COMESEP Deliverable D4.2

Report of SOHO event data collection and statistical investigation [Task 4.1 and 4.2]

Project Acronym: COMESEP

Project Title: COronal Mass Ejections and Solar Energetic Particles: forecasting the space weather impact

Grant Agreement Number: 263252

Project co-funded by the European Commission, Seventh Framework Programme Funding Scheme: FP7-SPACE-2010-1

Start date of the Project: Feb 1, 2011

Project Duration: 3 years

Coordinator: Norma B. Crosby (BIRA-IASB)

Lead Beneficiary for this Deliverable: ROB (L. Rodriguez)

Editors: M. Dierckxsens, L. Rodriguez

Authors: M. Dierckxsens, A. Devos, S. Dalla, M. Dumbović, K. Leer, N. Lygeros, O. Malandraki, M. Marsh, I.-A. Patsou, L. Rodriguez, D. Sudar, K. Tziotziou, S. Vennestrom

Work-Package (WP)	WP 4
Task(s)	Task 4.1 and 4.2
Deliverable	D4.2
Due Date of Deliverable: Month	24

Issue Record				
Version	Date	Author(s)	Reason for Modification	Status
1	31-JAN-2013	M. Dierckxsens, A. Devos, S. Dalla, M. Dumbović, K. Leer, N. Lygeros, O. Malandraki, M. Marsh, I.-A. Patsou, L. Rodriguez, D. Sudar, K. Tziotziou, S. Vennestrom	Version 1	Submitted

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

TABLE OF CONTENTS

1. Introduction

The objective of WP4 is to study the space weather impact of solar events, like Coronal Mass Ejections (CMEs) and solar flares. In particular, the space weather impacts investigated are geomagnetic storms and Solar Energetic Particle (SEP) events. This deliverable comprises Task 4.1: SOHO events, collection of database and Task 4.2: SOHO events, statistical analysis.

The main goal of Task 4.1 is to create a database of events during the SOHO era. Several lists of events have been compiled for this purpose, including CMEs, Interplanetary CMEs (ICMEs), SEPs, flares, filaments and geomagnetic storms. The lists of events are produced by different COMESEP teams and compiled within a database hosted at ROB. This database provides the ability to link events from different lists between each other. This allows the user to establish a chain of events linking solar with interplanetary and geomagnetic phenomena in a straightforward manner.

Task 4.2 aims to answer the following basic question: What is the probability that a given solar event will produce a major space weather event? For this purpose, different types of statistical analyses have been applied to the data collected for Task 4.1. In order to find the statistical significance of a certain solar event to produce a large geomagnetic storm or an important SEP event, it is essential to link the solar activity to the interplanetary and geomagnetic consequences. In particular, it was studied which CME parameters (such as velocity, angular width, source location, intensity, etc.) and which flare parameters (e.g. intensity, duration, location) are important in producing an eventual geomagnetic or energetic particle effect.

The report is organized as follows. Task 4.1 is described in Section 2, providing details of the database and the configuration of every event list used in the study, including some basic statistics and examples. In section 3, the studies performed for Task 4.2 are explained, comprising both the results found for the CME and the SEP analyses.

2. Task 4.1: SOHO events: collection of database

2.1 Description

A comprehensive database of event data constructed for the SOHO era was developed. It includes information of all the key observable parameters related to the solar environment, as well as geomagnetic and SEP impact in the near-Earth environment. The database uses comprehensive catalogues of solar and interplanetary events (CMEs, ICMEs, solar flares, geomagnetic storms, SEPs) as a starting point and is thus complementary to, and builds on, previous works starting from large space weather events.

It is built using a modular PHP-framework named Symphony. The framework supports templates, object database mapping, routing, authentication and authorization. The database is available online via http://www.sidc.be/comesep

The user may select several lists at the same time and relate the corresponding events from one to the other (for example link a CME with an ICME) and then create his or her own lists. Advanced search is possible by combining any parameter from all the available lists. All the results from searches and linking can be downloaded.

2.2 Incorporated lists

2.2.1 CME (ROB)

CME data comes from the SOHO LASCO CME Catalog that contains all CMEs manually identified since 1996 that is available at http://cdaw.gsfc.nasa.gov/CME_list/

The columns of the table contain the following information:

- Column 1, 2: Date and time of first appearance in the LASCO/C2 field of view.
- Column 3: Central position angle (CPA) (degrees).
- Column 4: sky-plane width of CMEs (degrees).
- Column 5: Linear speed obtained by fitting a straight line (km/s).
- Column 6: Quadratic speed obtained by fitting a parabola (km/s).
- Column 7: Quadratic speed evaluated when the CME is at a height of 20 solar radii (km/s).
- Column 8, 9, 10: Acceleration (m/s^2), mass (g), kinetic energy (erg).
- Column 11: Position angle at which the height-time measurements are made (MPA for measurement position angle) (degrees).
- Column 12: Remarks regarding the number of data points and other limitations, as well as links to the halo CME alerts from the LASCO operator.

2.2.2 ICME (ROB)

The ICME events (from 1996) come from the list compiled by Richardson and Cane (2010) and that is available at http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm

The columns contain the following information:

- Column 1, 2: Date and time of the associated disturbance arrival at Earth.
- Column 3, 4: Start date and time of the associated ICME.
- Column 5, 6: End date and time of the associated ICME.
- Column 7, 8: Start and end times of the associated interval of abnormal solar wind composition/charge state.
- Column 9, 10: Start and end times of the associated magnetic cloud.
- Column 11, 12: Evidence of BiDirectional suprathermal Electron strahls and of Bidirectional energetic Ion Flows (BIF) in 0.5 -4.0 MeV.
- Column 13: The "quality" of the boundary times (‘1’ indicating the most reliable) based on assessment of the various data sets, including plasma, magnetic field and solar wind composition/charge states. 'W' indicates that the overall ICME signatures are particularly weak.
- Column 14: Increase in solar wind speed at the upstream disturbance (shock/wave) estimated from 1 hour averaged solar wind data (km/s).
- Column 15: Mean ICME speed (km/s).
- Column 16: Maximum ICME speed (km/s).

- Column 17: Mean magnetic field strength in the ICME (nT).
- Column 18: '2' indicates that a magnetic cloud (MC) has been reported (in the WIND MC list or by Huttunen et al. (2005)) in association with the ICME. 'H' indicates an event reported by Huttunen et al. (2005) that is not present in the WIND list. '1' indicates that the ICME shows evidence of a rotation in field direction, but lacks some other characteristics of a magnetic cloud, for example an enhanced magnetic field. '0' indicates that the ICME is not a reported magnetic cloud.
- Column 19: Minimum value of the Disturbance Storm Time (Dst) index during the ICME (nT).
- Column 20: Mean 1 AU transit speed of the disturbance based on the CME association in the next column (km/s).
- Column 21, 22: Date and time of a CME associated with the ICME.

2.2.3 CMEs with corresponding flares (HVAR)

CME data comes from the SOHO LASCO CME Catalog (see subsection 2.2.1 for details), within the time period 10.01.1996 – 30.06.2011 (a total of 16824 CMEs). Using an automatic method, CMEs were associated with solar flares using temporal and spatial criteria (for details see Vršnak et al., 2005), where solar flare key parameters were taken from the NOAA X-ray solar flare list (available at ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/ or see section 2.2.8). In the first step, only a temporal criterion was used, i.e. a certain CME was associated with all flares within 1 hour of the liftoff. The liftoff time was derived using mean 1st order (linear) CME speed from the LASCO catalog and extrapolating it back to the solar surface. This gave a total of 3709 CMEs associated with 4644 flares. Next, a spatial criterion was used:

$$|\psi CME - \psi F| < \phi/2$$

ψ_{CME} – position angle of the CME motion, taken from the SOHO LASCO CME Catalog

ψ_F – position angle of the flare, calculated from the flare source position given by the NOAA X-ray solar flare list. A simplified expression $\tan\psi = \sin\lambda / \tan\beta$ was used to calculate ψ from the central meridian distance λ and the heliographic latitude β (for details see Roša et al., 1998)

ϕ – CME angular width measured at the height beyond which it remains roughly constant, taken from the SOHO LASCO CME Catalog

This criterion was not used for HALO CMEs (due to an apparent width of 360^o) and for solar flares without source position, where only the temporal criterion was applied. Since not all non-halo CMEs could be associated with solar flares using both spatial and temporal criteria, the sample was reduced to 1392 CMEs and 1617 associated flares. Finally, in case multiple flares pass the above criteria, the one with the strongest intensity was chosen. This results in a list of 1392 pairs with a one-to-one correspondence between the CME and solar flare. All but 38 pairs are given a source position on the Sun, meaning they are front sided events. Since these 38 events without source position might still be relevant, they were not discarded from the list (e.g. using spatial criteria could lead in discarding a type X flare for which a source position was not given in the NOAA solar flare list, but could be obtained from other sources, see section 2.2.5).

The columns of the table contain the following information:

- Column 1: First C2 appearance, the date and time of CME first appearance in the LASCO/C2 field of view (FOV) (also serves as an ID for each CME).
- Column 2: CME start time, the CME liftoff time obtained by extrapolating the linear fit to the solar surface (height = 1 solar radius).
- Column 3: N, number of measurements in CME height/time plot.
- Column 4: MPA (deg), the CME Measurement Position Angle, the position angle at which the CME height-time measurements are made.
- Column 5: CPA (deg), the CME Central Position Angle (some CMEs move non-radially so the CPA and MPA do not coincide).
- Column 6: WIDTH (deg), the sky-plane width of CMEs (typically measured in the C2 FOV after the width becomes stable).
- Column 7: v (km/s), the linear speed obtained by linear (first-order polynomial) fit to the CME height-time measurements.
- Column 8: h_beg (Rsun), the position of the first measurement in CME height/time plot.
- Column 9: h_end (Rsun), the position of the last measurement in CME height/time plot.
- Column 10: <h> (Rsun), the mean value between h_beg and h_end.
- Column 11: feat_qual, 1st quality index (by observer's remark), from 0.5 (ill defined) to 3 (excellent).
- Column 12: QI, 2nd quality index (by observer's remark), from 0 (ill defined) to 5 (excellent).
- Column 13: Remark, observer's remark on the quality of observation/measurement, Ill-defined (Ill), poor, fair, typical (Typ), good, and excellent (Exc).
- Column 14: Halo, 1 if halo (width=360 deg), 2 if partial halo (width>120 deg), 0 if width<120.
- Column 15: Mass, estimation of CME mass.
- Column 16: Energy, estimation of CME kinetic energy obtained from linear speed and mass.
- Column 17: Flare start time, the time when the X-ray flux starts to rise.
- Column 18: Flare max time, the time when the X-ray flux value reaches maximum.
- Column 19: Flare end time, the time when the X-ray flux value returns to half the peak value.
- Column 20: FDC (deg), Flare Distance from Center, angular distance of the flare position from the center of the solar disc (0 (center) < FDC < 90 (limb)).
- Column 21: FPA (deg), Flare Position Angle, measured clockwise from the north (0 < FPA < 360).
- Column 22: r (Rsun), distance from the center of the solar disc in solar radii (0 < r < 1).
- Column 23: Source region, the position of the flare in latitude and longitude, N or S for north or south latitude, E or W for east or west central meridian distance.
- Column 24: X-ray class, the order of magnitude of the peak burst intensity (B,C,M or X).
- Column 25: X-ray intensity, a number from 1.0 to 9.9 that multiplies the X-ray class.
- Column 26: Peak (W/m^2), the magnitude of the peak burst intensity.
- Column 27: Integrated Flux (W/m^2), the integrated flux from event start to end (beginning with January 1997 data).

2.2.4 CMEs with corresponding flares and Dst (HVAR)

CME and solar flare data come from the list described in subsection 2.2.3. In total 211 CME-solar flare pairs were selected with speed v>400 km/s in a way that all possible speeds are well represented by the sample. This resulted in a sample with speed distribution unrealistically shifted towards greater speeds, as all fast CMEs (v>1500 km/s) were taken, whereas CMEs with v<1500 km/s were randomly selected from the sample. The reasoning for such sampling was twofold. First, there is evidence that the faster CMEs are more geoeffective (Gopalswamy et al., 2007), so by choosing faster CMEs the list should encompass also more interesting events (i.e. large storms). Second, due to the fact that they are fast it is likely that they "pick up" slower CMEs on their way, therefore technically our sample contains much more than 211 CMEs, with a "true" distribution of speeds resembling the one for the whole sample. For 20 selected CME-solar flare pairs in the list where the source position was not available in the NOAA solar flare list, the association was made with the solar flare positions from the HALO CME SOHO LASCO catalog (http://cdaw.gsfc.nasa.gov/CME_list/halo/halo.html). For slower CMEs we discarded events with less than three CME height/time measurements, due to a possible uncertainty of the speed, and events very close to the limb (flare distance from the center >80^{o}), due to the uncertainty whether or not they are front sided. This criterion was not applied to very fast CMEs, because they are rare and could consequently be lost from the analysis.

Using composite figures of CME kinematical (measurement) curves and the Dst index from the SOHO LASCO catalog, we associated Dst events with CME-solar flare pairs (Figure 1). Starting from fastest CMEs, an approximate extrapolation to 214 R_{SUN} (the distance from the Sun to Earth) was made to derive a proxy of possible arrival time at Earth (black circle in Figure 1). A Dst event was then sought in the time window of 6 hours before and 48 hours after the proxy of the arrival time to account for possible errors in the speed measurements, influence of the drag and geometry effects like an ICME hitting with a flank (marked with purple ellipse in Figure 1). For very slow CMEs (v=400-600 km/s) we changed the time window to 24 hours ahead of and 36 hours after the arrival time proxy due to the possible acceleration by the drag effect. If there was no Dst event within a time window, the Dst index was measured from the most relevant variation within the time window. The Dst timing is not a reliable parameter for these event, but nevertheless a measurement of the Dst level was obtained. If several Dsts were observed within a time window, an association with different CMEs (or groups of CMEs) was made taking into account the kinematical properties, start time of CME, influence of the drag and possible CME-CME interaction.

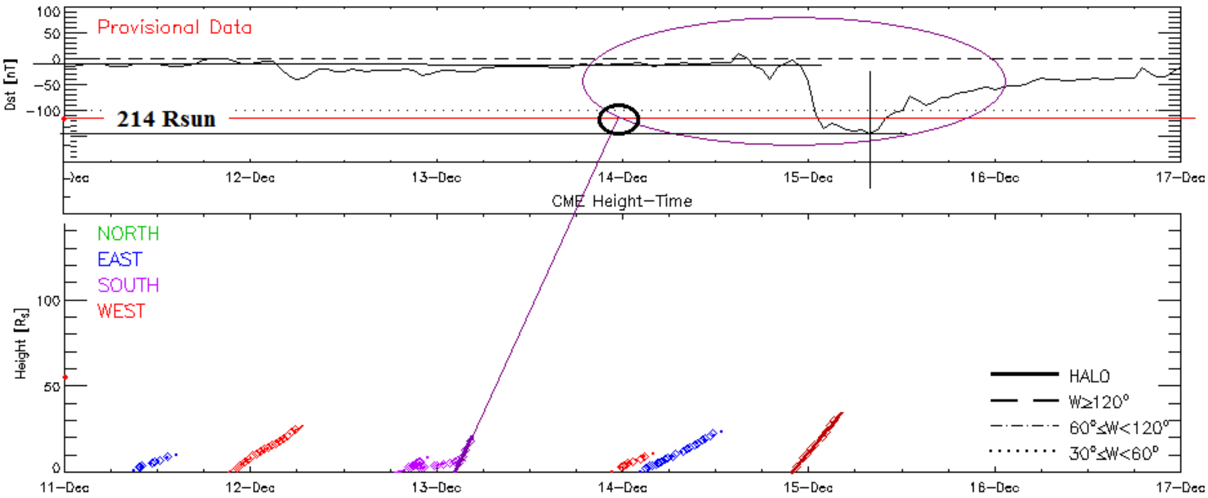


Figure 1: An example of a CME with no interaction between CMEs. Other CMEs in the figure are either not front sided or do not come from the close/neighboring region, therefore we denote this event as SOLO. The purple line represents the extrapolated kinematical curve to 214 Rsun and the small black circle marks the proxy of the arrival time. The time window for the Dst measurement is given by a purple ellipse and the Dst time and absolute minimum measurement are marked by black lines.

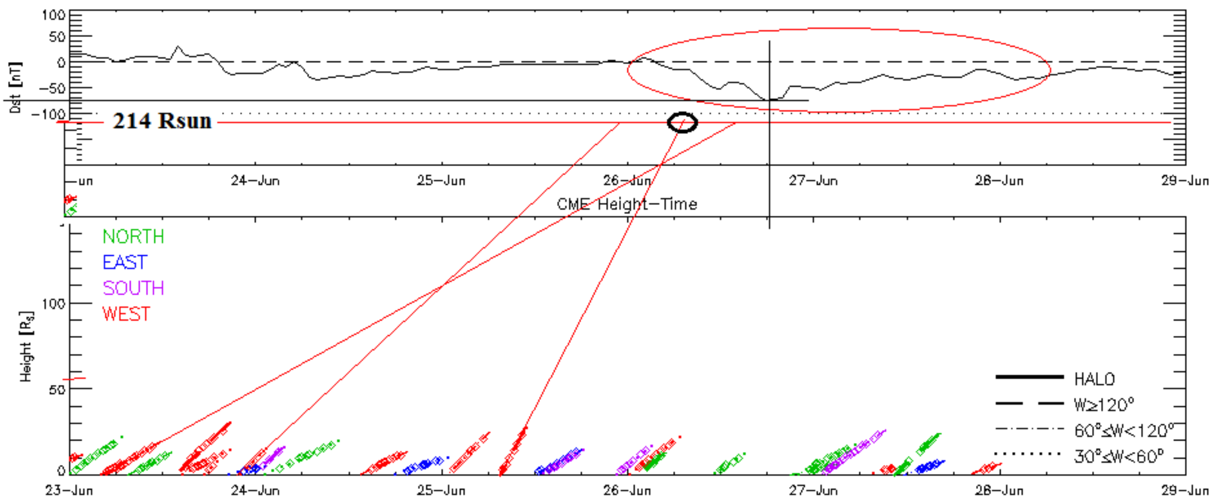


Figure 2: An example of a CME possibly interacting with other events. The three CMEs come from the same/neighboring region and from the extrapolated kinematical properties are likely to interact. However, due to the fact that the slowest one starts >2 days before the fastest one and that the two slower CMEs are narrow compared to the fastest one, it is somewhat questionable if they arrive at Earth together (e.g. it's possible they interact but we encounter only a glancing blow from the last ICME), therefore we denote them as TRAIN?. The small black circle marks the derived proxy of the arrival time, whereas a red ellipse marks the time window for the Dst measurement. The Dst measurement time and absolute minimum value are marked with black lines.

Using the complete CMEs with corresponding flares list as a representation of front sided events, we identified CMEs from the same/neighbouring region within a reasonable time window (~2 days) which might have interacted with the CME in question (i.e. their extrapolated kinematical curves cross or touch each other, marked red in Figure 2). The apparent CME width was also used as a subsidiary parameter, especially in case of a halo CME. It is reasonable to assume that the Earth directed halos are likely to pick up narrower CMEs on the way. This resulted in an association of Dst events with possible groups of interacting CMEs and also provided another interesting parameter for the analysis: the level of interaction. To be sure that the interaction indeed occurred, it has to be observed in the coronagraphic images. This is difficult even today with STEREO satellites observations, and almost impossible with the LASCO satellite observations (due to the limited point of view and small field of view). However, combining the kinematical properties, source region and width criteria with a Dst event observation, we obtained 4 levels of interaction: not interacting "SOLO (S) events", not likely interacting "SOLO? (S?) events", possibly interacting "TRAIN? (T?) events", and interacting "TRAIN (T) events". An example of a SOLO event can be seen in Figure 1, while a TRAIN? event is shown in Figure 2. Clearly this parameter bears some subjectivity of the observer. However, it was determined by a single observer and therefore possible errors are systematic. Furthermore, since it is a statistical sample, wrongly categorised events can be regarded as noise in the sample. As the key parameters for the interacting group of CMEs (TRAIN and TRAIN?) events, the parameters of the fastest CME were taken as relevant, assuming that the whole interacting group would be mostly influenced by the key parameters of the most prominent CME (i.e. the fastest one).

A detailed analysis of the *in situ* measurements was not done; however, we did try to associate *in situ* events with Dst events, to cross check how many of them are associated with ICMEs. It is important that there are no miss-associations of CMEs with Dst events caused by Corotating Interaction Regions (CIRs), at least for Dst events < -100 nT. Although CIR related geomagnetic storms with Dst<-100 nT are rare, there have been reports of such events (Zhang et al., 2003; Richardson et al., 2006). This also provided a way to determine if a CME actually arrived at Earth (so called false alarms). For that purpose, we used the list from Richardson and Cane (2010) and described in section 2.2.2, *in situ* data from the Advance Composition Explorer satellite (ACE; Stone et al. 1998a) Magnetometer (MAG; Smith et al., 1998) and Solar Wind Electron, Proton, and Alpha Monitor (SWEPAM; McComas et al., 1998) instruments (http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_MAG-SWEPAM.html), and *in situ* data from WIND satellite Magnetic Field Investigation (MFI; Lepping et al., 1995) and Solar Wind Experiment (SWE; Ogilvie et al., 1995) instruments (http://wind.gsfc.nasa.gov/mfi_swe_plot.php). A majority of the events (57%) were associated with ICMEs, whereas 41% of events were not associated with clear ICME signatures, and 2% could not be associated due to data gaps. Only one event not associated with an ICME had Dst < -100 nT, and looking at the *in situ* data we concluded that this is very likely a complex ejecta event. Therefore we did not discard it from the sample. Since all other not-associated events have Dst > -100 nT, we consider them as non-geoeffective or CMEs which missed the Earth. We note that some of them are in fact associated with CIRs, but for most of them the *in situ* signatures are not clear and would require a more detailed analysis. From the prediction point of view, it is only relevant that these ICMEs did not produce a geomagnetic storm with Dst < -100 nT, which is considered as the threshold for relevant geoeffectiveness.

The columns of the table contain the following information:

- Columns 1-27: as in the list described in section 2.2.3.
- Column 28: in-situ remark, the results of cross-checking with in-situ data and CME-ICME-Dst list from Richardson and Cane (2010).
- Column 29: Dst measurement time, time of measured Dst minimum.
- Column 30: Dst minimum (nT), minimum of Dst, measured from Dst = 0 nT.
- Column 31: Dst relative minimum (nT), minimum of Dst, measured from the Dst value at the onset.
- Column 32: CME SOLO (S)/TRAIN (T), the level of interaction (S, S?, T? or T).
- Column 33: comment, observer's remark on the measurements (especially on some uncertainty regarding the measurement).

2.2.5 CME-filaments association for events without flares (HVAR)

The CME data comes from the SOHO LASCO CME Catalog containing all CMEs manually identified since 1996 and described in section 2.2.2. Filament data are found here: ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/prominences-filaments/filaments/limb-disk-features_2000-2009/. Only four types of filaments were used: Disappearing Filament (DSF), Active Dark Filament (ADF), Eruptive Prominence on Limb (EPL), Active Prominence (APR). CMEs associated with flares were not considered (see section 2.2.3).

The procedure to select CME-filament pairs was similar as for the CME-flare associations (see section 2.2.3). The temporal window between the start of the CME and filament was limited to 0.08 days. After that, a spatial criterion was used:

$$|\psi_{CME} - \psi_F| < \phi/2$$

ψ_{CME} – position angle of the CME motion, taken from the SOHO LASCO CME Catalog

ψ_F – position angle of the filament, calculated from the filament position. A simplified expression $\tan \psi = \sin \lambda / \tan \beta$ was used to calculate ψ from the central meridian distance λ and the heliographic latitude β (for details see Roša et al., 1998)

ϕ – CME angular width measured at the height beyond which it remains roughly constant, taken from the SOHO LASCO CME Catalog

Finally, the list was purged by hand from duplicate entries (mostly due to the same filament being reported by different observatories). The list contains 336 events.

Columns description:

- Column 1 & 2: Start Date and Time of CME obtained by extrapolating the linear fit from height/time data to 1 solar radius.
- Column 3: JD, Start Date and Time of CME converted to Julian Days.
- Column 4: N, the number of measurements in the height/time plot.

- Column 5: fi (deg), the Measurement Position Angle, the position angle at which the height-time measurements are made.
- Column 6: w (deg), the sky-plane width of CMEs (typically measured in the C2 FOV after the width becomes stable).
- Column 7: vm (km/s), the linear speed obtained by a linear (first-order polynomial) fit to the height-time measurements.
- Column 8: h_f (Rsun), the position of the first measurement in the height/time plot.
- Column 9: h_l (Rsun), the position of the last measurement in the height/time plot.
- Column 10: JD_of_filament, the Julian Day of the filament.
- Column 11: FDC (deg), Filament Distance Centre, angular distance of the filament position from the centre of the solar disc (0 (centre) < FDC < 90 (limb)).
- Column 12: FPA (deg), Filament Position Angle, measured clockwise from the solar north pole (0 < FPA < 360).
- Column 13: fil_type, Type of the associated filament.

