

Connecting Coronal Mass Ejections to Their Solar Active Region Sources: Combining Results from the HELCATS and FLARECAST Projects

Sophie Murray, Jordan Guerra, Pietro Zucca, Sung-Hong Park, Eoin Carley, Peter T. Gallagher, Nicole Vilmer, Volker Bothmer

▶ **To cite this version:**

Sophie Murray, Jordan Guerra, Pietro Zucca, Sung-Hong Park, Eoin Carley, et al.. Connecting Coronal Mass Ejections to Their Solar Active Region Sources: Combining Results from the HELCATS and FLARECAST Projects. Solar Physics, 2018, 293 (4), pp.60. 10.1007/s11207-018-1287-4 . insu-01969976

HAL Id: insu-01969976
https://insu.hal.science/insu-01969976

Submitted on 12 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solar Physics
DOI: 10.1007/•••••-•••-•••-••••-•

Connecting Coronal Mass Ejections to their Solar Active Region Sources: Combining Results from the HELCATS and FLARECAST Projects

Sophie A. Murray[1] · Jordan A. Guerra[2] · Pietro Zucca[3] · Sung-Hong Park[4] · Eoin P. Carley[1,5] · Peter T. Gallagher[1] · Nicole Vilmer[5,6] · Volker Bothmer[7]

© Springer ••••

Abstract Coronal mass ejections (CMEs) and other solar eruptive phenomena can be physically linked by combining data from a multitude of ground-based and space-based instruments alongside models, however this can be challenging for automated operational systems. The EU Framework Package 7 HELCATS project provides catalogues of CME observations and properties from the *Heliospheric Imagers* onboard the two NASA/STEREO spacecraft in order to track the evolution of CMEs in the inner heliosphere. From the main HICAT catalogue of over 2,000 CME detections, an automated algorithm has been developed to connect the CMEs observed by STEREO to any corresponding solar flares and active region (AR) sources on the solar surface. CME kinematic properties, such as speed and angular width, are compared with AR magnetic field properties, such as magnetic flux, area, and neutral line characteristics. The resulting LOW-

✉ S. A. Murray
sophie.murray@tcd.ie

1 Astrophysics Research Group, School of Physics, Trinity College Dublin, Ireland

2 Department of Physics, Villanova University, Villanova PA, USA

3 ASTRON Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands

4 Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Japan

5 LESIA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universitès, UPMC Univ. Paris 06, Univ. Paris Diderot, Sorbonne Paris Citè, 5 place Jules Janssen, 92195 Meudon, France

6 Station de Radioastronomie de Nancay, Observatoire de Paris, PSL Research University, CNRS, Univ. Orleàns, 18330 Nancay, France

7 Institute of Astrophysics, University of Göttingen, Germany

arXiv:1803.06529v1 [astro-ph.SR] 17 Mar 2018

CAT catalogue is also compared to the extensive AR property database created by the EU Horizon 2020 FLARECAST project, which provides more complex magnetic field parameters derived from vector magnetograms. Initial statistical analysis has been undertaken on the new data to provide insight into the link between flare and CME events, and characteristics of eruptive ARs. Warning thresholds determined from analysis of the evolution of these parameters is shown to be a useful output for operational space weather purposes. Parameters of particular interest for further analysis include total unsigned flux, vertical current, and current helicity. The automated method developed to create the LOWCAT catalogue may also be useful for future efforts to develop operational CME forecasting.

Keywords: Active Regions, Magnetic Fields; Coronal Mass Ejections, Initiation and Propagation; Flares, Forecasting, Relation to Magnetic Field; Sunspots, Magnetic Fields

1. Introduction

Severe space weather has the potential to significantly impact a range of vital technologies, both on Earth and in near-Earth space. Adverse space weather is known to result from solar eruptions in the form of solar flares ,and coronal mass ejections (CMEs), as well as associated solar energetic particle events. These eruptions often originate from turbulent and highly complex active region (AR) magnetic fields. Understanding the processes involved in these solar eruptions, as well as their sources, is imperative to provide accurate space weather forecasts. CMEs are a main focus of current space weather operational systems considering their potentially severe impact on Earth, such as communication and navigation system disruption, power blackouts, and damage to spacecraft instrumentation (see *e.g.* Eastwood *et al.*, 2017, for an economic perspective). There are considerable research efforts underway to improve our understanding of CMEs to feedback to operational space weather forecasting efforts.

Catalogues of CME events have proven useful in order to better study their propagation. The NASA CDAW Data Center catalogue (Gopalswamy *et al.*, 2009) contains manually identified CMEs since 1996 from the *Large Angle and Spectrometric Coronagraph* (LASCO) onboard the *Solar and Heliospheric Observatory* (SOHO; Brueckner *et al.*, 1995) spacecraft. Automated catalogues such as CACTUS (Robbrecht and Berghmans, 2004) and CORIMP (Byrne *et al.*, 2012; Morgan, Byrne, and Habbal, 2012) have become increasingly popular in recent years, considering the difficult task of manually identifying the significant number of CME events that are observed in an age of improved availability of spacecraft observations. Much previous work has been undertaken to statistically analyze properties of the CMEs made available by these catalogues (see Webb and Howard, 2012, and references therein).

Most CME catalogues currently available have made use of the LASCO coronagraphs. However since 2007 the *Solar TErrestrial RElations Observatory* (STEREO; Kaiser *et al.*, 2008) mission has provided an unprecedented insight

into CME propagation, including a 3D perspective when the two STEREO A and B spacecraft were in the correct orbital position. Wide-angle imaging of the inner heliosphere has revolutionised our understanding of CMEs, solar wind, and co-rotating interaction regions (Harrison *et al.*, 2017). Heliospheric imaging provides the ability to track the evolution of these phenomena as they propagate through the inner heliosphere. Heliospheric Cataloguing, Analysis and Techniques Service (HELCATS) is a European Union Framework Package 7 project that provides the definitive catalogue of CME events observed by the STEREO *Heliospheric Imagers* (HI; Eyles *et al.*, 2009).

Thus far the results of the HELCATS project have been successful in enhancing our understanding of CMEs and their propagation (see *e.g.* Plotnikov *et al.*, 2016; Moestl *et al.*, 2017; Rouillard *et al.*, 2017). The results are particularly useful for improving CME arrival time predictions at Earth, and even at other planets in our solar system. However, forecasting when the CME itself will erupt remains elusive (Zheng, 2013). The more established field of flare forecasting now uses an extensive variety of methods to create predictions, from simple Poisson statistics (Bloomfield *et al.*, 2012) to more complex machine learning methods (Ahmed *et al.*, 2013). Unfortunately it is still unclear what methods are the best or even better than climatology (Barnes *et al.*, 2016), and many operational forecasting centers still rely on human forecasters to tailor predictions (Murray *et al.*, 2017a). Flare Likelihood And Region Eruption foreCASTing (FLARECAST) is a European Union Horizon 2020 project that aims to solve this uncertainty by creating an automated system that forecasts flares with unprecedented accuracy. In order to achieve such a goal, FLARECAST has created a large database of flaring activity predictors related to their AR sources and is undertaking a comprehensive evaluation of the many advanced prediction algorithms.

While there are a considerable number of different flare forecasting methods available, the vast majority of them use very similar predictors, namely some measure of the "complexity" of the solar surface AR magnetic field. It is natural to expect that AR magnetic fields may also prove to be key for CME forecasting purposes. Previous work has investigated correlations between AR photospheric magnetic field and CME properties (*e.g.* Venkatakrishnan and Ravindra, 2003; Guo, Zhang, and Chumak, 2007; Park *et al.*, 2012; Lee, Moon, and Lee, 2015; Tiwari *et al.*, 2015), as well as studied the locations of CME sources on the solar disk (Lara, 2008; Yashiro *et al.*, 2008; Wang *et al.*, 2011). The connection between CMEs and associated flares has also been extensively studied, particularly between the duration and intensity of the flares and CME speed (*e.g.* Harrison, 1995; Andrews, 2003; Youssef, 2012; Harra *et al.*, 2016). Statistical analyses of these AR properties are important to investigate the solar surface source of CMEs in order to make the first steps towards CME forecasting, just as the field of flare forecasting has done before it. However this work is limited by the observations available at the time of study and the considerable time it takes to manually associate the CMEs in these large databases with flares and ARs.

The main deliverables of the HELCATS project are focused on creating catalogue parameters related to STEREO CME observations in the heliosphere,

however the aim of HELCATS Deliverable 3.4 is to develop an automated algorithm to determine the solar surface source of the CME events. The algorithm is used to create the HELCATS LOWCAT catalogue, linking the STEREO/HI-observed CMEs to observations of associated solar flares and ARs, if any. This HELCATS catalogue provides an excellent opportunity for cross-collaboration with the FLARECAST project and their database of AR properties. It is worth noting that these projects focus on associations between CMEs and flares (and their AR sources). However CMEs can be also associated with other phenomena such as filament or prominence eruptions, with or without flares (see *e.g.* Moon *et al.*, 2002; Jing *et al.*, 2004).

This paper outlines the main algorithm steps in creating the LOWCAT catalogue, and the first initial scientific investigations using CME, flare, and AR data from both HELCATS and FLARECAST projects are also described. Section 2 describes the algorithm and resulting catalogue properties in detail, and highlights some initial results. Section 3 outlines the FLARECAST AR property database, and presents some example property comparisons between the two project databases. Discussion and conclusions are found in Section 4, particularly highlighting the availability of the data to the community and encouraging its use for future research efforts.

2. The LOWCAT Catalogue

2.1. Algorithm Development

With over 2,000 CME observations from mid 2007 until early 2017 available for analysis in the HICAT catalogue developed in HELCATS Work Package 2[1] (Barnes *et al.*, 2015), manually identifying each solar surface source for that many events was not feasible. To that end, an automatic algorithm has been developed that uses back-propagation to identify flare events and AR sources correlated with the CME events listed. The algorithm first associates the HI events to COR2 observations closer to the Sun, then to solar flare events, and finally to ARs on the solar surface. Note that while this simple method is sufficient for the purposes of this work, more complex methods could be used in future developments (see *e.g.* Davies *et al.*, 2013). Thomson scattering effects and 3D geometries need to be taken into account for events observed far away from the solar surface (Vourlidas and Howard, 2006; Davies *et al.*, 2012). This is particularly important for observations by HI-2, however only HI-1 observations from either STEREO A or B were used to create the HICAT catalogue.

