Publication Year	2018
Acceptance in OA@INAF	2021-04-20T09:35:42Z
Title	Halo Coronal Mass Ejections during Solar Cycle 24: reconstruction of the global scenario and geoeffectiveness
Authors	Scolini, Camilla; MESSEROTTI, Mauro; Poedts, Stefaan; Rodriguez, Luciano
DOI	10.1051/swsc/2017046
Handle	http://hdl.handle.net/20.500.12386/30812
Journal	JOURNAL OF SPACE WEATHER AND SPACE CLIMATE
Number	8

J. Space Weather Space Clim. 2018, 8, A09
© C. Scolini *et al.*, Published by EDP Sciences 2018
https://doi.org/10.1051/swsc/2017046

SWSC

Available online at:
www.swsc-journal.org

Flares, coronal mass ejections and solar energetic particles and their space weather impacts

RESEARCH ARTICLE OPEN ACCESS

Halo coronal mass ejections during Solar Cycle 24: reconstruction of the global scenario and geoeffectiveness

Camilla Scolini[1,2,*], Mauro Messerotti[3,4], Stefaan Poedts[1] and Luciano Rodriguez[2]

[1] Centre for mathematical Plasma Astrophysics (CmPA), KU Leuven, 3001 Leuven, Belgium
[2] Solar-Terrestrial Center of Excellence, SIDC, Royal Observatory of Belgium, 1180 Brussels, Belgium
[3] Department of Physics, University of Trieste, 34127 Trieste, Italy
[4] INAF-Astronomical Observatory of Trieste, 34149 Trieste, Italy

Received 1 June 2017 / Accepted 13 December 2017

Abstract – Coronal mass ejections (CMEs), in particular Earth-directed ones, are regarded as the main drivers of geomagnetic activity. In this study, we present a statistical analysis of a set of 53 fast ($V \geq 1000$ km·s^{-1}) Earth-directed halo CMEs observed by the SOHO/LASCO instrument during the period Jan. 2009–Sep. 2015, and we then use this CME sample to test the forecasting capabilities of a new Sun-to-Earth prediction scheme for the geoeffectiveness of Earth-directed halo CMEs. First, we investigate the CME association with other solar activity features such as solar flares, active regions, and others, by means of multi-instrument observations of the solar magnetic and plasma properties, with the final aim of identifying recurrent peculiar features that can be used as precursors of CME-driven geomagnetic storms. Second, using coronagraphic images to derive the CME kinematical properties at 0.1 AU, we propagate the events to 1 AU by means of 3D global MHD simulations. In particular, we use the WSA-ENLIL+Cone model to reconstruct the propagation and global evolution of each event up to their arrival at Earth, where simulation results are compared with interplanetary CME (ICME) in-situ signatures. We then use simulation outputs upstream of Earth to predict their impact on geospace. By applying the pressure balance condition at the magnetopause and the coupling function proposed by Newell et al. [*J Geophys Res: Space Phys* **113** (2008)] to link upstream solar wind properties to the global K_p index, we estimate the expected magnetospheric compression and geomagnetic activity level, and compare our predictions with global data records. The analysis indicates that 82% of the fast Earth-directed halo CMEs arrived at Earth within the next 4 days. Almost the totality of them compressed the magnetopause below geosynchronous orbits and triggered a minor or major geomagnetic storm afterwards. Among them, complex sunspot-rich active regions associated with X- and M-class flares are the most favourable configurations from which geoeffective CMEs originate. The analysis of related Solar Energetic Particle (SEP) events shows that 74% of the CMEs associated with major SEPs were geoeffective, i.e. they triggered a minor to intense geomagnetic storm ($K_p \geq 5$). Moreover, the SEP production is enhanced in the case of fast and interacting CMEs. In this work we present a first attempt at applying a Sun-to-Earth geoeffectiveness prediction scheme – based on 3D simulations and solar wind-geomagnetic activity coupling functions – to a statistical set of fast Earth-directed, potentially geoeffective halo CMEs. The results of the prediction scheme are promising and in good agreement with the actual data records for geomagnetic activity. However, we point out the need for future studies performing a fine-tuning of the prediction scheme, in particular in terms of the evaluation of the CME input parameters and the modelling of their internal magnetic structure.

Keywords: Coronal Mass Ejections (CMEs) / geomagnetic storms / geoeffectiveness predictions / Solar Cycle 24

1 Introduction

Coronal mass ejections (CMEs) are large-scale eruptions of magnetised plasma from the Sun, and are considered by the space physics community to be the main drivers of space weather (Gosling et al., 1991; Gosling, 1993; Koskinen & Huttunen, 2006). The phenomenon has been defined by Hundhausen et al. (1984) as an observable change in coronal structure that: (1) occurs on a time scale between a few minutes and several hours and (2) involves the appearance and outwards motion of a new, discrete, bright, white-light feature

*Corresponding author: camilla.scolini@kuleuven.be

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in the coronagraph field of view. CMEs usually originate in active regions (ARs) and appear in association with other solar activity signatures, mostly solar flares and filament activations/eruptions. It is now known that CMEs are extremely common events and occur at a rate which is highly dependent on the solar activity cycle. During solar minima, the average rate is ∼1 CME per day, while during the maxima of solar activity, it can exceed the value of 10 CMEs per day (Yashiro et al., 2004; Robbrecht et al., 2009). Among them, Earth-directed, fast ($V \geq 1000\,\mathrm{km \cdot s^{-1}}$) CMEs observed as halo events in coronagraphs along the Sun-Earth line are by far the most important CME class in terms of space weather implications and effects on Earth (Webb et al., 2000; Michalek et al., 2006). They constitute 3% of all CMEs and tend to originate close to the disk centre, even though about 10% of them originate close to the limb (Gopalswamy et al., 2015a). Considering a 73-month period in Solar Cycle 24, one of the weakest cycles ever recorded, a rate of 3.56 halo CMEs per month has been reported (Gopalswamy et al., 2015b). While halo CMEs are the most important ones in terms of impact on geospace, the determination of the kinematical and geometrical properties for such events is particularly difficult when using single-spacecraft observations only, due to the severe projection effects. Several approaches to cope with such limitations, based either on empirical relations derived from statistical analyses of limb (e.g. side-viewed) CME events observed from Earth (Gopalswamy et al., 2009), or on 3D-reconstruction methods based on multi-spacecraft observations (Mierla et al. (2010) and references therein), have been proposed. Statistical studies show that the average CME width (obtained from non-halo CMEs only) in the rising phase of Solar Cycle 24 is ∼55°, while the average CME speed is $\sim 650\,\mathrm{km \cdot s^{-1}}$ (Gopalswamy et al., 2014).

When observed in-situ, the interplanetary (IP) counterparts of CMEs are denoted as interplanetary CMEs (ICMEs). They are observed passing over Earth at an average rate of 1–2 per month (Richardson & Cane, 2010). Several definitions have been proposed in literature to identify and classify the variety of ICME signatures observed from in-situ measurements, making the interpretation of the original term somehow ambiguous and uncertain in its meaning. Typical in-situ ICME signatures are (Wimmer-Schweingruber et al. (2006) and references therein): (1) the presence of a forward shock followed by a turbulent region of highly distorted magnetic field resulting from the compression, deflection and heating of the ambient solar wind, known as sheath region; (2) enhanced He, O and Fe charged states; (3) different elemental composition compared to the surrounding solar wind (He/H and Mg/O ratios); (4) isotopic anomalies ($^3He^{2+}/^4He^{2+}$); (5) bidirectional electron streaming; (6) low proton temperature; (7) a magnetic structure classified as Magnetic Cloud (MC) (Burlaga, 1991). However, none of the above signatures or combination of them gives a foolproof ICME identifier (Wimmer-Schweingruber et al., 2006). For the sake of clarity and as we will deal with fast CMEs only, in the following analysis we will discuss the CME arrival at Earth position in terms of the ICME-driven forward shock.

Impacting CMEs affect technological systems and human activities in several ways (Cannon et al., 2013). The magnitude of a CME effect on geospace is strongly related to solar wind macroscopic parameters and magnetic properties, in particular the interplanetary magnetic field (IMF) vector $\boldsymbol{B}$ and the solar wind number density N and bulk speed V. While the density and speed determine the compression of the magnetosphere induced by the impinging solar wind, a southward magnetic field orientation is mainly responsible for facilitating the dayside magnetopause reconnection and the development of strong disturbances in the geomagnetic field referred to as magnetic storms (Gonzalez et al., 1994; Gopalswamy et al., 2007; Lugaz et al., 2015, 2016). Over the decades, several global geomagnetic activity indices have been defined in order to quantify the magnitude of geomagnetic storms recorded at ground level. Among them, the global Dst and K_p indices are both based on the measured variation of the horizontal component of the on-ground magnetic field as proxy of the geomagnetic activity level, and have been extensively used to evaluate the level of perturbation in the magnetospheric/ionospheric environment. The Dst index provides a measure of the strength of the equatorial ring current and it is used as reference for the classification of geomagnetic storms (Gonzalez et al., 1994). On the other hand, K_p is a mid-latitude index sensitive to contributions from both auroral and equatorial currents, which is used to set the various levels of alert in the NOAA Space Weather Scale for geomagnetic storms (http://www.swpc.noaa.gov/noaa-scales-explanation). The two indices are not related by a one-to-one correspondence, as they are sensitive to different current systems developing in the ionosphere and inner magnetosphere during geomagnetic storms (Kivelson & Russell, 1995; Huttunen et al., 2002; Huttunen & Koskinen, 2004).

Previous studies about the solar and IP sources of geomagnetic storms revealed that about 30% of all storms ($\mathrm{Dst} \leq -30\,\mathrm{nT}$) originate from ICMEs (Zhang et al., 2004). However, considering intense storms only ($\mathrm{Dst} \leq -100\,\mathrm{nT}$, Gonzalez et al., 1994) the fraction reaches up to ∼90% of the total (Zhang et al., 2004, 2007; see also Richardson & Cane (2012) for an analysis of the K_p index). Major disturbances are mostly caused by CMEs coming from the solar disk centre, although some moderate disturbances can be caused by events originated near or at the west limb (Huttunen et al., 2002; Rodriguez et al., 2009; Cid et al., 2012). On the other hand, CMEs originated at the disk centre can also be deflected away from the Sun-Earth line due to non-radial channelling caused by fast solar wind streams generated by adjacent coronal holes. Möstl et al. (2015) reported the case of a CME – occurred on 7 Jan. 2014 and included in our list as CME #37 (see Tab. 1) – which, despite being originated at the disk centre, was longitudinally deflected to the west due to a coronal hole on the east side of the source region, resulting in a weaker impact at Earth than expected. Previous studies have also found that ICME sheaths alone, i.e. not followed by any ICME ejecta, can also be source of intense storms (Tsurutani et al., 1988; Huttunen et al., 2002). The role of CME interactions as sources of strong geomagnetic storms has been investigated by Lugaz & Farrugia (2014) and Lugaz et al. (2017).

Another kind of CME-related phenomenon that is of interest for the space weather community is that of solar energetic particles (SEPs). They mainly consist of protons and electrons that are accelerated to quasi-relativistic speeds. Large gradual events triggered by CMEs are of particular interest to space travel as they can constitute a significant radiation hazard for astronauts and equipment, especially beyond the Earth's

Table 1. Complete list of the selected CME events. Columns 3 and 4 refer to the first appearance in LASCO C2 coronagraph. Column 5 lists the speed in 3D space calculated using equation (1). Columns 6 and 7 report the heliographic latitude and longitude of the source location. Column 8 reports the SXR class of the associated flares, when available; "–" indicates no associated flare or an association with a weak B- or A-class event.

Selected CMEs							
Event #	Set	Start date (YYYY-MM-DD)	Start time (UT)	Space speed ($km \cdot s^{-1}$)	Source location (HEEQ)		Associated flare (SXR class)
1	3	2010-04-03	10:33	939	−25	0	B7.4
2	3	2010-08-01	13:42	1030	20	−36	C3.2
3	2	2010-08-14	10:12	1280	17	52	C4.4
4	2	2011-03-07	20:00	2223	31	53	M3.7
5	1	2011-06-02	08:12	1147	−19	−25	C3.7
6	2	2011-06-07	06:49	1321	−21	54	M2.5
7	3	2011-08-03	14:00	785	16	30	M6.0
8	2	2011-08-04	04:12	1477	19	36	M9.3
9	1	2011-09-06	02:24	1232	14	7	M5.3
10	3	2011-09-06	23:05	830	14	18	X2.1
11	3	2011-09-22	10:48	1905	9	−89	X1.4
12	3	2011-09-24	12:48	2018	10	−56	M7.1
13	3	2011-09-24	19:36	1076	12	−42	M3.0
14	3	2011-10-22	01:25	666	35	40	–
15	3	2011-01-22	10:24	1011	25	77	M1.3
16	1	2012-01-19	14:36	1269	32	−22	M3.2
17	1	2012-01-23	04:00	2511	28	21	M8.7
18	3	2012-03-05	04:00	1627	17	−52	X1.1
19	1	2012-03-07	00:24	3146	17	−27	X5.4
20	1	2012-03-07	01:30	2160	25	−26	X1.3
21	1	2012-03-09	04:26	1229	15	3	M6.3
22	1	2012-03-10	18:00	1638	17	24	M8.4
23	3	2012-03-13	17:36	1931	17	66	M7.9
24	3	2012-04-23	18:24	769	14	17	C2.0
25	1	2012-06-14	14:12	1254	−17	−6	M1.9
26	3	2012-07-04	17:24	830	14	34	M1.8
27	2	2012-07-06	23:24	1907	−13	59	X1.1
28	1	2012-07-12	16:48	1405	−15	1	X1.4
29	2	2012-09-28	00:12	1093	6	34	C3.7
30	1	2013-03-15	07:12	1366	11	−12	M1.1
31	1	2013-06-28	02:00	1254	−18	19	C4.4
32	1	2013-08-17	19:12	1418	−5	30	M1.4
33	1	2013-09-29	22:12	1370	17	29	C1.3
34	3	2013-11-07	15:12	626	−13	−23	M2.4
35	2	2013-12-07	07:36	1165	−16	49	M1.2
36	3	2014-01-04	21:22	1166	−11	−34	M4.0
37	1	2014-01-07	18:24	2246	−15	11	X1.2
38	1	2014-02-16	10:00	1064	−11	−1	M1.1
39	3	2014-02-18	01:36	942	−24	−34	–
40	3	2014-02-20	03:12	1115	−14	−38	C3.3
41	3	2014-02-20	08:00	960	−15	73	M3.0
42	2	2014-04-18	13:25	1359	−20	34	M7.3
43	3	2014-06-04	12:48	555	−29	−40	–
44	1	2014-09-09	00:06	1080	12	−29	M4.5
45	1	2014-09-10	18:00	1652	14	−2	X1.6
46	3	2014-12-17	05:00	855	−20	−9	M8.7
47	1	2014-12-19	01:04	1513	−11	−15	M6.9
48	3	2015-03-15	01:48	932	−22	25	C9.1
49	3	2015-06-18	17:24	1398	15	−50	M3.0
50	3	2015-06-19	06:42	798	−27	−6	–
51	1	2015-06-21	02:36	1740	12	−13	M2.0
52	1	2015-06-22	18:36	1573	12	8	M6.5
53	2	2015-06-25	08:36	1805	9	42	M7.9

