A summary of solar 1 – 8 A measurements from the SMS and GOES satellites, 1977 – 1981. *NASA STI/Recon Technical Report N* **83**, 23263. ADS.

Cane, H.V.: 1985, The evolution of interplanetary shocks. *J. Geophys. Res.* **90**, 191. DOI. ADS.

Cane, H.V., Kahler, S.W., Sheeley, N.R. Jr.: 1986, Interplanetary shocks preceded by solar filament eruptions. *J. Geophys. Res.* **91**, 13321. DOI. ADS.

Cane, H.V., Richardson, I.G., St. Cyr, O.C.: 2000, Coronal mass ejections, interplanetary ejecta and geomagnetic storms. *Geophys. Res. Lett.* **27**, 3591. DOI. ADS.

Cane, H.V., Richardson, I.G., Wibberenz, G.: 1996, Energetic particles and solar wind disturbances. In: Winterhalter, D., Gosling, J.T., Habbal, S.R., Kurth, W.S., Neugebauer, M. (eds.) *Am. Inst. Phys. Conf. Ser.* **382**, 449. DOI. ADS.

Carrasco, V.M.S., Vaquero, J.M., Gallego, M.C., Trigo, R.M.: 2013, Forty two years counting spots: solar observations by D.E. Hadden during 1890 – 1931 revisited. *New Astron.* **25**, 95. DOI. ADS.

Chatfield, C., Collins, A.J.: 1980, *Introduction to Multivariate Analysis* Springer, New York. DOI.

Cliver, E.W.: 1995, Solar flare nomenclature. *Solar Phys.* **157**, 285. DOI. ADS.

Cliver, E.W., Crooker, N.U.: 1993, A seasonal dependence for the geoeffectiveness of eruptive solar events. *Solar Phys.* **145**, 347. DOI. ADS.

Cliver, E.W., Feynman, J., Garrett, H.B.: 1990, An estimate of the maximum speed of the solar wind, 1938 – 1989. *J. Geophys. Res.* **95**, 17103. DOI. ADS.

Cliver, E.W., Balasubramaniam, K.S., Nitta, N.V., Li, X.: 2009, Great geomagnetic storm of 9 November 1991: association with a disappearing solar filament. *J. Geophys. Res.* **114**, A00. DOI. ADS.

Colak, T., Qahwaji, R.: 2009, Automated solar activity prediction: a hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather* **7**, 6001. DOI. ADS.

Crooker, N.U., Siscoe, G.L., Shodhan, S., Webb, D.F., Gosling, J.T., Smith, E.J.: 1993, Multiple heliospheric current sheets and coronal streamer belt dynamics. *J. Geophys. Res.* **98**, 9371. DOI. ADS.

Crosby, N., Heynderickx, D., Jiggens, P., Aran, A., Sanahuja, B., Truscott, P., Lei, F., Jacobs, C., Poedts, S., Gabriel, S., Sandberg, I., Glover, A., Hilgers, A.: 2015, SEPEM: a tool for statistical modeling the solar energetic particle environment. *Space Weather* **13**, 406. DOI. ADS.

Denning, W.F.: 1873, Observations of solar spots – October 14th to November 15th, 1872. *Astron. Reg.* **11**, 73.

Dierckxsens, M., Tziotziou, K., Dalla, S., Patsou, I., Marsh, M.S., Crosby, N.B., Malandraki, O., Tsiropoula, G.: 2015, Relationship between solar energetic particles and properties of flares and CMEs: statistical analysis of solar cycle 23 events. *Solar Phys.* **290**, 841. DOI. ADS.

Dodson, H.W., Hedeman, E.R.: 1964, Problems of differentiation of flares with respect to geophysical effects. *Planet. Space Sci.* **12**, 393. DOI. ADS.

Dodson, H.W., Hedeman, E.R., Mohler, O.C.: 1979, Examples of problem flares or situations in past solar–terrestrial observations. In: Donnelly, R.F. (ed.) *NOAA Solar–Terrestrial Predictions Proc., Solar–Terrestrial Physics* **1**, 385. ADS.

Drake, S.A., Gurman, J.B.: 1989, The SMM UV observations of active region 5395. In: Winglee, R.M., Dennis, B.R. (eds.) *Developments in Observations and Theory for Solar Cycle 22*, 248. ADS.