2.2.6 Geomagnetic storms (DTU)

A geomagnetic storm list for the SOHO era was created as basis for the statistical studies of Task 4.2. The list contains entries for all geomagnetic storms occurring during the time-interval 1996-2011. The ap index was chosen as the defining index for generating the list because it is a non-log version of the most widely used index Kp. This index also shows the closest resemblance to the aa-index used in the project as the starting point for the study of the very large historic storms (COMESEP Task 4.3, Vennerstrom, 2012).

The list contains in total 585 storms, of which 307 were minor storms (i.e. Kp<6). Figure 3 illustrates the intensity distribution of the storms, and the time of occurrence of the stronger storms with respect to the solar cycle.

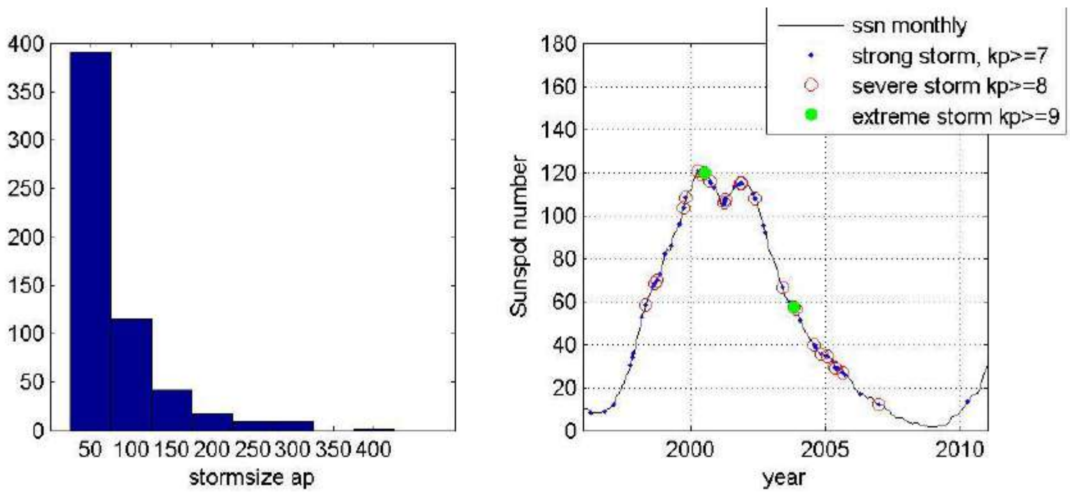


Figure 3: The distribution of the storm intensity (left), and the time of occurrence of the strong storms with respect to the solar cycle (right).

The defining entries for the list are thus the time and size of the peak ap of the storm. A number of other parameters describing each storm in considerable detail was collected, calculated and added to the list. The list includes the following parameters:

- Column 1: Storm number (in this list, chronologically listed)
- Column 2: Year (of ap peak) (YYYY)
- Column 3: Month (of ap peak) (MM)
- Column 4: Day (of ap peak) (DD)
- Column 5: t start (JD2000)
- Column 6: t peak in ap
- Column 7: t peak in Dst
- Column 8: Duration in ap (hours, 3h intervals)
- Column 9: ap peak
- Column 10: ap average
- Column 11: Dst peak
- Column 12: Number of local peaks in ap during storm
- Column 13: Auroral boundary index, lowest latitude value in CGM-coordinates
- Column 14: ULF pulsation index at ground, Tgr
- Column 15: AE peak value (nT)
- Column 16: AE average value (nT)
- Column 17: Peak of IMF magnitude Bpeak (nT) from OMNI data
- Column 18: Average IMF Bs (nT) from OMNI
- Column 19: Average solar wind speed v (km/s) from OMNI
- Column 20: Peak solar wind speed v (km/s) from OMNI
- Column 21: Richardson and Cane (2010) event list number (0 no id, negative uncertain id)

The storm period was defined using the same method used to generate the large historic storm list (Vennerstrom, 2012). A much lower threshold of ap=48 was used, corresponding to the threshold for a minor storm (Kp=5). The other parameters in the list related to the indices AE and Dst were then evaluated during this storm-period. While all three indices in general are elevated in storm periods there are exceptions to this rule. There is definitely no one-to-one correspondence between these different storm measures. Two examples of time series of the three indices during a storm are displayed in Figure 4.

Investigation of the role of multiple solar events on the geomagnetic impact is one of the objectives of Task 4.2. Therefore the number of local ap maxima during the storm interval was calculated. The number of local peaks was found to vary between 1 and 7. In total 29% of all the storms had more than 1 peak and 8% had more than 2 peaks. If only the strong storms (Kp>=7) were considered, these numbers increased to 59% and 25% respectively.

Two additional parameters were included in the list to further describe the character of the storm: the ULF pulsation index Tgr derived from ground based magnetometers (Kozyreva et al., 2007), and the auroral boundary index derived from the auroral particle precipitation measured by the DMSP satellites (Gussenhoven et al., 1981, 1982).

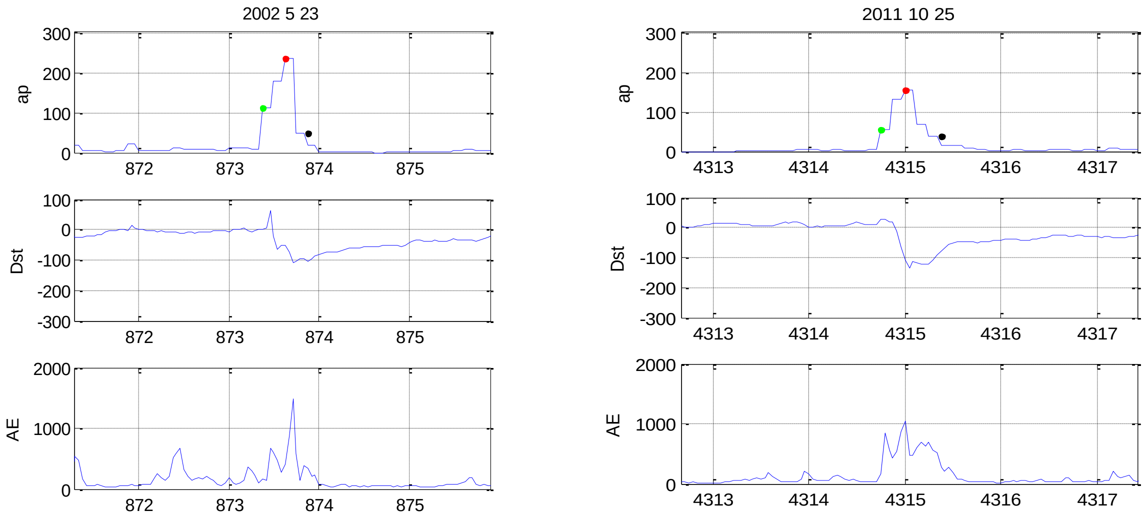


Figure 4: Two examples of the listed storms displayed as time series as a function of days since January 1st, 2000 (JD2000) in the three indices, ap, AE and Dst included as parameters in the list.

In order to link the geomagnetic storm occurrence to the interplanetary signatures, a set of interplanetary parameters during the storm was also included. These parameters were derived from the well-known "OMNI" data set derived by NGDC (http://omniweb.gsfc.nasa.gov/). The used data set consists of hourly averages of a collection of solar wind parameters measured by various satellites located in the solar wind near Earth. The chosen parameters were the interplanetary magnetic field (IMF) and the solar wind speed (v). Figure 5 shows the peak values of the solar wind speed (v) versus the peak value of the magnetic field intensity (Bpeak) during all the storms (black dots). It is seen that storms can be associated both with relatively weak IMF intensity and slow solar wind. However, this picture changes if the stronger events are singled out. The strong storms (Kp>=7) are shown as red crosses and the storms with Dst<-100nT are shown as blue circles. It is clear that all strong storms occur for relatively large IMF intensity. There is, however, a rather large fraction of moderate storms and minor storms with relatively large solar wind speeds. A large fraction of these are most likely associated with high-speed streams not driven by CMEs but rather associated with CIRs.

All the storms were also checked for possible association with the Richardson and Cane (2010) list of observed interplanetary ICMEs (see section 2.2.2), and the Richardson and Cane event list number was added as a parameter. Only 32% of the storms in the geomagnetic storm list were found to be associated with an ICME listed in the Richardson and Cane list. However, if only strong storms were considered, this number increased to 77%. Figure 6 illustrates the statistics of the association.

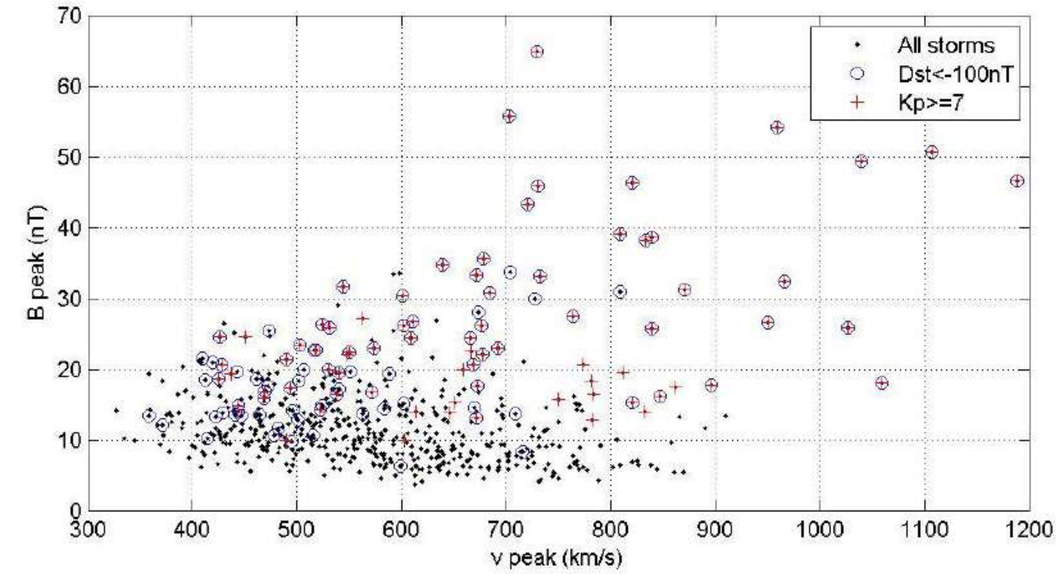


Figure 5: Solar wind parameters observed during the storm intervals in the storm list: the peak value of the magnetic field intensity (Bpeak) as a function of the solar wind speed (v).

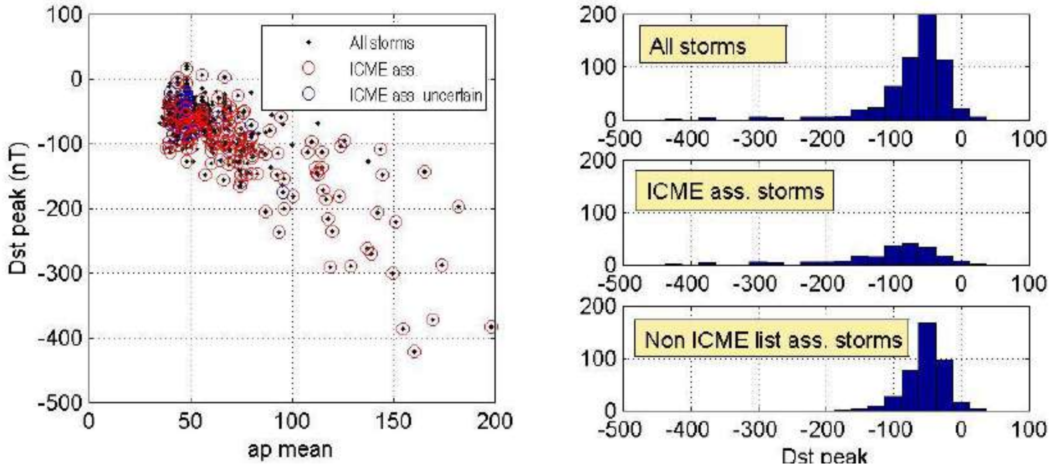


Figure 6: The peak Dst value versus the average ap during the storm interval for all the storms in the list (left). The storms which were found to be associated with an ICME were marked with a red circle or a blue circle (uncertain association). The distributions of the peak Dst value for all storms (top right), ICME associated storms (middle right), and storms not associated to ICMEs (bottom right).

2.2.7 SEP (BIRA-IASB/NOA/UClan)

The SEP list is based on the events at 1 AU compiled for Task 3.1 and fully exploits the software developed within this task. Details can be found in the Milestone 7 document (Crosby et al., 2012). The table contains the following columns:

- Column 1, 2: the COMESEP event and sub-event ID numbers.
- Column 3, 4: the start and end times of the (sub-)event.
- Column 5-14: the time of the peak flux in each SEPEM energy channel.
- Column 15-24: the peak flux value in each SEPEM energy channel (cm^{-2} s^{-1} sr^{-1} MeV^{-1}).
- Column 25-34: the fluence in each SEPEM energy channel (cm^{-2} sr^{-1} MeV^{-1}).
- Column 35-37: the fit parameters from a fit of a Weibull function to the energy spectrum of the flux at the time of the peak flux in the SEPEM reference energy channel.
- Column 38-40: the fit parameters from a fit of a Weibull function to the energy spectrum of the fluence.
- Column 41-43: the fit parameters from a fit of a Weibull function to the energy spectrum of the peak flux in each SEPEM energy channel.
- Column 44-52: the total fluence from the ACE/SIS data for the ions He, C, N, O, Ne, Mg, Si, S and Fe (cm^{-2} sr^{-1}).
- Column 53-56: the time of the first identified Energetic Storm Particle (ESP) during the event (if any), ESP onset time, peak flux value (cm^{-2} s^{-1} sr^{-1} MeV^{-1}) and total fluence (cm^{-2} sr^{-1}) from the ESP contribution.

A more detailed description and derivation of the values in these columns are given below. Figures will be shown for a few events as an example. However, all the plots made for each event can be obtained at ftp://ftp-ae.oma.be/D4.2/SEP_figures/.

2.2.7.1 Event and sub-event ID

The main COMESEP ID is based on the SEPEM reference event list (http://dev.sepem.oma.be/) and has the form of YYYYDDD, where YYYY is the year and DDD the day of year the event started. The sub-event id can be:

- = 0: corresponds to the entire SEPEM event and is always recorded
- > 0: the events are split up into sub-events in case there is more than one event in the list of Cane et al. (2010) during the SEPEM event period, or if the Cane et al. event starts after the SEPEM event. If there is only one sub-event, this is not recorded separately in the database as the derived parameters will be exactly the same as for the entire SEPEM event.

2.2.7.2 Start and end times

The start and end times for the entire SEPEM event (sub-id 0) are determined from the SEPEM reference event list. For sub-events, the start time of the next Cane et al. (2010) event determines the end time of the previous. The start time of the first and end time of the last sub-event correspond to the start and end time of the SEPEM event, respectively.

2.2.7.3 Peak flux times, peak flux values and fluences

The proton fluxes and total fluences are obtained from the SEPEM proton reference data, consisting of 10 energy channels. In each channel, the peak value is determined as the maximum flux before any identified ESP onset and the time at which this peak occurred. The fluence in each energy channel is calculated as the sum of the ESP subtracted flux from the start until the end of the event, multiplied by the sampling period of the data. More details about the ESP onset time and the subtraction of the contribution from the ESP enhancement are described in section 2.2.7.6.

2.2.7.4 Fit to energy spectra

Energy spectra are derived for the flux values at the time of the peak in the SEPEM reference channel (channel 2: 7.23 - 10.46 MeV), and the peak flux values and fluences in each channel as described above. The energy corresponding to each channel is taken as the square root of the product of the lower and upper limits of the channel. The spectra are fitted with a Weibull function which is shown to describe SEP spectra quite accurately up to high energies (Xapsos, 2000) and has the form:

$$\frac{d\phi}{dE} = f_0\, \kappa\, \alpha\, E^{\alpha-1} e^{-\kappa E^{\alpha}} ,$$

with E the energy of the particle measured in MeV/u, f0 a normalisation factor, and κ and α two parameters that determine the shape of the distribution. Examples of this fit can be seen in Figure 7 for the peak flux and fluence spectra.

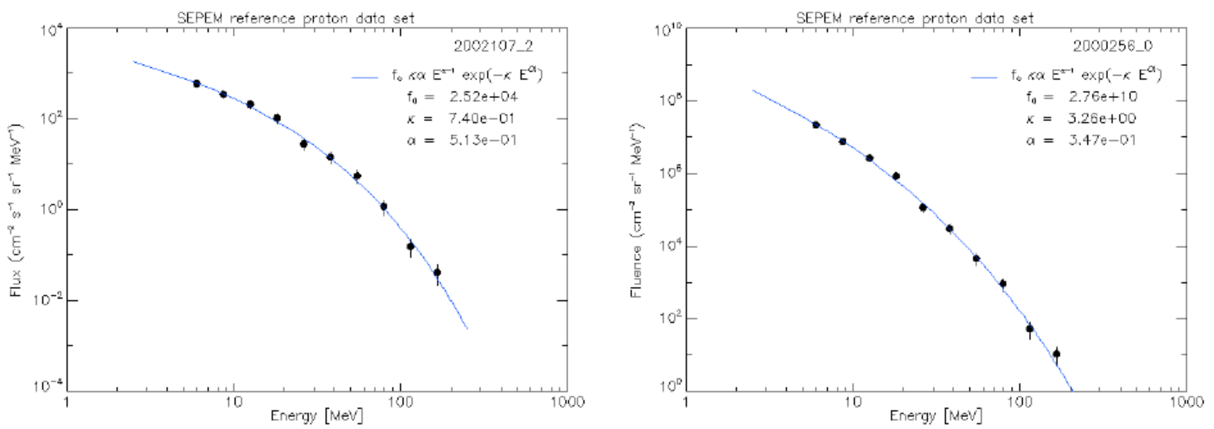


Figure 7: The energy spectrum of the SEP peak fluxes in each energy channel for sub-event 2002107_2 starting on April 21, 2002 (left) and of the fluence for event 2000256_0 starting on September 12, 2000 (right). The black dots show the flux measured in each energy channel, while the blue line represents the fit with a Weibull function. The resulting fit parameters are also shown.

2.2.7.5 Composition: fluences from ACE/SIS data

The hourly ion data of the SIS instrument on the ACE spacecraft (Stone et al. 1998b) is used to determine the total fluence during the event duration (SEPEM or sub-event) from the flux measurements for the following ions and energy ranges:

- Helium: 3.4 – 41.2 MeV/n
- Carbon: 6.1 – 76.3 MeV/n
- Nitrogen: 6.6 – 83.2 MeV/n
- Oxygen: 7.0 – 89.8 MeV/n
- Neon: 7.8 – 101.8 MeV/n
- Magnesium: 8.5 – 112.9 MeV/n
- Silicon: 9.0 – 123.2 MeV/n
- Sulphur: 9.5 – 132.9 MeV/n
- Iron: 10.5 – 167.7 MeV/n

The time profile of the flux from these ions in each energy channel can be seen in Figure 8 for one SEP event.

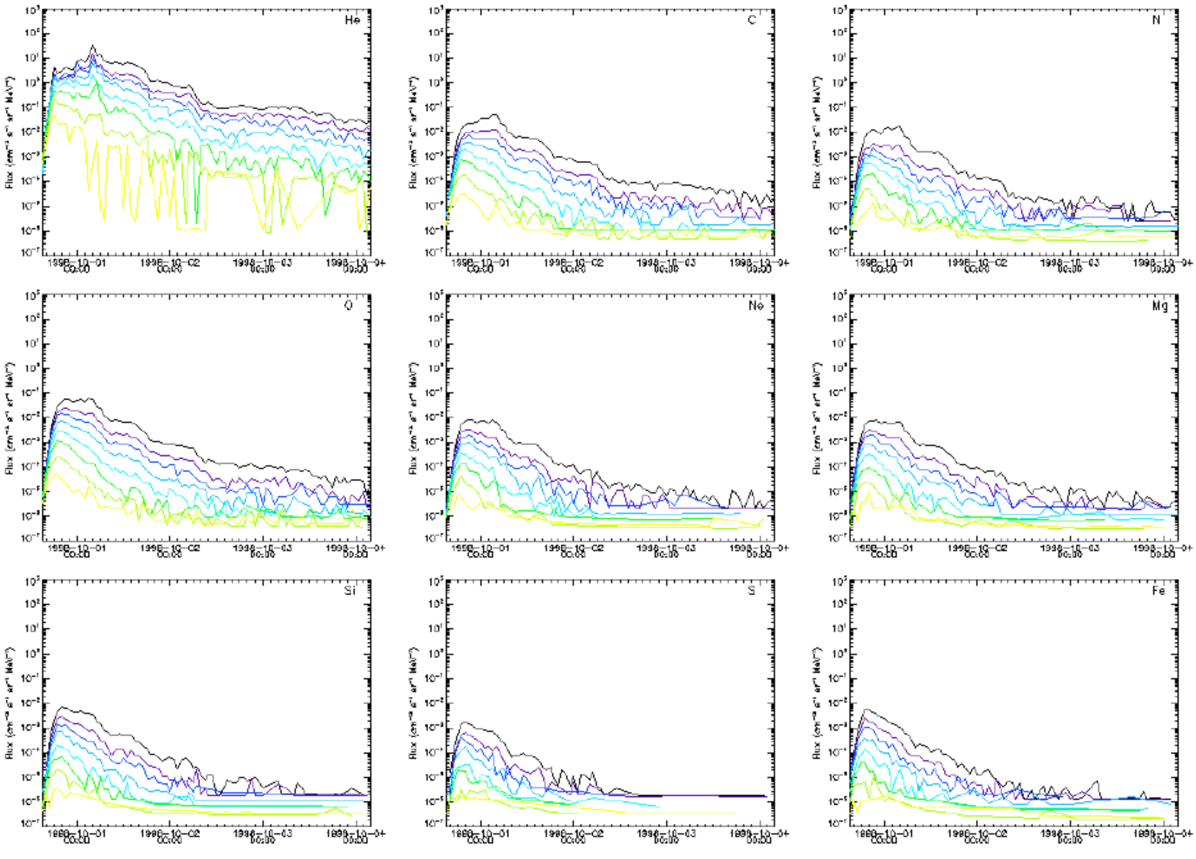


Figure 8: The time profile of the ion fluxes in all energy channels of the SIS instrument on board the ACE spacecraft during event 1998273_0 starting on September 30, 1998 for the elements He, C, N, O, Ne, Mg, Si, S and Fe.

2.2.7.6 ESP parameters

An ESP-like increase in the proton flux is an obvious additional intensity enhancement caused by a local interplanetary (IP) shock wave superposed on the decay phase of the main SEP event. For each event, the time profiles of the intensities in all energy channels were examined for ESP-like increases and recorded in a separate list. In parallel, lists of shocks observed at the ACE and WIND spacecraft (http://www.cfa.harvard.edu/shocks/) were used to verify whether a shock was observed at 1AU around the time of the ESP enhancement. Whenever such a local shock was associated with a particle flux increase, the time of each shock passage was recorded in the list, without applying any time shift to account for the distance between the ACE spacecraft and Earth (20-30 minutes). Accompanying plots of the intensity versus time were made, where the local shock passage time was marked with a vertical dotted line. In case a sub-event is selected, additional plots of the entire SEP event with all the local shock passages marked are provided for a comprehensive perspective. Two examples of events with an ESP enhancement are shown in Figure 9 and exhibit a good agreement between the time of shock passage and the period of the flux enhancement. In case more than one shock was observed during an SEP event, only the time of the first shock is recorded in the database.

The local shock associated with an ESP increase predominantly accelerates lower energetic particles more efficiently. Therefore, an ESP enhancement is usually much more apparent in the low energy channels (while the decay phase itself can be clearly noticed in the higher energy channels). The ESP onset refers to the time when the decay phase breaks and the intensity begins to rise instead of decreasing smoothly. Since it is not possible to predict when an IP shock passage will take place at 1 AU and whether it will affect the intensities of an SEP event that are observed locally, the ESP onset time cannot be predicted. This quantity has been derived by visually inspecting the intensity time profiles and determining the time when the lower energy channels exhibit an increase while being less pronounced in the higher energy channels.

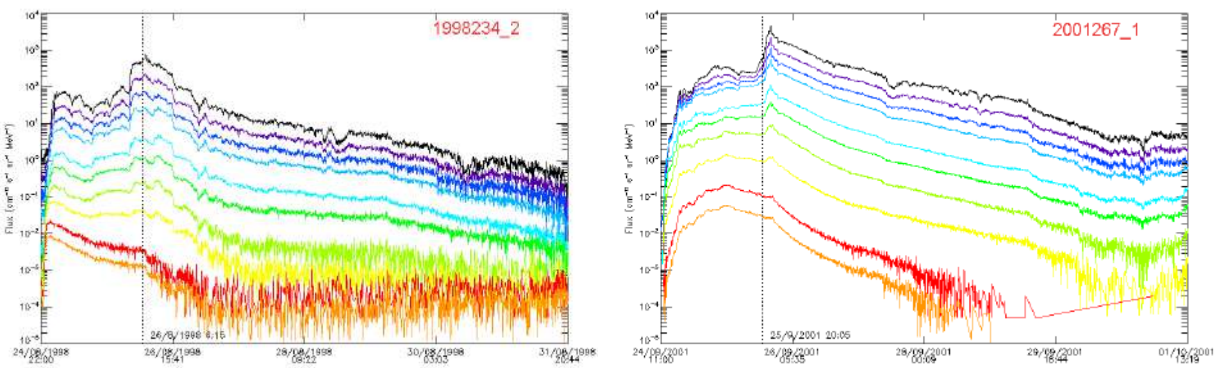


Figure 9: The time profile of the SEP flux in the SEPEM reference proton data of the sub-event 1998234_2, starting on August 8, 1998 (left) and the sub-event 2001267_1, starting on September 24, 2001 (right). The vertical dotted line indicates the time of the identified shock passing the ACE spacecraft and corresponds well to the time of the ESP-like enhancement.