magnetic field (Reames, 2013 and references therein). The SEP production is believed to take place at CME forward shocks, or in relation to CME-CME interaction phenomena (Cane & Lario, 2006; Reames, 2013; Gopalswamy et al., 2015c). Concerning the role of fast halo CME as major source of strong SEP events, Gopalswamy et al. (2015c) analysed the association of 37 major SEP events in the years 2010–2014 with CMEs, reporting that 97% of them were associated with fast Earth-directed full-halo events. More extensive statistical analyses addressing the relation between SEPs and CMEs were conducted by Dierckxsens et al. (2015) (>160 SEP events in Solar Cycle 23), Papaioannou et al. (2016) (314 SEP events from 1984 to 2013), and Paassilta et al. (2017) (176 SEP events in Solar Cycles 23 and 24), all evidencing the role of halo CMEs as major SEP generators. Considering the SEP generation as consequence of CME interactions (Gopalswamy et al., 2002), the most favourable condition for SEP production appears to be the case of a preceding, relatively slow CME which is caught up by a second, faster CME launched some hours later. The optimal time interval between two subsequent CME eruptions, in order to have their intersection close to the Sun ($\leq 20\,R_{\odot}$), has been found to be $\geq$7 h, so to have a maximum acceleration efficiency at $\sim 5-15 R_{\odot}$). This so-called twin-CME scenario, however, appears controversial and it has been debated by Kahler & Vourlidas (2014), who questioned the role of CME interactions in major SEP events finding that they can be explained by a general increase of both background seed particles and more frequent CMEs during times of higher solar activity. Overall, previous results have confirmed the major role of fast Earth-directed halo CMEs and interacting CMEs as sources of strong SEP events at Earth, and represent an additional reason to study these types of CMEs as the most potentially geoeffective.

For this reason, we consider of primary importance to further study the relationship between CMEs, their IP counterparts, and the triggered geomagnetic activity in the latest years as well. In the past years, several works have tried to construct storm prediction models using empirical relations linking real-time in-situ solar wind parameters to geomagnetic activity indices (O'Brien & McPherron, 2000; Newell et al., 2007, 2008) or statistical analyses of in-situ ICME properties and geomagnetic storms (Srivastava & Venkatakrishnan, 2004; Zhang et al., 2007). Although near-Earth solar wind parameters can yield a quite reliable prediction of geomagnetic storm events, for spacecraft located at the Lagrangian point L1 such kind of warnings can give only a one-hour notice. To overcome this forecasting limit, studies based on remote observations of CME parameters and their association with other kinds of solar activity features have also been carried out, proving to be a powerful complementary approach to solar wind-magnetospheric coupling functions and statistical ICME studies (Dumbović et al., 2015). A third approach, complementary to the other two, is the use of global 3D MHD simulations to model the CME propagation in the heliosphere from Sun to Earth, thus also providing a determinant support in studies of CME propagation and multiple CME interactions. In this sense, combining the three approaches mentioned before into a single methodological scheme analysis may lead to a more comprehensive space weather forecasting tool to be used not only at Earth location but possibly also at other spacecraft and planetary locations.

In this study we address the problem of predicting the geoeffectiveness of halo CMEs by reconstructing their Sun-to-Earth global evolution by means of 3D global MHD simulations, using then the simulation outputs upstream of Earth to predict their impact on geospace. Moreover, from the analysis of their solar source regions we look for recurrent features that can be used as precursors of CME-driven geomagnetic storms. We focus on Solar Cycle 24 because of the relatively limited number of statistical studies available compared to Solar Cycle 23 (see for example Richardson, 2013; Watari, 2017), as well as because of its peculiar characteristics – weak solar activity and mild space weather (Gopalswamy et al., 2015a).

We first select a set of fast halo CMEs observed by the SOHO/LASCO instrument over an 81-month period during Solar Cycle 24 (Jan. 2009–Sep. 2015). We use multi-instrument observations of the solar photosphere and low corona to investigate their association with other solar activity features. In the attempt of reconstructing the propagation and global evolution of each event up to its arrival at Lagrange point L1, we made use of the WSA-ENLIL+Cone model (Odstrcil, 2003) running at NASA/CCMC (https://ccmc.gsfc.nasa.gov), which represents the currently most widely used code for CME/ICME modelling; simulation outputs at Earth are then compared with in-situ measurements provided by the Wind spacecraft. Finally, using the coupling function proposed by Newell et al. (2008) to link upstream solar wind conditions to the global K_p index, we estimate the geomagnetic activity level and compare it with global data records.

In Section 2 we describe the CME selection procedure from the SOHO/LASCO halo CME catalogue and the complementary data used. CME modelling with WSA-ENLIL+Cone is described in Section 3. In Section 4 we present our geoeffectiveness prediction scheme. Results of the statistical analysis of the selected event properties, evolution, associated events and impact on geospace are then discussed in Section 5 (solar source regions), 6 (IP signatures and shock association), 7 (impact on geospace) and 8 (association with major SEP events). Conclusions are presented in Section 9.

2 Event selection and complementary data

Earth-directed halo CMEs observed from coronagraphs along the Sun-Earth line are the most geoeffective CME type. For this reason, the SOHO/LASCO halo CME catalogue (http://cdaw.gsfc.nasa.gov/CME_list/halo/halo.html, Gopalswamy et al., 2010a) has been the primary database used in this work to identify and select the CME events of interest.

To current knowledge, the most important parameters assessing the potential geoeffectiveness of a CME from coronagraphic observations are their source location and reconstructed speed in 3D space (Michalek et al., 2006; Cid et al., 2012; Dumbović et al., 2015). For this reason we have estimated the CME speed in 3D space V_{space} starting from coronagraphic observations of the speed projected on the sky plane V_{sky} by considering the early CME evolution as characterised by a fixed angular width (e.g. as in a cone model). To link $V_{\rm sky}$ to $V_{\rm space}$ we applied the relation proposed by Gopalswamy et al. (2010a)

$$V_{\text{space}} = \frac{\cos\omega + \sin\omega}{\cos\omega\cos\Theta + \sin\omega} V_{\text{sky}}, \quad (1)$$

where Θ is the angle between the cone axis and the sky plane and ω is the CME half-width angle. The CME source location was derived from the Solar and Geophysical Activity Summary (SGAS) listing (available at ftp://ftp.swpc.noaa.gov/pub/warehouse/), defined as the heliographic coordinates of the associated H_α flare. If no such data were available, source information was obtained from solar disk images such as Yohkoh/SXT, SOHO/EIT or from other data such as microwave images from the Nobeyama radioheliograph and H_α images. As an evaluation of the half width for halo CME events is not possible from single-spacecraft observations, and average half-width angle of the CME was derived from an empirical relation proposed by Gopalswamy et al. (2009), who reported a correlation of 0.69 between the sky plane speed and the angular width w of a CME.

$$V_{\text{sky}} = 360 + 3.62 \cdot w \quad [\text{km}\cdot\text{s}^{-1}], \quad (2)$$

for a set of 341 near-limb CMEs (for which the evaluation of the width is easier). Even though for halo CMEs a direct estimate of the width is extremely difficult, starting from the previous set of limb events an average half-width angle ω has been identified. Adapting this correlation to halo CMEs, $\omega = 32°$ for $V_{\text{space}} \leq 500\,\text{km}\cdot\text{s}^{-1}$, $\omega = 45°$ for $500\,\text{km}\cdot\text{s}^{-1} < V_{\text{space}} \leq 900\,\text{km}\cdot\text{s}^{-1}$ and $\omega = 66°$ for $V_{\text{space}} > 900\,\text{km}\cdot\text{s}^{-1}$. The errors associated to the various CME parameters have been estimated to be $dV_{sky}/V_{sky} = 10\%$, $d\omega = 10°$ and $d\Theta = 10°$ (Gopalswamy et al., 2010a; Jang et al., 2016).

Limiting our interest to the period Jan. 2009-Sep. 2015, we have selected all the fast ($V_{\text{space}} \geq 1000\,\text{km}\cdot\text{s}^{-1}$) halo CMEs that originated from a heliographic longitude between 30°E and 30°W from the solar central meridian. The application of these criteria to halo CMEs listed in the catalogue resulted in a set of 21 events hereafter denoted as "S1". As fast CMEs can be deflected towards the east due to the blocking effect of the solar wind background ahead of them (Wang et al., 2004), we have extended the source location condition towards the western hemisphere so to account for potentially geoeffective CMEs undergoing such kind of deflection. In particular, we have selected all the fast halo CMEs that originated from a heliographic longitude between 30° W and 60° W from the solar central meridian, resulting in a set of 9 events denoted as "S2". To be sure of including in our analysis all the geoeffective halo CME events that occurred during Solar Cycle 24, we have checked the K_p index data for the whole period considered (available at http://www.gfz-potsdam.de/en/kp-index/), and we have investigated the active periods characterised by a 3-hour $K_p \geq 5$. For all the geomagnetic activity periods above the threshold, assuming a CME propagation time from Sun to Earth of ~2–3 days, we have searched the catalogue for CME precursors not included in the S1 and S2 sets, which nevertheless resulted in a strong geomagnetic storm. In addition, we have also selected all the halo CMEs that originated from the same solar region and within 2 days before and after the events already included in the sets S1 and S2 (imposing no constraint on the CME speed in this case). This choice was made in order to take into account CME-CME interactions and pre-conditioning of the solar wind background due to the passage of previous CMEs, factors that may also affect the propagation and geoeffectiveness of the single events involved (Burlaga et al., 2002; Temmer et al., 2017). The application of these two criteria has resulted in a set of 23 additional CMEs named "S3", which includes: (a) all halo CMEs originated from the east (90° E < longitude < 30° E) and the west (60° W < longitude < 90° W) part of the visible disk that resulted in a geomagnetic storm ($K_p \geq 5$); (b) halo CMEs that took part in interactions/pre-conditioning of the solar wind background − in particular those recorded within ±2 days from the halo events in S2 and S3. It is worth noticing that all the S3 events satisfy one of the two geoeffectiveness conditions (source location or speed condition).

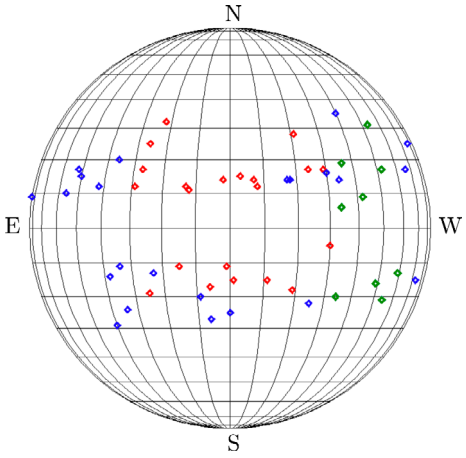

Fig. 1. Distribution of the source locations of the selected CME events in heliographic coordinates. Red, green and blue diamonds indicate CMEs belonging to S1, S2 and S3, respectively.

A total number of 53 CMEs composed the final set of events that have been analysed in this work. Table 1 presents a complete list of the selected events, together with their main observational properties as reported in the LASCO halo CME catalogue. Figure 1 shows the distribution of the source locations of the selected CMEs on the solar disk, in heliographic coordinates. The distribution appears symmetric with respect to the solar equator with the majority of the events (85%) originating within a latitude of ±(10° − 30°). Moreover, none of the selected events originated from a latitude higher than ±35°.

2.1 Reconstructing the global scenario by means of remote-sensing and in-situ data

To fully reconstruct the Sun-to-Earth evolution of the selected CME events, we have made use of complementary data archives containing both remote sensing observations of the Sun and in-situ measurements of the solar wind plasma and magnetic properties.

Searching for potential precursors of upcoming geoeffective CMEs, we have checked the association of all the selected CMEs with additional solar activity features observed by means of remote-sensing instruments. To investigate the properties of the CME solar source regions and their

association with solar flares, we have made use of the data provided by NOAA/SWPC and listed in the Solar region summary (SRS) and in the solar and geophysical activity summary (SGAS) (ftp://ftp.swpc.noaa.gov/pub/warehouse/).