Assuming constant radial velocity and constant CME width, a simple ballistic propagation model is used to search for candidates. First the algorithm identifies a time window during which a STEREO/HI-observed CME might have been observed by the STEREO/COR2 coronagraph. Here the initial distance can be taken as 12 $R_\odot$, final distance 2 $R_\odot$, and a typical range of CME speeds (150-1500 km) defined by Yurchyshyn *et al.* (2005) is used to constrain the search.

[1]see https://www.helcats-fp7.eu/catalogues/wp2_cat.html.

The CACTUS database, which contains CME events identified by an automated method, is searched for events occurring during this time window. The database was originally developed for LASCO/C2 observations, however was extended to include STEREO data during the HELCATS project (Pant *et al.*, 2016). The algorithm selects a COR2-observed CME with the closest angular width to the HI-observed CME, constrained by the north and south position angles.

Ballistic propagation is again used to define a time window of possible associated flares, using the identified COR2 CME event speeds where available. Here, the initial distance is taken as 2 $R_\odot$ and final distance can be set as high as 0.5 $R_\odot$ to take into account the non-constant speed of the initial CME phase (Byrne *et al.*, 2010). The algorithm searches for flare events in the National Oceanic and Atmospheric Administration Space Weather Prediction Center (NOAA/SWPC) Edited Solar Event Lists[2] by default. It will also check the NASA *Reuven Ramaty High Energy Solar Spectroscopic Imager* (RHESSI) flare list[3] if no events are found in the NOAA flare list, and for those events without any given location. The resulting flare list is further constrained by solar hemispheric location, ensuring the CME position angle corresponds to the quadrant where the potential flare peaks are located. If multiple flares are found within the search window the closest flare to the time window start is selected. The algorithm also limits the output to flares of *Geostationary Operational Environmental Satellite* (GOES) magnitude B-class and above, and for the RHESSI event list it allows only confirmed flare detections with a high-quality level defined in the list flags[4].

Finally, the algorithm associates any identified flare events with corresponding NOAA-numbered ARs. The flare peak location is used to search for nearby ARs on the solar disk if no number has been listed in the flare database. Some basic properties of these AR sources can be obtained from the NOAA/SWPC Solar Region Summaries, including the Hale (particularly the Modified Mount Wilson which includes δ spots; Künzel, 1965) and McIntosh (McIntosh, 1990) classifications.

2.2. Solar Monitor Active Region Tracker

In order to complete the HELCATS database, the Solar Monitor Active Region Tracker algorithm (SMART; Higgins *et al.*, 2011) is then run to obtain magnetic field properties of the identified ARs. SMART is an automated system for detecting, tracking, and cataloguing ARs throughout their evolution. The algorithm uses consecutive image differencing to remove quiet-Sun and transient magnetic features, then region-growing techniques to group flux concentrations into classifiable features. See Higgins *et al.* (2011) for more details about how the algorithm works. Figure 1 shows an example of a SMART detection on 21 June 2011. The dark blue lines in the Figure indicate the boundaries of the SMART AR detections, within which various magnetic field properties are calculated.

[2]see ftp://ftp.swpc.noaa.gov/pub/warehouse.

[3]see http://hesperia.gsfc.nasa.gov/hessidata/dbase/hessi_flare_list.txt.

[4]see http://sprg.ssl.berkeley.edu/$\sim$jimm/hessi/hsi_flare_list.html.

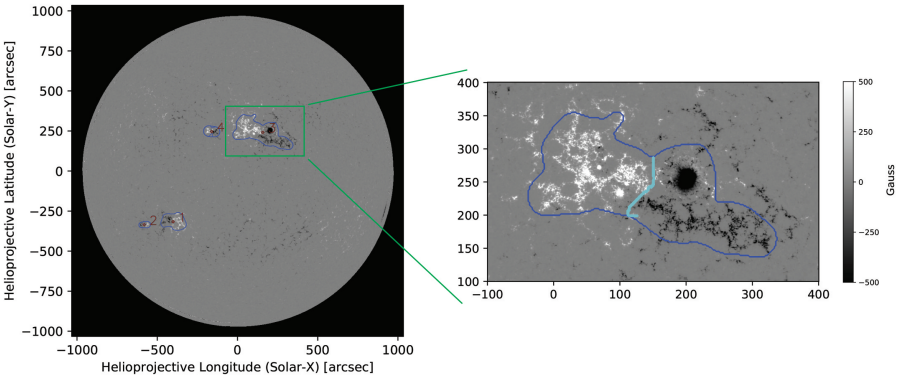


Figure 1. SMART detections for a HMI line-of-sight magnetogram on 21 June 2011 04:00 UTC (left). The zoomed-in region (right; green box on left) shows detection number 3, which corresponds to NOAA AR number 11236, and was the source of a C-class flare on 21 June 2011 as listed in the LOWCAT database. The dark blue line indicates the detection boundary, and light blue line the PIL.

SMART is run on regions using SOHO's *Michelson Doppler Imager* (MDI; Scherrer *et al.*, 1995) and *Solar Dynamics Observatory Helioseismic and Magnetic Imager* (SDO/HMI; Scherrer *et al.*, 2012) line-of-sight (LOS) magnetograms, depending on the date of the observations (HMI data being available from mid 2010). The flare peak time and location can be used to check for SMART regions even if no NOAA regions were identified. SMART calculates more complex magnetic properties related to the polarity inversion line (PIL), including PIL length, bipole separation, R value (Schrijver, 2007a) and $\mathrm{WL_{sg}}$ (Falconer, Moore, and Gary, 2008). Note that the gradient-weighted integral length of neutral line, $\mathrm{WL_{sg}}$, is defined as the line integral of the vertical field horizontal gradient over all PIL intervals where the value of potential horizontal field is greater than 150 gauss. This threshold is similarly used for the R value, which describes total unsigned flux near a strong-field, high-gradient PIL. A full list of AR information outputted by the algorithm is shown in Table 1.

Table 1. Magnetic source active region properties obtained from the SWPC Solar Region Summary and SMART algorithm outputs.

Source	Property	Unit	Description
NOAA	srs_no	-	SWPC-determined active region number of candidate region
NOAA	srs_mcintosh	-	SWPC-determined McIntosh classification of the group
NOAA	sr_hale	-	SWPC-determined Hale classification
NOAA	srs_area	millionths of a solar hemisphere	SWPC-determined active region total corrected area
NOAA	srs_ll	heliographic degrees	SWPC-determined longitudinal extent
NOAA	srs_nn	-	SWPC-determined total number of visible sunspots in the group
SMART	smart_totflx	maxwell	Total magnetic flux of SMART region
SMART	smart_posflx	maxwell	Total magnetic flux in positive polarity part of SMART region
SMART	smart_negflx	maxwell	Total magnetic flux in negative polarity part of SMART region
SMART	smart_frcflx	-	Flux fraction of SMART region, *i.e.* (posflx - negflx) / totflx
SMART	smart_totarea	millionths of a solar hemisphere	Total magnetic area of SMART region
SMART	smart_posarea	millionths of a solar hemisphere	Total magnetic area of positive polarity part of SMART region
SMART	smart_negarea	millionths of a solar hemisphere	Total magnetic area of negative polarity part of SMART region
SMART	smart_bmin	gauss	Total negative magnetic field strength of SMART region
SMART	smart_bmax	gauss	Total positive magnetic field strength of SMART region
SMART	smart_bmean	gauss	Mean magnetic field strength of SMART region
SMART	smart_psllen	megameters	Polarity inversion line length of SMART region
SMART	smart_bipolesep	megameters	Bipolar separation of SMART region
SMART	smart_rvalue	maxwell	R value of SMART region
SMART	smart_wlsg	gauss	WL_{sg} value of SMART region

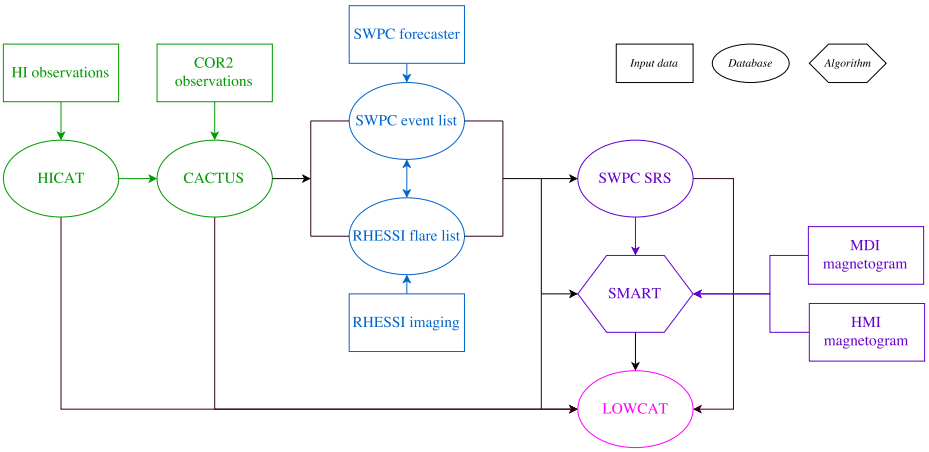


Figure 2. Flowchart outlining the steps involved in the automated algorithm used to create the LOWCAT catalogue. The event lists used are identified by ovals, input observations by rectangles, and the SMART algorithm by a diamond. Colours indicate steps in the main process, namely green for CME association, blue for flare association, purple for active region association, and magenta indicates the resulting LOWCAT catalogue.

2.3. Resulting Catalogue

All steps involved in the automated algorithm used to create the catalogue of properties are summarised by the flowchart in Figure 2. The resulting LOWCAT catalogue is freely available online at figshare.com[5] (Murray *et al.*, 2017b), which is also linked from the HELCATS website. The algorithm has been run with version 4 of the HICAT catalogue, which contains 2,020 HI-observed CMEs, the first of which launched on 15 April 2007 and the last on 26 February 2017. In total there are 1,591 COR2 events, 714 flare events, 552 NOAA regions, and 451 SMART regions in the LOWCAT catalogue besides the original 2,020 HI events. It is worth noting the SMART algorithm has also been run for the Framework Package 7 AFFECTS project KINCAT catalogue[6] as part of HELCATS Deliverable 3.3, which contains a small selection of manually identified CME events. The same properties listed in Table 1 are provided as well as an estimation of the tilt of the PIL.