The peak value in the SEPEM reference channel and the total fluence summed over all energy channels due to the ESP contribution only (i.e. SEP subtracted flux) are also recorded. In case there is only one sub-event, the ESP information will be recorded with the main SEPEM event (i.e. sub-event ID = 0). If there are multiple sub-events, the ESP parameters are only recorded for the sub-event that contains an identified ESP enhancement. However, the first occurring onset and shock time of these events are still added to the main SEPEM event to indicate there is at least one sub-event with an ESP.

In order to separate the contribution of the original SEP from the ESP enhancement, the decay phases of the SEP flux profile in each energy channel are modelled by a simple exponential decay superimposed on the background flux:

$$\frac{d\phi^i}{dt} = \phi_b^i + \phi_0^i \, e^{-(t-t_0)/\tau_i},$$

with ϕ_b^i representing the background flux, ϕ_0^i the flux at the start time t_0 of the event if the decay would have started from that time, and τ_i the decay time in the SEPEM energy channel i. The derivation of these parameters are described below and shown in Figure 10 for an event with clear separation between the ESP and ESP part and in Figure 11 for an event where the ESP enhancement starts right after the peak of the SEP part.

The criteria to build the SEPEM reference list (http://dev.sepem.oma.be/help/event_ref.html) required at least 24 hours between the end of the previous and the beginning of the next event (dwell time). The background flux ϕ_b has been derived by taking the average of the observed flux between 14 and 10 hours before the start of the main SEPEM event in each energy channel.

To determine the normalization of the decay phase of the SEP event, the flux is fixed to a certain value at the normalization time, a specific time before the ESP onset time. The pre-onset time is defined as the time difference between the time of the maximum (peak) flux before the onset time and the onset time itself. The normalization flux is calculated as the average of the flux during the period between 0.2 and 0.4 times the pre-onset time after the peak time. The normalization time has been set in the middle of this range, i.e. 0.3 times the pre-onset time after the peak. If the maximum before the onset time coincides with the onset time itself, then the flux at this time is taken as the normalization flux. The normalization flux was required to be at least the background level. When the background flux and decay time are determined, the flux ϕ_0 at the start time of the event can then be easily calculated from the formula above. The obtained normalization fluxes and times for each energy channel are indicated as diamonds in Figure 10 and Figure 11.

The decay time is determined by varying the parameter and visually observing which value best represents the decay phase of the flux after the first peak. To facilitate this process, the time profile of the flux is smoothed by convoluting it with a Gaussian function with a width of 5 sampling periods (25 minutes). However, for the lower energy channels of certain events, it is often impossible to derive the decay time, as the peak of the SEP contribution is followed very closely by the initial rise of the ESP enhancement and the start of an exponential decay is not observed (see Figure **11**). Further, it would be very impractical to visually determine all 10 decay times. Since the source of the SEP flux in the different channels is the same, there should be some correlation between the decay times in the different energy channels. This correlation is assumed to be energy dependent:

$$\tau_i = \tau_{10}\left(\frac{E_i}{E_{10}}\right)^{-\beta},$$

with E_i being the energy as derived for the spectra and β the exponent determining the energy dependence of the decay time τ_i in each energy channel i. The subscript 10 refers to the last and highest SEPEM energy channel (138.3–200.0 MeV). The decay time corresponding to this energy channel is determined first, as for most cases there are very little effects to be noticed due to the ESP enhancement. Then the exponent β is varied so the decay is well modelled in the medium energy channels. The assumption is that this energy dependence holds for the lowest energy channels as well. The resulting decay times vary from 0.13 days to 0.85 days, with an average of 7 hours, while the exponent only took values of 0.10, 0.15 or 0.20 with an average of 0.16 for all events that had an identified ESP. The decay flux can be seen superimposed on the smoothed flux in Figure 10 and Figure 11.

The SEP-only and ESP-only fluxes are constructed from the original (non-smoothed) SEP event flux using the results obtained from modelling the decay phase of the event. The SEP-only flux is identical to the original event flux up to the normalization time. From that point onwards, the SEP-only flux coincides with the derived decay flux. The ESP-only flux starts at the normalization time and is determined by subtracting the decay flux from the original event flux. Note that this flux is allowed to be negative, and should be more regarded as a change of flux, rather than an absolute flux. Both the SEP-only and the ESP-only fluxes are shown in Figure 10 and Figure 11.

The method described above might result in large uncertainties in the decay time for the lowest few energy channels, while the method works well at higher energies. The separation of the two contributions is used to derive the SEP fluence in each energy channel, and the ESP peak value and total ESP fluence. From the provided examples in Figure 10 and Figure 11, it can be clearly seen that even large uncertainties in the decay time for the lowest energy channels will have a small effect on the derived quantities as the decay phase only contributes a relatively small fraction to these quantities. However, care should be taken if this description of the decay phase of the SEP event is used for other purposes. The same approach should be applied to events without ESP enhancements to further verify and refine this method.

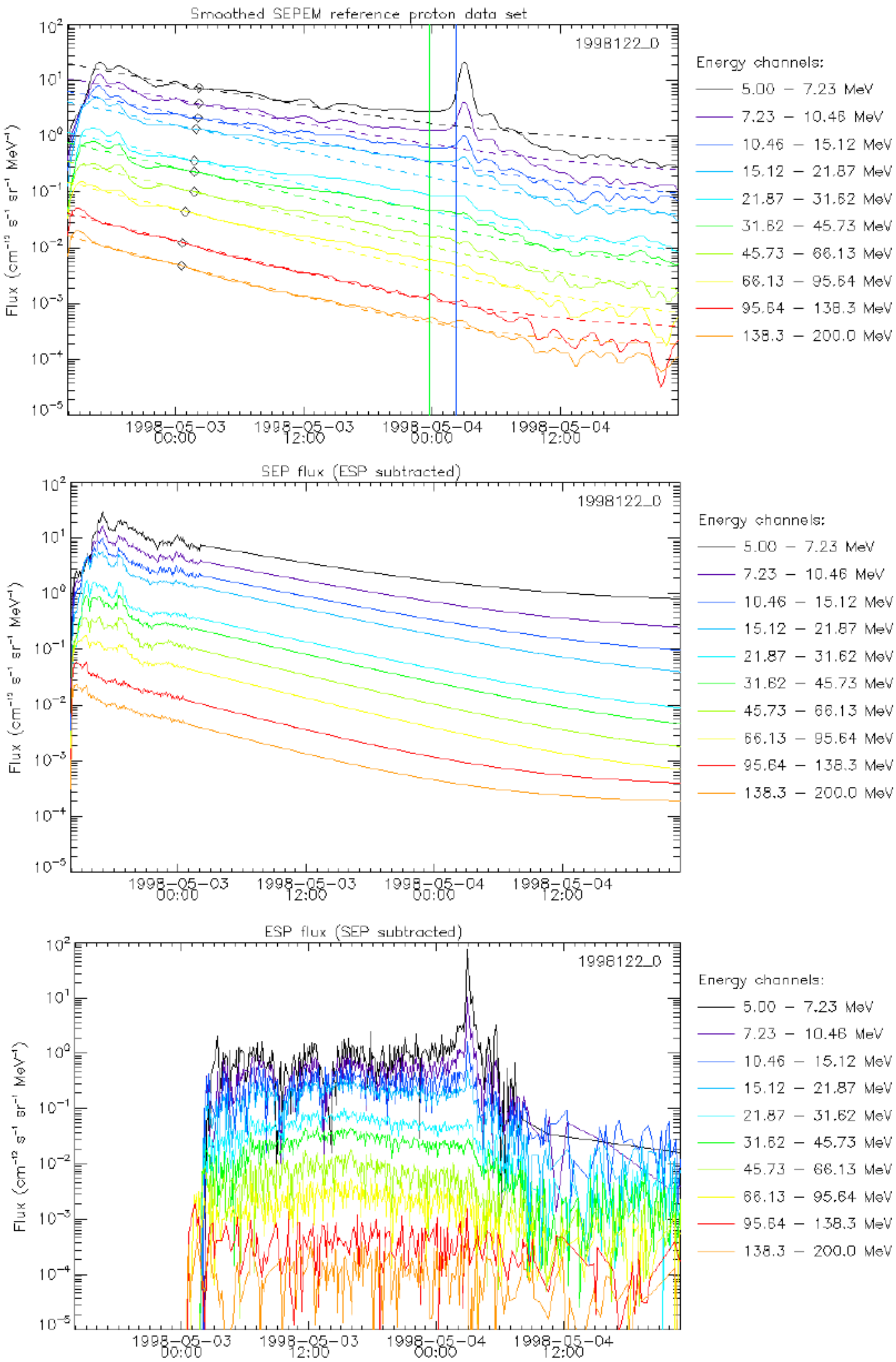


Figure 10: The separation of the SEP flux from the ESP enhancement for event 1998122_0 starting on March 2, 1998. Top: The smoothed time profiles of the SEP event flux (solid lines) compared to the derived decay flux (dashed lines) in the SEPEM energy channels. The blue vertical line represents the time of the identified shock passage at L1, while the green vertical line is the ESP onset time, and the diamonds indicate the normalization fluxes and times. Middle: the ESP subtracted SEP flux. Bottom: The SEP subtracted ESP flux. Note that the latter is allowed to have negative values.

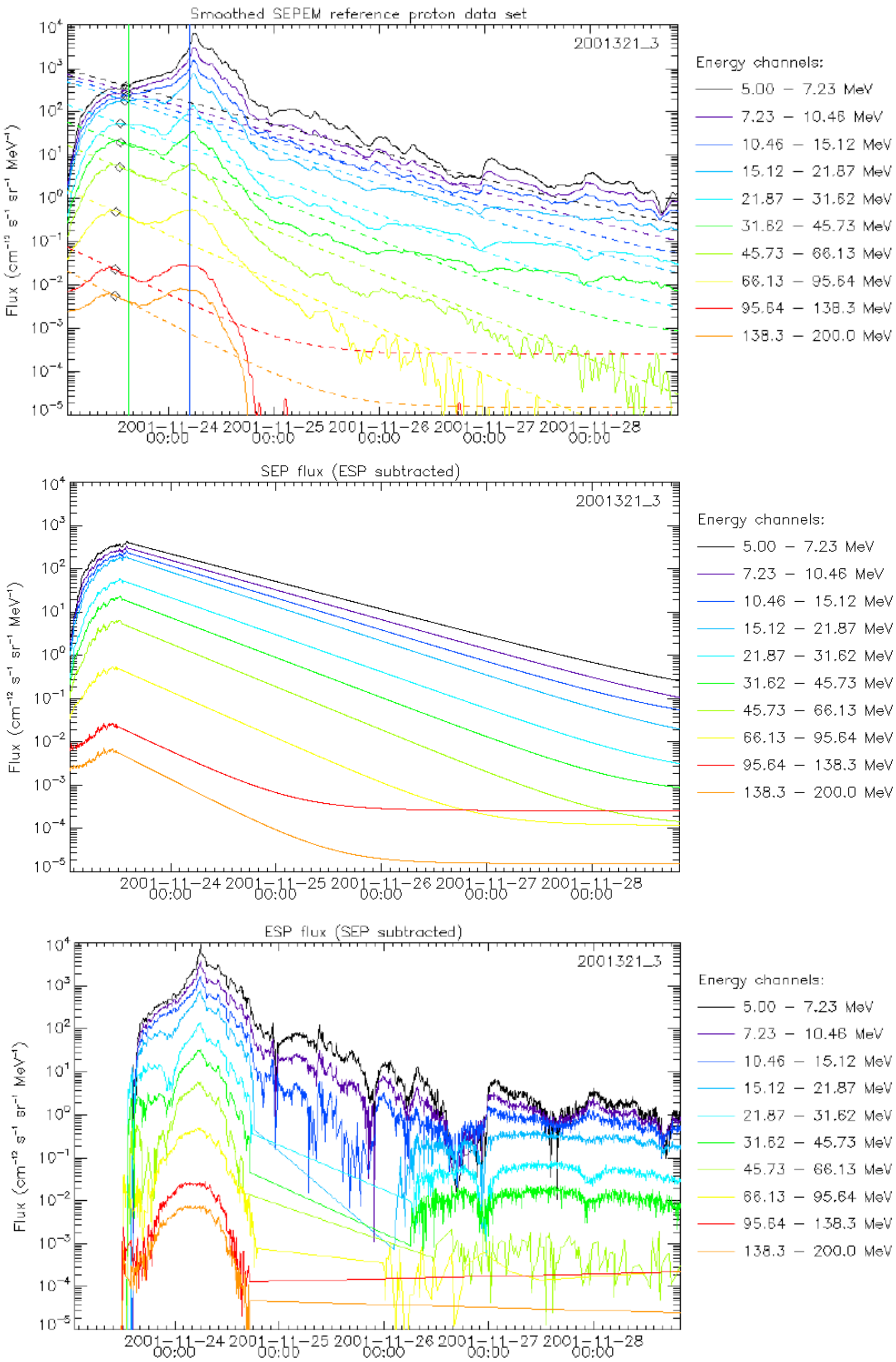


Figure 11: The separation of the SEP flux from the ESP enhancement for event 2001321_3 starting on November 24, 2001. Top: The smoothed time profiles of the SEP event flux (solid lines) compared to the derived decay flux (dashed lines) in the SEPEM energy channels. The blue vertical line represents the time of the identified shock passage at L1, while the green vertical line is the ESP onset time, and the diamonds indicate the normalization fluxes and times. Middle: the ESP subtracted SEP flux. Bottom: The SEP subtracted ESP flux. Note that the latter is allowed to have negative values.

2.2.8 GOES flares (ROB)

Flare data comes from NOAA and is available at http://www.swpc.noaa.gov/ftpmenu/warehouse/. The columns of the table contain the following information:

- Column 1: Flare date.
- Column 2: NOAA event number.
- Column 3, 4, 5: Begin, peak and end time, respectively. If the peak or end time is less than the begin time, then they correspond to the next UTC day. A single letter can precede these times: A=after, B=before, U=uncertain. For example, the begin time A0146 means the event began after 0146. The begin time of an x-ray event is defined as the first minute, in a sequence of 4 minutes, of steep monotonic increase in 0.1-0.8 nm flux. The maximum is taken as the minute when the peak occurs. The end time reached when the flux level decays to a point halfway between the maximum flux and the pre-flare background level.
- Column 6: X-ray Class.
- Column 7: NOAA solar region number from where the flare originates.
- Column 8, 9: Longitude and latitude from the flare origin (degrees).
- Column 10: Importance and brightness. Importance is the corrected flare area in heliospheric square degrees (sq. deg.) at maximum brightness, observed in the H-alpha line (656.3 nm), with the following classification: S - Subflare (area ≤ 2.0 sq. deg.), 1 (2.1 ≤ area ≤ 5.1 sq. deg.), 2 (5.2 ≤ area ≤ 12.4 sq. deg.), 3 (12.5 ≤ area ≤ 24.7 sq. deg.), 4 (area ≥ 24.8 sq. deg.). Brightness is the relative maximum flare brightness in H-alpha: F – faint, N – normal, B - brilliant.
- Column 11: Flare Characteristics:
  VWL = Visible in white light
  UMB = Greater than or equal to 20 percent umbral coverage
  PRB = Parallel ribbon
  LPS = Associated Loop Prominence (LPS)
  YSR = Y-shaped ribbon
  ERU = Several eruptive centers
  BPT = One or more brilliant points
  HSS = Associated high speed dark or bright surge
  DSD = Dark surge on the disk
  DSF = Flare followed the disappearance of a solar filament in the same region
  BLU = H-alpha emission greater in the blue wing than in the red wing.
- Column 12: Radio burst intensity on a relative scale: 1 (Minor), 2 (Significant), 3 (Major), and the type:
  Type II: Slow drift burst
  Type III: Fast drift burst
  Type IV: Broadband smooth continuum burst
  Type V: Brief continuum burst, generally associated with Type III bursts
  Type VI: Series of Type III bursts over a period of 10 minutes or more, with no period longer than 30 minutes without activity
  Type VII: Series of Type III and Type V bursts over a period of 10 minutes or more, with no period longer than 30 minutes without activity
  Type CTM: Broadband, long-lived, decametric continuum

3. Task 4.2: SOHO events: statistical analysis

The collected database described in section 2 is used to perform basic statistical analyses to quantify the relationship between the solar event parameters and their impact. The relevant question to be addressed is: what is the probability that a given solar event will produce a major space weather event? In section 3.1, the analyses related to CMEs are presented, while section 3.2 covers the SEPs. The output of these studies will be used as input for the COMESEP alert system.

3.1 Coronal Mass Ejections

3.1.1 Statistical analysis of CME-flare-Dst connection (HVAR)

The aim of this study is to analyze which CME and solar flare parameters based on remote sensing data can be used for predictions of major geomagnetic storms. Following the results of previous studies (e.g. Zhang et al., 2003; Srivastava & Venkatakrishnan, 2004; Gopalswamy et al., 2007; Zhang et al., 2007; Richardson & Cane, 2010; Richardson & Cane, 2011) we focus on CME speed and width, as well as solar flare intensity (or more precise solar flare type) and source region, as key parameters. In addition, the level of interaction was also analyzed. There have been indications that interaction of CMEs can influence their geoeffectivness (e.g. Farrugia & Berdichevsky, 2004) and that the most intensive storms are associated with observations of successive CMEs (Gopalswamy et al., 2007) and multiple ICMEs (Zhang et al., 2007). Besides mentioned parameters, the quality index was also considered as a rough approximation of brightness. However, no connection to geomagnetic storms was found. Finally, our sample of geomagnetic storms has shown to some extent a seasonal variation, as would be expected based on previous studies on this matter (e.g. Cliver & Crooker, 1993), and is presented in section 3.1.3.

The analysis is based on the table described in section 2.2.4. Distributions are used as a statistical tool for analysis with the following bins: $|Dst| < 100$ nT, 100 nT $< |Dst| <$ 200 nT, 200 nT $< |Dst| <$ 300 nT, and $|Dst| >$ 300 nT, $|Dst|$ representing the absolute value of the Dst index. This way we focus only on major geomagnetic storms (as explained in subsection 2.2.4) where obtained results can be used further as an empirical mean to predict the probability of major space weather events. One might argue that a more analytical approach would be more useful, however, due to our current knowledge and observation possibilities, the results of such an approach would be at least as imprecise as deriving probabilities. Furthermore, we found no significant correlations between CME/flare parameters and Dst, although in general, dependencies are visible (similar to e.g. Kim et al., 2011, but in contradiction to e.g. Srivastava & Venkatakrishnan, 2004).

In order to check how the Dst distributions are changing for a certain key parameter, the key parameters were binned as well. For some key parameters the binning was straightforward (e.g. interaction level and flare type) as they are already discrete parameters, but for continuous parameters such as CME speed and source position distance from the center, r, the binning was done in a way that all the bins have approximately the same number of events. The mean, skewness and kurtosis of the distributions were studied as relevant distribution parameters depicting the behavior of the Dst index with the change in the (discrete) CME/flare key parameter. The skewness and kurtosis are the 3rd and 4th order moments of the distribution and are a measure for the asymmetry and peakedness/flatness, respectively.

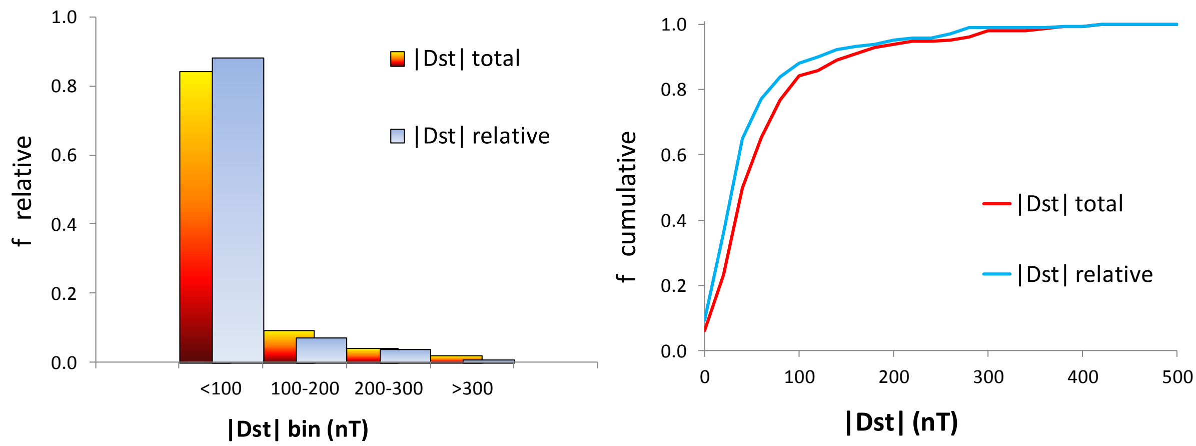


Figure 12: Relative (left) and cumulative (right) frequencies for absolute value of Dst index at the maximum intensity of the geomagnetic storm.

The statistical significance of the results was tested using two-sample t-test (2stt) at the 0.05 level (95% significance) for assuming dependence (equal variance assumed) and independence (equal variance not assumed) of the test samples. Due to the fact that 2stt is based on the normality assumption, i.e. requires certain sample sizes, nonparametric significance tests were also made, namely the Kolmogorov-Smirnov and Mann-Whitney U tests, but there were no notable changes. Therefore we present only 2stt results.

A geomagnetic strom is seen in the Dst-time plot as a decrease in the Dst index, where the intensity of the storm is given by the magnitude of this decrease (see Figure 4). The magnitude of the decrease in Dst index values was measured in two ways: the total magnitude (|Dst| total), measured from reference value 0, and relative magnitude (|Dst| relative), measured from the index value at the start of the storm. We examined how the measurement process can affect the results, i.e. is there a difference in looking at the total or relative magnitude. The statistical analysis shows that the distributions of |Dst| total and |Dst| relative are somewhat different, with |Dst| relative shifted to lower values (Figure 12). The mean values are 68 nT and 53 nT, respectively and are found to be significantly different at the 0.05 significance level with a two-sample t-test. This shows that true storm effects are somewhat smaller than those measured by the Dst stations, because effects such as pre-event variations and previous events are not subtracted. Therefore, we focus our study on |Dst| relative, as a more realistic measure of storm strength. It should be noted though that the same analysis was repeated for |Dst| total as well, with similar results. The distributions were systematically shifted towards somewhat greater values, however, with the same behaviour. Therefore we do not present them here.

3.1.1.1 CME Speed

The events were optimally categorized into different CME speed bins (linear speed, see subsection 2.2.4) so that all bins have approximately the same number of events, with samples big enough so that the results can be taken with maximal significancy. This resulted in 6 speed bins, with CME speed ranges: 400-600 km/s, 600-800 km/s, 800-1000 km/s, 1000-1200 km/s, 1200-1700 km/s, and v

> 1700 km/s. The number of events in each bin is 36, 34, 35, 35, 41, and 30, respectively. For events involved in (possible) interactions, the speed of the fastest CME was taken as relevant. For each CME speed bin, a |Dst| distribution was made using previously mentioned |Dst| binning. This resulted in 6 |Dst| distributions for 6 different CME speed ranges (Figure 13). The mean, skewness and kurtosis were calculated for each distribution to quantitatively examine the changes in the |Dst| distribution with different CME speed ranges (Figure 14). Furthermore to test the differences between |Dst| distributions (i.e. whether they represent statistically different samples) a two sample t-test was made between each pair of |Dst| distributions. The results are presented in Table 1.