With the aim of investigating pre- and post-eruptive conditions at the Sun, we have searched for associations with filaments/prominences, X-ray sigmoidal structures and global coronal perturbations. In performing this association check we have made use of the iSolSearch interactive tool provided by heliophysics events knowledgebase, a database collecting data acquired primarily by SDO/AIA and SDO/HMI instruments (Lemen et al., 2012; Scherrer et al., 2012) and available at http://www.lmsal.com/isolsearch. In determining whether an association exists, we imposed both temporal and spatial criteria, namely that a given activity feature was observed in the same NOAA AR of the reconstructed CME source location, and that it occurred within ±60 min from the CME onset time. The onset time of each CME event was estimated by back-extrapolating to the solar surface the height-time information contained in the LASCO catalogue, using a linear fitting.

To monitor the solar wind conditions right before the impact on the magnetosphere, we have made use of in-situ data obtained by the Wind spacecraft via two of its on-board instruments: the magnetic field investigation (MFI) and the solar wind experiment (SWE) (Lepping et al., 1995; Ogilvie et al., 1995). MFI and SWE solar Wind Data products and derived quantities were used as comparison with the ENLIL simulation results at Earth as discussed in Section 3. In particular, in the comparison phase we used 1-minute resolution data relative to the solar wind magnetic field strength B, solar wind (proton) bulk speed V, proton number density N, proton temperature T, and β factor.

In order to further characterise the effects of the selected CME events on Earth, we have analysed their association with major SEP events listed in the NASA (http://cdaw.gsfc.nasa.gov/CME_list/sepe/, containing data up to end 2014) and NOAA (http://umbra.nascom.nasa.gov/SEP/) catalogues (Reames, 1999). Major SEP events are those associated with intensities ≥10 pfu[1] in the ≥10 MeV proton energy channel, and are the most relevant events in terms of space weather effects (Schwenn et al., 2005; Gopalswamy et al., 2015c).

3 CME modelling with WSA-ENLIL + Cone

The ENLIL model is a global 3D ideal MHD code that models the evolution of the background solar wind plasma and magnetic field in the heliosphere up to 10 AU (Toth & Odstrcil, 1996; Odstrcil, 2003). In this work we have used ENLIL version 2.8f, currently running at NASA/CCMC and available for runs on request, in combination with the Wang-Sheeley-Arge (WSA) empirical coronal model (Arge & Pizzo, 2000), which takes as input synoptic magnetograms (Carrington maps) from the national solar observatory (NSO) and the global oscillation network group (GONG).

The WSA coronal model is made by the combination of the WSA potential field + current sheet (WSA PF + CS) and the WSA inner heliosphere (WSA-IH) models. The WSA PF + CS model combines a Potential Field Source Surface (PFSS) model with the Schatten Current Sheet model to model the magnetic field between the photosphere and the source surface, set at 2.5 $R_\odot$ (Owens & Forsyth, 2013). Starting from 2.5 $R_\odot$ outwards, the WSA-IH model then propagates the solar wind and magnetic field up to 21.5 $R_\odot$, where they are used as the inner boundary conditions for the heliospheric model. For the heliospheric model, we used a simulation domain between 21.5 $R_\odot$ and 2 AU in the radial direction, so to include the Earth orbit, with a latitudinal angle (Θ) going from 60° N to 60° S with respect to the solar equator, and an azimuthal angle (ϕ) spacing over 360°.

Run in combination with a cone model (Zhao et al., 2002; Xie et al., 2004), ENLIL can model the propagation of CMEs throughout the heliosphere. This kind of simplified model assumes a selfsimilar CME expansion, characterised by a constant angular width as result of the external magnetic pressure confinement, until the CME reaches the heliospheric inner boundary at 21.5 $R_\odot$ = 0.1 AU – as supported by coronagraphic observations (St. Cyr et al., 2000).

The WSA-ENLIL + Cone model takes as CME input parameters: the passage time at 21.5 $R_\odot$, the radial speed at 21.5 $R_\odot$, its direction of propagation, and its half width. The date and time of the CME passage at the ENLIL inner boundary have been obtained by de-projecting the linear V_{sky} contained in the LASCO catalogue by means of equation (1), and assuming that V_{space} was maintained constantfrom the CME onset to the boundary. Under this assumption, we have extrapolated the passage time at 21.5 $R_\odot$ starting from the estimated onset time contained in the height-time plots in the LASCO halo CME catalogue, assuming that the ENLIL cone originated at a distance of about 1 $R_\odot$ from the solar centre (e.g. near the solar surface or in the chromosphere). The obtained V_{space} has been used as input for the radial velocity parameter, assuming a completely radial direction of the CME cones at the inner boundary.

Assuming a radial propagation in the corona, the CME direction of propagation has been specified by the latitude and longitude of the CME source location as listed in the LASCO halo CME catalogue. The half-width angle has been estimated by means of the empirical relation proposed by Gopalswamy et al. (2009) and presented in Section 2. Note that the ENLIL+Cone model version used in this work did not take into account the CME internal magnetic structure, so that the CME blobs can only perturb the pre-existing IP magnetic field background once they are inserted in the heliospheric domain.

In addition to these major input parameters, the structure of a cone CME can also be adjusted by the user by specifying the shape of the cone base and that of the CME cloud on top of it. For the sake of simplicity and to avoid the introduction of too many parameters in the simulation runs, in this work a spherical shape has been used in all runs.

Additional parameters that can be set by the user involve the solar wind background properties. In particular, the density enhancement factor ("df factor") of the CME with respect to the solar wind background has been initially set equal to 4 (default value) for all the events. However, being aware that Taktakishvili et al. (2009) reported a better performance in the case of df = 2 when validating the model, the March 2012 (#18–23) and June 2015 (CMEs #49–53) sequences have been simulated under both conditions. For these two events, the

[1] pfu = particles·cm^{-2}·s^{-1}·sr^{-1}.

choice df = 2 has performed better in terms of arrival times and peak values at Earth, and therefore in the following discussion we have considered such runs only. Although further testing would be needed to confirm this point, we report that this result seems to affect in particular CMEs propagating through a pre-conditioned background, e.g. the case of CME sequences. To our knowledge, no publication exists investigating the effect of such parameter in the case of complex CME events, and although a better performance has been reported by Taktakishvili et al. (2009) in the case of single CMEs, no conclusive explanation of this effect in simulations has been reported so far.

4 Geoeffectiveness prediction scheme and forecast verification

To predict the geoeffectiveness of CME events starting from solar observations and heliospheric simulations, it is crucial to link ENLIL simulation outputs to the expected level of perturbation induced on geospace. To do this, we have assumed an operational definition of geoeffectiveness using both (1) the solar wind-geomagnetic activity coupling function approach and (2) the classical method based on the evaluation of the dayside magnetopause stand-off distance in the Sun-Earth direction. The chart reported in Figure 2 illustrates the prediction scheme discussed below.

Geomagnetic activity indices. An approach to estimate the geoeffectiveness of a CME is by means of empirical relations linking the solar wind plasma and magnetic field parameters upstream of Earth to its magnetospheric effects, considering for instance the geomagnetic activity indices most widely used. Newell et al. (2007) showed that all the most commonly used geomagnetic indices, including the K_p index, can be related to solar wind parameters by the use of the following coupling function, estimating the rate of magnetic field lines opened at the magnetopause

$$\frac{d\Phi_{MP}}{dt} = V^{4/3}B^{2/3}\sin^{8/3}(\theta_c/2), \quad (3)$$

where $\theta_c = \arctan(B_y/B_z)$ is called the *IMF clock angle* and indicates the direction of the IMF: B_z refers to the north-south component of the IMF relative to Earth and B_y refers to the component of the IMF perpendicular to both the Sun-Earth line and the north-south line in geocentric solar magnetospheric (GSM) coordinates. $\theta_c = 0$ corresponds to a completely northward oriented magnetic field, while $\theta_c = \pi$ corresponds to a completely southward oriented magnetic field. The K_p index can be predicted once the speed and the magnetic properties of the solar wind are known, by means of the following relation (with an r factor = 0.866)

$$K_p = 0.05 + 2.224\cdot10^{-4}\frac{d\Phi_{MP}}{dt} + 2.844\cdot10^{-6}N^{1/2}V^2, \quad (4)$$

where V is measured in km·s^{-1}, N is the solar wind number density and is measured in cm^{-3} and B (included in the reconnection rate term) is in nT (Newell et al., 2008). As the ENLIL+Cone model used in this work does not take into account the internal magnetic structure of CMEs/ICMEs, it is incapable of reproducing the magnetic features generally observed in association with ICMEs e.g. turbulent sheaths and magnetic clouds. For this reason, we have decided not to consider simulation outputs for the ***B*** field orientation as reliable enough to be used in our prediction scheme. To calculate the predicted K_p index values we have made use of equation (4) under the assumption of two orientation conditions for the ICME structure impacting on Earth:

- Completely southward magnetic field, consistent with a worst-case scenario in terms of CME geoeffectiveness (maximum impact on geospace): $\theta_c = \pi$ and therefore $\sin^{8/3}(\theta_c/2) = 1$.
- Randomly oriented magnetic field: assuming a uniform distribution of the clock angle θ_c, the $\sin^{8/3}(\theta_c/2)$ term entering equation (3) has an expectation value $\langle\sin^{8/3}(\theta_c/2)\rangle \sim 0.45$ (Emmons et al., 2013).

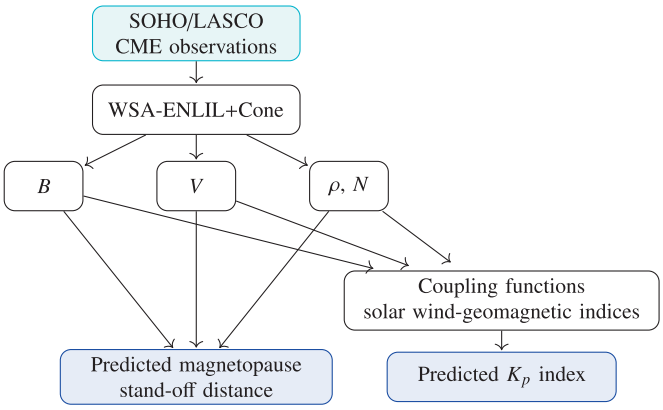

Fig. 2. Geoeffectiveness prediction scheme used to evaluate the expected impact at Earth of the selected CME events.

Magnetopause stand-off distance. Another important parameter assessing the geoeffectiveness of a CME impacting on Earth is the magnetospheric compression due to the dynamic pressure exerted by the solar wind on the magnetosphere. A first-order evaluation of this compression is given by the magnetopause stand-off distance, defined as the position of the magnetopause nose along the Sun-Earth direction. Assuming a dipolar shape of the magnetosphere on the dayside, a rough estimate of the magnetopause stand-off distance is given by (Taktakishvili et al., 2009)

$$d_{so} = \left(\frac{B_0^2}{2\cdot0.88\mu_0\rho V^2}\right)^{1/6}\cdot R_E, \quad (5)$$

In this case ENLIL outputs can be directly used to calculate the expected magnetopause stand-off distance.

Forecast verification. To quantitatively evaluate the performances of the geoeffectiveness prediction scheme discussed above, we have computed the 2×2 contingency tables for categorical forecasts.

(Weigel et al., 2006; Jolliffe & Stephenson, 2011) relative to two main observables: the CME arrival at Earth location and the CME induced perturbation on the geomagnetic field. Categorical forecasts are those associated with two possible outcomes: yes (event) and no (no event). Starting from the results of our prediction scheme, in this work we have

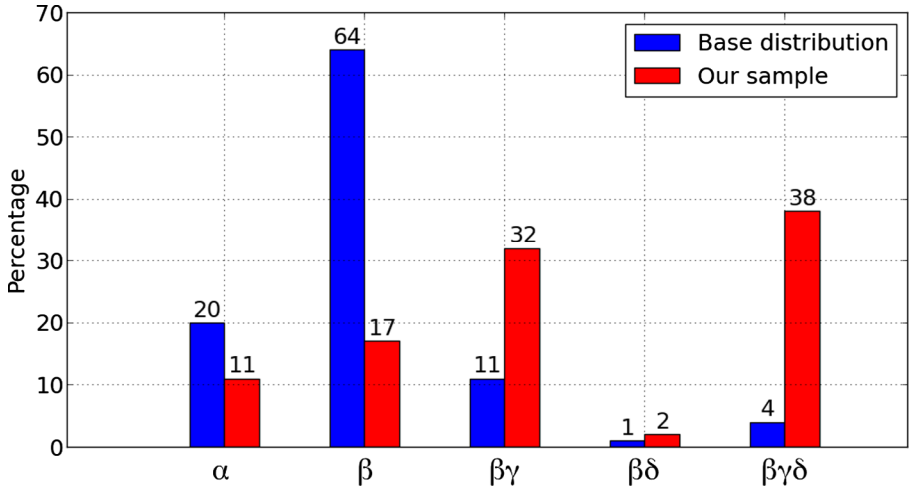

Fig. 3. Distribution of the Mount Wilson classes of the active regions associated with the selected CMEs. The base distribution is that reported by Jaeggli & Norton (2016) (calculated over the period 1992–2015).

considered the following two "yes" forecast conditions as independent conditions for CME geoeffectiveness: (a) CME arrival, marked in-situ by the detection of a ICME-driven forward shock; note that as shocks can extend much further than ICMEs, cases where the ICME missed the Earth but the forward shock was detected in-situ were considered as "yes" events (see previous works on "driverless shocks" by Gopalswamy et al. (2010b) and Janvier et al. (2014)). Note that in this work we have focused on the detection of IP shocks as signatures of ICME arrival as this was the only way to compare ENLIL results to in-situ measurements. In fact, as ENLIL+Cone treats CMEs as hydrodynamics plasma blobs inserted in the solar wind background field, it does not account for the CME internal magnetic structure and hence it is not able to provide information about the ICME/MC passage following a shock at Earth. (b) Geomagnetic storm condition defined as characterised by a $K_p \geq 5$ consistent with NOAA Space Weather Scale for geomagnetic storms. The contingency table for each forecast is characterised by the following entries, calculated by comparing forecasts to observations: *hits*, defined as events that were both predicted and observed to occur; *misses*, defined as events that were not predicted, but were observed to occur; *false alarms*, defined as events that were predicted to occur, but were observed not to occur; and *correct negatives*, events that were correctly predicted not to occur. Starting from the entries in the contingency tables for each observable, in order to provide a quantitative evaluation of the forecast performances we computed: (a) the correct rejection rate (CR rate), defined as the number of correct negatives divided by the number of observed "no" events and addressing the question *"what fraction of the observed "no" events was correctly predicted not to occur?"*; (b) the false alarm rate (FA rate), defined as the number of false alarms divided by the number of observed "no" events and addressing the question *"what fraction of the observed "no" events was incorrectly predicted to occur?"*; (c) the correct alarm ratio (CA ratio), defined as the number of hits divided by the number of predicted "yes" events and addressing the question *"what fraction of the predicted "yes" events did actually occur?"*; and (d) the false alarm ratio (FA ratio), defined as the number of false alarms divided by the number of predicted "yes" events and addressing the question *"what fraction of the predicted "yes" events actually did not occur?"*.