The entries in the LOWCAT catalogue for every 50th event in the HICAT catalogue were manually inspected for verification purposes, checking that the correct COR2 observation (85% correct), flare event (87.5% correct), and AR (95% correct), if any, were listed. Missed CME and flare entries were generally due to being outside the time search window. Limb events were the only ones missed by the AR detection part of the algorithm for the events examined, although these could not have been analysed even if they were included. The results of this basic validation are positive for the automated method, and highlights the feasibility of such a simple method for big data analysis. The

[5]see https://figshare.com/articles/HELCATS_LOWCAT/4970222 .

[6]see https://www.helcats-fp7.eu/catalogues/wp3_kincat.html .

CME detection accuracy is comparable to the validation efforts for the CACTUS database, which reproduced ≈ 75% of the catalogued CMEs (Berghmans, Foing, and Fleck, 2002), although more events were analysed in this study.

The first part of the LOWCAT catalogue contains the ID of the HELCATS HICAT event, for example, the HCME_B_20120305_01 event, which has been previously been studied by Magdalenić *et al.* (2014), as well as the position and launch time of CME. The automated algorithm identified a COR2 CME from the CACTUS database to be associated with this particular event, starting at 2012-03-05 02:54 UT, which matches well with the expected launch times from the fixed-phi (02:06 UT), self-similar expansion (02:18 UT), and harmonic mean models (02:25 UT) as listed in the HICAT event information[7]. The LOWCAT catalogue lists properties of this event such as the COR2 CME mean speed, position angle, and width. The automated algorithm associated the COR2 event with a GOES X1.1 flare starting at 2012-03-05 02:30 UT from NOAA AR 11429. Flare magnitude, location, and start, peak, and end times are listed in the catalogue, followed by the SWPC and SMART properties listed in Table 1 as well as observation time and location.

The locations on the solar disk of all flare events and SMART locations in the LOWCAT catalogue are shown in Figure 3. The transition of regions towards lower latitudes throughout the solar cycle is clear from the colour bar indicating observation date. A large majority of flare events and their AR sources are from around 2012 - 2014. This is unsurprising considering the period of deep solar minimum around 2009 during which there were considerably less ARs on the solar disk. Note that there is an approximately one year gap in STEREO-A data from 2014 October while the spacecraft was behind the Sun, and data from STEREO-B is not available after 2014 September since contact with the spacecraft had not yet been re-established at the time of the end of the HICAT catalogue development. Comparison between the left and right plots also highlights the difficulty the SMART algorithm has detecting regions on the limb (and magnetic field analysis of these regions would not be as accurate anyway). However, the automated method still identifies ≈ 45% of the COR2 CME events to be associated with flare events, and ≈ 80% of those flare events have SWPC regions associated with them.

2.4. Initial Analysis

Figures 4 and 5 show the percentage occurrence of all the main properties in the LOWCAT catalogue. A typical range of CME speeds is found as expected from previous studies, confirming the choice of values for the automated algorithm, with a large amount of the CMEs in this catalogue ranging between ≈ 200 - 600 kms^{-1}. The histogram of COR2 CME width shows that there is only a small number of halo events in the catalogue. Most of the SMART properties highlight typical ranges for these AR characteristics, particularly those associated with flaring activity. Interestingly, some parameters are distributed such that most

[7]see https://www.helcats-fp7.eu/catalogues/event_page.html?id=HCME_B_20120305_01 .

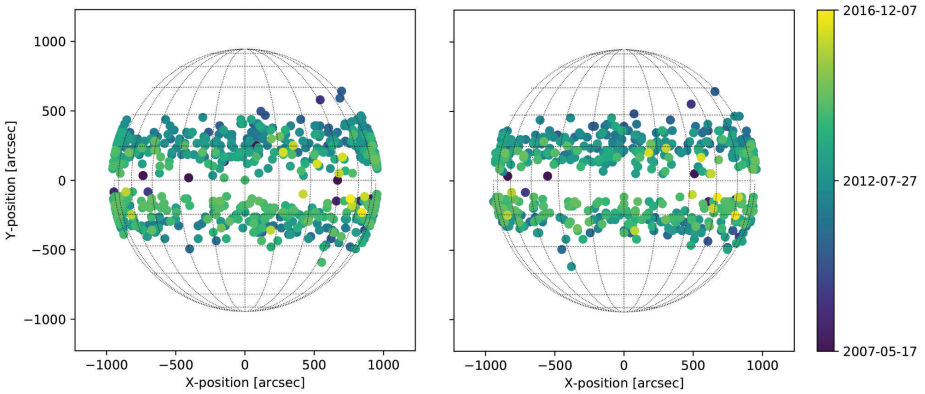


Figure 3. Locations of flare events (left) and SMART detections (right) identified in the LOWCAT catalogue. The colour bar indicates the day of flare peak (left) or magnetogram observation (right).

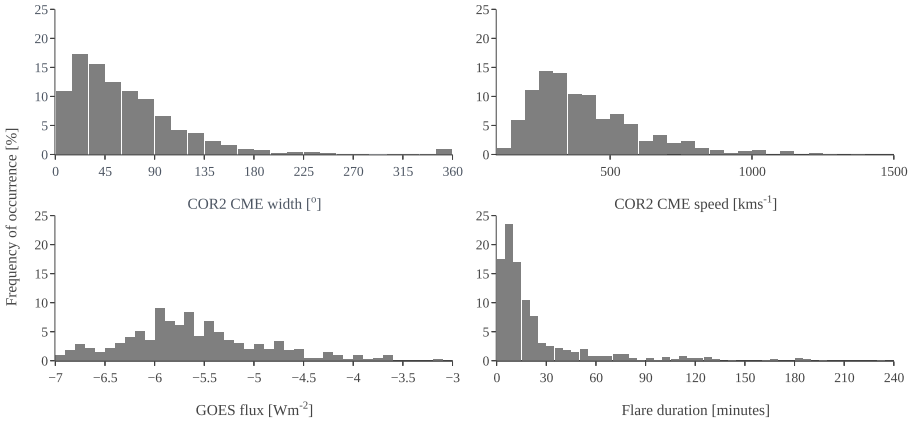


Figure 4. Percentage occurrence of a number of entries in the LOWCAT catalogue; the upper row shows COR2 CME properties, and lower row SWPC flare event properties. Bins left-to-right top-to-bottom are 15°, 50 kms^{-1}, 0.1 Wm^{-2}, and 5 minutes.

values exist above a certain value, for example the R and $\mathrm{WL_{sg}}$ values. This spurred some initial investigations into potential correlations between the CME, flare, and AR properties in the catalogue.

In order to confirm regions are being matched by the algorithm as expected (as a supplement to the manual checking), correlations between the flare event and NOAA numbers were examined. It is well established in the research community that large, complex ARs that have a history of flaring are likely to flare again, and AR properties are widely used for flare forecasting purposes (see *e.g.* Barnes *et al.*, 2016, and references within). Figure 6 shows the peak flare intensity *versus* the area of the LOWCAT regions as identified by SWPC forecasters, with colours indicating Hale class at the nearest time to the flare peak. The plot highlights that bigger flare events in the catalogue occur in bigger

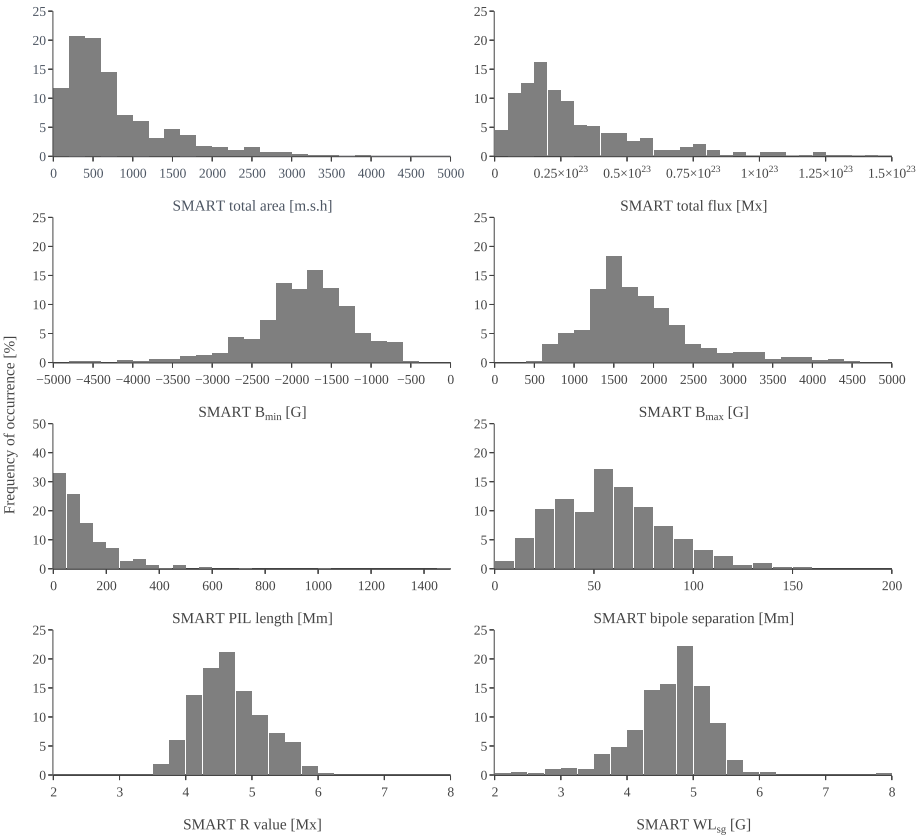


Figure 5. Percentage occurrence of the SMART properties available in the LOWCAT catalogue. Bins left-to-right top-to-bottom are 200 millionths of a solar hemisphere (m.s.h.), 5×10^{21} Mx, 200 G, 200 G, 50 Mm, 10 Mm, 0.25 Mx, and 0.25 GMm^{-1}.

regions with more complex Hale magnetic classification sunspot groups, and $\beta\gamma\delta$ regions produce the biggest flares in the catalogue. This is unsurprising, and is particularly reminiscent of Figure 2 of Sammis, Tang, and Zirin (2000), presenting the same axes and labels for 2,789 NOAA-numbered regions. It must be noted that in the LOWCAT catalogue several events may have the same source region, and the Sammis, Tang, and Zirin work plots the largest quantities at any time in the AR's lifetime (not necessarily at flare peak). Nevertheless, the expected results encouraged further investigations into other correlations that may be found between various LOWCAT catalogue properties.