Dumbović, M., Vršnak, B., Čalogović, J., Karlica, M.: 2011, Cosmic ray modulation by solar wind disturbances. *Astron. Astrophys.* **531**, A91. DOI. ADS.

Dumbović, M., Devos, A., Vršnak, B., Sudar, D., Rodriguez, L., Ruždjak, D., Leer, K., Vennerstrøm, S., Veronig, A.: 2015, Geoeffectiveness of coronal mass ejections in the SOHO era. *Solar Phys.* **290**, 579. DOI. ADS.

Erwin, E.H., Coffey, H.E., Denig, W.F., Willis, D.M., Henwood, R., Wild, M.N.: 2013, The Greenwich photoheliographic results (1874 – 1976): initial corrections to the printed publications. *Solar Phys.* **288**, 157. DOI. ADS.

Feynman, J.: 1980, Implications of solar cycles 19 and 20 geomagnetic activity for magnetospheric processes. *Geophys. Res. Lett.* **7**, 971. DOI. ADS.
Fritzová-Švestková, L., Švestka, Z.: 1966, Type IV bursts, II: in association with PCA events. *Bull. Astron. Inst. Czechoslov.* **17**, 249. ADS.
Garcia, H.A., Dryer, M.: 1987, The solar flares of February 1986 and the ensuing intense geomagnetic storm. *Solar Phys.* **109**, 119. DOI. ADS.
Gonzalez, W.D., Tsurutani, B.T., McIntosh, P.S., Clúa de Gonzalez, A.L.: 1996, Coronal hole-active region-current sheet (CHARCS) association with intense interplanetary and geomagnetic activity. *Geophys. Res. Lett.* **23**, 2577. DOI. ADS.
Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M.L., Howard, R.A.: 2001, Predicting the 1-AU arrival times of coronal mass ejections. *J. Geophys. Res.* **106**, 29207. DOI. ADS.
Gosling, J.T.: 1993, Coronal mass ejections – the link between solar and geomagnetic activity. *Phys. Fluids, B Plasma Phys.* **5**, 2638. DOI. ADS.
Gosling, J.T., Bame, S.J., McComas, D.J., Phillips, J.L.: 1990, Coronal mass ejections and large geomagnetic storms. *Geophys. Res. Lett.* **17**, 901. DOI. ADS.
Győri, L.: 1998, Automation of area measurement of sunspots. *Solar Phys.* **180**, 109. ADS.
Győri, L., Baranyi, T., Ludmány, A.: 2011, Photospheric data programs at the Debrecen Observatory. In: *IAU Symposium*, **273**, 403. DOI. ADS.
Győri, L., Baranyi, T., Muraközy, J., Ludmány, A.: 2005, Recent advances in the Debrecen sunspot catalogues. *Mem. Soc. Astron. Ital.* **76**, 981. ADS.
Hale, G.E.: 1929, No. 388. The spectroscope and its work. Part I. History, instruments, adjustments, and methods of observation. *Contributions from the Mount Wilson Observatory/Carnegie Institution of Washington* **388**, 1. ADS.
Hale, G.E.: 1931, The spectrohelioscope and its work. Part III. Solar eruptions and their apparent terrestrial effects. *Astrophys. J.* **73**, 379. DOI. ADS.
Howard, R.F.: 1992, The growth and decay of sunspot groups. *Solar Phys.* **137**, 51. DOI. ADS.
Howard, R.F.: 1993, How growth and decay of sunspot groups depend on axial tilt angles. *Solar Phys.* **145**, 95. DOI. ADS.