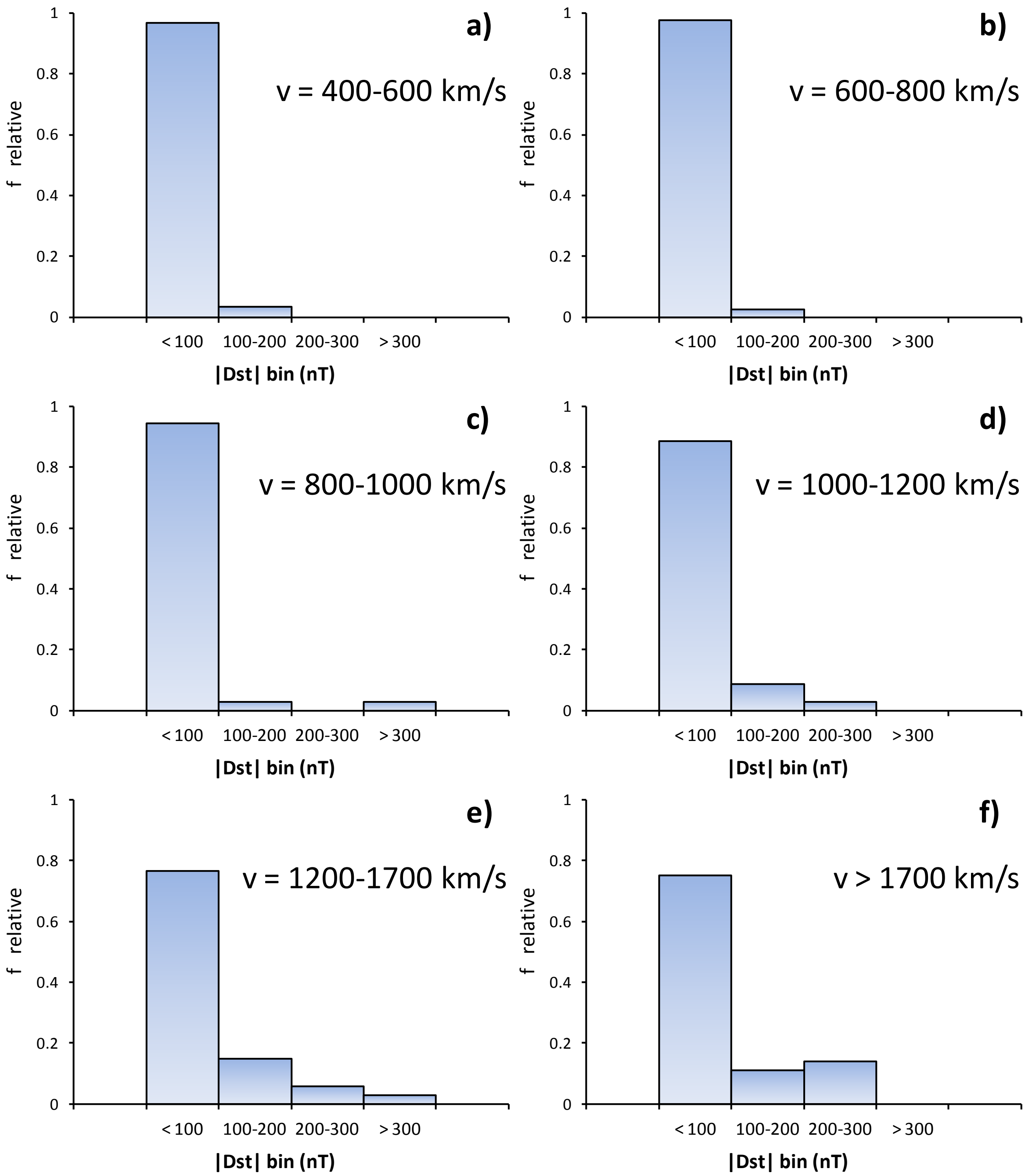


Figure 13: The |Dst| distributions normalized to one for different CME speed bins as shown in the figure.

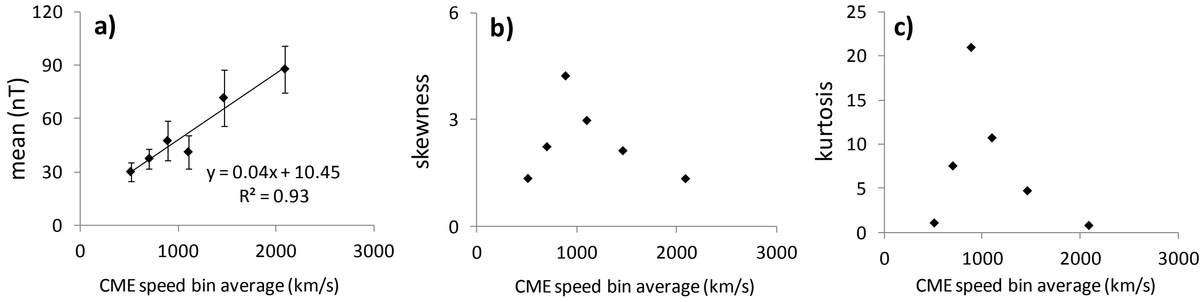


Figure 14: The |Dst| distribution mean (left), skewness (middle) and kurtosis (right) as a function of the average speed within a corresponding CME speed bin. Error bars in a) represent confidence intervals and the fit result and correlation coefficient R^2 are also shown.

It can be seen in Figure 13 a & b that the distribution is restricted to |Dst|<200 nT and is mostly contained within |Dst|<100 nT. Compared to Figure 13c these distributions are more symmetric, due to the fact that for the speed range v = 800-1000 km/s the distribution has a longer tail. In other words, as the speed range increases from lower speed bins (400-600 km/s) to middle speed bins (800-1000 km/s) the distribution becomes more asymmetric and most of the data are contained in the peak and the tail of the distribution (i.e. "the shoulders" of the distributions are missing). This is also seen in the behavior of distribution parameters, as the value of the mean becomes larger, as do skewness and kurtosis (Figure 14a). On the other hand, as the speed range increases from middle speed bins (800-1000 km/s) to upper speed bins (>1700 km/s) the distribution starts to lose in asymmetry and peakedness, as the values of skewness and kurtosis are decreasing, but the value of the distribution mean is still increasing. Increasing the CME speed from low to medium values leads to a change in the corresponding |Dst| distribution from a relatively symmetrical shape, grouped in the first two |Dst| bins, to a distribution with a tail, becoming very asymmetric. Going to even higher speeds, the tail starts to be filled in and the distribution loses its asymmetry. This shift of distribution towards larger |Dst| bins is also evident from the behavior of the distribution mean (Figure 14a), as it is shifted towards greater values of |Dst|. This behavior can be approximated with a linear function (R=0.96), although due to the small number of points this approximation should be taken more as an illustration.

2stt	>1700	1200-1700	1000-1200	800-1000	600-800	400-600	
>1700		0.43408	0.00509	0.02173	4.22E-04	3.20E-04	equal variance assumed
1200-1700	0.43632		0.09901	0.2126	0.03232	0.02142	
1000-1200	0.00503	0.10264		0.6621	0.72054	0.31743	
800-1000	0.02173	0.21545	0.66214		0.3981	0.1796	
600-800	9.34E-04	0.04855	0.72938	0.41961		0.34957	
400-600	1.77E-04	0.01714	0.29693	0.15894	0.33367		
equal variance NOT assumed							

Table 1: Two sample t-test for the |Dst| distribution mean between different CME speed bins (in km/s) with equal variance assumed (upper right values) and equal variance not assumed (lower left values). In the table, the values are given for the probabilities that the two sample means are different. Values which are not significant (i.e. the mean between two distributions is not significantly different) are marked red. The test value is 0.05.

This consequently means that faster CMEs have larger probabilities to produce strong geomagnetic storms and furthermore that slow CMEs (v < 600 km/s) cannot produce intense storms (|Dst|>200nT) unless they are involved in CME-CME interaction. The latter comes from the fact that interacting CMEs in the sample are related to the CME speed of the fastest CME in the train. It should be noted though that these distributions cannot be used to realistically quantify these probabilities because of other parameters that influence the |Dst| distribution (as will be seen in the following sections). Two sample t-test analysis reveals that there is no significant difference between two neighboring speed bins (or several, as we go to lower speed bins) indicating that the CME speed-|Dst| connection should be most prominent for very fast CMEs, whereas for slow CMEs it is "smeared" due to some (yet) unknown factors.

3.1.1.2 CME/flare source position

Several aspects of the source position were analyzed. First, the events were categorized by quadrant where the CME/flare source position is found: northeast (NE), northwest (NW), southeast (SE), and southwest (SW). However, no trend whatsoever was observed. It was confirmed by a two sample t-test that there is no difference in the samples coming from different quadrants. Next, it was investigated whether there is an asymmetry regarding the north/south and west/east CME/flare source position with the same result. Although a small difference in the |Dst| distributions was observed between the west and east solar regions, the differences were not confirmed with a two sample t-test.

Finally, the distance from the solar disc center (R) was investigated as a key parameter ranging from 0 to 1 (in units of solar radii, Rsun). Similarly as with CME speeds, the events were optimally categorized into different bins, resulting in 4 bins with R ranges: R < 0.4 Rsun, 0.4 – 0.6 Rsun, 0.6-0.8 Rsun, and R > 0.8 Rsun. The number of events in each bin is 45, 53, 53, and 60, respectively. For events involved in a (possible) interaction, the source region of the fastest CME was taken as relevant. A |Dst| distribution was made for each range of R, using the same |Dst| binning as with the CME speed analysis. This resulted in 4 |Dst| distributions for 4 different R ranges of CME/flare source region (Figure 15), with distribution parameters (mean, skewness and kurtosis) presented in Figure 16. Similar to the speed analysis desribed in section 3.1.1.1, a two sample t-test was made between each pair of |Dst| distributions, and the results are presented in Table 2.

It can be seen from Figure 15 that the distribution loses the tail and asymmetry as the distance from the solar disc center increases. For 0.6 Rsun < R < 0.8 Rsun, R is restricted to |Dst | < 200 nT. This is somewhat expected and in agreement with numerous previous studies, where CMEs closer to the center of the disc are more geoeffective (e.g. Zhang et al., 2003; Srivastava & Venkatakrishnan, 2004; Gopalswamy et al., 2007; Richardson & Cane, 2010). However, the distribution again gains in the tail and asymmetry for the near limb events, showing that limb CMEs can also be highly geoeffective, as pointed out in e.g. Schwenn et al. (2005) and Cid et al. (2012). The two sample t-test results show significant differences only between the bins around the solar disc center (R<0.4Rsun) and the other bins. However there is a decrease in the significance of the difference as we go towards near limb source positions (Table 2). We can see in Figure 16a how the distribution mean is decreasing with distance from the disc center following approximately a power law (although this is just illustrative, due to the small number of points), while other distribution parameters show no regularities.

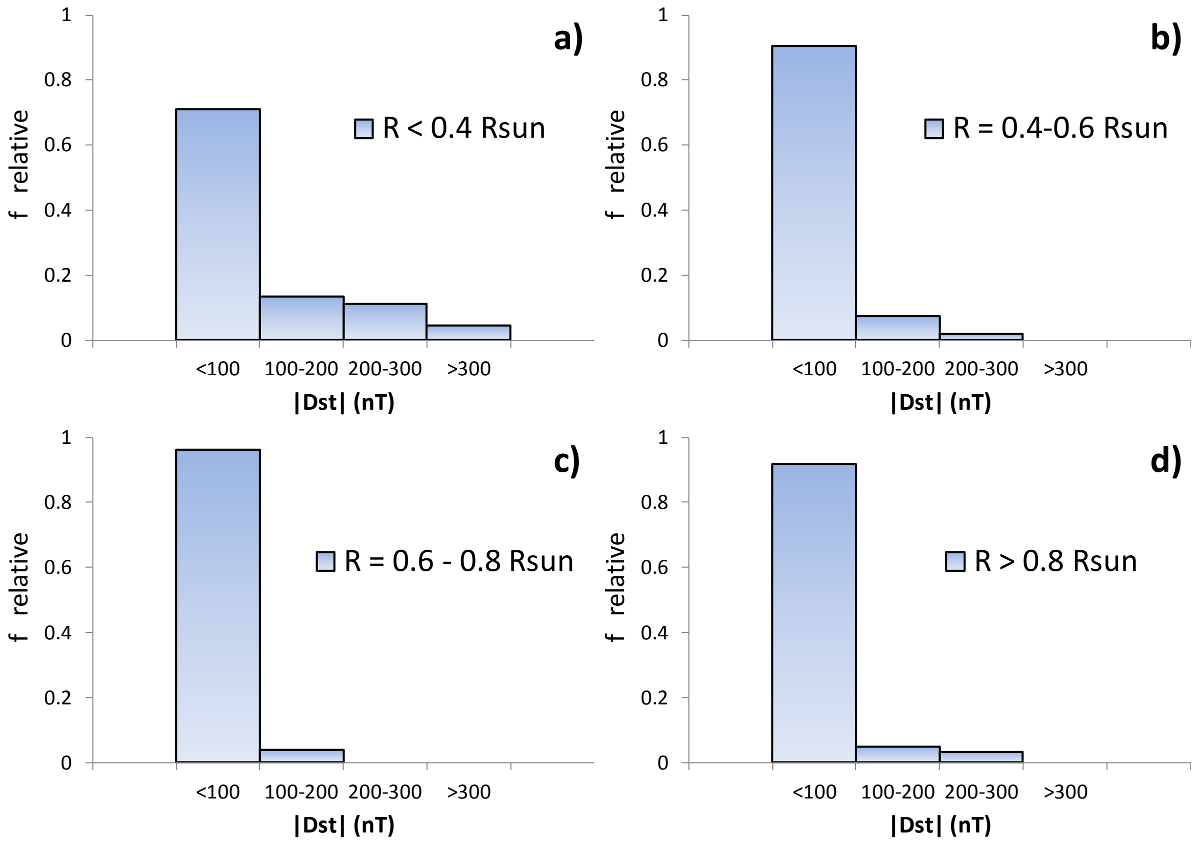


Figure 15: The |Dst| distributions normalized to one for different distance R from the solar disc center bins as shown in the figure.

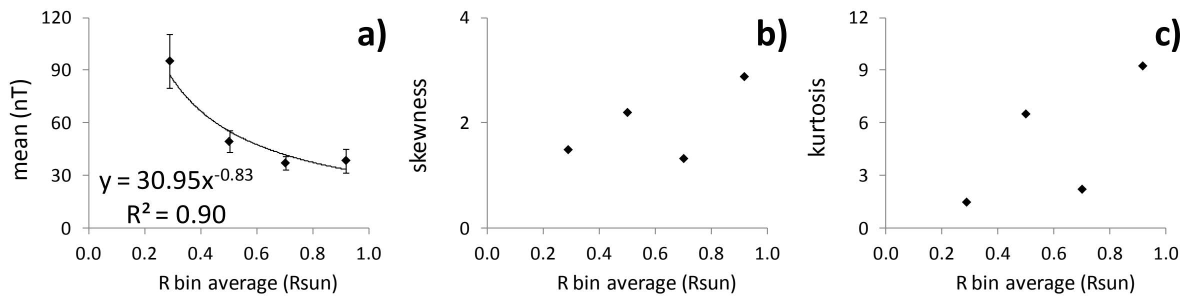


Figure 16: The |Dst| distribution mean (left), skewness (middle) and kurtosis (right) as a function of average distance R within a corresponding distance from the solar disc center bin. Error bars in a) represent the confidence intervals, and the fit result and correlation coefficient R^2 are also shown.

2stt	> 0.4	0.4-0.6	0.6-0.8	>0.8	equal variance assumed
> 0.4		0.00481	1.97E-04	0.000433	
0.4-0.6	0.00841		0.10199	0.24658	
0.6-0.8	0.000716	0.10252		0.85994	
>0.8	0.00147	0.24256	0.85535		
equal variance NOT assumed					

Table 2: Two sample t-test for the |Dst| distribution mean between different distances R from the solar disc center bins with equal variance assumed (upper right values) and equal variance not assumed (lower left values). In the table, the values are given for the probabilities that the two sample means are different. Values which are not significant (i.e. the mean between two distributions is not significantly different) are marked red. The test value is 0.05.

Due to the fact that CME-interactions were given as a single event (T? and T CMEs, see section 2.2.4) associated with a source region of the fastest event within the train, there might be some concerns whether or not this analysis outlines the true connection between the source location and geoeffectivity. Therefore the complete analysis was repeated on the S and S? samples only (a total of 132 CMEs). In all mentioned categorisations (quadrant, north/south, east/west and distance from the disc center) the results do not change significantly. Additionally, we inspected the central meridian distance, where the events were categorized as follows: -90 deg < CMD < -60 deg, -60 deg < CMD < -30 deg, -30 deg < CMD < 0 deg, 0 deg < CMD < 30 deg, 30 deg < CMD < 60 deg, and 60 deg < CMD < 90 deg, with number of events per bin 26, 36, 40, 51, 36, and 22, respectively. A small E-W asymmetry can be observed for 30 deg <|CMD|<60 deg, due to the fact that out of 36 east events in this bin, none had a |Dst|>100 nT. This bin is also significantly different from all other bins, however there is a lack of significant differences between the rest of the CMD samples. Therefore, contrary to studies that report east-west asymmetry (e.g. Zhang et al., 2003; Zhang et al., 2007; Gopalswamy et al., 2007) our analysis shows more or less symmetrical longitudinal distribution of geoeffective CMEs in agreement with e.g. Srivastava & Venkatakrishnan (2004).

3.1.1.3 Level of CME-CME interaction

As explained in section 2.2.4, the level of interaction between CMEs was determined in four categories: no interaction (SOLO CMEs, S), interaction not likely (SOLO? CMEs, S?), interaction likely (TRAIN? CMEs, T?), and interaction (TRAIN CMEs, T). The number of events in each bin is 98, 34, 28, and 51, respectively. For each level of interaction a |Dst| distribution was made, using the |Dst| binning explained in previous sections. This resulted in 4 |Dst| distributions for 4 different levels of interaction (Figure 17). The distribution parameters are represented in Figure 18 and results of the two sample t-test in Table 3. In Figure 18 numbers were associated to different interaction levels for quantitative reasons.

2stt	S	S?	T?	T	
S		0.05726	0.03294	3.85E-04	equal variance assumed
S?	0.1246		0.91516	0.41225	
T?	0.05666	0.91297		0.46517	
T	0.00128	0.4278	0.45169		
equal variance NOT assumed					

Table 3: Two sample t-test for the |Dst| distribution mean between different level of interaction bins with equal variance assumed (upper right values) and equal variance not assumed (lower left values). In the table, the values are given for the probabilities that the two sample means are different. Values which are not significant (i.e. the mean between two distributions is not significantly different) are marked in red. The test value is 0.05.

It can be seen in Figure 17 and Figure 18 that as the level of interaction increases, the skewness and kurtosis decrease, i.e. the distribution for low levels of interaction is more asymmetrical and shifted towards lower |Dst|. The distribution shifts towards larger |Dst| as the level of interaction increases and therefore loses asymmetry (although it's still highly asymmetrical). The mean of the distribution increases with the level of interaction, which can be approximated with a power law function (R=0.96). This dependence should be taken more as a picturesque view than a real dependence) due to the small number of points. The results of the two sample t-test are somewhat inconclusive, because there is only a significant difference between "S" and "T" samples. However, we see that the probability that the two samples are the same reduces with the level of interaction, and is highest for the neighboring bins. This implies that the effect comes from the mixing of the bins: "S?" and "T?" are actually mixtures of "S" and "T" events, with "S?" presumably dominated with "S" events and "T?" with "T" events, respectively. Therefore we conclude that there is indeed an influence of the CME-CME interaction on the probability of a certain |Dst|, where we can associate higher probabilities of (very) intense storms to CME trains. This does not mean that CME trains are more geoeffective (due to some physical mechanism), because we also observe SOLO CMEs producing super intense (|Dst|>300 nT) storms. Our results just show that they are less likely to.

For the "T" and "T?" events, the CME/flare parameters of the fastest CMEs were taken as representative parameters. This raises the question if our results for the level of interaction are just a byproduct of the dependence of the |Dst| distribution on the CME speed. There is a small, but significant difference between speed distribution for S and T events (with probability of 98.5%). However, this speed-interaction level connection is lost when we exclude all the fastest CMEs (v>1700 km/s), whereas the results for the |Dst|-interaction level connection stay the same. Furthermore, the dependence of the distribution mean on interaction level improves and can be approximately fitted with a linear function (R=0.97). Therefore we can conclude that the connection between the CME speed and CME interaction level is not the source of the connection between the interaction level and geoeffectiveness.

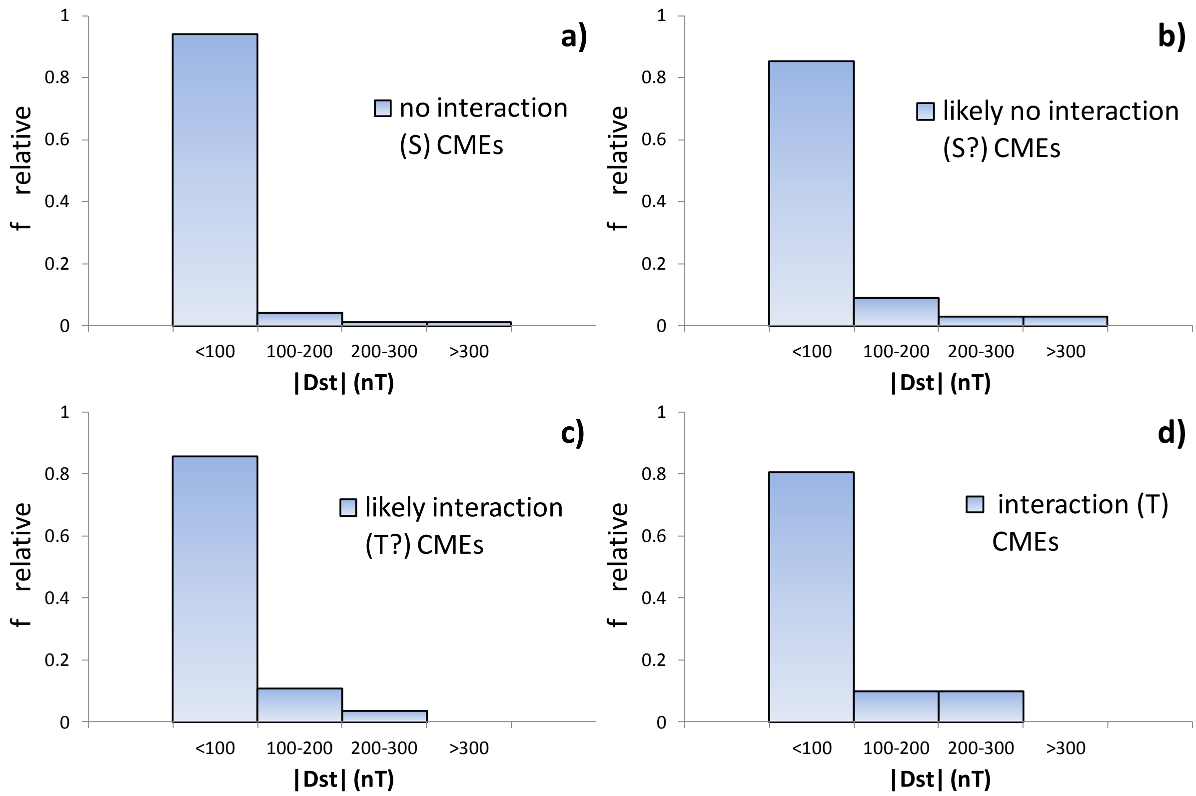


Figure 17: The |Dst| distributions normalized to one for different levels of interaction as shown in the figure.

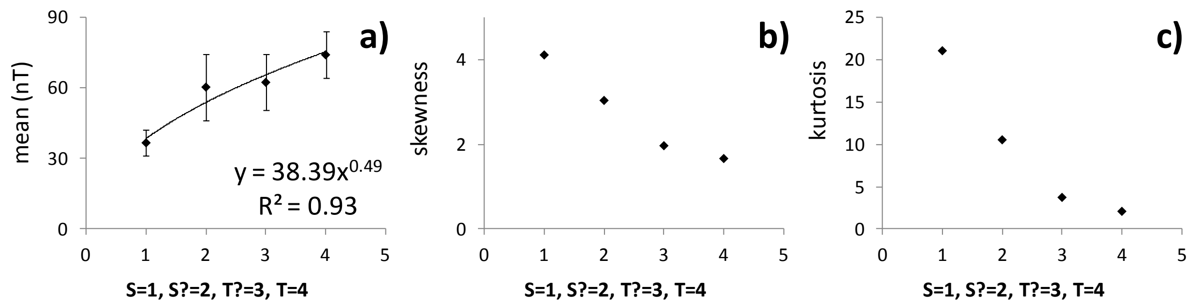


Figure 18: The |Dst| distribution mean (left), skewness (middle) and kurtosis (right) as a function of the numbers associated with different levels of interactions. Error bars in a) represent confidence intervals and the fit result and correlation coefficient R^2 are also shown.

3.1.1.4 CME width

The binning for CME (apparent) width was quite straightforward, as it is already given as a discrete parameter via the non-halo, partial halo and halo categorization (see section 2.2.4). Since the CME-interactions were given as a single event (T? and T CMEs, see section 2.2.4), each event was associated with the largest apparent width of the CMEs involved (i.e. halo or partial halo, if present). The number of events within a certain width bin for non-halo, partial halo and halo CMES are 59, 35, and 117, respectively. Using |Dst| binning explained in the previous sections, 3 |Dst| distributions were made (Figure 19 a) to c)), and the distribution parameters were obtained (Figure 19 d) to f)). The results of the two sample t-test are presented in Table 4. In Figure 19 d) to f), the numbers were associated to different width bins for quantitative reasons. Table numbers for width were not used due to the fact that all halo CMEs have apparent width of 360 degrees.

In Figure 19 a) to c) we see an obvious progression in the |Dst| distribution towards larger |Dst| as the apparent width of the CME increases. For non-halos we see a single bin distribution with no |Dst|>100 nT, for partial halos the distribution gains a small tail, whereas for halos a long tail is observed. This is visible from the distribution parameters as well (Figure 19 d) to f)), where the asymmetry or skewness monotonically increases. There is no notable change in the kurtosis between partial halos and halos, due to the fact that although distribution for halo CMEs has a longer tail it is also more “filled up” with events (i.e. has “shoulders”). The distribution mean has an obvious increasing trend with larger widths (Figure 19 d)), which can be illustrated with a quadratic function (R=1). These results are confirmed with the two sample t-test, showing that non-halo, partial halo and halo CME associated |Dst| distributions are significantly different (Table 4).