Figure 7 presents comparisons between COR2 CME speeds (y-axes) and a selection of other LOWCAT properties (x-axes). The upper row in particular shows GOES flare peak intensities (left) and SMART region WL$_{sg}$ (right), with colour indicating COR2 CME width, such that yellow highlights halo CME events. It is clear that there is trend of larger flares associated with faster CMEs, although there is some amount of scatter. There is even more scatter found for WL$_{sg}$, however both plots show wider CMEs for larger property values with the

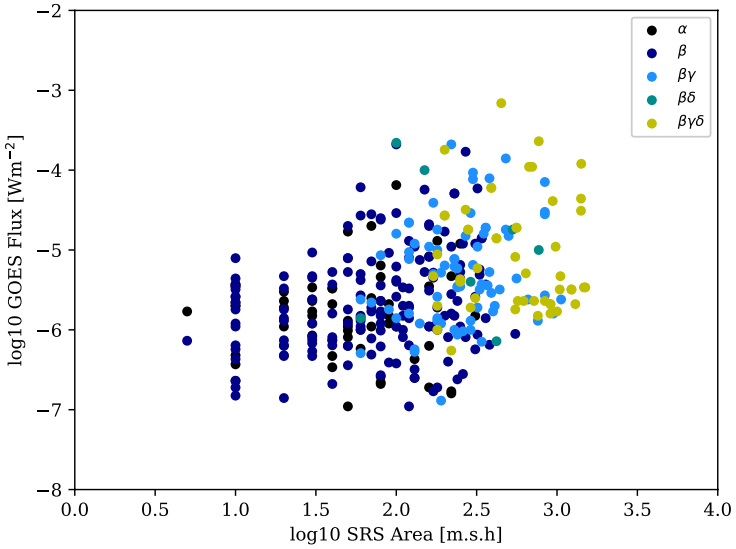


Figure 6. GOES flux *versus* NOAA/SRS sunspot group area, with colours indicating Hale class.

halo events tending towards the top right corner of the plot. This suggests the faster wider CMEs are associated with bigger flares and ARs with high WL_{sg} values. For example, if forecasters were only interested in CMEs with speeds above 600 kms^{-1}, the LOWCAT catalogue suggests flares greater than B class and a WL_{sg} value greater than 4 would be of interest. The other rows of Figure 7 show more SMART properties, with similar scatter to the upper row, however the colour now highlights the flare magnitude. X-class flares in yellow mainly tend towards the top right corner of the plots similar to the halo events but with more spread in values.

Previous work has suggested thresholds of importance for flare forecasting and CME forecasting alike may exist for these sorts of flaring AR properties (see *e.g.* Falconer, Moore, and Gary, 2008; Higgins, 2012). Although one must be careful interpreting these results considering the small number of halo CME and X-class flare events relative to the rest of the catalogue (see Appendix), the results of this initial investigation comparing the LOWCAT CME and AR properties suggests some of these parameters may indeed be useful for CME warning efforts. However, scatter plots can be difficult to interpret, particularly if hoping to draw conclusions regarding what parameters might be most useful for forecasting efforts. To better investigate this idea of warning thresholds, the FLARECAST project database was examined using different analytical techniques to expand upon the AR properties available for comparison.

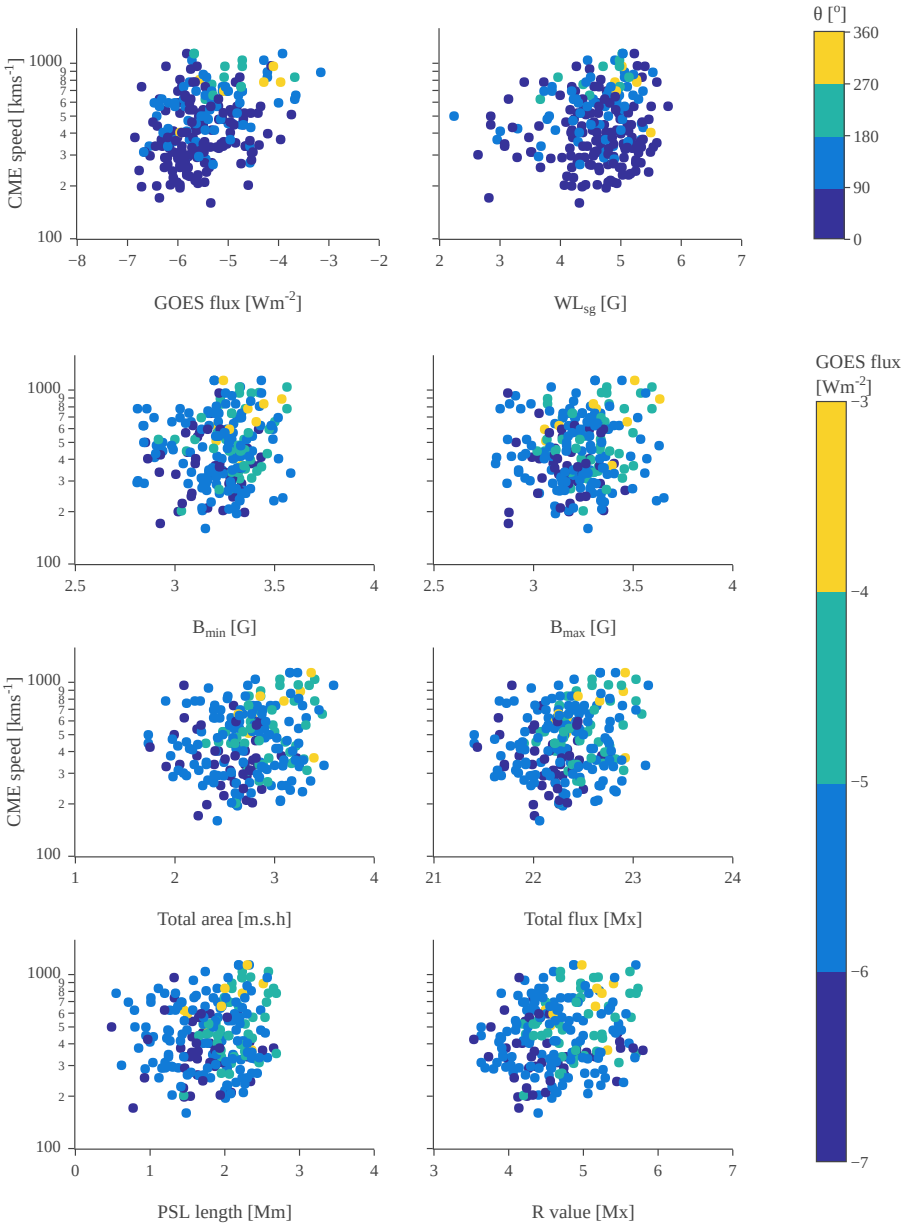


Figure 7. Selection of LOWCAT flare and active region properties *versus* CME speed. The upper row colourbar highlights CME width, whilst the colours in the other rows show flare peak intensity (B-class in dark blue, C-class in light blue, M-class in green, and X-class in yellow). Note that all y-axes are in log space, and x-axes in log10 space.

3. FLARECAST Active Region Properties

FLARECAST (flarecast.eu aims to create an automated system for forecasting solar flares with unprecedented accuracy by combining advanced machine learning algorithms with a comprehensive database of flaring activity predictors. The AR property database[8] consists of AR photospheric magnetic properties extracted from the Spaceweather HMI Active Region Patches (SHARP, Bobra *et al.*, 2014a). SHARP data products contain AR data maps obtained from SDO/HMI (Scherrer *et al.*, 2012) full-disk data. Since the FLARECAST system under development is planned to be operational, all magnetic properties are calculated from the near-real-time (NRT) version of the Cylindrical Equal Area (CEA) de-projection of SHARP data (series hmi.sharp_cea_720s_nrt). See Bobra *et al.* (2014a) for further details on the SHARP data products.

Table 2 lists the AR property groups included in the FLARECAST properties database. Each property group contains several properties. Property groups in Table 2, in turn, appear grouped according to the data source they are derived from (SWPC AR catalogues, SHARP vector, LOS magnetograms, and intensity maps). The property database encompasses different AR physical phenomena and areas, for example, magnetic polarity inversion lines, (multi-) fractal distribution of field, photospheric proxies for coronal connectivity, magnetic helicity, and photopsheric flows, among others. The majority of properties included in the FLARECAST property database have been reported as having a relevant association with the occurrence of flares (see references in Table 2) and others have been developed and tested in the framework of FLARECAST.

[8]see http://api.flarecast.eu.