5 Analysis of solar conditions

5.1 Association with active regions

Over a set of 53 CMEs under study, 47 of them (89%) originated from an active region. Previous studies on slower CME samples, reported lower CME-AR association rates: on the one hand, a 63% association was found by Chen et al. (2011) over a set of 224 CMEs observed by the LASCO instrument in the years 1997–1998, using data relative to 108 ARs obtained by the SOHO michelson doppler imager (MDI); on the other hand, Subramanian & Dere (2001) found an 85% association studying a set of 32 front-side halo (full and partial) CMEs observed by LASCO over the same period, primarily using AR data from SOHO/EIT and SOHO/MDI.

Mount Wilson classification. Considering the Mount Wilson classification of ARs, the most common classes were $\beta\gamma\delta$ and $\beta\gamma$ (18 and 15 events respectively), followed by simple β topologies (8 events) and simple α topologies (5 events). There was only one AR showing a β topology and none representing other classes. No information on the AR associated with CME #27 was available, so this event has not been considered in the following discussion. As shown in Figure 3, comparing our distribution with the base distribution reported by Jaeggli & Norton (2016) we found a significant enhancement in the fraction of $\beta\gamma\delta$ and $\beta\gamma$ regions observed, while α and β classes resulted significantly depressed in their occurrence frequency. The fraction of $\beta\delta$ regions remained almost unchanged.

These results show that $\beta\gamma\delta$ and $\beta\delta$ configurations are characterised by the highest rate of production for fast CMEs, confirming the interpretation that AR with complicated magnetic configurations have a higher probability to undergo an energetic eruption that releases the magnetic energy stored over a short time period. Such kind of analysis has been previously conducted on solar flares by Qu (2008), who

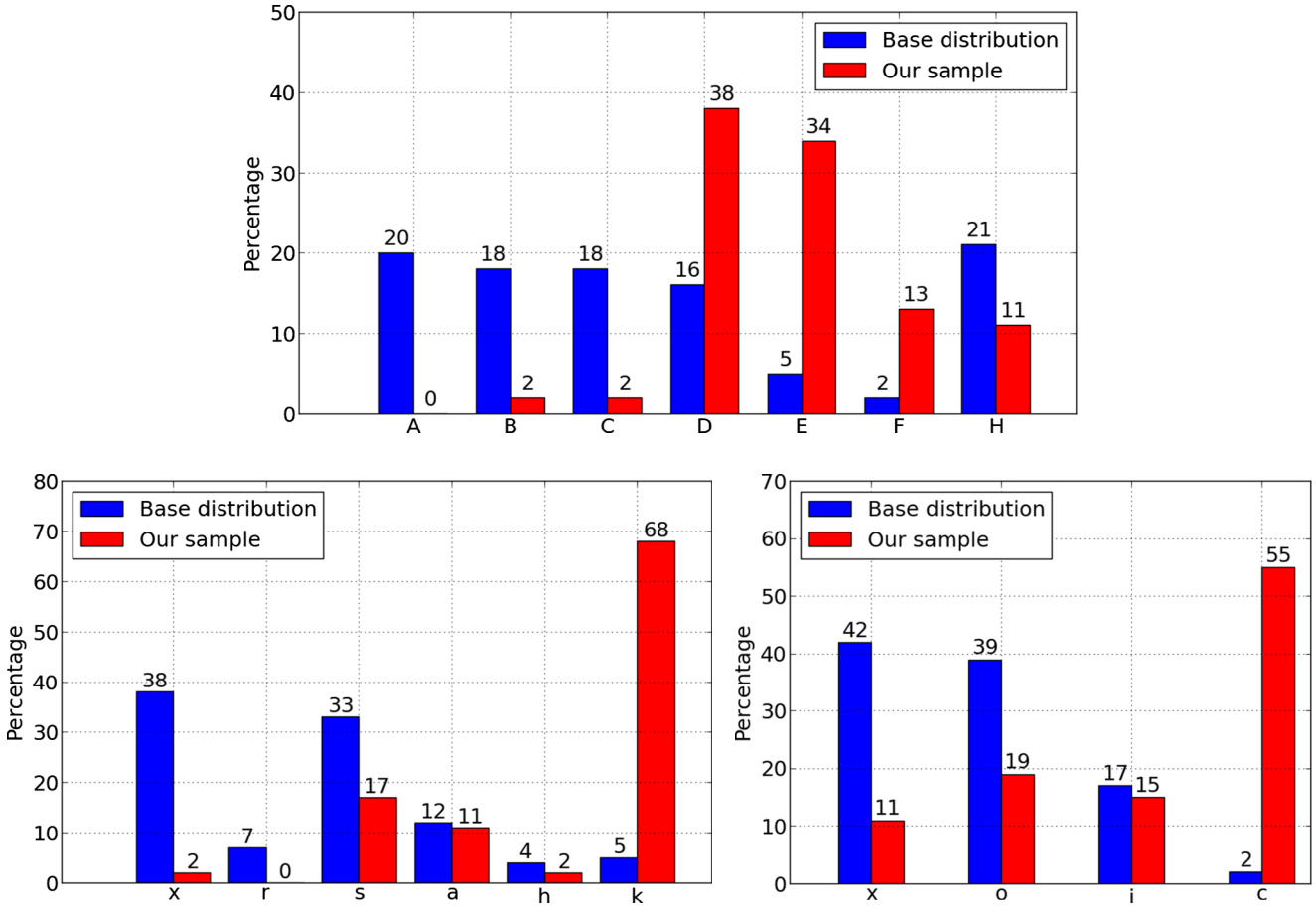

Fig. 4. Distribution of the McIntosh classes of the active regions associated with the selected CMEs. *Top*: Z code, *bottom left*: p code, *bottom right*: c code. The base distribution is that reported by McIntosh (1990) (calculated over the period 1969–1976).

reported a similar result in terms of flare production rates. However, to our knowledge, no previous literature exists in the case of CMEs, not for Cycle 24 nor for previous ones.

Considering the evolution of sunspot group magnetic classifications, we found that a significant fraction of active regions (16 out of 47) underwent a change in their magnetic topology on the day of the CME eruption, compared to the day before. Moreover, 12 of the 16 evolution patterns observed involved changes from or to $\beta\gamma\delta$ configurations.

McIntosh classification. Considering the Modified Zürich class of ARs (McIntosh Z code), the most common classes were the *D* and *E* classes (18 and 16 events respectively), followed by *F* topologies (6 events) and *H* topologies (5 events). The *A* class, corresponding to small unipolar sunspots, is not present in our sample. Comparing this distribution with the base distribution reported in McIntosh (1990), we find a significant enhancement in the fraction of *F*, *D* and *E* classes in our sample, while *B*, *C* and *H* classes result depressed in their occurrence frequency.

Turning to penumbral and sunspot distribution classes of ARs (McIntosh p and c codes), we found that 33 out of 47 active regions (70%) showed a large penumbra (*h* and *k* classes), while 32 out of 47 (68%) showed a large penumbra which was asymmetric (classified as *k*). 26 out of 47 regions (55%) showed a compact sunspot distribution, classified as *c*. Combining together these two parameters, we obtained the fraction of active regions which are classified as "complex regions" by McIntosh (1990). We found that 25 out of 47 regions (53%) were classified as *kc* in their combined penumbral/sunspot distribution. On average, only ∼2% of the regions show this degree of complexity. All these results differ significantly from the base distribution, as reported in Figure 4. In particular, we observe that the presence of complex magnetic topologies in ARs strongly increases the productivity rate of fast CMEs and represents one of the most favourable condition identified in this work's analysis of solar active regions sources of Earth-directed halo CMEs during Solar Cycle 24. This work's results are in good agreement with those found by Michalek & Yashiro (2013) studying a set of 68 ARs associated with halo CMEs observed by LASCO during the years 2001–2004, even though in our sample we observe an even higher tendency for ARs to show *kc* topologies. Figure 4 summarises the above results.

CME-productive ARs. The overall number of ARs involved is 31, 8 of which gave origin to more than one CME belonging to the sample: NOAA AR 11261 (Aug. 2011), 11283 (Sep. 2011), 11302 (Sep. 2011), 11402 (Jan. 2012), (Mar. 2012), 11944 (Jan. 2014), 12158 (Sep. 2014), 12371 (Jun. 2015). These 8 ARs alone produced 23 CMEs out of the 53 events considered. The most productive one were AR 11429, which gave origin to 6 CMEs (#18–23) and AR 12371, which gave origin to 4 CMEs (#49, 51–53).

Table 2. Association of CMEs with different classes of flares. Column 3 shows the percentage of CMEs associated to each class with respect to the total number of CMEs.

Solar flares: CME association		
SXR class	No. of CMEs	Percentage
X	9	17.0%
M	30	56.6%
C	9	17.0%
$\leq B$	5	9.4%
Total	53	100%

Table 3. Association with additional solar activity features. Column 3 shows the percentage of CMEs associated to each activity feature with respect to the total number of CMEs. Note that the association with filaments includes quiet filaments, eruptive filaments, filament activations, surges and sprays.

Other activity features: CME association		
	No. of CMEs	Percentage
Filaments	19	35.8%
X-ray sigmoid configurations	23	43.5%
Coronal waves	38	71.7%
Total	53	100%

On average, CME-productive ARs were observed to contain more sunspots than other ARs, and to be larger in terms of their sunspot area. CME-productive ARs were also extremely flare-productive: all the CMEs originated from this kind of ARs were in fact associated with energetic flares of class ≥M2.0. Moreover, 7 out of the 8 X-class flares associated with the selected 52 CMEs for which AR data were available, were actually associated to CMEs originated from CME-productive ARs. On the days associated with the eruptions of the selected CMEs, all these CME-productive regions were classified as *D*, *E* or *F* according to McIntosh classification scheme, and they were all characterised by a large asymmetric penumbra (classified as *k*). The only exception was AR 11283 on 6 September 2011, which was classified as *a* (small asymmetric penumbra). Most of them (6 out of 8) also showed a compact sunspot distribution (classified as *c* in the McIntosh c class). Overall, 16 out of 23 CMEs originated from CME-productive regions (70%) were associated with ARs classified as complex (*kc*) regions on the days of the eruptions.

5.2 Association with solar flares

Over a set of 53 CMEs under study, 48 of them (91%) were associated with C+ solar flares observed in the Soft X-Ray domain within 2 h before or after the first CME appearance in the LASCO C2 coronagraph. 39 out of 53 CMEs (74%) were associated with either X- or M-class flares. This result is in good agreement with previous work results; for example, Wang et al. (2002) found an association of 70% on a set of 132 frontside (full and partial) halo CMEs in the period 1996–2000.

As shown in Table 2, among the 48 CMEs associated with flares of importance greater than B, 9 were associated with X-class flares, 30 with M-class flares and 9 with C-class flares. Interestingly, of the 8 X-class flares of which data concerning the source AR were available, all were associated with *D*, *E* or *F* regions. Moreover, 7 of them showed a *k* (large asymmetric) penumbral configuration and 6 of them a *c* (compact) spot distribution. Combining together these parameters, we find that 6 out of 8 X-class flares were originated in complex *kc* active regions.

Finally, considering the duration of the associated flares, we find that 31 out of 48 flares (65%) are impulsive events decaying to half of the peak intensity within 60 min from their start, while 17 events (35%) are long duration flares. The average duration of flares is 67 min, in a range spacing from 247 min to 9 min. These results agree with those found in a statistical study of 69 flares observed in association with fast CMEs during Solar Cycle 23 (Lakshmi & Umapathy, 2013).

5.3 Association with other activity features

In this work we have focused on filaments/prominences, X-ray sigmoidal structures and global coronal perturbations with the aim of investigating pre- and post-eruptive conditions at the Sun to fully characterise the large-scale eruptions considered. Table 3 summarises the association rate found in the case of all these additional features.

We observe low association rates compared to previous works, in particular in the case of filaments/prominences. For comparison, St. Cyr & Webb (1991) found an association of 76% with filaments/prominences over a set of 73 CMEs observed by Solar Maximum Mission in the period 1984–86. One explanation to our low association rates can be traced back to the fact that AR filaments are also often not well defined. A second factor could be the fact that, since we are dealing mostly with CMEs originated near the solar disk centre, the identification of eruptive filaments might have been difficult and some eruptions might have passed unnoticed (Palmerio et al., 2017).