Howard, R.A.: 2006, A historical perspective on coronal mass ejections. In: Gopalwamy, N., Mewaldt, R., Torsti, J. (eds.) *Solar Eruptions and Energetic Particles*. *AGU, Geophys. Mon. Ser.* **165**, 7. ADS.
Howard, T.A., Tappin, S.J.: 2005, Statistical survey of earthbound interplanetary shocks, associated coronal mass ejections and their space weather consequences. *Astron. Astrophys.* **440**, 373. DOI. ADS.
Hoyt, D.V., Schatten, K.H.: 1998, Group sunspot numbers: a new solar activity reconstruction. *Solar Phys.* **181**, 491. ADS.
Hudson, H.S.: 1991, Solar flares, microflares, nanoflares, and coronal heating. *Solar Phys.* **133**, 357. DOI. ADS.
Joselyn, J.A., McIntosh, P.S.: 1981, Disappearing solar filaments – a useful predictor of geomagnetic activity. *J. Geophys. Res.* **86**, 4555. DOI. ADS.
Kahler, S.W.: 1982, The role of the big flare syndrome in correlations of solar energetic proton fluxes and associated microwave burst parameters. *J. Geophys. Res.* **87**, 3439. DOI. ADS.
Koskinen, H.E.J., Huttunen, K.E.J.: 2006, Geoeffectivity of coronal mass ejections. *Space Sci. Rev.* **124**, 169. DOI. ADS.
Krivský, L.: 1973, Trends of development of the proton active region of 24 January 1971. *Bull. Astron. Inst. Czechoslov.* **24**, 96. ADS.
Lefèvre, L., Clette, F.: 2011, A global small sunspot deficit at the base of the index anomalies of solar cycle 23. *Astron. Astrophys.* **536**, L11. DOI. ADS.
Lefèvre, L., Clette, F.: 2014, Survey and merging of sunspot catalogs. *Solar Phys.* **289**, 545. DOI. ADS.
Lundstedt, H., Persson, T., Andersson, V.: 2015, The extreme solar storm of May 1921: observations and a complex topological model. *Ann. Geophys.* **33**, 109. DOI.
Maričić, D., Vršnak, B., Stanger, A.L., Veronig, A.M., Temmer, M., Roša, D.: 2007, Acceleration phase of coronal mass ejections, II: synchronization of the energy release in the associated flare. *Solar Phys.* **241**, 99. DOI. ADS.
Maunder, E.W.: 1904, Greenwich, Royal Observatory, the "great" magnetic storms, 1875 to 1903, and their association with sun-spots. *Mon. Not. Roy. Astron. Soc.* **64**, 205. ADS.
Mayaud, P.N.: 1980, Derivation, meaning and use of geomagnetic indices. *AGU* **22**, 538. DOI.
McAllister, A.H., Dryer, M., McIntosh, P., Singer, H., Weiss, L.: 1996, A large polar crown coronal mass ejection and a "problem" geomagnetic storm: April 14 – 23, 1994. *J. Geophys. Res.* **101**, 13497. DOI. ADS.
McIntosh, P.S.: 1990, The classification of sunspot groups. *Solar Phys.* **125**, 251. DOI. ADS.
Menvielle, M., Marchaudon, A.: 2007, Geomagnetic indices in solar–terrestrial physics and space weather. In: Lilensten, J. (ed.) *Space Weather: Research Towards Applications in Europe 2nd European Space Weather Week (ESWW2)*, *Astrophys. Space Sci. Library* **344**, 277. DOI. ADS.