The analysis was repeated for S and S? CMEs only, obtaining the same results and minor loss in significance (probably due to the smaller number of events). Furthermore, the same analysis was applied to the original sample, i.e. where the width of the fastest CME was assigned to the CME train (T and T? events) and similar results were obtained. In the latter, both distributions for non-halo and partial halo CMEs is restricted to |Dst|<200 nT, but the distribution mean for partial halos is somewhat larger (although not significantly). These results confirm a widely held view that halo CMEs are more geoeffective. They clearly show that halo CMEs have larger probabilities to lead to intense storms, in agreement with numerous previous studies (e.g. Zhang et al., 2003; Srivastava & Venkatakrishnan, 2004; Gopalswamy et al., 2007). In addition, we can conclude that non-halo CMEs cannot produce major storms (|Dst|>100nT) unless they are involved in CME-CME interaction.

2stt	NH	PH	H	
NH		0.03768	2.55691E-06	equal variance assumed
PH	0.06132		0.0036	
H	8.47817E-10	9.25368E-06		
equal variance NOT assumed				

Table 4: Two sample t-test for the|Dst| distribution mean between different bins of CME apparent width (non-halo, NH, partial halo, PH, and halo, H) with equal variance assumed (upper right values) and equal variance not assumed (lower left values). In the table, the values are given for the probabilities that the two sample means are different. Values which are not significant (i.e. the mean between two distributions is not significantly different) are marked in red. The test value is 0.05.

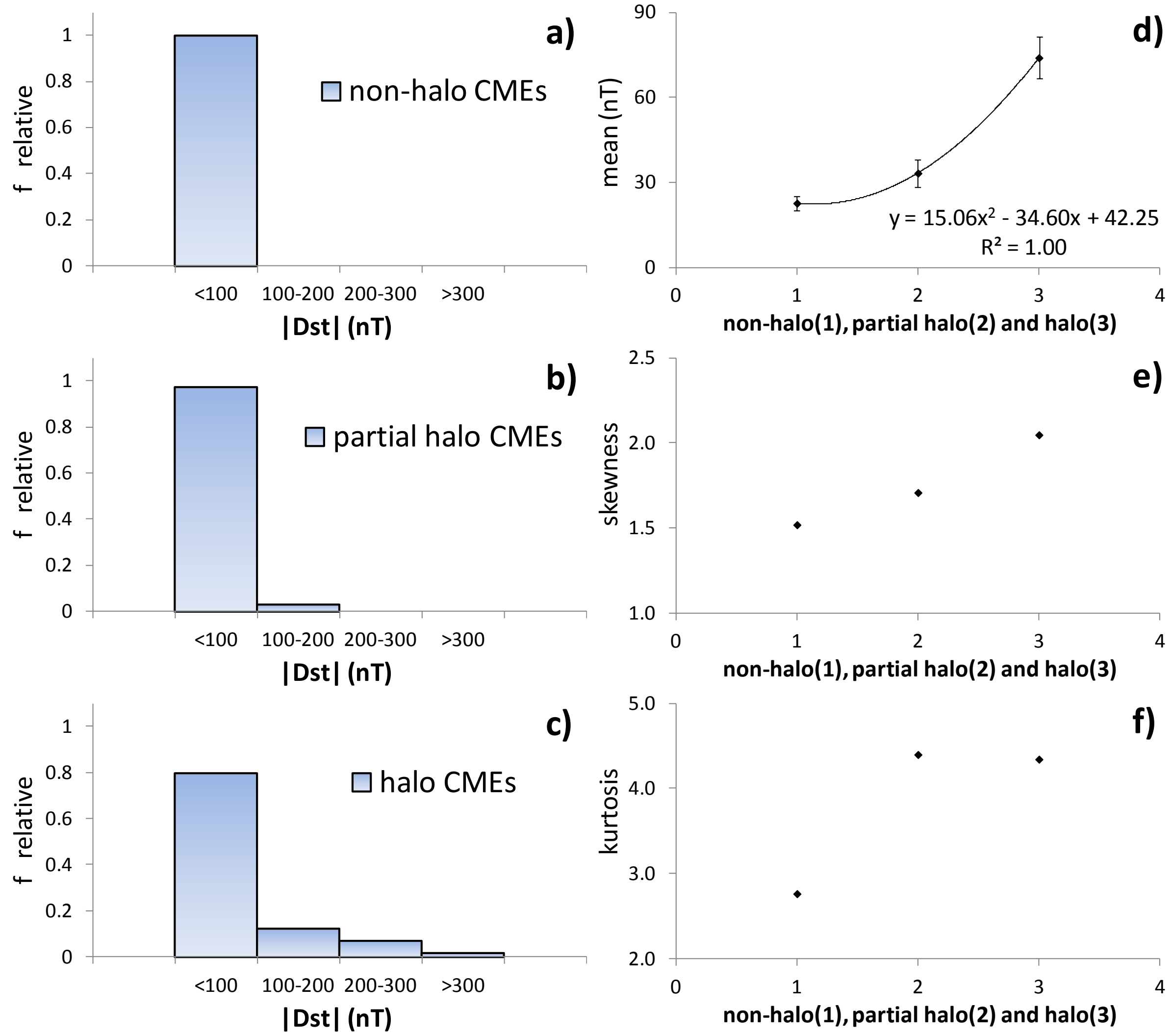


Figure 19: The |Dst| distributions normalized to one (left) and the |Dst| distribution parameters (right) for CMEs binned by apparent width, where the number 1 represents the non-halo CMEs, 2 the partial halo CMEs and 3 the halo CMEs. Error bars in d) represent the confidence intervals and the fit result and correlation coefficient R^2 are also shown.

3.1.1.5 Solar flare type

Similar as with the CME apparent width, the binning for solar flare type was straightforward, since its classification is already given as a discrete parameter. Due to a lack of events associated with B flares, they were put in the same bin with C flares. Therefore three flare categories were made, namely B & C type flares, M flares, and X flares. The number of events in each class are 98, 74, and 39, respectively. Using the |Dst| binning explained in previous sections, 3 |Dst| distributions were made (Figure 20 a) to c)), and the distribution parameters were obtained (Figure 20 d) to f)). The results of the two sample t-test are presented in Table 5. Similarly as in previous sections, in Figure 20 d) to f) numbers were associated with the different solar flare bins for quantitative reasons.

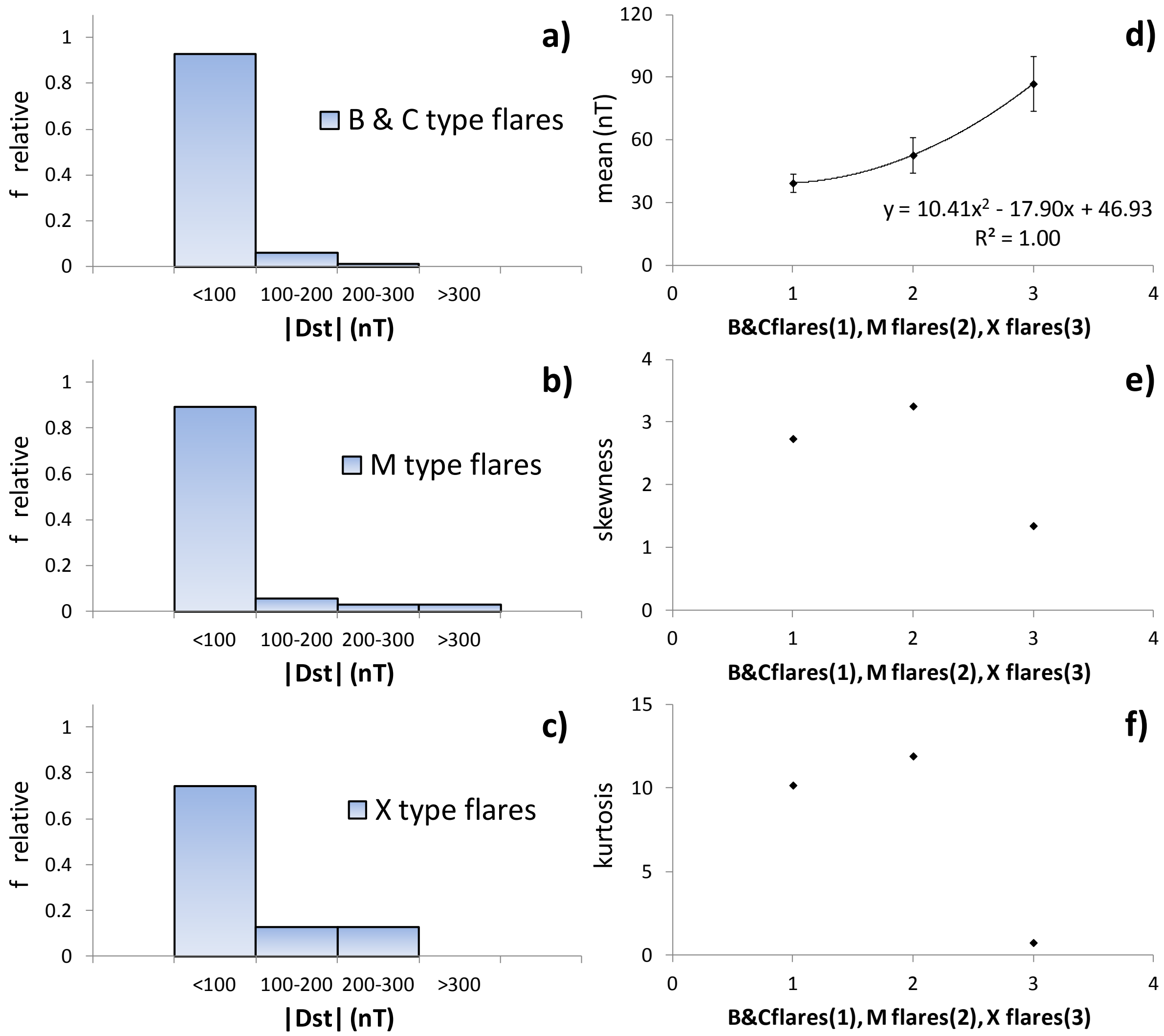


Figure 20: The |Dst| distributions normalized to one (left) and the |Dst| distribution parameters (right) for different types of solar flares, where the number 1 represents the B&C flares, 2 the M flares and 3 the X flares. Error bars in d) represent confidence intervals and the fit result and correlation coefficient R^2 are also shown.

2stt	B & C	M	X	
B & C		0.14404	2.35E-05	equal variance assumed
M	0.17238		0.02629	
X	1.20E-03	3.20E-02		
equal variance NOT assumed				

Table 5: Two sample t-test for the |Dst| distribution mean between different solar flare types (B&C, M, and X) with equal variance assumed (upper right values) and equal variance not assumed (lower left values). In the table, the values are given for the probabilities that the two sample means are different. Values which are not significant (i.e. the mean between two distributions is not significantly different) are marked in red. The test value is 0.05.

In Figure 20 a) to c) we can observe minor differences in the |Dst| distribution between B&C type flares and M type flares. There is a clear difference in regards to X type flares which contains substantially more events in its tail than the other two distributions. This is visible from the distribution parameters as well (Figure 20 d) to f)), where the skewness and kurtosis drops for the X type flare distribution. The distribution mean has an increasing trend with flare type (Figure 20 d)), which can be illustrated with a quadratic function (R=1), similarly as with the apparent width analysis. The same results are reflected by the two sample t-test: B&C type and M type flares are not significantly different samples, but both are significantly different from X flare type sample (Table 5).

The analysis was repeated for S and S? CMEs only, with similar results, but loss in significance (only 19 X flare events). Nevertheless, we can conclude that geoeffective CMEs are associated with stronger flares, in agreement with previous studies (e.g. Zhang et al., 2007; Srivastava & Venkatakrishnan, 2004).

3.1.1.6 Summary

We presented an analysis to derive the key CME and solar flare parameters and their influence on the probability of observing major, intense, and super storms. Our results reflect some previous well known connections between remote solar properties and geomagnetic storms, namely the importance of CME speed, apparent width, source position and associated solar flare type. It also offers a quantification of these connections. Furthermore, it is clearly shown that the CME-CME interaction is associated with a larger probability to produce intense storms. The results of this statistical analysis can be used for the prediction of the probability that a given event, observed with coronagraph, Xray and EUV imagers on a satellite located at the L1 point (or even ground based instruments), will produce a major, intense or very intense geomagnetic storm at Earth. Although the prediction of this kind will be quite rough, the advantage is that it provides an early warning.

It should be noted that although we are referring to the geoeffectiveness throughout the study, our results do not really reflect the geoeffectiveness of the CME, but only a probability that the event will be seen as geoeffective. In other words, the probability that the CME will reach the Earth is also indirectly implemented. This is due to the sampling, where the general idea was to observe solar sources and Earth effects, without an interplanetary component. This way nongeoeffective CMEs were basically treated equally as the CMEs which never reached the Earth. Although physically incorrect, we find this view appropriate regarding the statistical analysis, where physical variables are treated as random variables. Another useful aspect of this approach is that the false alarms are also implemented within this study, since they are present in the sample, although this view does not provide the means to predict them directly. Therefore they will be studied separately.

3.1.2 Relating CME parameters to the Dst value for geomagnetic events (ROB)

A statistical analysis was performed to relate CME parameters to the Dst index of geomagnetic storms. The list of 211 front sided flare-related CMEs from the SOHO period was used, which is described in section 2.2.4. All CMEs in the database had a speed larger than 400 km/s and were selected such that all possible speeds were well represented. As variables we considered positional and physical parameters, like the distance r from the source region to the center of the Sun (in units of radius of the solar disk R_{sun}) , the CME speed v, the CME width, the level of interaction with other CMEs and the intensity of the associated flare. Their bivariate and multivariate relationships to the Dst index were tested. The goal was to identify key solar parameters that are directly measurable and determine the strength of a geomagnetic storm. A geomagnetic storm is classified according to the Dst index as moderate (-50nT > Dst > -100nT), severe (-100nT > Dst > -250nT) or super-storm (Dst < -250nT). The output of the analysis could deliver some insight and define the input parameters for a geomagnetic storm forecast system. A bivariate analysis has been performed between Dst and CME parameters, below visualized in a few figures. A few key observations for the list of 211 events are listed below.

The variability is higher among CMEs with a source region close to the disk center (see Figure 21). CMEs with a source region at a distance to the center lower than 0.5 R_{Sun}, have Dst values from 0 to -400nT, while for CMEs further than 0.75 R_{Sun} from the center most CMEs have a Dst > -200nT. The smoothed curve indicates a slight positive relationship between Dst and r, with a correlation of 0.34.

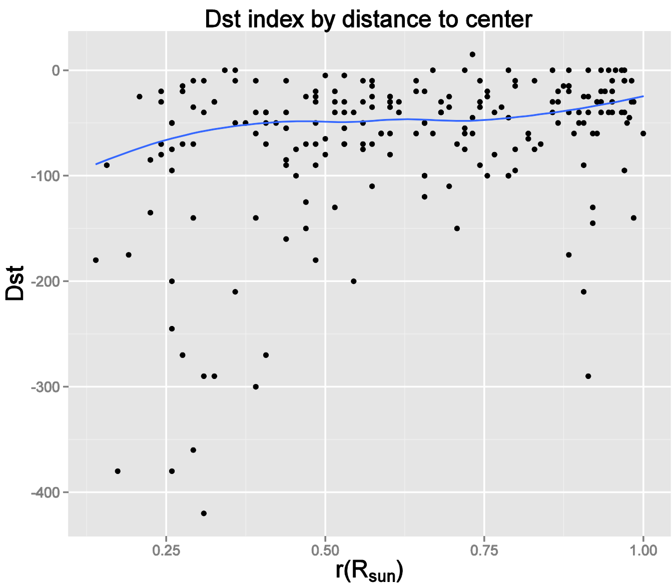


Figure 21: Relating the Dst index to the distance to the centre of the Sun.

The CME speed increases the chances for a geomagnetic storm, though even at v < 1000 km/s these cannot be excluded. Dividing the sample in different groups according to the speed provides the following quartiles for each subgroup (see Table 6). The three quartiles separate the sample in four equal parts; the first quartile separates one fourth with the lowest values from the rest of the sample. The second quartile or median splits the sample in half; a part with lower values and one with higher values. The third quartile separates the quarter with the highest values from the remainder with lower values. The largest range in Dst occurs at a CME speed of 2000 to 2500 km/s. The correlation between Dst and the CME speed is calculated as -0.33, which is among the strongest correlations.

From a similar analysis using the CME width it can be concluded that a geomagnetic index Dst < -200 nT only occurs for full halo CMEs. The range between the 1st and 3rd quartile, also called the interquartile range, is much broader and skewed to low Dst values for full halo CMEs (see Table 7).

CMEs associated with a large flare (M, X-class flare) have a higher chance to cause a large storm. This might be related to the larger energy release. However, even CMEs associated with a C-class flare could produce severe geomagnetic storms (see Figure 22).

Lower bound v ($\frac{km}{s}$)	Upper bound v ($\frac{km}{s}$)	1st quartile Dst	Median Dst (2nd quartile)	3rd quartile Dst	Number of CMEs
0	500	-78	-45	-29	12
500	1000	-60	-33	-20	94
1000	1500	-75	-40	-30	52
1500	2000	-93	-60	-34	36
2000	2500	-225	-120	-68	12
2500	4000	-100	-80	-75	5

Table 6: Quartiles of the Dst value for different subgroups defined for the CME speed.

Halo/non-halo	1st quartile Dst	Median Dst (or 2nd quartile)	3rd quartile Dst	Number of CMEs
Non-halo CME	-45	-30	-15	65
Full halo CME	-110	-70	-36	110
Partial halo CME	-60	-40	-25	36

Table 7: Quartiles of the Dst value for full halo CMEs, partial halo CMEs and non-halo CMEs.

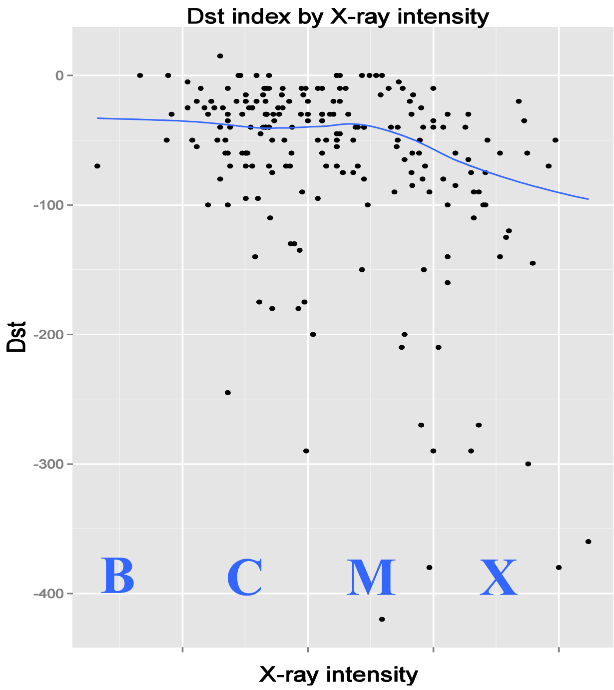


Figure 22: Relating the Dst index to the intensity of the associated flare.

On average, a sequence of interacting CMEs (TRAIN event) produces stronger storms than SOLO CMEs (see section 2.2.4 for the definition of the levels of interaction with other CMEs). No strong statements can be made on this relation, since for several CMEs (62 out of 211) it was not fully clear to assign the level of interaction (SOLO?: not likely interacting and TRAIN?: likely interacting). These CMEs could also produce very strong storms, as illustrated in Table 8.

Combinations of multiple CME variables were also related to the Dst index, by using heatmaps (see Figure 23). Moderate storms (-50nT > Dst > -100nT) can occur at a small CME width of 60 to 240°. But all super-storms (Dst < -250nT) in the database are caused by halo CMEs and can occur at any CME speed above 500 km/s.

Level of interaction	1st quartile Dst	Median Dst (2nd quartile)	3rd quartile Dst	Number of CMEs
SOLO	-60	-35	-16	98
SOLO?	-120	-45	-26	34
TRAIN?	-95	-60	-30	28
TRAIN	-100	-60	-40	51

Table 8: Quartiles of the Dst value depending on the level of interaction with other CMEs (SOLO: not interacting, SOLO?: not likely interacting, TRAIN?: likely interacting, TRAIN: interacting).

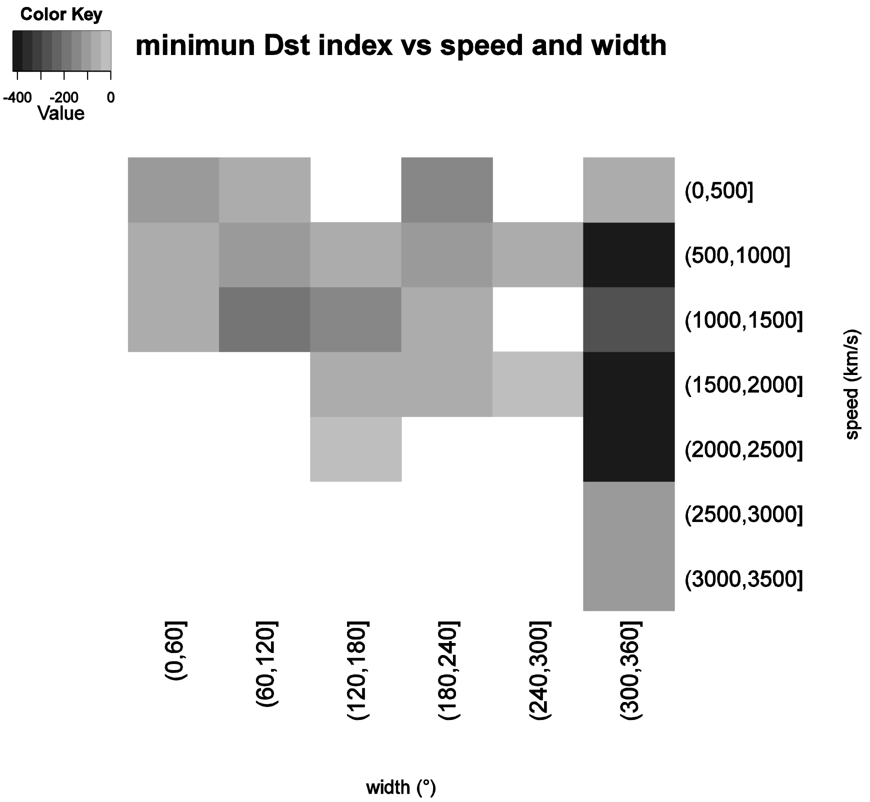


Figure 23: Heatmap of the Dst index depending on the combination of speed and width. The color indicates the minimum Dst value for each combination of CME speed and CME width. Black corresponds to a minimum Dst value of -400nT and light grey means the lowest Dst value in the subgroup is about 0.

The CMEs resulting in the strongest geomagnetic storms originate from a source region near the disk center. Though less likely, CMEs from our sample closer to the limb (with a distance to the center of more than 0.8 R_{Sun}) can also cause super-storms at v > 1000km/s.

The main results can be summarized as follows:

- In general, the risk for a geomagnetic storm increases in case of a fast CME, a source location close to the disk center or a halo CME.
- Slow CMEs do not exclude the possibility of large storms to occur. Examples were found if the slow CME is a full halo CME, has a source region close to the center, is part of a train of CMEs or is associated with a strong flare.
- Magnetic field and density certainly play a role, but were not available in the database. These parameters could be studied in the future.

3.1.3. Semiannual variations of geoefficiency (DTU)

It has long been known that the disturbance of the geomagnetic activities varies semi-annually. The variations have been studied in many indices, like aa and Dst (Clua de Gonzalez et al., 2002; Cliver et al. 2002; Vennerstrom, 2000; Emery et al., 2011, just to mention a few). The disturbances are largest around March and September and have minima around June and January. There are three established theories for the semiannual variations:

1. The axial hypothesis where the Earth's orbit out of the Suns equator plane is given as the explanation for the semiannual variations. The angle between the Sun's equator plane and the ecliptic plane is 7.3^0. The Earth is furthest away from the Sun's equator plane on March 5^{th} (DOY 64) and September 5^{th} (DOY 248). This hypothesis was proposed by (Cortie, 1912).
2. The equinoctial hypothesis states that the annual variation of the orientation of Earth's magnetic axis is the driver of the semiannual variations of geo-magnetic disturbances. When the angle between the Earth's magnetic axis and the Earth-Sun line is 90^0 the coupling efficiency with the magnetosphere is maximized because the Earth's magnetic field is weakest in the direction of the solar wind. This effect is at maximum during equinoxes, March 21^{st} and September 21^{st}. The hypothesis was proposed by (Bartels, 1932).
3. The Russell-McPherron mechanism is a combination of the two other hypotheses. The controlling parameter is the angle φ between the z-axis of the GSM coordinates and the solar equatorial plane. Geomagnetic disturbances are expected to be largest when φ is minimum, which happens on April 5^{th} and October 8^{th}. Proposed by (Russell and McPherron, 1973).

It should be mentioned that it is not known exactly why there are semiannual variations. It could be any of the above hypotheses, a combination or even a complete other mechanism. For a forecasting perspective it is not mandatory to know why, just how the semiannual variations behave.

3.1.3.1 SOHO era

The SOHO era from 1997 until today offers large amounts of data and observations of the Sun. The LASCO instrument makes more or less constantly coronagraphic observations of the Sun's corona making it possible to study and statistically analyze CMEs. One could ask the simple question: Is the semiannual variation of geomagnetic disturbances driven by some kind of variation in the CME rate? Or is there some kind of selection effect in the way we observe CMEs?