Considering X-ray sigmoid configurations, we find a 43% association with fast Earth-directed halo CMEs, a result that is much lower than that reported in previous studies. For example, Canfield et al. (1999) found a 65% association rate comparing flare-productive ARs with X-ray sigmoids by considering a sample of 79 eruptive ARs catalogued by NOAA in the years 1993 and 1997. We suspect that an observational bias may have affected our result by lowering the association rate for close-to-limb CMEs, for which recognition of sigmoidal features is expected to be more difficult due to projection effects.

Large-scale coronal waves appear to have the highest correlation (>70%) with fast CMEs. This result is in agreement with the association rate found by Cliver et al. (2005) for fast Earth-directed CMEs; however, in a scenario in which CMEs are expected to trigger global coronal disturbances in most cases, it is difficult to envision the use of coronal waves as indicators of a potentially geoeffective CMEs as this kind of signature would be not distinctive of this class of CMEs only.

6 Signatures at Earth: comparing simulation results to Wind data

In order to reconstruct the propagation and evolution of the selected CMEs/ICMEs in the heliosphere up to their arrival at Earth, we have compared ENLIL simulation outputs to in-situ measurements obtained by the Wind spacecraft, focusing on the identification of ICME-driven forward shocks to assess the performance of our prediction scheme and to investigate the properties of CME transit times.

Shock identification and transit times. To assess the arrival rate of the selected CMEs at Earth, we have applied the forward-shock identification criteria used by the Heliospheric Shock Database developed and maintained at the University of Helsinki (http://ipshocks.fi/, Kilpua et al., 2015) to 1-minute resolution data relative to magnetic field strength, plasma bulk speed and proton number density from the Wind/MFI and Wind/SWE instruments. In particular, the following criteria have been applied:

$$\frac{B_{down}}{B_{up}} \geq 1.2, \tag{6}$$

$$\frac{N_p^{down}}{N_p^{up}} \geq 1.2, \tag{7}$$

$$V_{down} - V_{up} \geq 20 \mathrm{km \cdot s^{-1}}, \tag{8}$$

where upstream and downstream values were calculated over a fixed time interval $\Delta t_{up} = \Delta t_{down} = 10$ min before and after the shock. For all the events, we have checked the data to make sure that an identified shock was driven by an ICME and not by other kind of transient events such as corotating interaction regions (CIRs) (Jian et al., 2006). Note that no Wind data were available for the days following the LASCO observation of CMEs #34, #46 and #47, so we have dropped out these event from the following analysis.

In this case we have uniquely identified a total of 36 forward shocks; among those, 5 could be linked to two or more CMEs at the Sun, leading to a total of 41 out of 50 (82%) CMEs driving a forward shock that arrived at Earth. This result is in agreement with other recent studies on Earth-directed halo CMEs (Shen et al., 2014).

ENLIL time series at a certain position in space are characterised by a variable time step, providing higher time resolution in conjunction with highly variable solar wind conditions. The typical time resolution at Earth for a $256 \times 30 \times 90$ grid was between 3 and 6 min. To identify forward shocks from ENLIL simulation outputs we have adapted the above shock identification conditions (Eqs. (6)–(8)) by considering Δt_{up} and Δt_{down} as composed of 60 data points each, for a typical $\Delta t \sim 3 - 6$ h. By applying these conditions to all the ENLIL simulation outputs at Earth, we have been able to uniquely identify a total of 34 forward shocks. Among the identified shocks, 4 could be linked to multiple CMEs at the Sun that merged within 1 AU, leading to a total of 38 out of 50 (76%) CMEs that were predicted to arrive at Earth.

For all the events that arrived at Earth we computed the *measured transit time*, defined as the time interval between the passage of the ICME forward shock observed by Wind and the estimated CME onset time, and the *simulated transit time*, defined as the difference between the ENLIL-estimated arrival time of the ICME at Earth and the estimated CME onset time. The measured transit times range from 34 h to 76 h, with an average value of 55 h and a median of 51 h. From simulation outputs we find simulated transit times spacing from 28 h to 79 h, with average and median values equal to 51 h. Assuming a typical 12-hour interval between the forward shock arrival and the ICME passage, the measured transit times presented here are consistent with those reported in Gopalswamy et al. (2001) for a set of 47 CMEs/ICMEs occurred during Solar Cycle 23 in the years 1996–2000. The mean error, corresponding to the average difference between the measured and forecasted arrival times, calculated over the whole set of events, was of 0.2 h; the median of the error distribution was −0.4 h. The mean absolute error (average the absolute values of the difference between the measured and forecasted arrival times) was 10.8 h. Considering single events, only 52% of them were predicted to arrive within ±10 h from the actual arrival time. However, as in this work we are mainly concerned with categorical forecasts, we leave a detailed investigation on the arrival time performance for further studies.

CME arrival forecast performance. The CME/shock association rate found by analysing ENLIL results is consistent with that observed in Wind data, despite the presence of some misses and false alarms implying that some CMEs that were expected to impact Earth according to simulations, did not actually arrive at Earth location, and vice versa. To assess the forecast performance for CME arrivals, from the results of the ICME-driven forward shock analysis we considered the event CME/shock arrival condition at Earth position. A "yes" forecast was hence marked in-situ by the identification of a forward shock; a "no" forecast was associated to a case when the CME did not arrive at Earth or a forward shock was not detected in Wind data. Considering the contingency tables for the CME arrival prediction, we have 34 correct arrival predictions, 6 correct negatives, 4 false alarms and 6 misses, resulting in a CR rate of 60% an FA rate of 40%, a CA ratio of 89% and an FA ratio of 11%. In this case the forecast performed poorly in the case of the FA rate (40%, with respect to a perfect score = 0%). On the other hand, the FA ratio score (11%) was the best among the different forecast scores considered in this work. Such results indicate that the CME arrival forecast performed particularly well in terms of the number of predicted hits (which dominated the forecast sample), while it gave no useful information in the case of predicted "no" events.

To identify the cause of the missing detections, we checked the ENLIL movies finding that 3 events (#11, #14 and #15) were predicted as impacting Earth from the flank, while events #49 and #50 were predicted to partially merge with event #51 within 1 AU. This suggests that the reason of the missing detection in ENLIL data may be due to a weak shock signal in the timeseries that led to a failure in the shock detection algorithm applied. On the other hand, in the case of false alarms we consider the discrepancy as primarily due to the CME width values used as input conditions in our simulations, as not all halo CMEs are characterised by average widths such as those described in Section 2, as well as by the errors related

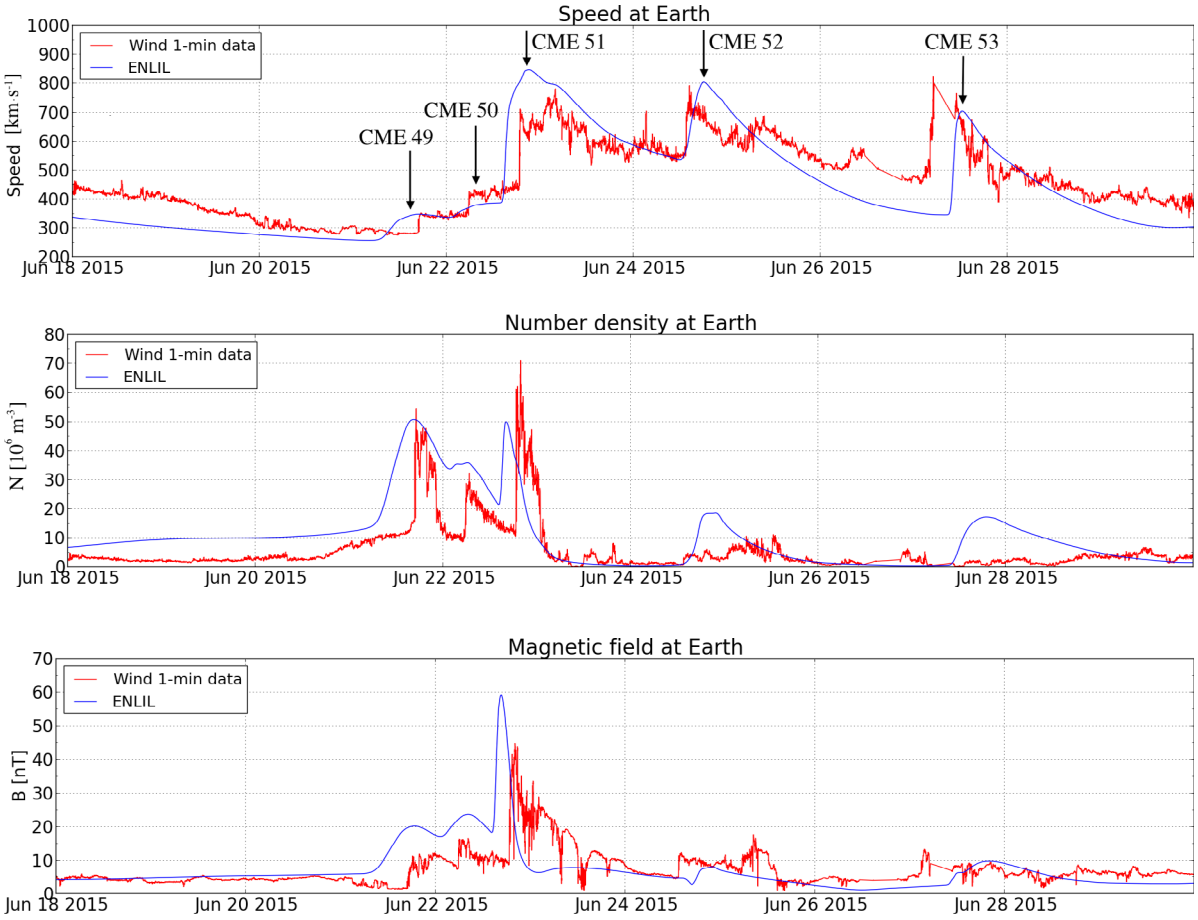

Fig. 5. Comparison between Wind data and ENLIL simulation results at Earth for a set of 5 CMEs (#49–53) observed by LASCO in the days 18–25 June 2015.

Table 4. Summary of the performances for the CME arrival and K_p index forecasts, including the correct rejection (CR) rate, false alarm (FA) rate, correct alarm (CA) ratio and false alarm (FA) ratio scores defined in Section 4.

Forecast performances				
Forecast type	CR rate	FA rate	CA ratio	FA ratio
CME arrival	60%	40%	89%	11%
K_p index (random)	75%	25%	80%	20%
K_p index (southward)	29%	71%	68%	32%

to the propagation direction in case of deflected events. In this sense, a major improvement would come from the analysis of CME stereoscopic images taken by the SECCHI coronagraphs on-board the STEREO spacecraft, as in this case Earth-viewed halo CMEs would appear as quasi-limb events from the STEREO spacecraft and a more precise evaluation of the actual width of single events could be performed (Jang et al., 2016 and references therein). We plan to do a detailed study of the effects that different CME input parameters (width, passage time at 0.1 AU, direction of propagation, etc.) have on the predicted arrival time at Earth and other spacecraft locations in future works. Figure 5 shows the comparison between Wind data and ENLIL with cone simulation results at Earth for a set of 5 CMEs (#49–53) observed by LASCO in the days 18–25 June 2015. The performances of the CME arrival and K_p forecasts are reported in Table 4.

7 Prediction of geoeffectiveness

7.1 K_p index

As shown in Figure 6, 34 out of 50 CMEs (68%) for which Wind data were available, resulted in a 3-hour $K_p \geq 5$ over the 96 h after their first appearance in LASCO C2 field-of-view. 30 out of 50 events (60%) triggered a $K_p \geq 6$ storm while 15 out of 50 (30%) triggered a $K_p \geq 7$ storm. Among the non-geoeffective CME, we found 9 events that were marked in-situ by a shock, but did not trigger any geomagnetic storm (CMEs #11, 17, 21, 32, 36, 37, 44, 49, 53). On the other hand, we found 2 cases in which a geomagnetic storm was triggered within 3–4 days from the CME observation even though no shock was observed at L1 (CMEs #24 and 35). Checking the in-situ data, we found CIR signatures before the storm onsets, meaning that in such cases the storms were not triggered by CMEs. We therefore classified CMEs #24 and 35 as non-geoeffective.

By considering $K_p = 5$ as minimum geoeffectiveness threshold, about one-fourth of the CMEs has not resulted in a significant K_p impact. A fraction of 68% geoeffective CMEs

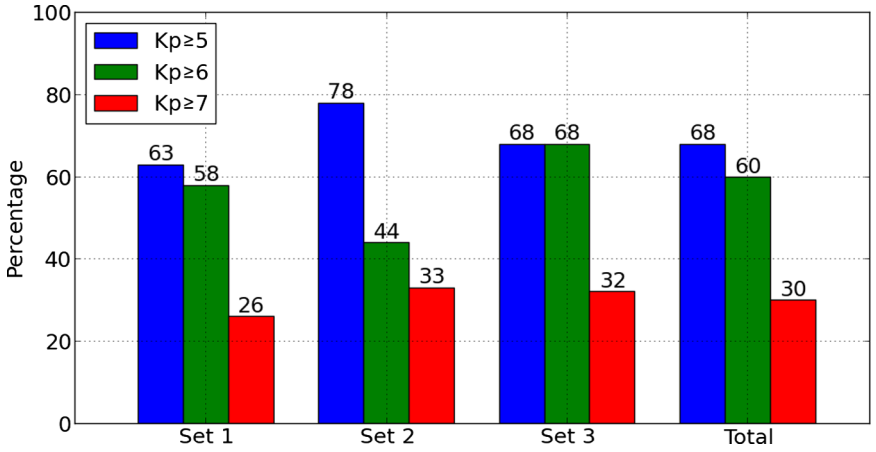

Fig. 6. Fraction of CMEs that triggered a K_p geomagnetic storm over the 96 h after their first appearance in LASCO C2 coronagraph, by set.