Moon, Y.-J., Choe, G.S., Wang, H., Park, Y.D., Cheng, C.Z.: 2003, Relationship between CME kinematics and flare strength. *J. Korean Astron. Soc.* **36**, 61. ADS.
Neidig, D.F., Cliver, E.W.: 1983, A catalog of solar white-light flares, including their statistical properties and associated emissions, 1859 – 1982. *NASA STI/Recon Technical Report N* **84**, 24521. ADS.
Newton, H.W.: 1943, Solar flares and magnetic storms. *Mon. Not. Roy. Astron. Soc.* **103**, 244. ADS.
Newton, H.W.: 1944, Solar flares and magnetic storms (second paper). *Mon. Not. Roy. Astron. Soc.* **104**, 4. ADS.
Newton, H.W.: 1950, A significant time-distribution of great solar flares and great geomagnetic storms. *Observatory* **70**, 233. ADS.
Pearson, K.: 1895, Notes on regression and inheritance in the case of two parents. *Proc. Roy. Soc. London* **58**, 500. DOI. ADS.
Priest, E.R., Forbes, T.G.: 2002, The magnetic nature of solar flares. *Astron. Astrophys. Rev.* **10**, 313. DOI. ADS.
Proctor, R.A.: 1871, *The Sun: Ruler, Fire, Light, and Life of the Planetary System*, Longman & Green, London.
Qahwaji, R., Colak, T.: 2007, Automatic short-term solar flare prediction using machine learning and sunspot associations. *Solar Phys.* **241**, 195. DOI. ADS.
Qahwaji, R., Colak, T., Al-Omari, M., Ipson, S.: 2008, Automated prediction of CMEs using machine learning of CME – flare associations. *Solar Phys.* **248**, 471. DOI. ADS.
Reames, D.V.: 1999, Particle acceleration at the Sun and in the heliosphere. *Space Sci. Rev.* **90**, 413. DOI. ADS.
Richardson, R.S.: 1944, Solar flares versus bright chromospheric eruptions: a question of terminology. *Publ. Astron. Soc. Pac.* **56**, 156. DOI. ADS.
Richardson, I.G., Cane, H.V.: 2010, Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996 – 2009): catalog and summary of properties. *Solar Phys.* **264**, 189. DOI. ADS.
Ruždjak, V., Vršnak, B., Brajša, R., Schroll, A.: 1989, A comparison of H-alpha and soft X-ray characteristics of spotless and SPOT group flares. *Solar Phys.* **123**, 309. DOI. ADS.
Sakurai, K.: 1970, On the magnetic configuration of sunspot groups which produce solar proton flares. *Planet. Space Sci.* **18**, 33. DOI. ADS.
Sammis, I., Tang, F., Zirin, H.: 2000, The dependence of large flare occurrence on the magnetic structure of sunspots. *Astrophys. J.* **540**, 583. DOI. ADS.
Scherrer, P.H., Bogart, R.S., Bush, R.I., Hoeksema, J.T., Kosovichev, A.G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T.D., Title, A., Wolfson, C.J., Zayer, I., MDI Engineering Team: 1995, The solar oscillations investigation – Michelson Doppler imager. *Solar Phys.* **162**, 129. DOI. ADS.
Schrijver, C.J., Beer, J., Baltensperger, U., Cliver, E.W., Güdel, M., Hudson, H.S., McCracken, K.G., Osten, R.A., Peter, T., Soderblom, D.R., Usoskin, I.G., Wolff, E.W.: 2012, Estimating the frequency of extremely energetic solar events, based on solar, stellar, lunar, and terrestrial records. *J. Geophys. Res.* **117**, 8103. DOI. ADS.
Schwenn, R., dal Lago, A., Huttunen, E., Gonzalez, W.D.: 2005, The association of coronal mass ejections with their effects near the Earth. *Ann. Geophys.* **23**, 1033. DOI. ADS.
Secchi, P.: 1872, Sur les taches et le diametre solaires. *C. R. Acad. Sci., Paris* **75**, 1581.
Secchi, P.: 1879, *El Sol*, Madrid Administracion de la Biblioteca Cientifico-Literaria, Libreria de Victoriano Suarez, Sevilla.
Silverman, S.M., Cliver, E.W.: 2001, Low-latitude auroras: the magnetic storm of 14 – 15 May 1921. *J. Atmos. Solar-Terr. Phys.* **63**, 523. DOI. ADS.
Simnett, G.M., Harrison, R.A.: 1985, The onset of coronal mass ejections. *Solar Phys.* **99**, 291. DOI. ADS.
Spearman, C.: 1904, The proof and measurement of association between two things. *Am. J. Psychol.* **15**(15), 500. DOI. ADS.
Spoerer, G.: 1876, *Beobachtungen der Sonnenflecken – Part II*, Der Astronomisches Gesellschaft, Leipzig.
Srivastava, N., Venkatakrishnan, P.: 2002, Relationship between CME speed and geomagnetic storm intensity. *Geophys. Res. Lett.* **29**, 1287. DOI. ADS.
Srivastava, N., Venkatakrishnan, P.: 2004, Solar and interplanetary sources of major geomagnetic storms during 1996-2002. *J. Geophys. Res.* **109**(A18), 10103. DOI. ADS.
Srivastava, N., Gonzalez, W.D., Gonzalez, A.L.C., Masuda, S.: 1997, On the characteristics of solar origins of geoeffective CMEs observed during August 1992 – April 1993. In: Wilson, A. (ed.) *Correlated Phenomena at the Sun, in the Heliosphere and in Geospace*, *ESA SP* **415**, 443. ADS.
Srivastava, N., Gonzalez, W.D., Gonzalez, A.L.C., Masuda, S.: 1998, On the solar origins of intense geomagnetic storms observed during 6 – 11 March 1993. *Solar Phys.* **183**, 419. DOI. ADS.
Švestka, Z.: 1966, Proton flares before 1956. *Bull. Astron. Inst. Czechoslov.* **17**, 262. ADS.
Švestka, Z.: 1969, The optical flare. In: de Jager, C., Svestka, Z. (eds.) *Solar Flares and Space Research*, North-Holland, Amsterdam, 16. ADS.