To test this, 17513 CMEs observed by LASCO from 1997-2011 were analyzed by counting how many CMEs were detected in each month of the year. Furthermore the list of geomagnetic storms described in task 4.1. (section 2.2.6) was used to calculate the monthly storm rate. The results are shown in Figure 24.

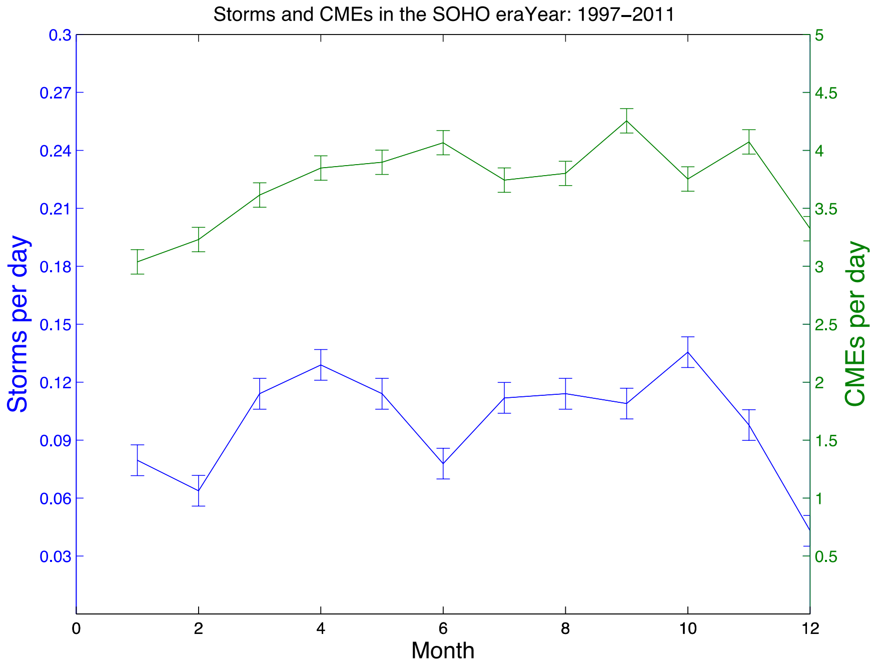


Figure 24: The daily CME rate and the daily storm rate as a function of month based on all storms and CMEs from 1997-2011. Based on a chi square test it was evaluated that the CMEs are evenly distributed over the year, whereas the storms are *not* evenly distributed over time. The error bars represent the standard deviation of the mean.

The average number of storms per day was found to be 0.10 and the average number of CMEs per day was found to be 3.7.

To test whether the CMEs and storms are evenly distributed over the year they were tested by a chi square assuming that each month would have the same rate per day. It was found that:

Prob (c^2 =4.0) = 96% for CMEs

Prob (c^2 =142.8) = 0% for storms

5% level of significance would be a standard choice of confidence and in that case the test shows that CMEs are evenly distributed over the year, whereas the storms have a variation. This is not surprising but it is nice to know that we should expect the same CME rate during the year.

3.1.3.2 Semiannual variation in Dst index

The Dst index from 1957-2011 has been used to study semiannual variations of geomagnetic disturbances. The data set thus covers 5 solar cycles from cycle 19 until the beginning of cycle 24. The storms in the Dst period have been identified and grouped by the month they occurred. From a forecasting perspective it is interesting to know when CMEs are most likely to generate a geomagnetic storm. However, it is well known that not only CMEs, but also Corotating Interaction Regions (CIRs) can generate storms. CIRs occur when the fast solar wind encounters the slow solar wind and thus increases the dynamic pressure of the slow solar wind. CIRs can generate geomagnetic storms which are usually not as intense as storms generated by CMEs can be. To exclude storms generated by CIRs, only storms that produced a Dst index of less than -100 nT are therefore included.

The time from 1957 to 2010 has been divided into solar maximum and solar minimum intervals in order to study the difference in the solar cycle. One could take many different approaches to the division into solar activity. Here a solar cycle is divided evenly so that half the cycle is considered maximum and the other half minimum. One could also only consider the maximum as a few years and minimum as a few years, this would however give fewer storms for the statistic. The division is based on sunspot number. Table 9 gives an overview of the division. It has then been counted how many storms there were in a given month depending on the solar cycle. The results are shown in Figure 25.

Solar max	Solar min
1957-1961	1962-1967
1968-1973	1974-1978
1979-1983	1984-1988
1989-1993	1994-1998
1999-2004	2005-2010

Table 9: The division of solar maximum and solar minimum periods.

It was expected to see that there is a variation in the storm rate of about a factor of 3 during the year, both in solar minimum and solar maximum. It is however surprising to see that the peak shifts during solar cycle, so that solar maximum has the highest storm rate in March and September, while solar minimum has the highest rate in April and October. This could indicate that the Russell-McPherron effect is strong in solar minima. This study doesn’t offer any explanation for this observation. Finally it was also considered how many storm hours there were in the different months. The results are shown in Figure 26.

The durations of the storms as seen in Figure 26 show the same pattern as the number of storms in Figure 25. It is a noteworthy observation that the storms are also longer in spring and fall compared to the rest of the year. During solar maximum it is found that the mean duration of a storm is about 50 hours in January and about 85 hours in March.

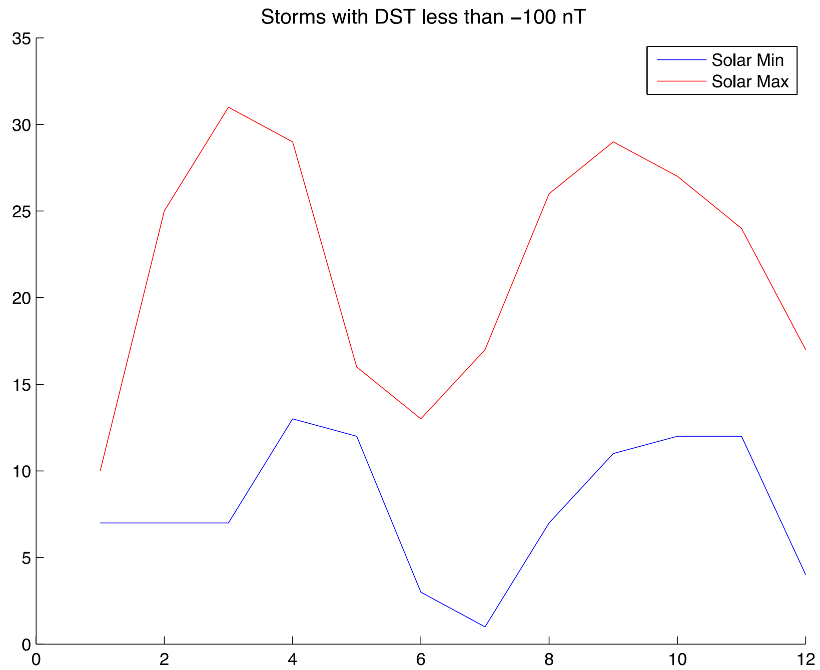


Figure 25: The number of strong storms in each month depending on solar cycle. It is interesting to see that during solar maximum the number of storms peak in March and September, while during solar minimum the peaks are in April and October.

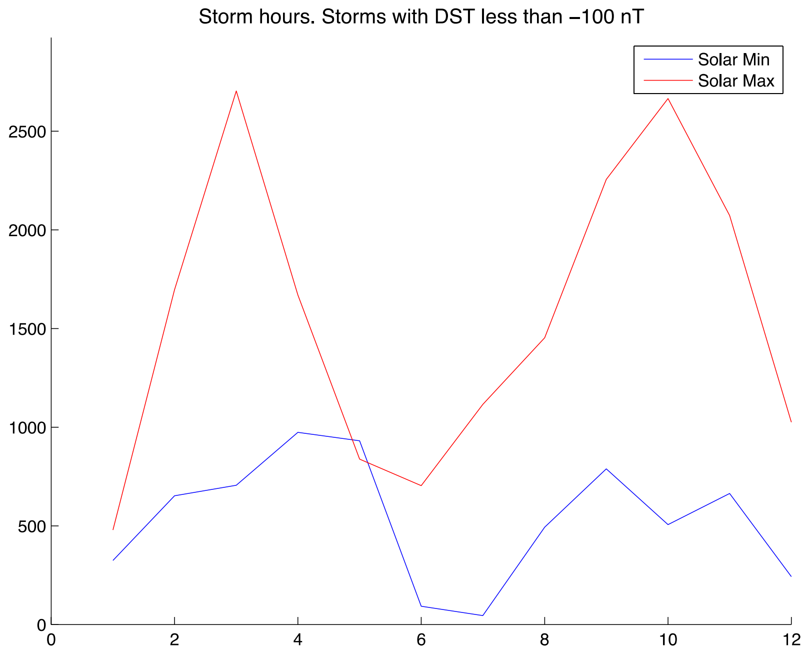


Figure 26: The total amount of storm hours in each month (numbered 1 to 12) during solar minimum and solar maximum.

3.1.3.3 Conclusions

It has been evaluated how the seasonal variations of storms should be treated when forecasting geomagnetic activities:

- There are no seasonal variations of CME rate from the Sun during the year.
- Strong storms (Dst<-100 nT) generated by CMEs vary both semiannually and during the solar cycle. During solar maximum the storms peak in March and September, while the peak is found in April and October during solar minimum.
- The storms are also found to be longer in spring and fall, compared to winter and summer.

3.2 Solar Energetic Particles

3.2.1 Probability of SEP occurrence (NOA)

In the following subsections we present the derived probabilities for SEP occurrence based on a statistical analysis of flare and CME characteristics and their association with SEPs. We are investigating possible association of SEP occurrence with two flare characteristics, flare intensity and flare location, and two CME characteristics, CME velocity and CME width, by using and combining information from the available flare and CME databases that have already been described in detail in section 2. As it is known from literature (e.g. Reames, 1999) these four variables define both the SEP profile and its physical characteristics.

The methodology for deriving the probability of SEP occurrence is as follows. If N_{sep} is the number of SEP productive flares and/or CMEs, within a certain subset defined from the aforementioned flare and/or CME characteristics (e.g. flares within a certain intensity range, CMEs within a certain velocity range etc.) and N_{total} the total size of this subset (irrespective of SEP association) then the probability of SEP occurrence $\boldsymbol{p}$ within this subset is defined as

$$\boldsymbol{p} = N_{sep} / N_{total}$$

where $\boldsymbol{p}$ is a number between [0,1]. Multiplication of this number by 100 gives the corresponding percentage probability. Respective errors $\boldsymbol{p_{err}}$ to this probability are derived according to the Binomial proportion confidence interval defined as

$$\boldsymbol{p_{err}} = \pm\, \mathbf{z_{1-\alpha/2}} \sqrt{\frac{\boldsymbol{p}\,(1-\boldsymbol{p})}{\mathbf{N}}}$$

where $\boldsymbol{p}$ is the probability, $\boldsymbol{N}$ (=N_{total}) is the sample size, $\mathbf{z_{1-\alpha/2}}$ is the 1-α/2 percentile of a standard normal distribution and α is the error percentile. We have used a value $\mathbf{z_{1-\alpha/2}} = 1$ for which the derived uncertainties correspond to the 68% confidence level intervals. We should point out that if the sample size $\boldsymbol{N}$ is small then the sample itself is not described well by a normal distribution and derived errors can become quite large.

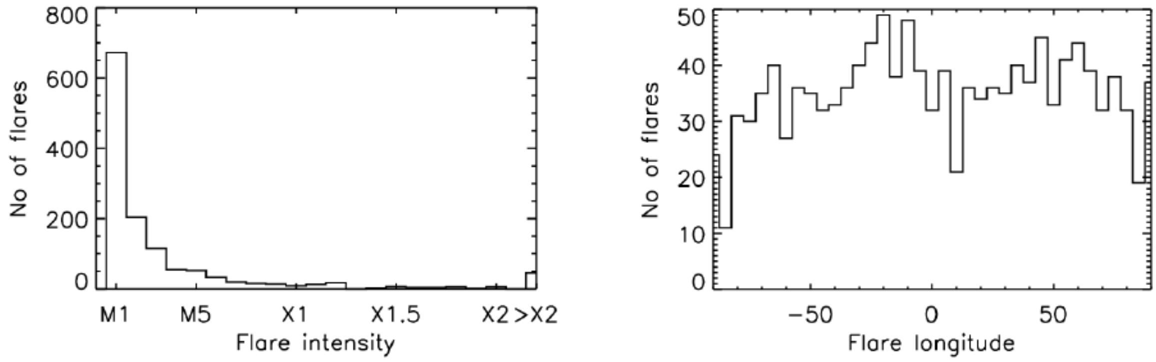


Figure 27: Histograms of flare intensity (left) and flare longitude (right) for the flare sample under study.

3.2.1.1 *Probabilities of SEP occurrence using only flare characteristics*

Probabilities of SEP occurrence have been derived through a statistical analysis of a flare sample that includes 1298 M and X class flares of solar cycle 23. This sample is a subset of all M and X class flares of the NOAA flare catalog of solar cycle 23, compiled from GOES data, that a) covers only the period 1997-2006 which coincides with the SOHO era and b) the identified solar longitude of the flare is between [-90, 90] degrees (visible earthward solar disc). Cross-checking of both flare characteristics, intensity and location, with all available corresponding information in literature (e.g. Cane et al. 2010) has been performed during compilation of the aforementioned list. The final derived sample includes 1181 M-class flares and 117 X-class flares. The full flare list under study is available in the COMESEP database described in section 2.2.8.

Figure 27 shows histograms of the flare intensity and location (longitude) of our sample. As expected within a solar cycle, the number of flares is larger for lower intensity flares and drops considerably for larger intensity flares. There is no clear longitude dependence for flare occurrence as expected.

The SEP events and association between solar flares are taken from the Cane et al. (2010) list. Unfortunately there are only 160 SEP events associated with this 1298-flare sample; 104 of them are associated with M-class flares and the remaining 56 with X-class flares. Given the relatively small number of SEPs and in order to improve derived probabilities, the statistical analysis has been performed in a small number of flare intensity and flare location (longitude) bins. Since we are dealing with observations from near-Earth satellites located on the ecliptic, no analysis concerning the latitude of the flare is necessary, since as it is already known from literature it is mostly longitude that defines the SEP profile and characteristics in this case (Hollebeke et al., 1975, Cane et al., 1988, Dalla & Agueda, 2010).

We are investigating probabilities of SEP occurrence in five (5) flare intensity bins, namely [M1-M3.9], [M4-M6.9], [M7-M9.9], [X1-X4.9] and [≥X5] and five (5) longitude bins, namely [-90,-71], [-70, -31], [-30,29], [30,69] and [70,90] degrees. We examine two distinct cases a) when only flare intensity is known, and b) when both flare intensity and location is available. Cases with only flare location do not need to be examined separately, since the flare intensity is always available when the flare location is known. Derivation of SEP occurrence probability for these two aforementioned cases stems from the characteristics of the alert system itself, since flare intensity information usually precedes the associated, if any at all, flare location information. Hence flare intensity alone can be used for deriving a first 0-level SEP occurrence alert, well before any location information becomes available that can be used to further refine this alert.

Table 10 shows the derived probabilities when only flare intensity information is available. Probabilities (in the range of 0 to 1) and respective errors (in the range of 0 to 1) have been derived after splitting our sample only in the five pre-defined intensity bins; the number size of each of these five sub-samples is also presented in the Table. The derived probabilities and errors are also presented as a plot in Figure 28. We notice that the probability of SEP occurrence increases considerably with flare intensity. The derived curve is in accordance with similar previous studies in literature (e.g. Figure 10 of Belov et al. 2005).

When both flare intensity and location (longitude) are available, the five pre-defined intensity and location bins define a 5×5 matrix of probabilities for SEP occurrence. In Table 11 and Figure 29 we present the derived probabilities (in the range of 0 to 1) and respective errors (in the range of 0 to 1) for these 25 sub-samples of the original flare sample. The number size of each sub-sample is also shown in Table 11. Apart from the increase in probabilities of SEP occurrence with flare intensity, we also notice that there is a clear western-hemisphere preference for SEP-productive flares in accordance with literature.

Flare intensity	**M1-M3.9**	**M4-M6.9**	**M7-M9.9**	**X1-X4.9**	**≥X5**
N_{flares}	*991*	*140*	*50*	*99*	*18*
SEP probability	0.062±0.008	0.2± 0.034	0.3± 0.065	0.444± 0.05	0.667± 0.111

Table 10: The SEP occurrence probabilities and respective errors as a function of flare intensity. The total number of flares per intensity bin is also shown. Percentage probabilities can be derived by multiplying the derived values by a factor of 100.

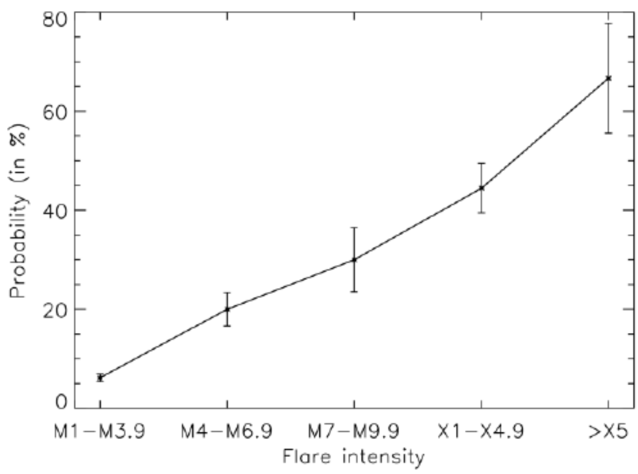
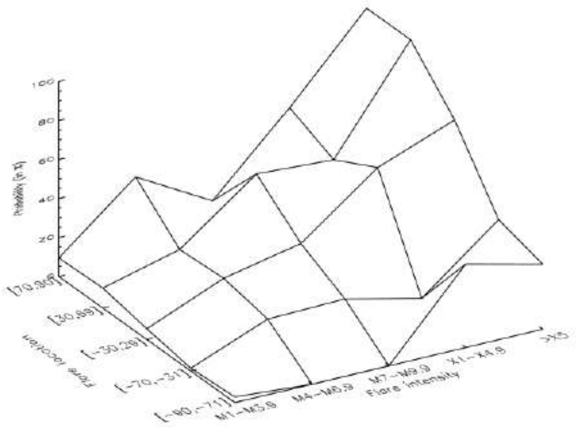


Figure 28: The SEP percentage probabilities and respective errors as a function of flare intensity.

Location		Flare intensity				
		M1-M3.9	M4-M6.9	M7-M9.9	X1-X4.9	≥X5
[-90,-71]	N_{flares}	*70*	*15*	*1*	*7*	*3*
	Probability	**0.029±0.02**	**0**	**0**	**0.429±0.187**	**0.333±0.272**
[-70,-31]	N_{flares}	*207*	*29*	*11*	*22*	*5*
	Probability	**0.014±0.008**	**0.172±0.07**	**0.182±0.116**	**0.091±0.061**	**0.4±0.219**
[-30,29]	N_{flares}	*344*	*50*	*23*	*35*	*4*
	Probability	**0.055±0.012**	**0.22±0.059**	**0.304±0.096**	**0.6±0.083**	**0.75±0.217**
[30,69]	N_{flares}	*244*	*34*	*10*	*23*	*3*
	Probability	**0.102±0.019**	**0.206±0.069**	**0.5±0.158**	**0.478±0.104**	**1**
[70,90]	N_{flares}	*126*	*12*	*5*	*12*	*3*
	Probability	**0.095±0.026**	**0.417±0.142**	**0.2±0.179**	**0.583±0.142**	**1**

Table 11: SEP occurrence probabilities and respective errors as a function of flare intensity and flare location. The total number of flares per intensity and location bin is also shown. Percentage probabilities can be derived by multiplying the derived values by a factor of 100.



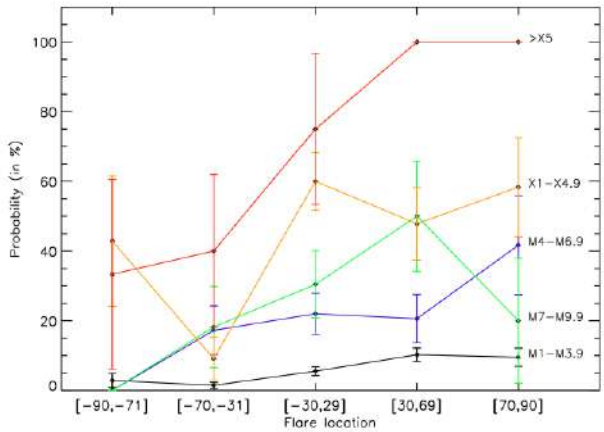


Figure 29: SEP percentage probabilities and respective errors as a function of flare intensity and location.

3.1.1.2 *Probabilities of SEP occurrence using both flare and CME characteristics*

The list described in section 2.2.3 provides the CME/flare associations and is completely independent of SEPs. Hence, the flare sample used previously in our flare/SEP statistical analysis is compiled and merged with the CME/flare list in order to define a flare/CME/SEP association. The derived list has been careful cross-checked with other available lists such as the Cane et al. (2010) list. The final derived sample has only 438 flares, out of the initial 1298, which are associated with CMEs; 359 are M-class flares while the rest 79 are X-class flares. This 438 CME-associated flare sample includes only 118 SEP productive flares, 78 SEPs associated with M-class flares and 40 with X-class flares. The full flare/CME/SEP list under study is available at the COMESEP database described in section 2.

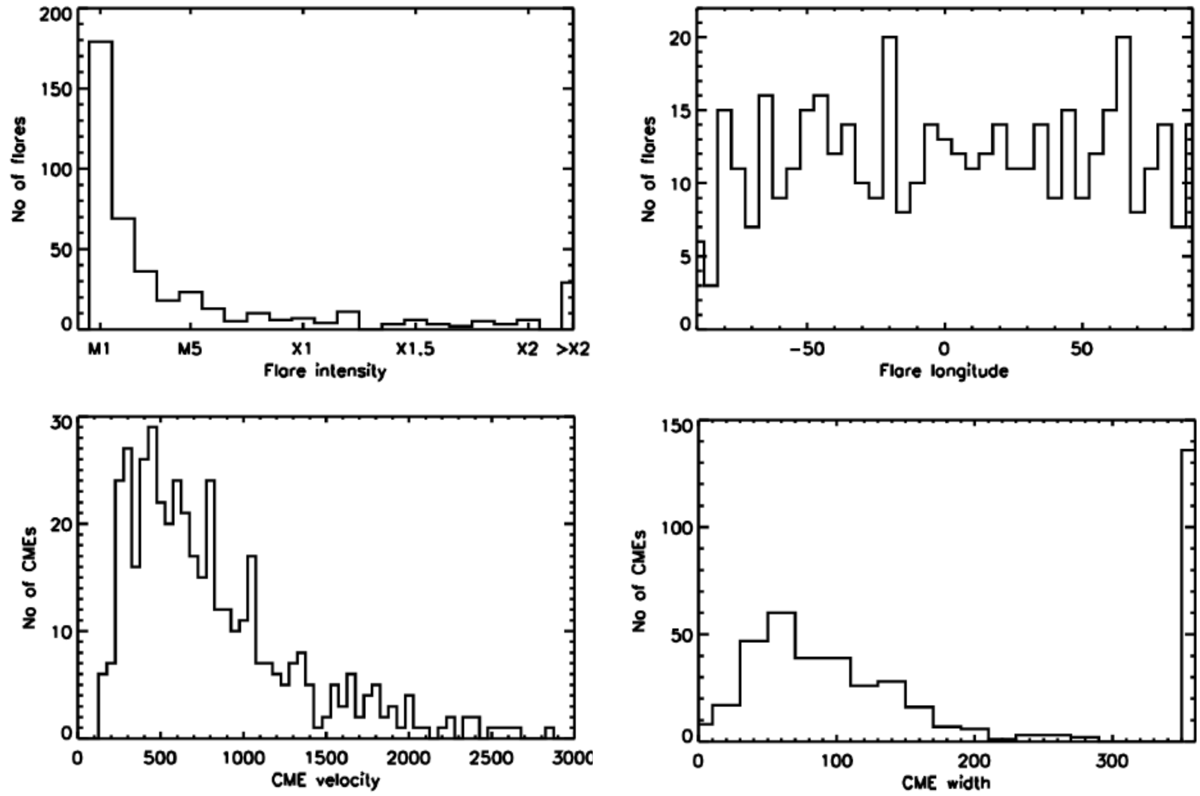


Figure 30: Histograms of flare intensity (top left panel, flare width (top right panel), CME velocity (bottom left panel) and CME width (bottom right panel) for the flare/CME sample under study.

Figure 30 shows histograms of the two flare characteristics (intensity and longitude) and the two CME characteristics (CME velocities and CME widths) of our sample. The number of flares is larger for lower intensity flares and drops considerably for larger intensity flares while there is no clear longitude dependence for flare occurrence. Similarities in the behaviour of intensity and longitude between this limited sample and the large 1298-flare sample studied above, denote that this flare/CME sample is quite representative of the whole 1298-flare sample. The majority of CME velocities are in the range of 0 to 1500 km/s, however there is a large number of CMEs with velocities larger than 1500 km/s. As for CME widths there seem to be two distinct populations, the first with CME widths in the range of 0 to 180 degrees and then a large population of halo CMEs (360 degrees).