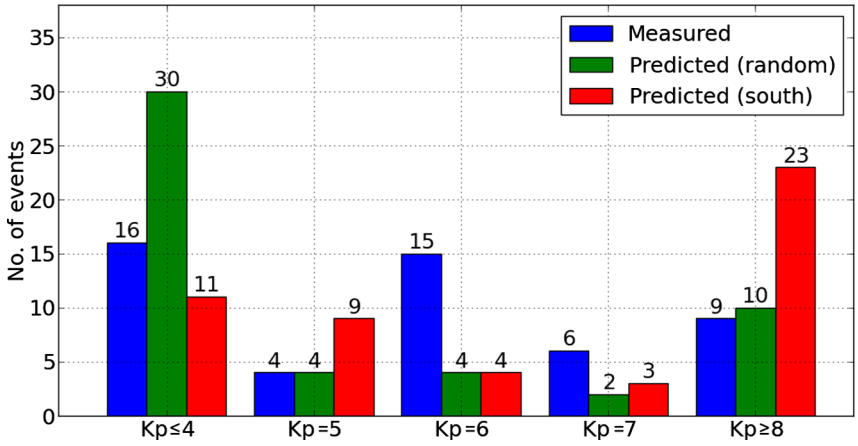

Fig. 7. Comparison between the measured and predicted geomagnetic storm intensities, based on maximum K_p index.

among all the selected Earth-directed halo CMEs is significantly higher than that found, for example, by Wang et al. (2002) using the same geoeffectiveness criteria. They reported that only 45% of Earth-directed halo CMEs are geoeffective, by considering a set of 132 frontside (full and partial) halo CMEs in the period 1996–2000. This discrepancy is reasonably due to the fact that in this work we have considered only fast CMEs and it confirms, *a posteriori*, our choice of fast CMEs as the most geoeffective type. As reported in Figure 6, the impact on Earth in terms of the K_p varies with the set of events and the K_p threshold considered. The association with $K_p \geq 5$ was 63–78% depending on the set considered, with the highest association rate (78%) for events belonging to S2. This result confirms the existence of an "east-west asymmetry" in the source locations of geoeffective CMEs, with regions close to the disk centre and in the western hemisphere being the most favourable sources of potentially geoeffective events (Michalek et al., 2006). The association with $K_p \geq 6$ was 44–68% depending on the set considered, with the highest association rate (68%) for events belonging to S3. The association with $K_p \geq 7$ storms was lower than 35% for all subsets. Finally, the inverse-checking procedure allowed to associate the majority of geomagnetic storms ($K_p \geq 5$) of Cycle 24 with Earth-directed halo CMEs observed at the Sun. However, a significant fraction of geomagnetic storms showed no temporal association with any halo CME observed by LASCO coronagraphs. In such cases, the source of the observed geomagnetic activity is expected to be a partial-halo CME or a CIR.

K_p predictions. Figure 7 reports the measured and predicted number of storms associated with CME/ICME pairs, by K_p intensity. The predicted K_p indices are reported in the case of both southward and random orientations. As expected, completely southward $\boldsymbol{B}$ fields would lead to much stronger effects on Earth, due to the enhanced magnetic reconnection rate triggered at dayside magnetopause. Overall, this worst-case scenario tends to overestimate the maximum K_p value in most of the events; the number of events associated with a $K_p \leq 7$ is significantly depressed, while that of the events associated with a $K_p \geq 8$ is greatly enhanced (23 $K_p \geq 8$ events compared to the 9 observed events). However, the estimated number of 39 events resulting in a $K_p \geq 5$ is very close to the actually observed number of 36 events. Using the expectation value for the clock-angle term (random case), the predicted maximum K_p distribution closely resembles actual measurements in the number of highly geoeffective ($K_p \geq 8$) events. On the other hand, it highly overestimates the number of non-geoeffective ($K_p \leq 4$) events. Such results emphasise the crucial importance of developing models capable to

reliably predict the ***B*** orientation of Earth-impacting ICME structures, as cone CME models and artificial ***B*** orientations such as those considered in this work cannot fully capture the intensity distribution of geomagnetic storms.

K_p forecast performance. To test the K_p index prediction performance, we have considered the "event" to be the "geomagnetic storm" defined by a $K_p \geq 5$ condition. A "yes" forecast was then labelled by a $K_p \geq 5$ prediction, while a "no" condition was associated to a $K_p \leq 4$. The contingency table relative to the average-case scenario of a randomly orientated IMF was characterised by a total of 16 hits, 18 misses, 4 false alarms and 12 correct negatives, leading to a CR rate of 75%, an FA rate of 25%, a CA ratio of 80% and an FA ratio of 20% (see Tab. 4). In this case the forecast performed generally well, with good scores for both the FA ratio and the FA rate. As observable in Table 4, the FA ratio for this particular forecast was the second best predicted one among all the different forecasts considered in this work. However, in this case there was a high number of misses, i.e. many events that arrived at Earth were not predicted to impact. For the worst-case scenario of a totally southward oriented IMF the contingency table gave 26 hits, 7 misses, 12 false alarms and 5 correct negatives, giving a CR rate of 29%, an FA rate of 71%, a CA ratio of 68% and an FA ratio of 32%. In this case the forecast performed poorly especially in terms of the number of false alarms (FA rate = 71%, with respect to a perfect score = 0%), meaning that in this case the forecast significantly overestimated the number of events that arrived at Earth. On the other hand, the FA ratio score was still good (32%) although not as good as in the randomly-oriented case. The number of misses in this case was reduced by a factor of 2 with respect to the previous case. In conclusion, the forecast performance analysis for the K_p index indicates that the randomly-oriented K_p forecast performed better than the completely southwardly-oriented one in the case of false alarms, but it performed worse in the case of misses.

As a final note, we point out the case of CME #37, which was a false alarm in our prediction scheme for both the clock angle orientations considered. ENLIL in this case did not forecast correctly the event, predicting a face-on impact on Earth and a strong geomagnetic storm ($K_p = 9$) for both the orientations considered. A complete event analysis was presented by Möstl et al. (2015), who attributed the unexpected low geomagnetic impact of the event as due to a non-radial deflection away from the Sun-Earth line to the west due to a nearby coronal hole on the east side of the CME source region. Such "channelling" effect has been observed to act $<21.5\,R_\odot$ from the Sun, hence within the ENLIL inner heliospheric boundary. In this sense, we believe that ENLIL missed this channelling effect due to the fact that CMEs are assumed to have a totally-radial motion during insertion in the heliosphere at $<21.5\,R_\odot$. Therefore, by taking as input condition the position of the associated AR for the CME direction of propagation, we ignored all the deflection effects that may have acted on the CME between the photosphere and $21.5\,R_\odot$. To try to balance out these effects one could re-calculate the CME propagation direction at $21.5\,R_\odot$ by using more sophisticated reconstruction methods such as the forward-modelling technique (Thernisien et al., 2009).

7.2 Magnetopause stand-off distance

When considering the geoeffectiveness of CME/ICME pairs in terms of the induced magnetospheric compression, no continuous monitor is available for comparison. On average, the magnetopause stand-off distance along the Sun-Earth direction is $\sim 10\,R_E$ for unperturbed solar wind conditions. By applying equation (5) to ENLIL-at-Earth data series, we have found an average and median values of the distribution of the minimum magnetopause stand-off distance of $5.3\,R_E$, with a maximum compression of $4.4\,R_E$.

According to our estimates, the compression driven by the identified IP forward shocks is sufficient to push the magnetopause within the region of geosynchronous orbits, located at about $6.6\,R_E$ from the Earth centre, in the case of 41 out of 50 CMEs (82%). The minimum magnetopause stand-off distance gives an idea of the geoeffectiveness of CME/ICME events in terms of their hazard to satellite operations, as spacecraft designed to survive in a region of space normally shielded by the Earth's magnetic field become subjected to a much harsher environment once they have entered the magnetosheath. Moreover, 17 events (43% of the ENLIL-identified forward shocks) are expect to compress the magnetopause at altitudes $\leq 4.9\,R_E$, causing disturbances to navigation systems such as GPS satellites orbiting in MEO at ~26 000 km above the Earth surface.

Figure 8 shows a comparison of the predicted and measured geoeffectiveness caused by the impact of CMEs #18–23 in March 2012.

8 Association with major SEP events

In order to fully characterise the impact of the selected CME events on geospace, we have analysed their association with major SEP events (peak proton particle flux above 10 MeV > 10 pfu), starting from the SEP event lists maintained by the NASA CDAW Data Center (http://cdaw.gsfc.nasa.gov/CME_list/sepe/) and by NOAA (http://umbra.nascom.nasa.gov/SEP/). In the years 2009–2015, a total of 41 major SEP events have been observed and the majority of them (92%) was caused by a halo CME. Among those 41 major SEP events, 19 (44%) were associated to the CMEs included in our analysis. Addressing the problem the other way around, we found that 19 of the 53 CMEs considered in the analysis (36%) triggered a major SEP within few hour from their onset, and 4 of them were extremely intense events associated with a peak proton particle flux above 10 MeV > 1000 pfu, three orders of magnitude above the minimum threshold for major SEPs.

Considering the source locations of SEP-associated CMEs, we observed a strong "east-west asymmetry" in their heliographic distribution, with the majority of them originated in the west part of the visible solar disk. In fact, the 13 out of 19 SEP-associated CMEs (68%) originated from the western hemisphere of the Sun, with an average source solar longitude equal to 20° W. For comparison, the average solar longitude for SEP-less CMEs was equal to 3° E, while considering all the CMEs regardless of their association with SEP events, the average source longitude was 5° W.

Investigating the association of SEP-CME events with solar flares, we found that all the 19 SEP-associated CMEs were also associated with a C+ flare (C:16%-M:52%-X:32%). For comparison, in the case of the 34 SEP-less CMEs, the flare class distribution was no/B-class flare:15%-C:18%-M:59%-X:9%. The flare class distribution for the whole CME set,

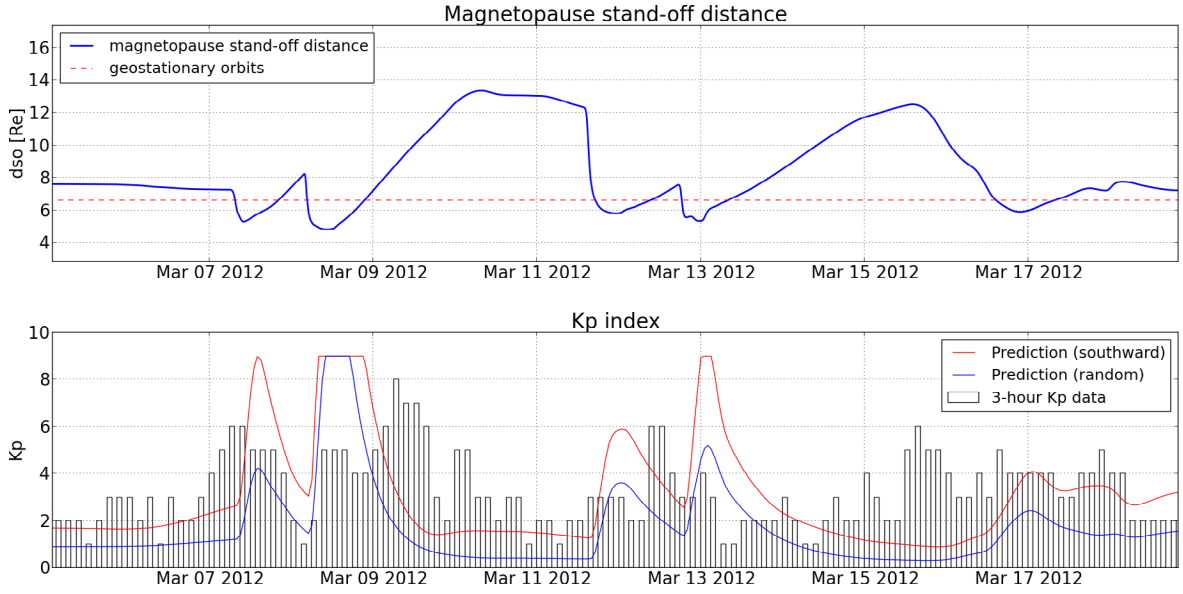

Fig. 8. Predicted and observed effects at Earth for geomagnetic storm of 5–13 March 2012, caused by CMEs #18–23. *Top*: magnetopause stand-off distance. *Bottom*: 3-hour K_p index measured on ground (bars) and predicted from ENLIL outputs at Earth (red and blue).

regardless of their association with major SEPs, was no/B-class flare:10%-C:17%-M:57%-X:17%. This result suggests that strong (X-class) flares are observed more often than usual in the case of CMEs triggering major SEPs (32% compared to 9% for SEP-less CMEs), but C-class flares account for a significant fraction of the total distribution in both cases (16% compared to 30% for SEP-less CMEs). In our set, association of SEP-CME events with large (X or M) flares is 84%, slightly higher than that reported by Gopalswamy et al. (2003) for a set of 48 major SEP events associated to CMEs occurred during Solar Cycle 23 (1997–2001). The correlation between the flare class and the intensity of the associated SEP event seems to be very weak (0.36), comparable to that reported by Gopalswamy et al. (2003).