Ternullo, M., Contarino, L., Romano, P., Zuccarello, F.: 2002, A statistical analysis on sunspot-groups correlated to M and X flares. In: Wilson, A. (ed.) *Solar Variability: From Core to Outer Frontiers*, *ESA SP* **506**, 1045. ADS.

Tousey, R., Bartoe, J.D.F., Bohlin, J.D., Brueckner, G.E., Purcell, J.D., Scherrer, V.E., Sheeley, N.R. Jr., Schumacher, R.J., Vanhoosier, M.E.: 1973, A Preliminary study of the extreme ultraviolet spectroheliograms from skylab. *Solar Phys.* **33**, 265. DOI. ADS.

Vaquero, J.M., Valente, M.A., Trigo, R.M., Ribeiro, P., Gallego, M.C.: 2008, The 1870 space weather event: geomagnetic and auroral records. *J. Geophys. Res.* **113**, 8230. DOI. ADS.

Vaquero, J.M., Trigo, R.M., Gallego, M.C., Dominguez-Castro, F.: 2012, Improving sunspot records: solar drawings of the late 19th century from the Royal Astronomical Observatory of Lisbon. *Observatory* **132**, 376. ADS.

Vennerstrøm, S., Lefevre, L., Dumbovic, M., Crosby, N., Malandraki, O., Patsou, I., Clette, F., Veronig, A., Vrsnak, B., Leer, K., Moretto, T.: 2016, Extreme geomagnetic storms – 1868 – 2010. *Solar Phys.* DOI.

Vršnak, B., Sudar, D., Ruždjak, D.: 2005, The CME-flare relationship: are there really two types of CMEs? *Astron. Astrophys.* **435**, 1149. DOI. ADS.

Wagner, W.J., Hildner, E., House, L.L., Sawyer, C., Sheridan, K.V., Dulk, G.A.: 1981, Radio and visible light observations of matter ejected from the sun. *Astrophys. J. Lett.* **244**, L123. DOI. ADS.

Wang, Y.M., Ye, P.Z., Wang, S., Zhou, G.P., Wang, J.X.: 2002, A statistical study on the geoeffectiveness of Earth-directed coronal mass ejections from March 1997 to December 2000. *J. Geophys. Res.* **107**, 1340. DOI. ADS.

Willis, D.M., Coffey, H.E., Henwood, R., Erwin, E.H., Hoyt, D.V., Wild, M.N., Denig, W.F.: 2013b, The Greenwich photo-heliographic results (1874 – 1976): summary of the observations, applications, datasets, definitions and errors. *Solar Phys.* **288**, 117. DOI. ADS.

Willis, D.M., Henwood, R., Wild, M.N., Coffey, H.E., Denig, W.F., Erwin, E.H., Hoyt, D.V.: 2013a, The Greenwich photo-heliographic results (1874 – 1976): procedures for checking and correcting the sunspot digital datasets. *Solar Phys.* **288**, 141. DOI. ADS.

Yashiro, S., Akiyama, S., Gopalswamy, N., Howard, R.A.: 2006, Different power-law indices in the frequency distributions of flares with and without coronal mass ejections. *Astrophys. J. Lett.* **650**, L143. DOI. ADS.

Zhang, J., Dere, K.P., Howard, R.A., Kundu, M.R., White, S.M.: 2001, On the temporal relationship between coronal mass ejections and flares. *Astrophys. J.* **559**, 452. DOI. ADS.

Zhang, J., Dere, K.P., Howard, R.A., Bothmer, V.: 2003, Identification of solar sources of major geomagnetic storms between 1996 and 2000. *Astrophys. J.* **582**, 520. DOI. ADS.

Zhang, J., Richardson, I.G., Webb, D.F., Gopalswamy, N., Huttunen, E., Kasper, J.C., Nitta, N.V., Poomvises, W., Thompson, B.J., Wu, C.-C., Yashiro, S., Zhukov, A.N.: 2007, Solar and interplanetary sources of major geomagnetic storms (Dst ≤ -100 nT) during 1996 – 2005. *J. Geophys. Res.* **112**(A11), 10102. DOI. ADS.

Zirin, H., Liggett, M.A.: 1987, Delta spots and great flares. *Solar Phys.* **113**, 267. DOI. ADS.