Given the relatively small number of both SEPs (118) and CME-associated flares (438), the statistical analysis cannot be performed in a large number of multi-dimensional bins in the 4 parameters (flare intensity, flare location longitude, CME velocity and CME width). The number of events in each bin would be mostly very small, leading to poor statistics. Hence we investigate probabilities of SEP occurrence in three (3) flare intensity bins, namely [M1-M3.9], [M4- M9.9] and [≥X1], two (2) longitude bins, namely [-90,-1] and [0, 90] degrees, three (3) CME velocity bins, namely [0,499], [500,999] and [≥1000], and two (2) CME width bins, namely [0,179] and [180,360] degrees. Since both CME velocity and CME width is always available when a CME is identified by existing satellites that continuously monitor the Sun, we examine here two distinct cases a) when only flare intensity and both CME characteristics are known, and b) when all flare (intensity and location) and CME (velocity and width) characteristics are available. There exists a third case when a CME occurs without any flare association at all, not even intensity (e.g. CME from a large filament eruption), which has however to be further investigated and hence will not be presented here; unfortunately there is so far no such CME/SEP list available. Although one expects that such CMEs will not be very SEP productive, it is a case that probably has to be also integrated in our final alert system.

Table 12 shows the derived probabilities when only flare intensity and both CME characteristics (velocity and width) are available. Probabilities (in the range of 0 to 1) and respective errors (in the range of 0 to 1) have been derived after splitting our sample in the pre-defined flare intensity, CME velocity and CME width bins; the number size of each of these five sub-samples is also presented in the Table. The derived probabilities and errors are also presented as a plot in Figure 31. We do notice that probabilities of SEP occurrence are always higher for CME widths in the range of [180,360] degrees which mostly correspond to halo CMEs (360 degrees) as the respective histogram of CME widths indicates.

			CME velocity (km/s)		
Flare intensity	**CME width**		**0-499**	**500-999**	**≥1000**
M1-M3.9	**[0,179]**	N_{flares}	*100*	*100*	*19*
		Probability	**0.03±0.017**	**0.08±0.027**	**0.421±0.113**
	[180,360]	N_{flares}	*13*	*25*	*27*
		Probability	**0.077±0.074**	**0.52±0.1**	**0.519±0.096**
M4-M9.9	**[0,179]**	N_{flares}	*8*	*21*	*7*
		Probability	**0**	**0.19±0.086**	**0.143±0.132**
	[180,360]	N_{flares}	*2*	*12*	*25*
		Probability	**1**	**0.5±0.144**	**0.72±0.09**
≥X1	**[0,179]**	N_{flares}	*10*	*8*	*7*
		Probability	**0**	**0.25±0.153**	**0.571±0.187**
	[180,360]	N_{flares}	*2*	*11*	*41*
		Probability	**1**	**0.545±0.15**	**0.634±0.075**

Table 12: SEP occurrence probabilities and respective errors as a function of flare intensity and CME characteristics (velocity and width). The total number of CME-associated flares per flare intensity, CME velocity and CME width bin is also shown. Percentage probabilities can be derived by multiplying the derived values by a factor of 100.

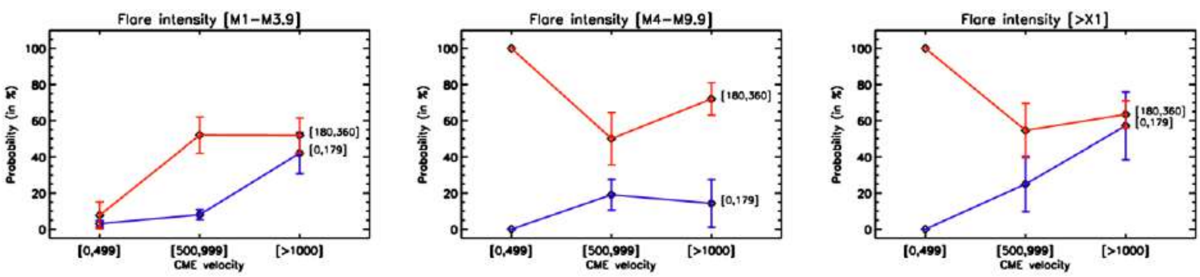


Figure 31: SEP percentage probabilities and respective errors as a function of flare intensity and CME characteristics.

Table 13 and Table 14 show the derived probabilities when both flare (intensity and location) characteristics and both CME characteristics (velocity and width) are available. Probabilities (in the range of 0 to 1) and respective errors (in the range of 0 to 1) have been derived after splitting our sample in the pre-defined flare intensity, flare location, CME velocity and CME width bins. The number size of each of these five sub-samples is also presented in the Tables. The derived probabilities and errors are also presented as a plot in Figure 32 and Figure 33. Again, the probabilities of SEP occurrence are usually higher for CMEs with widths in the range of [180,360] degrees.

Flare location [-90,-1]					
			CME velocity (km/s)		
Flare intensity	CME width		0-499	500-999	≥1000
M1-M3.9	[0,179]	N_{flares}	*48*	*53*	*7*
		Probability	**0**	**0**	**0.143±0.132**
	[180,360]	N_{flares}	*3*	*14*	*11*
		Probability	**0**	**0.357±0.128**	**0.182±0.116**
M4-M9.9	[0,179]	N_{flares}	*4*	*8*	*7*
		Probability	**0**	**0.125±0.117**	**0.143±0.132**
	[180,360]	N_{flares}	*0*	*5*	*15*
		Probability	**0**	**0.2±0.179**	**0.6±0.126**
≥X1	[0,179]	N_{flares}	*6*	*5*	*3*
		Probability	**0**	**0.2±0.179**	**0.333±0.272**
	[180,360]	N_{flares}	*0*	*3*	*14*
		Probability	**0**	**0**	**0.286±0.121**

Table 13: SEP occurrence probabilities and respective errors as a function of flare intensity and CME characteristics (velocity and width) for the [-90,-1] flare longitude bin. The total number of CME-associated flares per flare intensity, CME velocity and CME width bin for this flare location bin is also shown. Percentage probabilities can be derived by multiplying the derived values by a factor of 100.

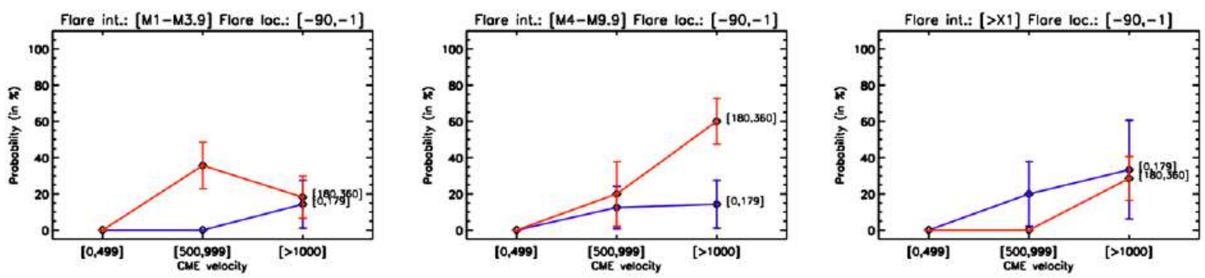


Figure 32: SEP percentage probabilities and respective errors as a function of flare intensity and CME characteristics for the [-90,-1] flare location bin.

Flare location [0,90]					
			CME velocity (km/s)		
Flare intensity	CME width		0-499	500-999	≥1000
M1-M3.9	[0,179]	N_{flares}	*52*	*47*	*12*
		Probability	**0.058±0.032**	**0.17±0.055**	**0.583±0.142**
	[180,360]	N_{flares}	*10*	*11*	*16*
		Probability	**0.1±0.095**	**0.727±0.134**	**0.75±0.108**
M4-M9.9	[0,179]	N_{flares}	*4*	*13*	*0*
		Probability	**0**	**0.231±0.117**	**0**
	[180,360]	N_{flares}	*2*	*7*	*10*
		Probability	**1**	**0.714±0.171**	**0.9±0.095**
≥X1	[0,179]	N_{flares}	*4*	*3*	*4*
		Probability	**0**	**0.333±0.272**	**0.75±0.217**
	[180,360]	N_{flares}	*2*	*8*	*27*
		Probability	**1**	**0.75±0.153**	**0.815±0.075**

Table 14: SEP occurrence probabilities and respective errors as a function of flare intensity and CME characteristics (velocity and width) for the [0, 90] flare longitude bin. The total number of CME-associated flares per flare intensity, CME velocity and CME width bin for this flare location bin is also shown. Percentage probabilities can be derived by multiplying the derived values by a factor of 100.

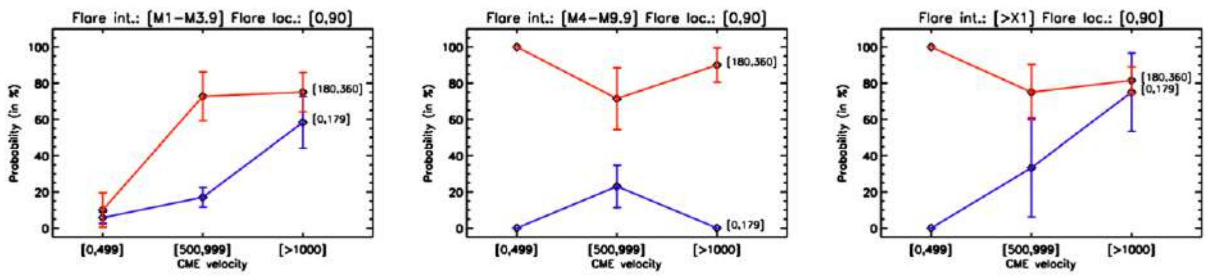


Figure 33: SEP percentage probabilities and respective errors as a function of flare intensity and CME characteristics for the [0, 90] flare location bin.

3.1.2 Magnitude of SEP events (BIRA-IASB)

This section describes the correlations derived between the flare and CME parameters on the one hand, and the impact of the SEP event on the other hand. The analysis is based on the sub-events recorded in the database described in section 2.2.7 and the solar event associations as determined for COMESEP Task 3.1 (see Crosby 2012). These correlations will be used to predict the strength of a potential SEP event. In the next sections, only a few example plots will be shown but the full set of figures can be obtained at ftp://ftp-ae.oma.be/D4.2/SEP_figures/.

3.2.2.1 Impact parameters

The impact of the SEP event is defined in different ways, as it is useful for different applications. The impacts studied here are:

- The peak flux and fluence in each of the 10 SEPEM proton energy channels.
- The integrated peak flux and fluence of protons above 10 MeV.
- The integrated peak flux and fluence of protons above 60 MeV.
- The integrated fluences for the ions He, C, N, O, Ne, Mg, Si, S and Fe.

This makes a total of 33 impact parameters for which the correlations are derived. The integrated peak fluxes and fluences are derived from the parameters obtained from the fit to their respective energy spectra (section 2.2.7.4). The distribution of the peak flux in the SEPEM reference channel (7.23 - 10.45 MeV) and the integrated flux above 60 MeV can be seen in Figure 34 for all 141 events.

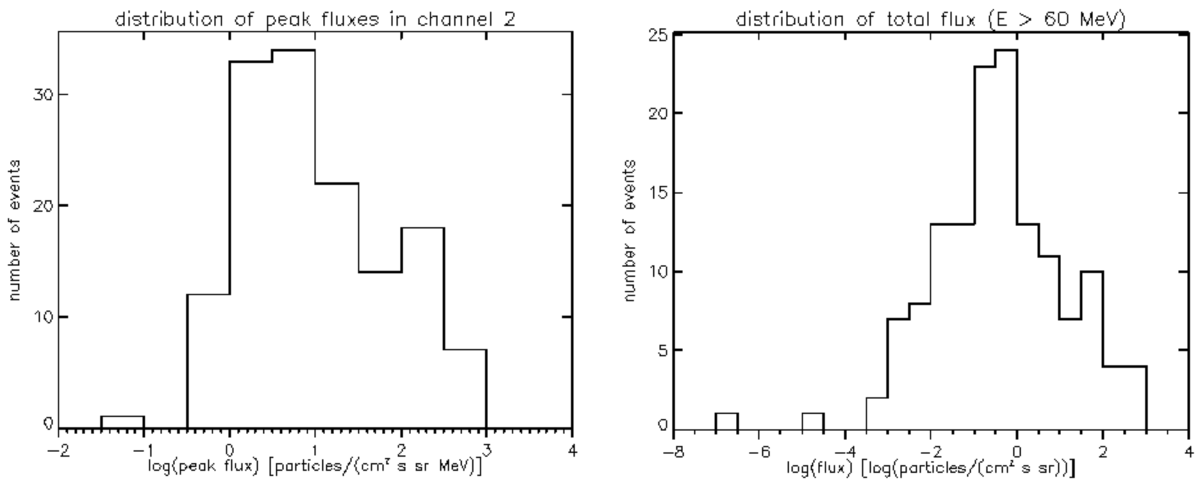


Figure 34: The distribution of the SEP peak flux in the SEPEM reference channel (left) and the integrated flux above 60 MeV (right) for all SEP events.

3.2.2.2 Solar event parameters

The following input parameters have been used to derive any possible correlation with the impact parameters. In order to better quantify the correlations, the data will be binned in these solar parameters. The distribution of these parameters for all events can be seen in Figure 35.

- Flare strength. In total 124 events have an associated flare with properly measured intensity. Due to the large variation, the logarithm of the flare intensity will be taken to derive the correlations with the impact parameters. Events with flare strength below C-class will not be considered. There is one event with flare intensity below this threshold (B4.7) which is behind the limb and not included in the statistics when including the flare strength. The peak flux in the SEPEM reference channel of that event is small, about 4/(s cm^2 sr MeV). The following 6 bins will be used: [-6.0,-5.5], [-5.5,-5.0], [-5.0,-4.5], [-4.5,-4.0], [-4.0,-3.5], [> -3.5] in units of log(W/m^2).
- Flare longitude. The flare location has been determined for 136 events. For events behind the limb, Cane et al. (2010) tried to estimate the flare location from the evolution of the active region. For western events, the maximum was set at 120°, hence the peak in the distribution. The longitude will be binned as follows: [< -90°], [-90°,-30°], [-30°,30°], [30°,90°], [> 90°].
- CME speed. CME associations were done for 119 events. Note there are no CMEs with speed below 400 km/s. To derive the correlations, the speed will be binned as follows: [400,800], [800,1200], [1200,1600], [1600,2000], [2000,2400], [2400,3000] in units of km/s.
- CME width. CME association with proper opening angle determination is available for 118 events. These values are from Cane el al. (2010) where they re-evaluate the opening angle for halo events and assign smaller values for part of them. The width will be binned as follows: [0°,90°], [90°,180°], [180°,359°], [> 359°]. The last bin corresponds to halo events.

3.2.2.3 Derived quantities

First, a scatter plot of the logarithm of the impact parameters versus the solar parameters is plotted, and the correlation coefficient ρ is calculated. The data is then binned in the solar parameter according to the bin sizes given above, and the average of both the solar and (logarithm of the) impact parameter is calculated for all the events in each bin. These values are then plotted again as the average impact parameter versus the average solar parameter. As uncertainties, the Root Mean Square (RMS) of the impact parameters in each bin is calculated.

The scatter and binned plots for the integrated flux above 10 MeV can be seen in Figure 36 for the correlation with the solar flare parameters, and in Figure 37 for the correlation with the CME parameters. There is a fairly good correlation with the flare intensity and the peak width, while there is only a small correlation between the flare longitude and CME width. The full list of correlations for all the impact parameters are given in Table 15, while the averages and RMS in each bin are given in Table 16 for the flare intensity and in Table 17 for the CME speed.

The above correlation and dependencies are derived for each parameter independently. In order to be useful for predictive purposes within an alert system, the correlations between each input parameter should be taken into account properly. Such a study is under way, following an approach of binning in multiple dimensions similar to what is used for the SEP probabilities. Another potential technique that might be explored is a Principle Components Analysis, where independent parameters are constructed from linear combinations of the input parameters.

The database also contains information about the ESP enhancement, namely the peak flux and fluence contributing to the SEP event. It is planned to also perform the statistical study described above to investigate if there is any correlation between these ESP characteristics and the associated solar flare and CME parameters. Further, it would be interesting to see if these ESP events can be linked with the events in the ICME list described in section 2.2.2 and if correlations can be found.

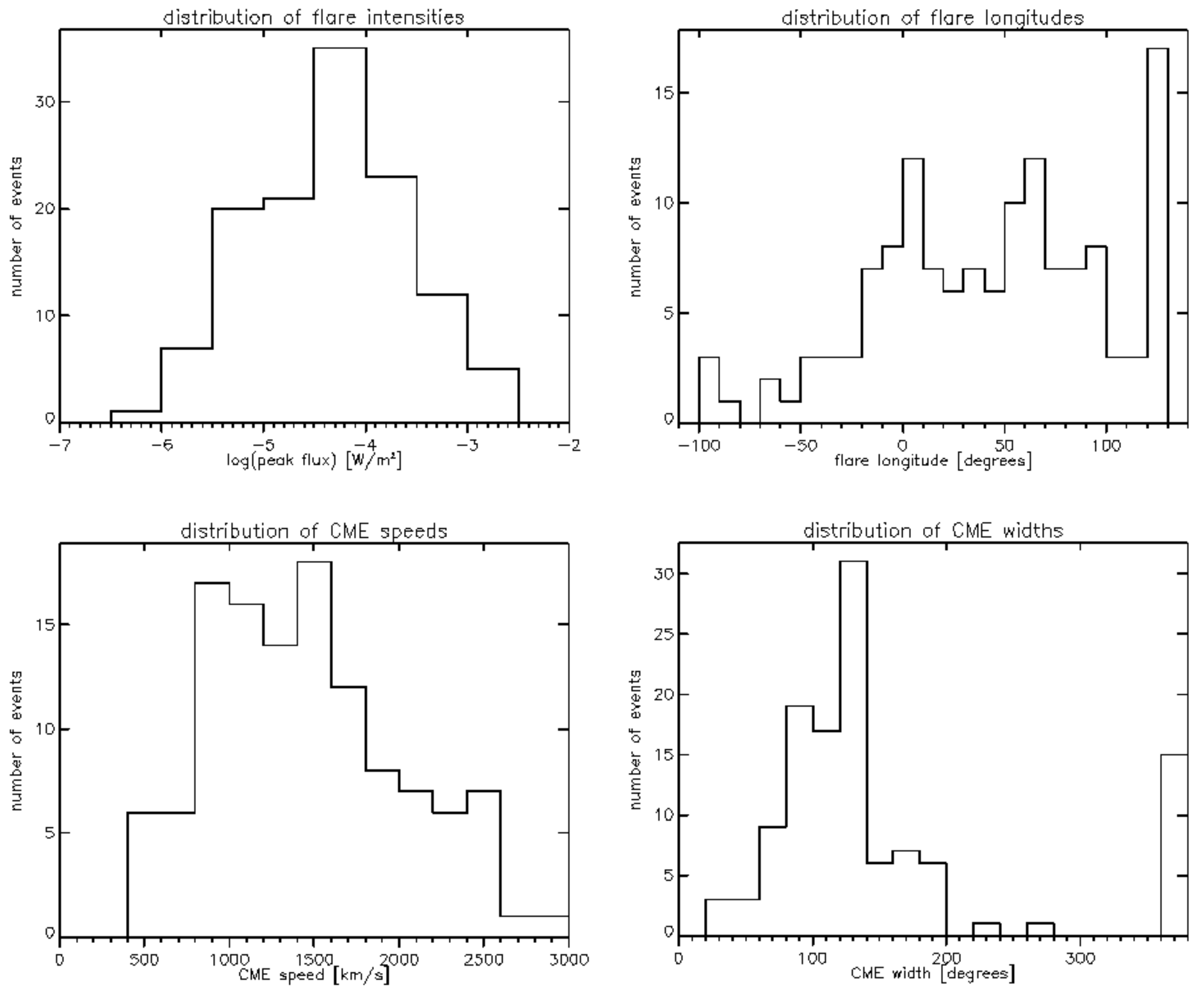


Figure 35: Distribution of solar parameters associated with SEP events: flare intensity (top left) and location (top right), and CME speed (bottom left) and width (bottom right).

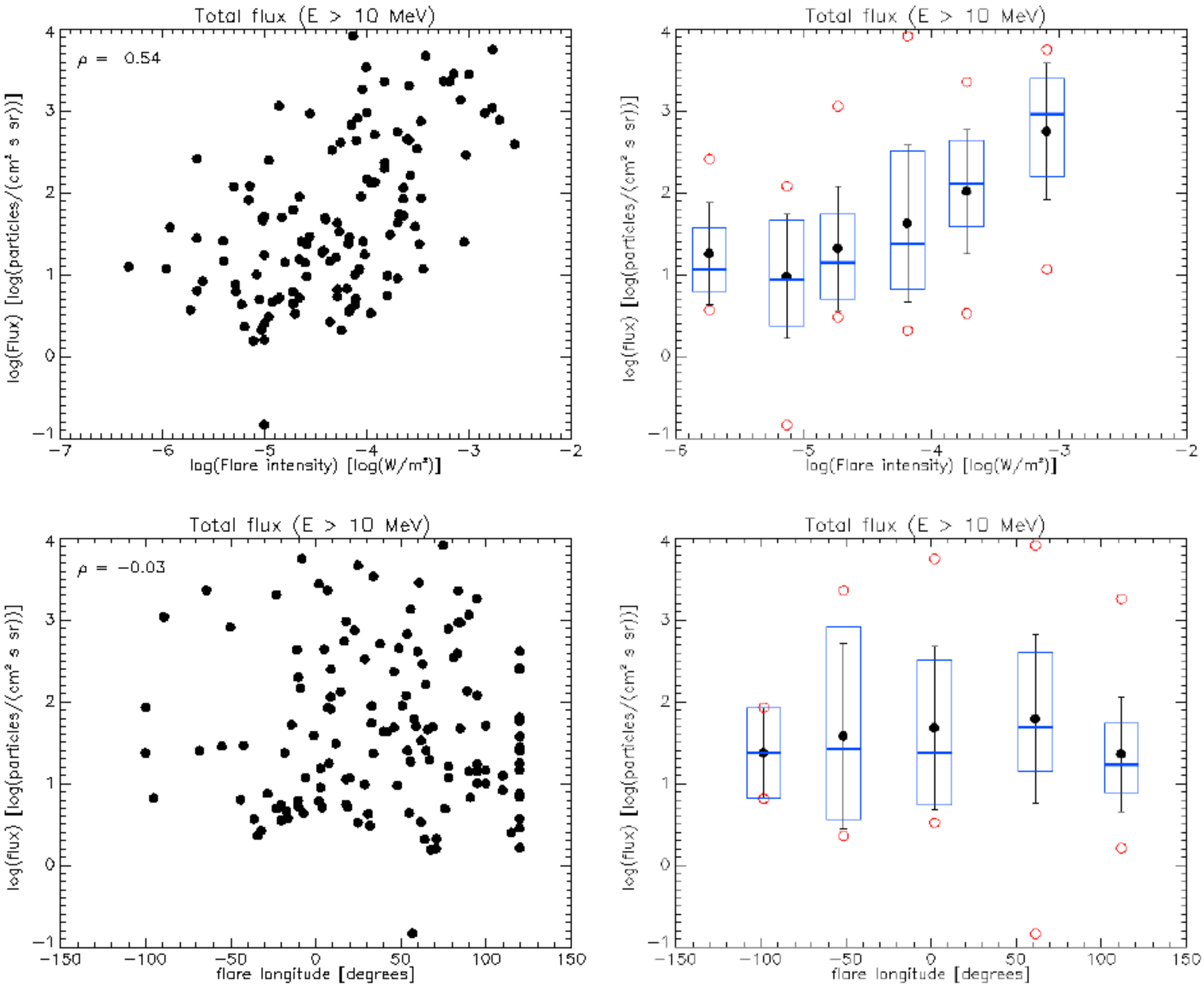


Figure 36: The scatter plots (left) and binned plots (right) for the logarithm of the SEP peak flux > 10 MeV as a function of the logarithm of the flare intensity (top row) and longitude (bottom row). The dots in the binned plots represent the average in that bin, while the error bars represent the RMS. The horizontal blue line shows the median, and the box shows where the second quartile starts, and the third quartile ends, so 50% of the events lie in that box. In case of an even number of events, the median (and the quartiles, which are the medians of the 2 half samples), is calculated as the mean of the two middle values. The red open circles show the outermost data points (outliers) in each bin.