Considering the speed properties of SEP-associated CMEs, they were found to be significantly faster than SEP-less ones. They were characterised by an average speed in space equal to 1692 $km \cdot s^{-1}$, while SEP-less CMEs had an average speed of 1162 $km \cdot s^{-1}$. Correlating the CME speed to the intensity of the associated SEP-CME events, we find a correlation coefficient of 0.69, slightly higher than that reported by previous studies (Kahler, 2001; Gopalswamy et al., 2003).

Finally, 74% of the CMEs that triggered a major SEP event were found to be geoeffective in terms of magnetopause compression and/or K_p index. For comparison, 71% of the SEP-less CMEs and 72% of all the CMEs considered (regardless of their association with major SEPs) were geoeffective according to at least one of the geoeffectiveness conditions considered. Figure 9 shows a comparative plot of the CME events, reporting their association with major SEP events and their maximum K_p index, with a color code for the K_p bar based on the CME speed. Such results show that while the CME SEP-production efficiency correlates very well with the CME speed, the CME geoeffectiveness is the result of the interplay of multiple factors.

Being aware of the importance of CME-CME interactions for particle acceleration (Gopalswamy et al., 2002), we have paid particular attention to CMEs that originated from the same AR within a period of some days. As seen in Table 1, among the 53 CMEs under study, 23 were originated from CME-productive regions, for a total of 8 ARs involved. Among these 8 sequences of CMEs, 7 (88%) triggered a major SEP event in coincidence of one of the CMEs involved. The only exception was represented by CMEs #9 and #10, which were launched within a 20-hour interval but in which case the first CME was faster than the second one. Considering the remaining 4 double-CME events, the SEP peak flux was observed to be coincident with the launch of the second CME in all cases: this result supports the idea that the particle acceleration efficiency increases as consequence of CME-CME interactions and that the most probable configuration is the case of a fast CME that reaches up a slower CME launched a few hours before from the same solar region.

9 Conclusion

In this work, we have presented a statistical analysis of a set of 53 fast ($V \geq 1000\ km \cdot s^{-1}$) Earth-directed full halo CMEs observed by the SOHO/LASCO instrument during the period Jan. 2009-Sep. 2015, and we have then used this CME sample to test the forecasting capabilities of a Sun-to-Earth prediction scheme – based on 3D simulations and solar wind-geomagnetic activity coupling functions – for the geo-effectiveness of Earth-directed halo CMEs.

We have first analysed the solar conditions associated with each CME by considering their association with other solar activity features such as active regions, solar flares, filaments/prominences, X-ray sigmoids and global coronal disturbances, with the final aim of identifying recurrent peculiar features that can be used as precursors of CME-driven geomagnetic storms. We have found that the solar regions that most likely generate geoeffective CMEs are active regions showing bipolar magnetic topologies with large asymmetric penumbrae and

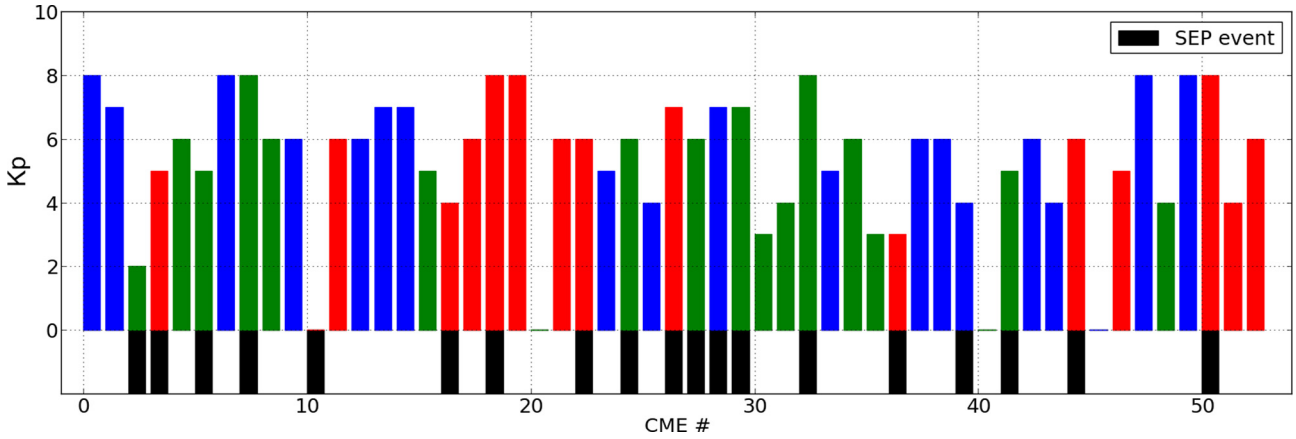

Fig. 9. Comparative plot of the CME events, showing their association with major SEP events (black bars) compared to their maximum K_p index. The color code for the K_p bar is based on the CME speed: blue for $V_{space} < 1100\,km{\cdot}s^{-1}$, green for $1100\,km{\cdot}s^{-1} \leq V_{space} < 1500\,km{\cdot}s^{-1}$, and red for $V_{space} \geq 1500\,km{\cdot}s^{-1}$.

a compact sunspot distribution – usually referred to as "complex" topologies (Figs. 3 and 4). The presence of energetic flares, in particular X- and M-class ones, is also a highly favouring factor (Tab. 2). Among the other activity features considered, coronal waves have been the only ones observed in more than one half of the events (Tab. 3). Moreover, this study reports a high level of correlation between CME speed and source location properties and the resulting geoeffectiveness, with fast events originating from the central and western regions of the solar disk having the highest probability of resulting in a geomagnetic storm (Fig. 6).

To reconstruct the global evolution into IP space up to Earth location, we have propagated all the CME events into the heliosphere up to 2 AU by means of the WSA-ENLIL+Cone model running at NASA/CCMC. From the comparison of ENLIL-at-Earth results with in-situ solar wind measurements at L1 obtained by the Wind spacecraft, we have been able to link ICME-driven IP forward shocks to the majority of CMEs, confirming the fact that Earth-directed halo CMEs or their associated shocks do arrive at Earth in most cases. From Wind data series we were able to uniquely identify IP forward shocks for 82% of the selected CMEs, and a similar association rate was found in the case of ENLIL-at-Earth results (76%). Over the totality of the simulated events, the majority (78%) were correctly forecasted. Among the 11 incorrectly forecasted events, a further analysis showed that most of these events in simulations were predicted to impact Earth from the flank, leading to a weak shock signature in ENLIL timeseries that passed unnoticed by the detection algorithm. The CME input conditions used may have also played a role in limiting the forecast accuracy.

In the last part of our analysis we have used simulation outputs upstream of Earth to predict the geoeffectiveness of each CME event in terms of the expected geomagnetic activity level and magnetospheric compression. Our prediction of the induced geomagnetic activity in terms of the 3-hour planetary K_p index was primarily affected by the unreliable ENLIL prediction of the ICME magnetic field orientation at L1. For this reason, we envisioned a worst-case scenario with a completely southward IMF and an average-case scenario of a randomly-oriented IMF. We found that the worst-case scenario tends to overestimate the single-event K_p value, but it generally well represents the fraction of $K_p \geq 5$ events over the total (78%, compared to 68% for actual measurements). On the other hand, a random orientation of the $\boldsymbol{B}$ field heavily underestimates the fraction of CMEs causing a mild geomagnetic activity, while it well reproduces the number of strong storms (20%, compared to 18% for actual measurements) (Fig. 7). Improvements on this point would come from the use of an MHD model that includes the CME internal magnetic structure e.g. modelled as a flux-rope CME. Such kind of models are under development (Shiota & Kataoka, 2016) but none is currently available for extensive modelling such as that needed in the case of the statistical analysis presented in this work.

Our predictions of the minimum magnetopause stand-off distance suggest that the dayside magnetopause is compressed below the altitude of geosynchronous (GEO) satellites whenever one of the identified ICMEs impacts on Earth. In fact, in our study, all the CMEs that arrived at Earth compressed the magnetopause below GEO orbits. In a significant number of cases, it was compressed even below MEO orbits (42% of the CMEs that arrived at Earth). This result confirms the importance of studying fast Earth-directed CMEs as they can severely affect human activities by generating major disruptions to satellites serving as telecommunication and navigation systems.

The analysis of related SEP events shows that 74% of the CMEs associated with major SEPs were geoeffective i.e. triggered a minor to intense geomagnetic storm ($K_p \geq 5$). Analysing the association of the selected CMEs with major SEP events, we found that CMEs characterised by higher speeds and originating close to the solar disk centre or from the western hemisphere are much more likely to trigger a major SEP event at Earth. Moreover, the SEP production resulted enhanced in the case of fast CMEs, with a correlation coefficient between CME speeds and SEP peak fluxes of 0.69. Under these conditions, we found that the likelihood for a multiple-CME event to originate a strong SEP as consequence of CME-CME interactions was considerably high.

The results of our prediction scheme appear promising as the forecast performances for the CME arrival and K_p index in terms of the scores reported in Table 4 show good agreement with in-situ observations and actual data records for geomagnetic activity; improvements to this prediction scheme could come from an extension of the CME sample considered. Moreover, as already pointed out, the internal magnetic structure of ICMEs represents a critical issue in determining their impact on geospace and future models will have to take this point into account to provide more reliable space weather predictions. Finally, in the case of halo CME simulations, a big improvement would be represented by the use of stereoscopic images taken by SECCHI coronagraphs on-board the STEREO spacecraft when they are in quadrature with the Earth to determine the kinematical CME properties and hence provide better input parameters for heliospheric CME models; in this case, Earth-viewed halo CMEs would appear as limb CMEs and a more precise evaluation of the actual width and speed and direction of propagation could be performed. In this sense, the prediction scheme presented in this work and the open questions arisen from our analysis represent a promising starting point for future studies.

Acknowledgements. Part of this work was carried out at the University of Trieste, Department of Physics. C. S. was partially funded by the Research Foundation − Flanders (FWO) SB PhD fellowship number 1S42817N. The paper uses data from the SOHO/LASCO CME catalogue, generated and maintained at the CDAW Data Center by NASA and the Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international collaboration between ESA and NASA. The Wind solar wind plasma and magnetic field data were obtained at NASA's CDAWeb (cdaweb.gsfc.nasa.gov). The Dst geomagnetic index data were obtained from the World Data Center for Geomagnetism in Kyoto, Japan. Final definitive K_p indices are from the Helmholtz Center Potsdam GFZ German Research Centre for Geosciences. Simulation results have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their public Runs on Request system (http://ccmc.gsfc.nasa.gov). The CCMC is a multi-agency partnership between NASA, AFMC, AFOSR, AFRL, AFWA, NOAA, NSF and ONR. The WSA-ENLIL+Cone Model was developed by the D. Odstrcil at the Univ. of Colorado at Boulder. The authors thank two anonymous referees for constructive comments and suggestions. The editor thanks two anonymous referees for their assistance in evaluating this paper.

References

Arge CN, Pizzo VJ. 2000. Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J Geophys Res* **105**: 10465–10480.

Burlaga LFE. 1991. Magnetic Clouds. In R. Schwenn, and E. Marsch, eds., Physics of the Inner Heliosphere II. p. 152.

Burlaga LF, et al. 2002. Successive CMEs and complex ejecta. *J Geophys Res: Space Phys* **107**: 1266. DOI:10.1029/2001JA000255.

Cane HV, Lario D. 2006. An introduction to CMEs and energetic particles. *Space Sci Rev* **123**: 45–56.

Canfield RC, et al. 1999. Sigmoidal morphology and eruptive solar activity. *Geophys Res Lett* **26**: 627–630.

Cannon P, et al. 2013. Extreme space weather: impacts on engineered systems and infrastructure. Tech. rep., Royal Academy of Engineering, London, UK.

Chen C, et al. 2011. Statistical study of coronal mass ejection source locations: 2. Role of active regions in CME production. *J Geophys Res: Space Phys* **116**: A12108.

Cid C, et al. 2012. Can a halo CME from the limb be geoeffective? *J Geophys Res* **117**.

Cliver EW, et al. 2005. On the origins of Solar EIT waves. *Astrophys J* **631**: 604–611.

Dierckxsens M, et al. 2015. Relationship between solar energetic particles and properties of flares and CMEs: statistical analysis of solar cycle 23 events. *Sol Phys* **290**: 841–874. DOI:10.1007/s11207-014-0641-4. arXiv: 1410.6070 [astro-ph.SR].

Dumbović M, et al. 2015. Geoeffectiveness of coronal mass ejections in the SOHO Era. *Sol Phys* **290**: 579–612. DOI:10.1007/s11207-014-0613-8. arXiv: 1410.3303 [astro-ph.SR].

Emmons D, et al. 2013. Ensemble forecasting of coronal mass ejections using the WSA-ENLIL with CONED model. *Space Weather* **11**: 95–106. DOI:10.1002/swe.20019.

Gonzalez WD, et al. 1994. What is a geomagnetic storm? *J Geophys Res* **99**: 5771–5792.

Gopalswamy N, et al. 2001. Predicting the 1-AU arrival times of coronal mass ejections. *J Geophys Res* **106**: 29207–29218.

Gopalswamy N, et al. 2002. Interacting coronal mass ejections and solar energetic particles. *Astrophys J Lett* **572**: L103–L107.

Gopalswamy N, et al. 2003. Large solar energetic particle events of cycle 23: a global view. *Geophys Res Lett* **30**: 8015. DOI:10.1029/2002GL016435.

Gopalswamy N, et al. 2007. Geoeffectiveness of halo coronal mass ejections. *J Geophys Res* **112**: A06112. DOI:10.1029/2006JA012149.

Gopalswamy N, et al. 2009. The expansion and radial speeds of coronal mass ejections. *Cent Eur Astrophys Bull* **33**: 115–124.