Table 2. FLARECAST Active Region Magnetic Properties

Data source	Property group	Reference
SWPC catalogues	Solar Region Summary (SRS) properties GOES soft X-ray flare events	
Line-of-sight magnetograms	Effective connected magnetic field strength	Georgoulis and Rust (2007), Georgoulis (2010, 2013)
	Fractal and multifractal parameters	Conlon *et al.* (2008)
	Fourier and wavelet power spectral indices	Guerra *et al.* (2015), Abramenko *et al.* (2002), Hewett *et al.* (2008)
	Decay index	Liu (2008), Zuccarello *et al.* (2014)
	Magnetic polarity inversion line properties	Falconer *et al.* (2012), Schrijver (2007b), Mason and Hoeksema (2010)
	3D magnetic null points	Haynes and Parnell (2007), Pontin, Priest, and Galsgaard (2013)
	Ising energy	Ahmed *et al.* (2010)
	Magnetic helicity injection rate proxy	Park, Chae, and Wang (2010), Park *et al.* (2012)
Vector magnetograms	SHARP properties	Bobra *et al.* (2014b)
	Magnetic helicity injection rate	Berger and Field (1984)
	Magnetic energy injection rate	Kusano *et al.* (2002)
	Non-neutralized currents	Georgoulis, Titov, and Miki (2012)
	Diverging/converging/shear flows	Yang *et al.* (2004), Wang *et al.* (2014), Deng *et al.* (2006)
Intensity images	Flow field properties	Korsós, Baranyi, and Ludmny (2014)

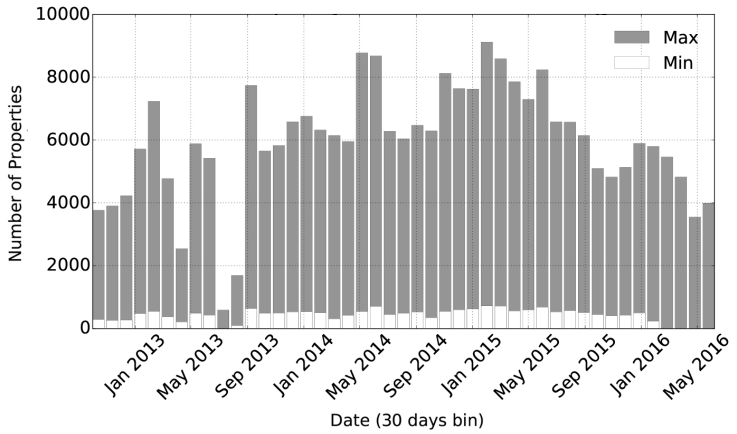


Figure 8. FLARECAST active region property database coverage. Property database covers the entire lifespan of the SHARP data repository (September 2012 - April 2016) with properties calculated every hour. Here the number of property values *per* 30 days is shown. Maximum (grey bars) and minimum (white bars) values for any 30-days bin are a consequence of different properties having longer computation times.

At present, the FLARECAST property database spans from the beginning of the SHARP data availability (September 2012) to April 2016[9]. Although SHARP data is produced at a 12-minute cadence, properties are calculated only hourly. Figure 8 displays the property data coverage. The number of property values present in the database is shown in time with a 30-days bin. Grey and white bars correspond to maximum and minimal numbers of properties present in each 30-days bin. Minimum and maximum values are due to the differences in computation time among different property-calculating algorithms. For properties that require longer computation times, their values can (at the moment) be found at cadences of 6, 12, and 24 hours. It is found that the maximum number of property values is quite small in June and July of 2013, mainly because many of SHARP data, in the FLARECAST HMI/SHARP service, have insufficient metadata pointing information for the property calculation.

3.1. Catalogue Comparisons

In order to cross-correlate the LOWCAT catalogue to the FLARECAST property database, an algorithm was developed that matches the source regions reported in the former to the regions listed in the latter. Source regions in the LOWCAT catalogue are labelled by their NOAA/SWPC AR number, if available, and/or SMART solar disk location. Regions in the FLARECAST database are primarily identified by HARP number, location, and NOAA number, if available. For each event in the LOWCAT catalogue registered after 1 September 2012, the algorithm searches for all entries in the FLARECAST

[9]Properties between this date and September 2017 need to be re-calculated due to defective NRT SHARP data.

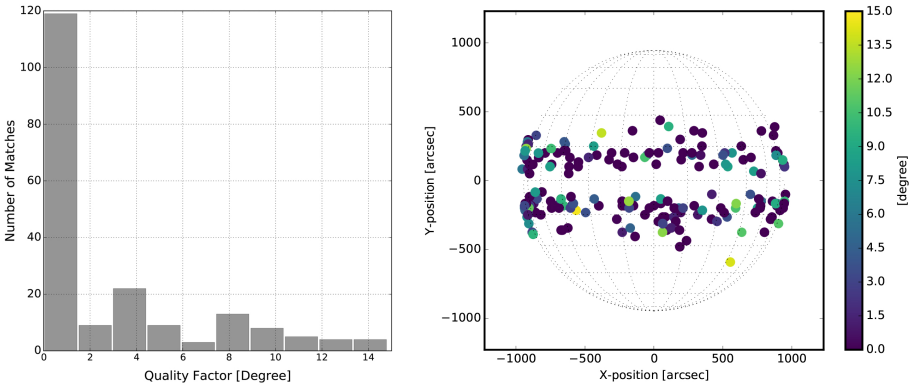


Figure 9. Quantitative measure for the location-based matching between LOWCAT CME source regions and FLARECAST property database regions. Left and right panels show the histogram and heliographic-spatial distribution of the quality factor. Majority of regions (81%) are matched within a 5-degree radius from each other and there is no apparent systematic bias with the heliographic location of regions.

database with time closest to the associated flare start or peak time, rounded to the hour before. Subsequently, the region matching is done by using the NOAA number (if available) or position. In the case of matching regions by position, a tolerance of 15 degrees in total position ($= [\text{longitude}^2 + \text{latitude}^2]^{1/2}$) is used. This implies that for a LOWCAT source region to be matched to a FLARECAST region, the former must be located within a circle of radius 15 degrees centred at the latter location. Once the LOWCAT source region has been matched to the FLARECAST region, any photospheric magnetic field property present in the database can then be associated to that LOWCAT event.

It is important to consider that SHARP regions can on some occasions include more than one NOAA region. Such cases are excluded from our analysis since properties can then be overestimated due to the presence of multiple ARs in the field of view. Between 1 September 2012 and 31 April 2016 a total of 812 HI and 665 COR2 CME events are registered in the LOWCAT catalogue. A total of 196 events are matched to FLARECAST regions, according to the criteria stated above. For each LOWCAT source region matched to a FLARECAST region, a matching quality factor is assigned. If the matching was done by means of the NOAA number, a zero value is assigned. On the other hand, if the source region is matched by location, the difference in distance between the regions is used as quality factor. he lower the quality factor the better, with zero values meaning the matching of regions was done exactly. Figure 9 shows the statistics of region matching quality. On the left, a histogram of the quality factor is displayed, while on the right the quality factor is displayed as a function of the heliographic (HG) position of the LOWCAT source regions. More than half ($\approx$ 120 or 60%) of matched regions are done so exactly or within 1 degree (Figure 9, left), although, 90% of the regions are matched with a quality factor less than ten degrees. On the other hand, no apparent systematic bias of the quality factor with the HG location is seen in Figure 9.

Table 3. Photospheric active-region properties. This table briefly describe the properties from the FLARECAST property database studied in this investigation. Keywords are used throughout as short names for properties.

Property	Description	Units	Keyword
$I_{z,\mathrm{tot}}$	Total unsigned vertical current ($= \sum \lvert J_z \rvert dA$)	A	usiz_tot
$I_{z,\mathrm{max}}$	Max unsigned vertical current ($= \max\{\lvert J_z \rvert dA\rvert\}$)	A	usiz_max
$\mathrm{WL_{SG}}(B_r)$	Gradient-weighted integral length of neutral line	G	wlsg_br
$R(B_r)$	Schrijver's R value from B_r	G	r_values_br
Φ_{tot}	Total unsigned flux density ($= \sum \lvert B_z \rvert dA$)	Mx	usflux_total
$H_{c,\mathrm{ave}}$	Average unsigned current helicity, B_z contribution ($= \frac{1}{N} \sum \lvert B_z J_z \rvert$)	$\mathrm{G^2\ m^{-1}}$	ushz_ave
$H_{c,\mathrm{tot}}$	Total unsigned current helicity, B_z contribution B_z contribution ($= \sum \lvert B_z J_z \rvert$)	$\mathrm{G^2\ m^{-1}}$	ushz_tot
$H_{c,\mathrm{max}}$	Max unsigned current helicity, B_z contribution B_z contribution ($= \max\{\lvert B_z J_z \rvert\}$)	$\mathrm{G^2\ m^{-1}}$	ushz_max
$E_{\mathrm{Ising}}(B_{\mathrm{los}})$	Ising energy from B_{los} ($= -\sum_{ij} \frac{S_i(B_{\mathrm{los}}) S_j(B_{\mathrm{los}})}{d^2}$)	$\mathrm{pixel^{-2}}$	ising_energy_blos
H_{m}	Magnetic helicity injection energy	$\mathrm{Mx^2\ hr^{-1}}$	abs_tot_dhdt
$(L/h_{\mathrm{min}})_{\mathrm{tot}}$	Sum of PIL-length over minimal height of critical decay index ratios		di4_br
$J_{z,\mathrm{max}}$	Max vertical current density ($\propto \{\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\}$)	$\mathrm{mA\ m^{-2}}$	jz_max
$\nabla_h(B_h)_{\mathrm{max}}$	Max horizontal gradient of horizontal field	$\mathrm{G\ Mm^{-1}}$	hgrad_bh_max
$\alpha(B_r)$	Fourier power spectral index		alpha_fft_br

Using the FLARECAST property database allows the investigation of a great variety of photospheric properties in relation to CME production. Of particular interest are properties that are extracted from vector magnetic field data, since only LOS magnetograms are analysed by SMART for the LOWCAT catalogue. Due to the large amount of the properties in the FLARECAST property database (over 100 at the time of publication), this investigation focused on a subset of properties which have been found to be the top predictors for flare occurrence, tested using four different machine learning methods for prediction (Campi and Benvenuto, 2017). Table 3 lists and describes the subset of studied properties selected. The resulting distribution of values for these properties in Table 3 are shown in Figures 10 and 11.

3.2. Superposed Epoch Analysis

This section presents an example of how the FLARECAST property database can be used along with the LOWCAT catalogue for understanding photospheric conditions leading to eruptive flares. One powerful resource of the FLARECAST property database is that time series of AR properties can be constructed to observe the evolution of a property before the flare-CME event. Due to the relatively high number of CME events matched to AR properties, empirical average evolution of properties can be determined by performing a superposed epoch analysis.