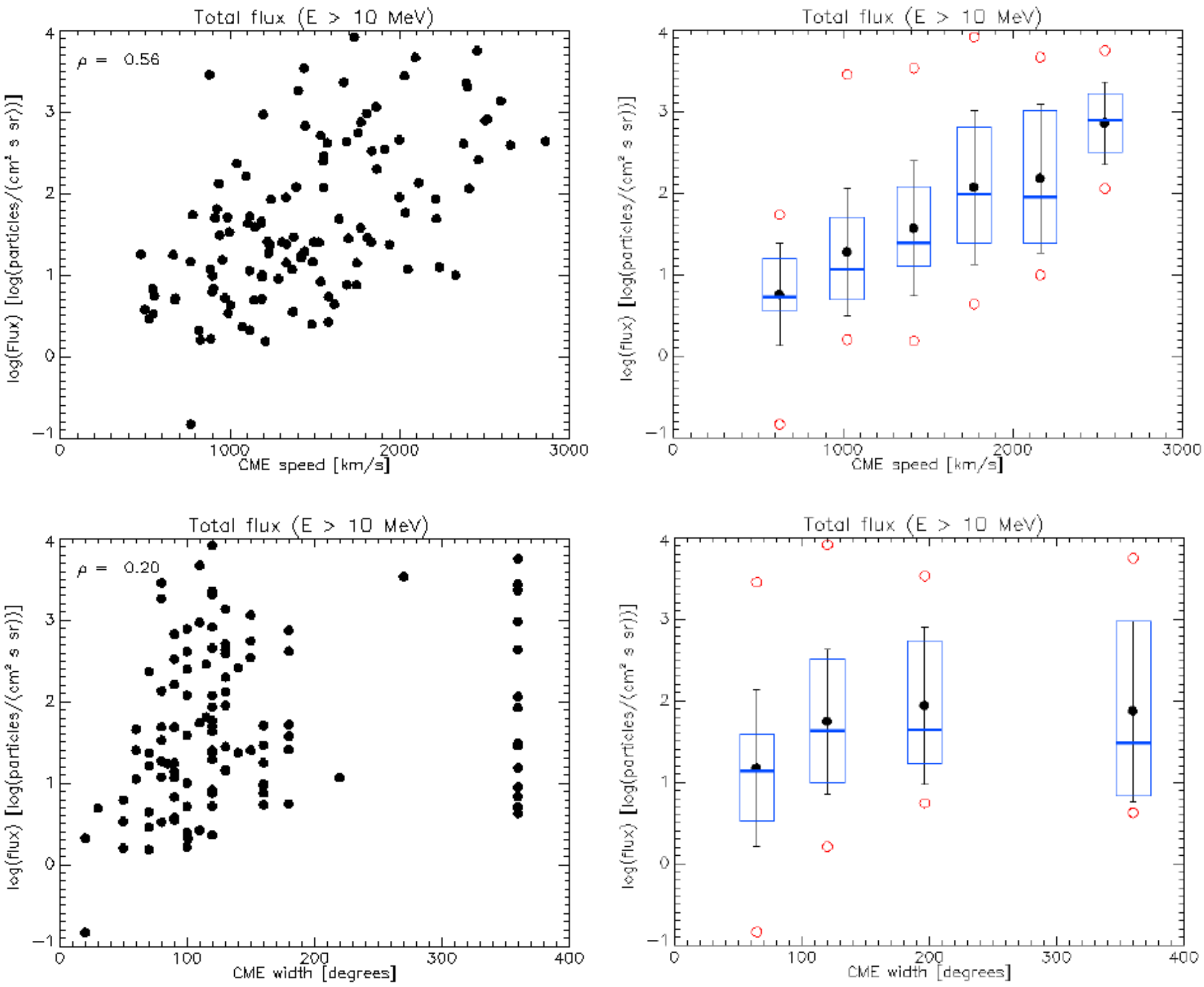


Figure 37: The scatter plots (left) and binned plots (right) for the logarithm of the SEP peak flux > 10 MeV as a function of the CME speed (top row) and width (bottom row). The dots in the binned plots represent the average in that bin, while the error bars represent the RMS. The horizontal blue line shows the median, and the box shows where the second quartile starts, and the third quartile ends, so 50% of the events lie in that box. In case of an even number of events, the median (and the quartiles, which are the medians of the 2 half samples), is calculated as the mean of the two middle values. The red open circles show the outermost data points (outliers) in each bin.

SEP impact parameter	Flare		CME	
	intensity	longitude	speed	width
Peak flux SEPEM channel 1	0.41	-0.13	0.52	0.20
Peak flux SEPEM channel 2	0.47	-0.08	0.54	0.21
Peak flux SEPEM channel 3	0.51	-0.03	0.55	0.21
Peak flux SEPEM channel 4	0.55	0.01	0.54	0.21
Peak flux SEPEM channel 5	0.57	0.04	0.53	0.20
Peak flux SEPEM channel 6	0.56	0.05	0.52	0.19
Peak flux SEPEM channel 7	0.57	0.05	0.46	0.14
Peak flux SEPEM channel 8	0.55	0.07	0.43	0.09
Peak flux SEPEM channel 9	0.54	0.05	0.35	0.16
Peak flux SEPEM channel 10	0.51	0.05	0.33	0.14
Fluence SEPEM channel 1	0.40	-0.18	0.53	0.20
Fluence SEPEM channel 2	0.45	-0.13	0.56	0.21
Fluence SEPEM channel 3	0.48	-0.10	0.57	0.22
Fluence SEPEM channel 4	0.51	-0.06	0.57	0.23
Fluence SEPEM channel 5	0.54	-0.02	0.57	0.21
Fluence SEPEM channel 6	0.55	-0.02	0.54	0.21
Fluence SEPEM channel 7	0.55	-0.01	0.51	0.18
Fluence SEPEM channel 8	0.54	-0.06	0.49	0.12
Fluence SEPEM channel 9	0.52	-0.05	0.44	0.21
Fluence SEPEM channel 10	0.44	-0.01	0.36	0.13
Total flux E > 10 MeV	0.54	-0.01	0.54	0.21
Total flux E > 60 MeV	0.57	0.12	0.44	0.15
Total fluence E > 10 MeV	0.50	-0.09	0.56	0.23
Total fluence E > 60 MeV	0.53	0.06	0.48	0.19
Total fluences He	0.48	-0.18	0.60	0.23
Total fluences C	0.48	-0.15	0.61	0.22
Total fluences N	0.49	-0.14	0.60	0.22
Total fluences O	0.50	-0.13	0.61	0.20
Total fluences Ne	0.52	-0.12	0.59	0.19
Total fluences Mg	0.51	-0.12	0.60	0.18
Total fluences Si	0.52	-0.09	0.59	0.18
Total fluences S	0.53	-0.05	0.55	0.14
Total fluences Fe	0.56	-0.07	0.49	0.13

Table 15: The correlations between the SEP impact parameters and the flare intensity and longitude, and the CME speed and width.

SEP impact parameter	Flare intensity bin [log(W/m^2)]					
	[-6.0,-5.5]	[-5.5,-5.0]	[-5.0,-4.5]	[-4.5,-4.0]	[-4.0,-3.5]	[> -3.5]
Peak flux SEPEM channel 1	1.14±0.49	1.04±0.80	1.45±0.58	1.35±0.74	1.68±0.78	2.14±0.71
Peak flux SEPEM channel 2	0.70±0.51	0.53±0.79	0.92±0.64	0.94±0.81	1.32±0.77	1.90±0.74
Peak flux SEPEM channel 3	0.30±0.56	0.11±0.72	0.41±0.74	0.59±0.88	1.01±0.75	1.66±0.78
Peak flux SEPEM channel 4	-0.16±0.72	-0.46±0.85	0.13±0.86	0.21±0.98	0.70±0.76	1.45±0.85
Peak flux SEPEM channel 5	-0.91±0.90	-1.24±0.82	-0.93±0.93	-0.48±1.05	0.06±0.79	0.89±0.93
Peak flux SEPEM channel 6	-1.32±0.97	-1.77±0.77	-1.38±0.90	-0.88±1.05	-0.36±0.78	0.57±1.02
Peak flux SEPEM channel 7	-1.50±0.80	-2.32±0.55	-2.09±1.05	-1.43±1.13	-0.89±0.83	0.11±1.13
Peak flux SEPEM channel 8	-2.05±0.98	-2.80±0.51	-2.32±0.87	-1.86±1.02	-1.50±0.86	-0.44±1.19
Peak flux SEPEM channel 9	-2.55±0.83	-3.30±0.44	-3.13±0.76	-2.50±0.95	-2.25±0.88	-1.09±1.28
Peak flux SEPEM channel 10	-3.08±1.02	-3.75±0.54	-3.40±0.79	-2.93±0.96	-2.71±0.91	-1.58±1.34
Fluence SEPEM channel 1	5.83±0.49	5.65±0.99	6.08±0.68	5.99±82	6.31±0.91	6.94±0.73
Fluence SEPEM channel 2	5.41±0.53	5.11±1.05	5.53±0.77	5.58±0.87	5.94±0.90	6.68±0.73
Fluence SEPEM channel 3	5.02±0.62	4.66±0.98	5.01±0.86	5.21±0.91	5.62±0.89	6.41±0.75
Fluence SEPEM channel 4	4.58±0.81	4.02±1.26	4.47±0.98	4.83±0.99	5.32±0.90	6.17±0.79
Fluence SEPEM channel 5	3.73±1.14	3.16±1.30	3.64±1.06	4.13±1.05	4.65±0.91	5.59±0.84
Fluence SEPEM channel 6	3.16±1.52	2.51±1.48	3.15±1.07	3.72±1.06	4.23±0.90	5.25±0.90
Fluence SEPEM channel 7	2.82±0.99	1.71±1.26	2.09±1.53	3.05±1.24	3.61±0.97	4.75±1.01
Fluence SEPEM channel 8	1.94±1.73	1.40±1.03	1.80±1.47	2.58±1.16	3.00±1.00	4.16±1.09
Fluence SEPEM channel 9	1.37±1.54	0.64±1.31	1.04±1.41	1.84±1.20	2.23±1.04	3.33±1.42
Fluence SEPEM channel 10	0.83±1.79	0.45±1.33	0.68±1.47	1.26±1.47	1.72±1.19	2.77±1.59
Total flux E > 10 MeV	1.25±0.62	0.97±0.76	1.32±0.76	1.58±0.93	2.02±0.76	2.77±0.86
Total flux E > 60 MeV	-1.02±1.85	-1.65±1.46	-1.20±1.19	-0.33±1.22	0.26±0.86	1.38±1.22
Total fluence E > 10 MeV	5.59±0.67	5.52±1.12	5.94±1.89	6.21±0.94	6.62±0.89	7.50±0.78
Total fluence E > 60 MeV	2.90±2.73	1.69±2.99	2.70±1.81	4.02±1.45	4.73±1.03	5.91±1.15
Total fluences He	4.61±0.73	4.69±0.94	5.00±0.76	5.14±1.16	5.52±1.17	6.55±0.85
Total fluences C	1.89±0.85	1.83±0.98	2.11±0.94	2.37±1.34	2.72±1.23	4.03±0.91
Total fluences N	1.29±0.80	1.23±0.93	1.50±0.97	1.78±1.39	2.16±1.25	3.54±0.99
Total fluences O	2.02±0.87	2.00±0.90	2.26±0.97	2.53±1.31	2.94±1.22	4.25±0.89
Total fluences Ne	1.04±0.91	1.12±0.82	1.37±0.97	1.61±1.32	2.10±1.21	3.42±0.93
Total fluences Mg	1.16±0.87	1.11±0.84	1.39±0.99	1.65±1.28	2.08±1.22	3.41±0.93
Total fluences Si	0.97±0.86	0.85±0.90	1.15±1.03	1.44±1.28	1.91±1.15	3.25±0.94
Total fluences S	0.19±0.84	0.17±0.75	0.36±1.11	0.70±1.26	1.21±1.09	2.55±0.99
Total fluences Fe	0.71±0.93	0.55±0.77	0.96±1.01	1.19±1.12	1.76±0.93	2.97±0.99

Table 16: The average of the logarithm of the impact parameter and RMS in each flare intensity bin.

SEP impact parameter	CME speed bin [km/s]					
	400-800	800-1200	1200-1600	1600-2000	2000-2400	2400-3000
Peak flux SEPEM channel 1	0.71±0.64	1.14±0.64	1.38±0.69	1.77±0.73	1.86±0.73	2.22±0.43
Peak flux SEPEM channel 2	0.22±0.60	0.70±0.66	0.97±0.74	1.42±0.79	1.51±0.80	1.90±0.52
Peak flux SEPEM channel 3	-0.16±0.50	0.32±0.70	0.61±0.79	1.09±0.87	1.18±0.85	1.60±0.63
Peak flux SEPEM channel 4	-0.65±0.70	-0.14±0.83	0.20±0.88	0.74±1.01	0.85±0.93	1.37±0.74
Peak flux SEPEM channel 5	-1.35±0.64	-0.87±0.95	-0.52±0.95	0.04±1.13	0.21±0.98	0.78±0.84
Peak flux SEPEM channel 6	-1.85±0.78	-1.24±0.98	-0.95±0.97	-0.40±1.22	-0.21±1.02	0.42±0.89
Peak flux SEPEM channel 7	-2.28±0.74	-1.76±1.09	-1.52±1.02	-1.07±1.49	-0.78±1.08	-0.07±0.95
Peak flux SEPEM channel 8	-2.90±0.91	-2.08±1.09	-2.04±0.96	-1.34±1.33	-1.36±1.04	0.73±1.00
Peak flux SEPEM channel 9	-3.31±0.73	-2.62±1.07	-2.72±0.87	-2.20±1.39	-2.05±1.03	-1.62±1.26
Peak flux SEPEM channel 10	-3.67±0.82	-2.96±1.12	-3.28±1.02	-2.41±1.33	-2.50±1.02	-1.84±0.97
Fluence SEPEM channel 1	5.23±0.93	5.74±0.73	6.02±0.78	6.50±0.77	6.52±0.86	6.98±0.45
Fluence SEPEM channel 2	4.71±0.99	5.27±0.73	5.60±0.80	6.14±0.84	6.16±0.91	6.65±0.53
Fluence SEPEM channel 3	4.31±0.88	4.85±0.74	5.22±0.84	5.80±0.93	5.81±0.96	6.34±0.63
Fluence SEPEM channel 4	3.73±1.23	4.37±0.87	4.82±0.91	5.44±1.06	5.50±1.02	6.09±0.72
Fluence SEPEM channel 5	3.00±1.10	3.56±1.09	4.08±0.98	4.73±1.17	4.86±1.06	5.48±0.80
Fluence SEPEM channel 6	2.39±1.37	3.11±1.24	3.62±1.06	4.24±1.30	4.41±1.09	5.09±0.86
Fluence SEPEM channel 7	1.65±1.21	2.44±1.40	2.80±1.28	3.44±1.71	3.76±1.17	4.56±0.92
Fluence SEPEM channel 8	0.90±1.52	2.07±1.41	2.22±1.35	3.15±1.34	3.16±1.08	3.91±0.94
Fluence SEPEM channel 9	0.16±1.47	1.56±1.32	1.39±1.33	2.22±1.63	2.37±1.13	3.08±1.01
Fluence SEPEM channel 10	0.02±1.59	1.14±1.43	0.74±1.62	1.86±1.60	1.91±1.18	2.57±1.01
Total flux E > 10 MeV	0.75±0.63	1.27±0.78	1.56±0.83	2.09±0.97	2.17±0.91	2.68±0.70
Total flux E > 60 MeV	-1.50±1.07	-0.80±1.69	-0.45±1.13	0.03±1.65	0.43±1.03	0.98±1.13
Total fluence E > 10 MeV	5.16±1.09	5.81±0.80	6.20±0.86	6.79±1.00	6.81±1.01	7.40±0.69
Total fluence E > 60 MeV	1.86±1.91	2.96±2.78	3.72±1.55	4.43±1.93	4.92±1.09	5.66±0.97
Total fluences He	3.98±0.94	4.70±0.86	5.11±1.00	5.77±0.08	5.82±1.04	6.63±0.62
Total fluences C	1.30±1.04	1.77±0.87	2.37±1.08	3.12±1.30	3.16±1.18	4.12±0.79
Total fluences N	0.73±1.03	1.19±0.87	1.77±1.13	2.53±1.41	2.59±1.22	3.64±0.89
Total fluences O	1.49±1.01	1.96±0.88	2.54±.08	3.25±1.32	3.37±1.17	4.34±0.75
Total fluences Ne	0.60±1.07	1.14±0.90	1.61±1.09	2.33±1.37	2.52±1.20	3.46±0.76
Total fluences Mg	0.65±1.02	1.14±0.87	1.64±1.07	2.35±1.37	2.50±1.18	3.49±0.76
Total fluences Si	0.49±1.02	0.93±0.92	1.43±1.09	2.11±1.39	2.31±1.18	3.33±0.75
Total fluences S	-0.01±0.90	0.27±0.89	0.70±1.11	1.30±1.48	1.57±1.16	2.57±0.77
Total fluences Fe	0.41±1.05	0.94±0.96	1.16±1.08	1.64±1.45	1.90±1.11	2.87±0.68

Table 17: The average of the logarithm of the impact parameter and RMS in each CME speed bin.

References

Bartels, J., (1932), Terrestrial-magnetic activities and its relation to solar phenomena, *Terr. Magn. Atmos. Electr., 37*, p1.

Belov, A., Garcia, H., Kurt, V., Mavromichalaki, H., and Gerontidou, M. (2005), Proton enhancements and their relation to the X-ray flares during the three last solar cycles, *Sol. Phys., 229*, p. 135.

Cane, H. V., Reames, D. V., and von Rosenvinge, T. T. (1988), The role of interplanetary shocks in the longitude distribution of solar energetic particles*, J. Geophys. Res., 93*, pp. 9555-9567.

Cane, H.V., Richardson, I.G., and Rosenvinge, T.T. (2010), A study of solar energetic particle events of 1997-2006: Their composition and associations*, J. Geophys. Res., 115*, A08101.

Cid, C., Cremades, H., Aran, A., Mandrini, C., Sanahuja, B., Schmieder, B., Menvielle, M., Rodriguez, L., Saiz, E., Cerrato, Y., Dasso, S., Jacobs, C., Lathuillere, C., and Zhukov, A. (2012), Can a halo CME from the limb be geoeffective?, *J. Geoph. Res., 117*, A11102.

Cliver, E.W. and Crooker, N.U. (1993), A seasonal dependence for the geoeffectiveness of eruptive solar events*, Solar Physics, 145*, pp. 347-357.

Cliver, E. W., Kamide, Y., and Ling, A. G. (2002), The semiannual variation of geomagnetic activity: phases and profiles for 130 years *aa* data, *J. Atmos. Sol.-Terr. Phys., 64*, pp. 47-53.

Clua de Gonzalez, A. L., Silbergleit, V.M., Gonzalez, W.D., and Tsurutani, B. T. (2002), Irregularities in the semiannual variation of the geomagnetic activity*, Adv. Space Res., 30*, pp. 2215-2218.

Cortie, A.L. (1912), Sunspots and terrestrial magnetic phenomena, 1898-1911: The cause of the annual variation in magnetic disturbances, *Mon. Not. R. Astron. Soc., 73*, p. 52.

Crosby, N., Dierckxsens, M., Malandraki, O., Patsou, I.-A., Papaioannou, A., Tzioziou, K., Dalla, S., and Marsh, M. (2012), Designing and set-up of the in situ data repository finished in Task 3.1, *COMESEP Milestone MS7 report*.

Dalla, S., and Agueda, N. (2010), Role of latitude of source region in Solar Energetic Particle events, *Twelfth International Solar Wind Conference, AIP Conference Proceedings, 1216*, pp. 613-616.

Emery, B. A., Richardson, I. G., Evans, D. S., Rich, F. J., and Wilson, G. R. (2011), Solar rotational periodicities and the semiannual variation in the solar wind, radiation belt, and aurora, *Solar Phys., 274*, pp. 399-425.

Farrugia, C.J. and Berdichevsky, D.B. (2004), Evolutionary signatures in complex ejecta and their driven shocks, *Annales Geophysicae, 22*, pp. 3679–3698.

Gopalswamy, N., Yashiro, S., and Akiyama, S., (2007), Geoeffectiveness of halo coronal mass ejections, *J. Geophys. Res., 112*, A06112.

Gussenhoven M.S., D. A. Hardy, and W. J. Burke (1981), DMSP/F2 electron observations ofequatorward auroral boundaries and their relationship to magnetosphereic electric fields, *J. Geophys. Res., 86*, p. 768.

Gussenhoven, M. S., Hardy, D. A., Heinemann, N. and Holeman, E. (1982), Diffuse auroral boundaries and a derived auroral boundary index, *Rep. AFGL-TR-82-0398, Air Force Geophys. Lab., Hanscom, AFB, Mass.*

Hollebeke, M. A. I. van, Ma Sung, L. S., McDonald, F. B. (1975), The variation of slar proton energy spectra and size distribution with heliolongitude, *Solar Physics, 41*, pp. 189-233.

Huttunen, K. E. J., Schwenn, R., Bothmer, V., and Koskinen, H. E. J. (2005), Properties and geoeffectiveness of magnetic clouds in the rising, maximum and early declining phases of solar cycle 23, *Annales Geophysicae, 23*, pp. 625–641.

Kim, R.-S., Cho, K.-S., Moon, Y.-J., Dryer, M.; Lee, J., Yi, Y., Kim, K.-H., Wang, H., Park, Y.-D., and Kim, Y.-H (2010), An empirical model for prediction of geomagnetic storms using initially observed CME parameters at the Sun, *J. Geophys. Res., 115*, p. 12108.

Kozyreva, O., Pilipenko, V., Engebretson, M. J., Yumoto, K., Watermann, J., and Romanova, N. (2007), In search of a new ULF wave index: Comparison of Pc5 power with dynamics of geostationary relativistic electrons, *Planet. Space Sci., 55*, pp. 755-769.

Lepping, R. P., Acũna, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., Mariani, F., Ness, N. F., Neubauer, F. M., Whang, Y. C., Byrnes, J. B., Kennon, R. S., Panetta, P. V., Scheifele, J., and Worley, E. M. (1995), The Wind Magnetic Field Investigation, *Space Sci. Rev., 71*, pp. 207-229.

McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., and Griffee, J. W. (1998), Solar Wind Electron Proton Alpha Monitor (SWEPAM) for the Advanced Composition Explorer*, Space Sci. Rev., 86*, pp. 563-612.

Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., Miller, G., Scudder, J. D., Sittler, E. C., Jr., Torbert, R. B., Bodet, D., Needell, G., Lazarus, A. J., Steinberg, J. T., Tappan, J. H., Mavretic, A., and Gergin, E. (1995), SWE, A Comprehensive Plasma Instrument for the Wind Spacecraft, *Space Sci. Rev., 71*, pp. 55-77.

Reames, D.V. (1999), Particle Acceleration at the Sun and in the Heliosphere*, Space Sci. Rev., 90*, p. 413.

Richardson, I.G. and Cane, H.V. (2010), Near-Earth Interplanetary Coronal Mass Ejections During Solar Cycle 23 (1996 – 2009): Catalog and Summary of Properties*, Solar Physics, 264*, pp. 189–237.

Richardson, I.G. and Cane, H.V. (2011), Geoeffectiveness (Dst and Kp) of interplanetary coronal mass ejections during 1995–2009 and implications for storm forecasting, *Space Weather, 9*, S07005.

Richardson, I.G., et al. (2006), Major geomagnetic storms (Dst≤-100 nT) generated by corotating interaction regions, *J. Geophys. Res., 111*, A07S09.

Roša, D., Vršnak, B., Božić, H., Brajša, R., Ruždjak, V., Schroll, A., and Wohl, H. (1998), A method to determine the solar synodic rotation rate and the height of tracers, *Solar Physics, 179,* pp. 237–252.

Russell, C.T., and McPherron, R. L., (1973), Semiannual variation of geomagnetic activity*, J. Geophys. Res., 78*, p. 92.

Schwenn, R., Dal Lago, A., Huttunen, E., and Gonzalez, W.D. (2005), The association of coronal mass ejections with their effects near the Earth, *Annales Geophysicae, 23*, pp. 1033–1059.

Smith, C. W., L'Heureux, J., Ness, N. F., Acuña, M. H., Burlaga, L. F., and Scheifele, J. (1998), The ACE Magnetic Fields Experiment, *Space Sci. Rev., 86,* pp. 613-632.

Srivastava, N. and Venkatakrishnan, P., (2004), Solar and interplanetary sources of major geomagnetic storms during 1996–2002, *J. Geoph. Res., 109*, A10103.

Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., and Snow, F. (1998a), The Advanced Composition Explorer, *Space Sci. Rev., 86,* pp. 1-22.

Stone, E. C et al. (1998b), The Solar Isotope Spectrometer for the Advanced Composition Explorer, *Space Sci. Rev., 86,* pp. 357-408.

Vennerstrom, S. (2000), Long-term rise in geomagnetic activity-A close connection between quiet days and storms, *Geoph. Res. Lett., 27*, pp. 69-72.

Vennerstrom, S. (2012), List of the largest magnetic storms [Task 4.3], *COMESEP Deliverable 4.1 report.*

Vršnak, B., Sudar, D., and Ruždjak, D. (2005), The CME-flare relationship: Are there really two types of CMEs*?, Astron. Astrophys., 435*, pp. 1149–1157.

Xapsos, M. A., Barth, J. L., and Stassinopoulos, E. G. (2000), Characterizing solar proton energy spectra for radiation effects applications, *IEEE Trans. on Nucl. Sci., 47*, pp. 2218-2223.

Zhang, J., Dere, K.P., Howard, R.A., and Bothmer, V. (2003), Identification of solar sources of major geomagnetic storms between 1996 and 2000, *Astroph. J., 582*, pp. 520–533.

Zhang, J., Richardson, I.G., Webb, D.F., Gopalswamy, N., Huttunen, E., Kasper, J.C., Nitta, N.V., Poomvises, W., Thompson, B.J., Wu, C.-C., Yashiro, S., and Zhukov, A.N. (2007), Solar and interplanetary sources of major geomagnetic storms (Dst 100 nT) during 1996–2005, *J. Geophys. Res., 112,* A10102.