Gopalswamy N, et al. 2010. Solar sources of driverless interplanetary shocks. 12th International Solar Wind Conference **1216**, pp. 452–458. DOI:10.1063/1.3395902.

Gopalswamy N, et al. 2010. A catalog of halo coronal mass ejections from SOHO. *Sun Geosph* **5**: 7–16.

Gopalswamy N, et al. 2014. Anomalous expansion of coronal mass ejections during solar cycle 24 and its space weather implications. *Geophys Res Lett* **41**: 2673–2680. DOI:10.1002/2014GL059858. arXiv: 1404.0252 [astro-ph.SR].

Gopalswamy N, et al. 2015a. The mild space weather in solar cycle 24. ArXiv e-prints: arXiv: 1508.01603 [astro-ph.SR].

Gopalswamy N, et al. 2015b. The peculiar behavior of halo coronal mass ejections in solar cycle 24. *Astrophys J Lett* **804**: L23.

Gopalswamy N, et al. 2015c. High-energy Solar Particle Events in cycle 24. *J Phys Conf Ser* **642**: 012012.

Gosling JT. 1993. The solar flare myth. *J Geophys Res* **98**: 18937–18950.

Gosling JT, et al. 1991. Geomagnetic activity associated with earth passage of interplanetary shock disturbances and coronal mass ejections. *J Geophys Res* **96**: 7831–7839. DOI:10.1029/91JA00316.

Hundhausen AJ, et al. 1984. Coronal mass ejections observed during the Solar Maximum Mission − Latitude distribution and rate of occurrence. *J Geophys Res* **89**: 2639–2646.

Huttunen K, Koskinen H. 2004. Importance of post-shock streams and sheath region as drivers of intense magnetospheric storms and high-latitude activity. *Ann Geophys* **22**: 1729–1738. DOI:10.5194/angeo-22-1729-2004.

Huttunen KEJ, et al. 2002. Variability of magnetospheric storms driven by different solar wind perturbations. *J Geophys Res: Space Phys* **107**: 1121. DOI:10.1029/2001JA900171.

Jaeggli SA, Norton AA. 2016. The magnetic classification of solar active regions 1992–2015. *Astrophys J Lett* **820**.

Jang S, et al. 2016. Comparison between 2D and 3D parameters of 306 front-side halo CMEs from 2009 to 2013. *Astrophys J* **821**: 95. DOI:10.3847/0004-637X/821/2/95.

Janvier M, et al. 2014. Mean shape of interplanetary shocks deduced from in situ observations and its relation with interplanetary CMEs. *Astron Astrophys* **565**: A99. DOI:10.1051/0004-6361/201423450.

Jian L, et al. 2006. Properties of Stream Interactions at 1 AU During 1995-2004. *Sol Phys* **239**: 337–392. DOI:10.1007/s11207-006-0132-3.

Jolliffe IT, Stephenson DB. 2011. Forecast verification: a practioner's guide in atmospheric science. 2nd. ed., Wiley, online library.

Kahler SW. 2001. The correlation between solar energetic particle peak intensities and speeds of coronal mass ejections: Effects of ambient particle intensities and energy spectra. *J Geophys Res* **106**: 20947–20956. DOI:10.1029/2000JA002231.

Kahler SW, Vourlidas A. 2014. Do Interacting Coronal Mass Ejections Play a Role in Solar Energetic Particle Events? *Astrophys J* **784**: 47. DOI:10.1088/0004-637X/784/1/47.

Kilpua EKJ, et al. 2015. Properties and drivers of fast interplanetary shocks near the orbit of the Earth: 1995-2013. *J Geophys Res: Space Phys* **120**: 4112–4125. DOI:10.1002/2015JA021138.

Kivelson MG, Russell CT. 1995. Introduction to space physics, Cambridge University Press, Cambridge, UK, p. 586.

Koskinen HEJ, Huttunen KEJ. 2006. Geoeffectivity of coronal mass ejections. *Space Sci Rev* **124**: 169–181. DOI:10.1007/s11214-006-9103-0.

Lakshmi MA, Umapathy S. 2013. Coronal mass ejections associated with short and long duration X-ray flares. *Astron Soc India Conf Series* **10**.

Lemen JR, et al. 2012. The atmospheric imaging assembly (AIA) on the solar dynamics observatory (SDO). *Sol Phys* **275**: 17–40.

Lepping RP, et al. 1995. The wind magnetic field investigation. *Space Sci Rev* **71**: 207–229.

Lugaz N, Farrugia CJ. 2014. A new class of complex ejecta resulting from the interaction of two CMEs and its expected geoeffectiveness. *Geophys Res Lett* **41**: 769–776. DOI:10.1002/2013GL058789. arXiv: 1402.1075 [astro-ph.SR].

Lugaz N, et al. 2015. Extreme geomagnetic disturbances due to shocks within CMEs. *Geophys Res Lett* **42**: 4694–4701. DOI:10.1002/2015GL064530.

Lugaz N, et al. 2016. Factors affecting the geoeffectiveness of shocks and sheaths at 1 AU. *J Geophys Res: Space Phys* **121**: 10. DOI:10.1002/2016JA023100.

Lugaz N, et al. 2017. The interaction of successive coronal mass ejections: a review. *Sol Phys* **292**: 64.

McIntosh PS. 1990. The classification of sunspot groups. *Sol Phys* **125**: 251–267.

Michalek G, Yashiro S. 2013. CMEs and active regions on the sun. *Adv Space Res* **52**: 521–527.

Michalek G, et al. 2006. Properties and geoeffectiveness of halo coronal mass ejections. *Space Weather* **4**: S10003.

Mierla M, et al. 2010. On the 3-D reconstruction of Coronal Mass Ejections using coronagraph data. *Ann Geophys* **28**: 203–215. DOI:10.5194/angeo-28-203-2010.

Möstl C, et al. 2015. Strong coronal channelling and interplanetary evolution of a solar storm up to Earth and Mars. *Nat Commun* **6**: 7135. DOI:10.1038/ncomms8135. arXiv: 1506.02842 [astro-ph.SR].

Newell PT, et al. 2007. A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *J Geophys Res: Space Phys* **112**.

Newell PT, et al. 2008. Pairs of solar wind-magnetosphere coupling functions: Combining a merging term with a viscous term works best. *J Geophys Res: Space Phys* **113**.

O'Brien TP, McPherron RL. 2000. An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay. *J Geophys Res* **105**: 7707–7720.

Odstrcil D. 2003. Modeling 3-D solar wind structure. *Adv Space Res* **32**: 497–506.

Ogilvie KW, et al. 1995. SWE, A comprehensive plasma instrument for the wind spacecraft. *Space Sci Rev* **71**: 55–77.

Owens MJ, Forsyth RJ. 2013. The heliospheric magnetic field. *Living Rev Sol Phys* **10**.

Paassilta M, et al. 2017. Catalogue of 55–80 MeV solar proton events extending through solar cycles 23 and 24. *J Space Weather Space Clim* **7**: A14. DOI:10.1051/swsc/2017013. arXiv: 1707.00498 [physics.space-ph].

Palmerio E, et al. 2017. Determining the Intrinsic CME Flux Rope Type Using Remote-sensing Solar Disk Observations. *Sol Phys* **292**: 39. DOI:10.1007/s11207-017-1063-x. arXiv: 1701.08595 [astro-ph.SR].

Papaioannou A, et al. 2016. Solar flares, coronal mass ejections and solar energetic particle event characteristics. *J Space Weather Space Clim* **6**: A42. DOI:10.1051/swsc/2016035.

Qu ZQ. 2008. Prediction of solar flares from a statistical analysis of events during solar cycle 23. *ArXiv e-prints*.

Reames DV. 1999. Particle acceleration at the Sun and in the heliosphere. *Space Sci Rev* **90**: 413–491. DOI:10.1023/A:1005105831781.

Reames DV. 2013. The two sources of solar energetic particles. *Space Sci Rev* **175**: 53–92.

Richardson IG. 2013. Geomagnetic activity during the rising phase of solar cycle 24. *J Space Weather Space Clim* **3**: A08. DOI:10.1051/swsc/2013031.

Richardson IG, Cane HV. 2010. Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): catalog and summary of properties. *Sol Phys* **264**: 189–237. DOI:10.1007/s11207-010-9568-6.

Richardson IG, Cane HV. 2012. Solar wind drivers of geomagnetic storms during more than four solar cycles. *J Space Weather Space Clim* **2**: A01. DOI:10.1051/swsc/2012001.

Robbrecht E, et al. 2009. Automated LASCO CME catalog for solar cycle 23: are cmes scale invariant? *Astrophys J* **691**: 1222–1234. DOI:10.1088/0004-637X/691/2/1222. arXiv: 0810.1252.

Rodriguez L, et al. 2009. Three frontside full halo coronal mass ejections with a nontypical geomagnetic response. *Space Weather* **7**: S06003. DOI:10.1029/2008SW000453.

Scherrer PH, et al. 2012. The helioseismic and magnetic imager (HMI) investigation for the solar dynamics observatory (SDO). *Sol Phys* **275**: 207–227.

Schwenn R, et al. 2005. The association of coronal mass ejections with their effects near the earth. *Ann Geophys* **23**: 1033–1059.

Shen C, et al. 2014. Full-halo coronal mass ejections: arrival at the earth. *J Geophys Res* **119**: 5107–5116.

Shiota D, Kataoka R. 2016. Magnetohydrodynamic simulation of interplanetary propagation of multiple coronal mass ejections with internal magnetic flux rope: SUSANOO-CME. *Space Weather* **14**: 56–75. DOI:10.1002/2015SW001308.

Srivastava N, Venkatakrishnan P. 2004. Solar and interplanetary sources of major geomagnetic storms during 1996–2002. *J Geophys Res: Space Phys* **109**: A10103. DOI:10.1029/2003JA010175.

St Cyr OC, Webb DF. 1991. Activity associated with coronal mass ejections at solar minimum – SMM observations from 1984–1986. *Sol Phys* **136**: 379–394.

St Cyr OC, et al. 2000. Properties of coronal mass ejections: SOHO LASCO observations from January 1996 to June 1998. *J Geophys Res* **105**: 18169–18186. DOI:10.1029/1999JA000381.

Subramanian P, Dere KP. 2001. Source regions of coronal mass ejections. *Astrophys J* **561**: 372–395.

Taktakishvili A, et al. 2009. Validation of the coronal mass ejection predictions at the earth orbit estimated by ENLIL heliosphere cone model. *Space Weather* **7**. issn: 1542–7390.

Temmer M, et al. 2017. Preconditioning of interplanetary space due to transient CME disturbances. *Astrophys J* **835**: 141. DOI:10.3847/1538-4357/835/2/141. arXiv: 1612.06080 [astro-ph.SR]. 32

Thernisien A, et al. 2009. Forward modeling of coronal mass ejections using STEREO/SECCHI data. *Sol Phys* **256**: 111–130. DOI:10.1007/s11207-009-9346-5.

Toth G, Odstrcil D. 1996. Comparison of some flux corrected transport and total variation diminishing numerical schemes for hydrodynamic and magnetohydrodynamic problems. *J Comput Phys* **128**: 82–100.

Tsurutani BT, et al. 1988. Origin of interplanetary southward magnetic fields responsible for major magnetic storms near solar maximum (1978–1979). *J Geophys Res* **93**: 8519–8531. DOI:10.1029/JA093iA08p08519.

Wang YM, et al. 2002. A statistical study on the geoeffectiveness of Earth-directed coronal mass ejections from March 1997 to December 2000. *J Geophys Res: Space Phys* **107**: 1340. DOI:10.1029/2002JA009244.

Wang Y, et al. 2004. Deflection of coronal mass ejection in the interplanetary medium. Sol Phys **222**: 329–343. DOI:10.1023/B:SOLA.0000043576.21942.aa.

Watari S. 2017. Geomagnetic storms of cycle 24 and their solar sources. *Earth Plan Space* **69**: 70. DOI:10.1186/s40623-017-0653-z.

Webb DF, et al. 2000. Relationship of halo coronal mass ejections, magnetic clouds, and magnetic storms. *J Geophys Res* **105**: 7491–7508.

Weigel RS, et al. 2006. Decision theory and the analysis of rare event space weather forecasts. *Space Weather* **4**: 05002. DOI:10.1029/2005SW000157.

Wimmer-Schweingruber RF, et al. 2006. Understanding interplanetary coronal mass ejection signatures. Report of working group B. *Space Sci Rev* **123**: 177–216. DOI:10.1007/s11214-006-9017-x.

Xie H, et al. 2004. Cone model for halo cmes: application to space weather forecasting. *J Geophys Res: Space Phys* **109**: A03109. DOI:10.1029/2003JA010226.

Yashiro S, et al. 2004. A catalog of white light coronal mass ejections observed by the SOHO spacecraft. *J Geophys Res: Space Phys* **109**: A07105. DOI:10.1029/2003JA010282.

Zhao XP, et al. 2002. Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. *J Geophys Res: Space Phys* **107**: 1223. DOI:10.1029/2001JA009143.

Zhang J, et al. 2004. A statistical study of the geoeffectiveness of magnetic clouds during high solar activity years. *J Geophys Res: Space Phys* **109**: A09101. DOI:10.1029/2004JA010410.

Zhang J, et al. 2007. Solar and interplanetary sources of major geomagnetic storms (Dst ≤ -100 nT) during 1996-2005. *J Geophys Res: Space Phys* **112**: A10102. DOI:10.1029/2007JA012321.

Cite this article as: Scolini C, Messerotti M, Poedts S, Rodriguez L. 2018. Halo coronal mass ejections during Solar Cycle 24: reconstruction of the global scenario and geoeffectiveness. *J. Space Weather Space Clim.* **8**: A09