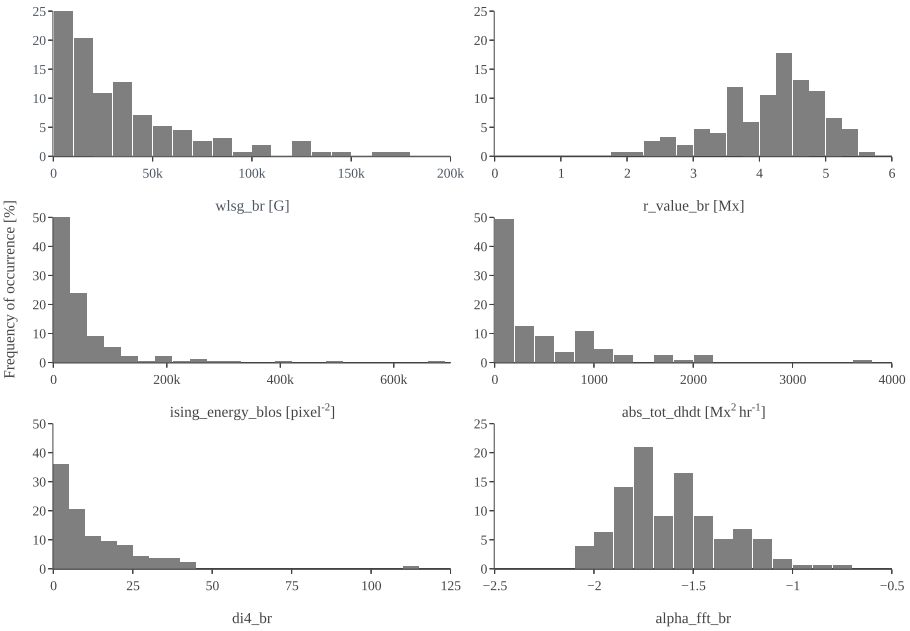


Figure 10. Percentage occurrence of the FLARECAST database active region properties used to compare with the LOWCAT catalogue. Bins left-to-right top-to-bottom are 10,000, 0.25, 30,000, 200, 5, and 0.1.

For all 196 LOWCAT events matched to FLARECAST regions, the 24-hour evolution prior to the eruption start time is saved for each studied property. No substantial difference was found using the peak instead of start time, since the cadence at which properties are registered is an hour and the difference between start and peak time in a solar soft-Xray flare is usually few to tens of minutes (Drake, 1971; Shibata and Magara, 2011). The superposed epoch analysis is performed separately for three groups according to the CME speed value: i) slow ($v_{\rm CME} < 420$ km^{-1}), ii) average-speed (420 km^{-1} $\leq v_{\rm CME} <$ 630 km^{-1}), and iii) fast ($v_{\rm CME} \geq 630$ km^{-1}). These boundary CME-speed values correspond to median($v_{\rm CME}$) and median($v_{\rm CME}$) + $\sigma(v_{\rm CME})$ values derived from the CME speed distribution (Figure 4 in Appendix section). In all related plots (*e.g.* Figures 12, 13, and 14), green, blue, and red curves are associated to the slow, average-speed, and fast CMEs groups, correspondingly.

For this analysis only events with $v_{\rm CME} \neq 0$ are considered from the HELCATS catalogue. These zero-speed events correspond to events that have an initial HI-identified CME but no associated COR2 events are found. Out of 196 matched CME-AR cases, 150 have non-zero CME speed. From these cases, 67 are slow CMEs, 50 average-speed CMEs, and 33 fast CMEs. It must be noted that not all cases are included in each CME-speed group, depending on the specific AR property in consideration. This variation is caused by incompleteness of the property time series – in some cases not all 24 hourly values are present in the property database, and therefore null values can be present. In order to make

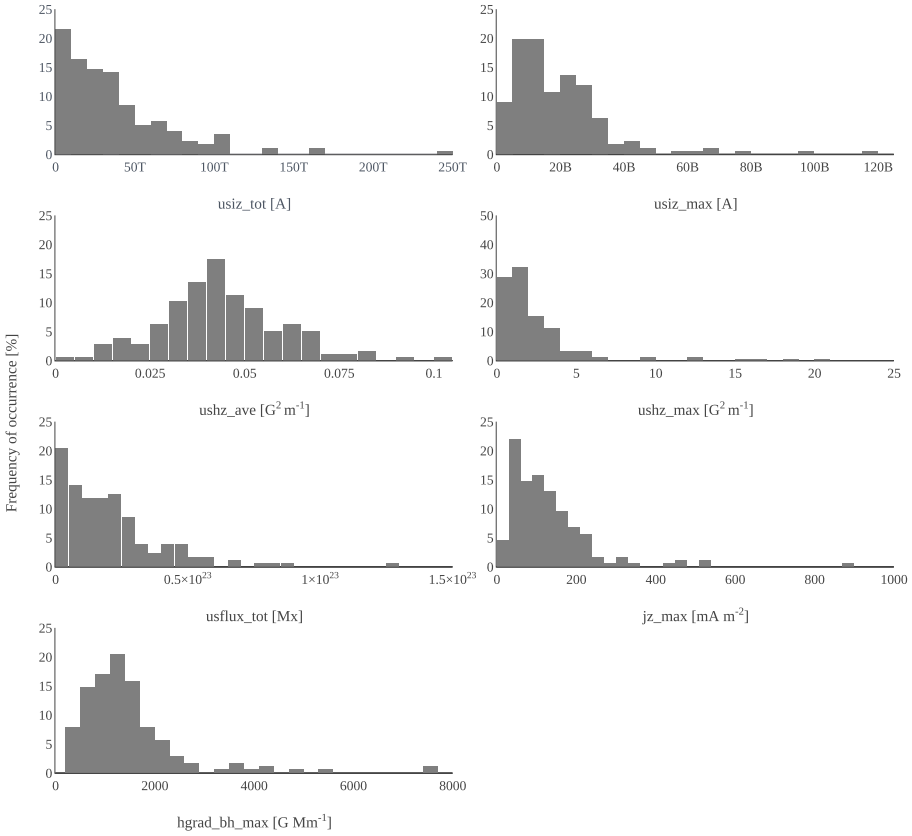


Figure 11. Percentage occurrence of the FLARECAST database active region properties, particularly SHARP properties, used to compare with the LOWCAT catalogue. Bins left-to-right top-to-bottom are 10^{11} A, 5×10^9 A, 0.005 $\mathrm{G^2 m^{-1}}$, 1 $\mathrm{G^2 m^{-1}}$, 5×10^{21} Mx, 30 $\mathrm{mA m^{-2}}$, and 300 $\mathrm{GM m^{-1}}$.

the present analysis and its results meaningful, a filtering of the time series is performed. In this investigation a similar procedure to Singh and Badruddin (2006) is followed, which suggests the F test can be used to determine the statistical significance of the average time series. In this test, the sample is assumed to have a F-distribution, and the null hypothesis is rejected at a particular level of confidence. See Singh and Badruddin (2006) and references therein for more details.

The average time series depends on the number of samples, k, included in each v_{CME} group. In this case, k is determined by the maximum number of null values allowed in any time series, n_{null}. As a first approximation to this selection procedure, $n_{\mathrm{null}} < 12$, is used. That is, any sample with more than half of their epochs as valid values, is considered. For any null value in the selected time series, interpolated values are calculated. Once k is determined in each CME-speed group for each property, the average time series and its associated F-value is calculated. Finally, F_α, the tabulated F-distribution value, is calculated using

Table 4. Photospheric properties showing statistically significant average temporal behaviour from 24 hours prior to CME-associated flares. Properties are listed according to their results in the F-test. From top to bottom the number of significant average time series *per* property *per* CME-speed groups decreases.

Property	CME-speed group	F	# of curves	F-test result
usflux_total	Slow	2.44	52	Significant
	Medium	1.36	39	Significant
	Fast	4.99	32	Significant
usiz_tot	Slow	1.23	54	Insignificant
	Medium	1.36	44	Significant
	Fast	1.99	33	Significant
ushz_tot	Slow	0.85	52	Insignificant
	Medium	4.82	39	Significant
	Fast	2.54	33	Significant
wlsg_br	Slow	1.28	52	Significant
	Medium	1.77	39	Significant
	Fast	0.64	33	Insignificant
r_values_br	Slow	1.13	54	Insignificant
	Medium	1.92	44	Significant
	Fast	0.93	33	Insignificant
usiz_max	Slow	0.77	52	Insignificant
	Medium	1.70	39	Significant
	Fast	0.91	33	Insignificant
alpha_fft_br	Slow	1.68	54	Significant
	Medium	0.96	44	Insignificant
	Fast	0.61	33	Insignificant

$n-2$ and $k(n-2)$ degrees of freedom at $\alpha = 95\%$ confidence, where $n = 24$ is the number of epochs. For those properties/CME-speed groups where $F > F_\alpha$, the average time series is then statistically significant at α level. Table 4 summarizes the results from the F test, for all properties that showed significant results in at least one CME-speed group.

Figures 12, 13, and 14 show the superposed epoch plots for the top three performing properties from Table 4, namely usflux_total, usiz_tot, and ushz_tot. Figure 12 (left panels) shows that the initial value of total unsigned flux ($\Phi_{\rm tot}$, usflux_tot), which is measured 24 hours before the eruption start time, has a dependence on the $v_{\rm CME}$-group. The minimal initial value in each group increases with the CME-speed groups: slow CMEs show a minimal initial value of $\approx 1.0 \times 10^{20}$ Mx cm^{-2}, medium-speed CMEs show $\approx 1.0 \times 10^{21}$ Mx cm^{-2}, and fast CMEs show $\approx 3.0 \times 10^{21}$ Mx cm^{-2}. This finding is consistent with the well-known concept that AR containing large amounts of magnetic flux are prone to eruptive activity. However, and according to Figure 12, the higher the total unsigned flux 24 hours before eruption, the bigger the chance of this eruption to be associated to fast CME speeds. On the other hand, Figure 12 (right) shows that in all three CME-speed groups usflux_total average time series seem

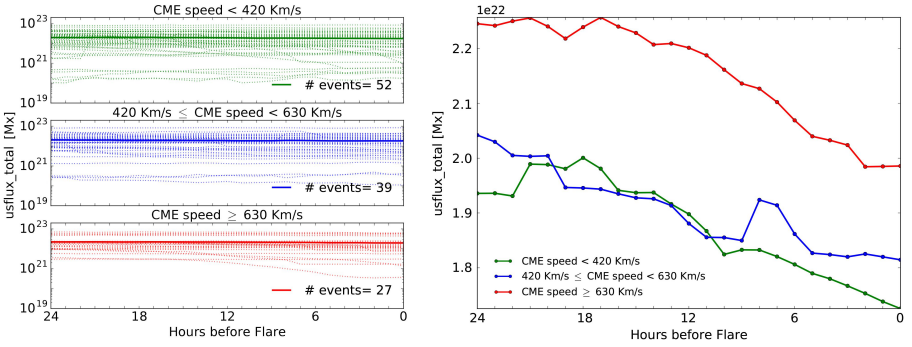


Figure 12. Superposed epoch analysis for the total unsigned magnetic flux ($\Phi_{\rm tot}$, usflux_total). In the left plot, three panels show the superposed time series for each CME-speed group: slow (top, green), medium (middle, blue), and fast (bottom, red). The plot on the right displays the average time series for each group in the left panels. Included time series and averages are determined to have statistical significance according to the F test.

to systematically decrease over the 24-hour period prior to the eruption. This trend could be a manifestation of decaying ARs. Average time series for slow and average-speed CMEs (green and blue) show close values between hours 21 and 9 before eruption, but for any other epoch AR that produced average-speed CMEs show higher $\Phi_{\rm tot}$ values. Values of $\Phi_{\rm tot}$ for fast-CMEs associated ARs (red) are well above the two other categories.

Corresponding superposed epoch analysis for the total unsigned vertical current ($I_{z,\rm tot}$, usiz_tot) is shown in Figure 13. As in the previous case, the minimal initial value of $I_{z,\rm tot}$ seems to increase with the CME-speed group: $2{\times}10^{11}$ A, $5{\times}10^{11}$ A, and $5{\times}10^{12}$ A for slow, average-speed, and fast CMEs, respectively. In comparison to $\Phi_{\rm tot}$, more variations of $I_{z,\rm tot}$ (up to several order of magnitudes) can be seen in all three panels. This variation could be spurious, a consequence of the interpolation done for null values. Although, their appearance on the average time series (Figure 13, right) for average-speed and fast CMEs (the statistically significant cases for this property) suggests that rapidly changing vertical currents could be present in regions before erupting. This effect, real or artefact, could be better studied by constructing higher-cadence time series, particularly making use of the highest SHARP data cadence of 12 min. It is important to remark that statistically significant time series in Figure 13 (right, blue and red curves) decrease in value with increasing $v_{\rm CME}$.

Figure 14 displays the superposed epoch analysis for the total unsigned current helicity due to the B_z component ($H_{c,\rm tot}$, ushz_tot). As in previous cases, time series of $H_{c,\rm tot}$ *per* CME-speed group (Figure 14, left) show the spread over the property value's range decreasing from slow to medium-speed to fast CME speeds. In particular, minimal initial values (24 hours before eruption) of total B_z-contribution unsigned helicity are observed to increase with the CME-speed group: $\lesssim$ 10 G^2 m^{-1} for slow CMEs, $\approx$ 20 G^2 m^{-1} for medium-speed CMEs, and 200 G^2 m^{-1} for fast CMEs. The level of hourly variations within the 24-hour study period appear to be in between those of $I_{z,\rm tot}$ and $\Phi_{\rm tot}$. This last finding becomes clearer by inspection of the average time series/CME-speed

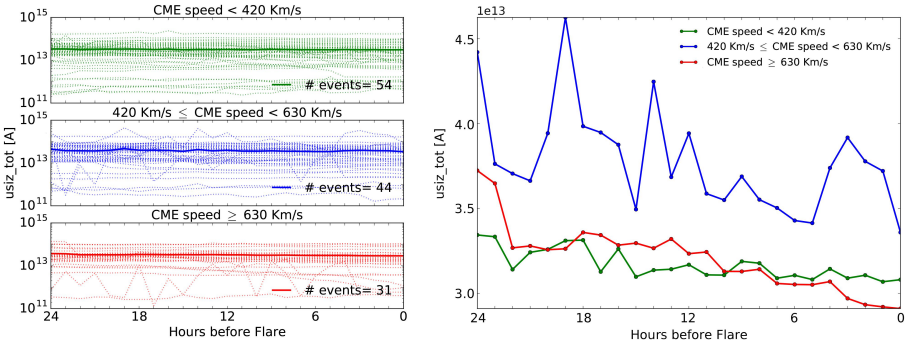


Figure 13. Same as Figure 12 but for total unsigned vertical current ($I_{z,\mathrm{tot}}$, usiz_tot).

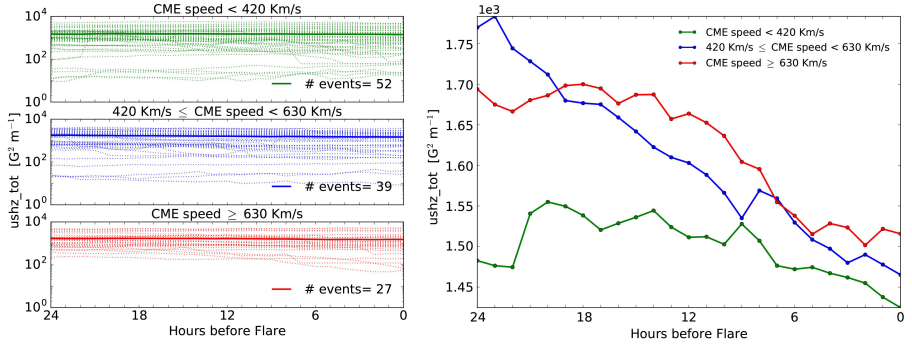


Figure 14. Same as Figure 12 and 14 but for total unsigned B_z-contribution current helicity ($H_{c,\mathrm{tot}}$, ushz_tot).

group (Figure 14, right). The average time series for medium-speed CMEs (blue) is also observed to systematically decrease over a day before the eruption. On the other hand, fast and slow CMEs (red and green, respectively) appear to have a seemingly flat period, between hour 21 and 9, then a decrease over two hours, follow by a less steep decrease from 7-6 hours before eruption. Although, only red and blue curves are statistically significant according Table 4. More AR properties could display statistically significant results in the superposed epoch analysis by performing a better filtering of the timeseries sample.

Time series of photospheric AR properties associated to eruptive events allows us to investigate statistical associations between CME characteristics (speed and/or angle) with parametrizations of the 24-hour distribution of values. It is clear from Figure 7 that scatter plots display large levels of dispersion to claim any (linear or non-linear) correlation between the variables. As an alternative, box-and-whisker plots allow a better defined association between CME characteristics and photospheric properties. Figure 15 presents the box-and-whisker plots for CME speed, v_{CME}, *versus* 24-hours Φ_{tot} distribution mean, median, and standard deviation (top row) and minimum, maximum, and initial values (bottom row). Box-and-whisker plots present groups of numerical data through

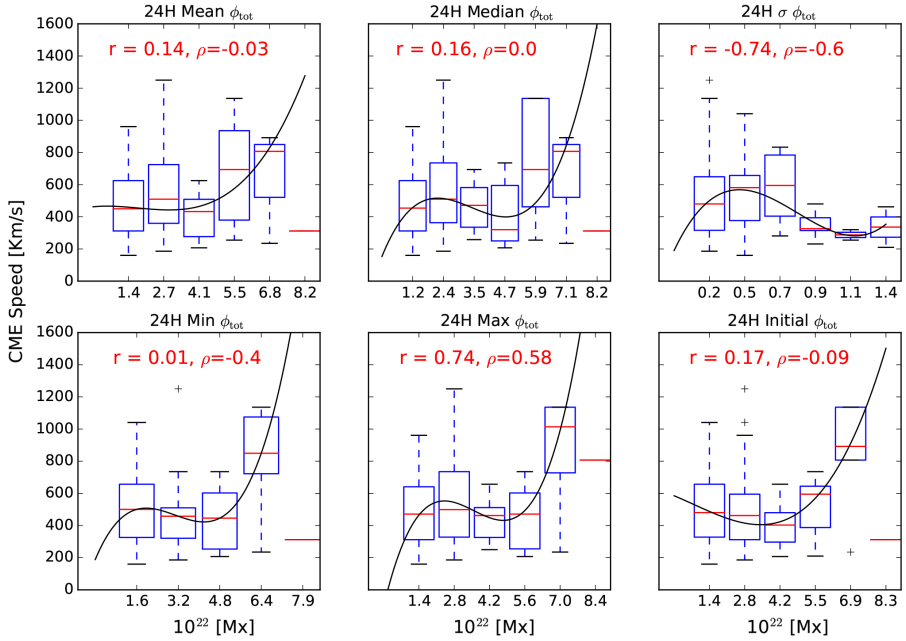


Figure 15. Box-and-whiskers plots of CME speed *versus* the 24-hour Φ_{tot} mean, median, standard deviation (top panels), minimum, maximum, and initial value (bottom panels). Solid black line corresponds to a fit of the median values (red horizontal lines) using a third-degree polynomial. Values in the horizontal axis are binned so the residual between the data and fit is minimized.

their distribution quartiles. In order to do so, the data is binned in the abscissa variable thus providing a distribution of values in the ordinate variable *per* bin. For a bin in any panel of Figure 15, the red horizontal bar corresponds to the median value (second quartile), and the top and bottom of the blue box mark the third and first quartile, correspondingly. The whiskers (dashed lines) represent the maximum and minimum values of the distribution, excluding outliers (black crosses). In this way, the CME speed is represented by the median value of the distribution in each Φ_{tot} bin.

In order to test for a linear or non-linear association between the variables, for each box-and-whiskers plot in Figure 15 a third polynomial curve (solid black line) is estimated by fitting the median values. In addition, linear (Pearson) and non-linear (rank Spearman) correlation coefficients are calculated and displayed. The polynomial degree and number of bins in the abscissa coordinate were selected to minimize the residuals between the binned data and fit. By visual inspection of the data-fit curves, median values of CME speeds seem to show an increasing relation with all 24-hour Φ_{tot} distribution-related quantities in Figure 15 but one, the 24-hour Φ_{tot} standard deviation. At least two relationships show moderate to strong correlations (value $s = 0.5 - 0.75$): maximum value and standard deviation of the 24-hour Φ_{tot}. In the case of v_{CME} and the 24-hour $\sigma(\Phi_{tot})$ (although the number of data points in each plot is small; between 5 and 7), correlation coefficients show the CME speed having a moderate, negative,

and slightly more linear relation with the 24-hour $\Phi_{\rm tot}$ standard deviation. This statistical association might suggest that fast CMEs can originate from regions in which the total unsigned flux varies little over the 24-hour window leading to the eruption. On the other hand, $v_{\rm CME}$ and the maximum value of $\Phi_{\rm tot}$ 24 hours before eruption, show a moderate, positive, and linear correlation association, implying that the higher the maximum measured value of $\Phi_{\rm tot}$, a faster CME could be expected. It should be kept in mind these statistical results correspond to CMEs associated to flares and ARs. Therefore similar analyses using non-eruptive flaring or quiet active regions should be performed for completeness. Such comparative analyses could then provide precursor signals and/or probabilistic relationships for CME characteristics inferred from the evolution of AR photospheric properties.

4. Conclusions and Future Work

An automatic back propagation algorithm has been developed to identify source regions associated with CME events. The algorithm traces HI-identified events back to COR-identified events, associates flares with these COR-identified events, and then associates the flares with their AR sources. SMART is used to obtain complex photospheric magnetic field properties of these identified ARs. This algorithm (as summarised in Figure 2) has been run on the HELCATS project HICAT catalogue, producing the LOWCAT catalogue.

The LOWCAT catalogue is available open-access online (Murray *et al.*, 2017a), containing properties of the $\approx$ 2,000 HICAT CME events, $\approx$ 720 flares ($\approx$ 45% of the CME events), and $\approx$ 550 NOAA/SWPC ARs sources. Although some events may be missed due to the strict algorithm constraints, the automated method enables easier study of such a large number of events, unlike the significant time effort involved with manual association. The solar cycle period during which observations were taken also limits associations, however the resulting numbers do compare well with similar large catalogue studies. Andrews (2003) found all X-class flares and $\approx$ 55% of the M-class flares were associated with 311 LASCO-observed CMEs from 1996 to 1999, supporting the conclusion of Harrison (1995) that particular flare intensities and durations do not guarantee CME association. Yashiro *et al.* (2005) found 20% C-class flares associated with 1,301 CMEs in a larger statistical study, and half of the CMEs associated with the C-class flares were invisible in LASCO. Regarding eruption location on the solar disk, Lara (2008) previously found that CME activity has a slight preference for the northern and southern hemispheres during the first and second halves of a solar cycle, respectively. The LOWCAT events seem generally evenly distributed in Figure 3, although the data gaps do not allow any more conclusions to be drawn regarding spatial correlations.

The simple correlations between the properties available within LOWCAT presented in Section 2.4 are comparable to previous investigations involving other catalogues of LASCO CME observations (rather than STEREO). Yashiro *et al.* (2005) found CMEs associated with X-class flares were faster, with wider extents, than those associated with C-class flares. In a study of 86 flare-associated CMEs,

Guo, Zhang, and Chumak (2007) found faster flare-associated CMEs tend to be accompanied by more intense flares. Further, for CMEs associated with 55 ARs, Guo, Zhang, and Chumak found that fast CMEs initiated from ARs with large total flux and effective distance (a parameter quantifying magnetic complexity). Similar trends between CME, flare, and AR properties are found in Figure 7, with the fast halo CMEs being associated with X-class flares and big, complex ARs. However, it is apparent that the traditional scatter plot method generally does not clearly highlight potential correlations, and the more detailed analysis used for the LOWCAT-FLARECAST comparisons has proven to be more insightful in this case into what particular AR properties may be most useful for CME warning purposes.

The initial exploration of the FLARECAST AR property database in relation to eruption/CME characteristics demonstrate the importance of investigating full-vector field photospheric properties, in accordance with previous studies (Tiwari *et al.*, 2015; Venkatakrishnan and Ravindra, 2003). It is made clear in this investigation that the 24-hour time evolution of photospheric properties provides more insight than single values at the eruption onset and improve the statistical quality of results. According to the results presented here, three properties show significant statistical association with the CME speed: total unsigned flux, total unsigned vertical current, and the total unsigned B_z-contribution helicity current. In all three cases, the minimal property value 24 hours before eruption showed an increasing tendency from slow, to medium-speed, to fast CMEs. The evolution during the study period showed two interesting behaviours of the properties: for eruption-associated ARs, on average i) the total unsigned magnetic flux decreases relatively steadily, ii) total unsigned vertical current shows short-time oscillations. On the other hand, using alternative data representations to traditional scatter plots, empirical relations between the CME speed and the 24-hour property distribution descriptors for total unsigned magnetic flux were found: i) a negatively linear correlation between CME speed and the standard deviation, and ii) a positively linear correlation between CME speed and the maximum measured values in 24 hours.

The results presented in this paper highlight some initial investigations with the properties listed in the new LOWCAT catalogue and FLARECAST database. It must be noted that the analyses presented here serve to show examples of what can be done with the datasets in the hope they will be used extensively in the future. With both data sets freely available online the authors encourage the community to continue these initial investigations. Comparison with other catalogues may prove interesting in the future, such as associating the CME events with filament eruptions (*e.g.* the SDO catalogue of McCauley *et al.*, 2015). The HELCATS project has also developed numerous other catalogues (such as ARRCAT, CIRCAT, ICMECAT, LINKCAT, and RADCAT) which are all also open access and could be used for more detailed studies of CMEs as they propagate out from their source into interplanetary space (some impacting Earth).

Ultimately these statistical analyses on large datasets provide important insight into the link between CME and flare events, as well as characteristics of eruptive active regions. The correlation algorithms presented here have been

written such that other catalogues could be easily integrated, for example they could be run with LASCO CME catalogues in the future. The code is freely available to anyone who may wish to use it, and may prove useful for future operational CME forecasting efforts.

Acknowledgments The FLARECAST database is available at http://api.flarecast.eu/, HICAT catalogue at https://www.helcats-fp7.eu/catalogues/wp2_cat.html, and LOWCAT at https://figshare.com/articles/HELCATS_LOWCAT/4970222. The authors wish to acknowledge the use of the Overleaf to prepare the manuscript, and also the following Python libraries and packages used when creating the figures in this paper: Astropy, Matplotlib, NumPy, pandas, Plotly, SciPy, and SunPy. The STEREO/SECCHI data used here are produced by an international consortium of the Naval Research Laboratory (USA), Lockheed Martin Solar and Astrophysics Laboratory (USA), NASA Goddard Space Flight Center (USA), Rutherford Appleton Laboratory (UK), University of Birmingham (UK), Max-Planck-Institut für Sonnensystemforschung (Germany), Centre Spatial de Liège (Belgium), Institut d'Optique Théorique et Appliqué (France), and Institut dAstrophysique Spatiale (France). SDO is a mission for NASA's Living With a Star (LWS) program, with the SDO/HMI data provided by the Joint Science Operation Center (JSOC). E.C., P.Z., and S.A.M were supported by the European Union Seventh Framework Program under grant agreement No. 606692 (HELCATS project). J.G.A., S.A.M., and S.-H.P were supported by the European Union Horizon 2020 research and innovation program under grant agreement No. 640216 (FLARECAST project). VB acknowledges support of the CGAUSS (Coronagraphic German and US Solar Probe Plus Survey) project for WISPR by the German Space Agency DLR under grant 50 OL 1601. S.A.M. acknowledges the IRC Postdoctoral Fellowship Scheme and AFOSR award FA9550-17-1-039. The authors would like to thank the anonymous referee for their suggestions to improve the paper.

Disclosure of Potential Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

Abramenko, V.I., Yurchyshyn, V.B., Wang, H., Spirock, T.J., Goode, P.R.: 2002, Scaling Behavior of Structure Functions of the Longitudinal Magnetic Field in Active Regions on the Sun. *Astrophys. J.* **577**, 487. DOI. ADS.

Ahmed, O.W., Qahwaji, R., Colak, T., DudokDe Wit, T., Ipson, S.: 2010, A new technique for the calculation and 3d visualisation ofmagnetic complexities on solar satellite images. *The Visual Computer* **26**(5), 385. DOI. https://doi.org/10.1007/s00371-010-0418-1.

Ahmed, O.W., Qahwaji, R., Colak, T., Higgins, P.A., Gallagher, P.T., Bloomfield, D.S.: 2013, Solar Flare Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection. *Solar Phys.* **283**, 157. DOI. ADS.

Andrews, M.D.: 2003, A Search for CMEs Associated with Big Flares. *Solar Phys.* **218**, 261. DOI. ADS.

Barnes, D., Byrne, J., Davies, J., Harrison, R., Helcats, E.U.: 2015, HELCATS HCME_WP2_V02. DOI. https://figshare.com/articles/HELCATS_HCME_WP2_V02/1492351.

Barnes, G., Leka, K.D., Schrijver, C.J., Colak, T., Qahwaji, R., Ashamari, O.W., Yuan, Y., Zhang, J., McAteer, R.T.J., Bloomfield, D.S., Higgins, P.A., Gallagher, P.T., Falconer, D.A., Georgoulis, M.K., Wheatland, M.S., Balch, C., Dunn, T., Wagner, E.L.: 2016, A Comparison of Flare Forecasting Methods. I. Results from the 'All-Clear' Workshop. *Astrophys. J.* **829**, 89. DOI. ADS.

Berger, M.A., Field, G.B.: 1984, The topological properties of magnetic helicity. *Journal of Fluid Mechanics* **147**, 133148. DOI.

Berghmans, D., Foing, B.H., Fleck, B.: 2002, Automated detection of CMEs in LASCO data. In: Wilson, A. (ed.) *From Solar Min to Max: Half a Solar Cycle with SOHO*, *ESA Special Publication* **508**, 437. ADS.

Bloomfield, D.S., Higgins, P.A., McAteer, R.T.J., Gallagher, P.T.: 2012, Toward Reliable Benchmarking of Solar Flare Forecasting Methods. *Astrophys. J. Lett.* **747**, L41. DOI. ADS.

Bobra, M.G., Sun, X., Hoeksema, J.T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G., Leka, K.D.: 2014a, The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs - Space-Weather HMI Active Region Patches. *Solar Phys.* **289**, 3549. DOI. ADS.

Bobra, M.G., Sun, X., Hoeksema, J.T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G., Leka, K.D.: 2014b, The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs - Space-Weather HMI Active Region Patches. *Solar Phys.* **289**, 3549. DOI. ADS.

Brueckner, G.E., Howard, R.A., Koomen, M.J., Korendyke, C.M., Michels, D.J., Moses, J.D., Socker, D.G., Dere, K.P., Lamy, P.L., Llebaria, A., Bout, M.V., Schwenn, R., Simnett, G.M., Bedford, D.K., Eyles, C.J.: 1995, The Large Angle Spectroscopic Coronagraph (LASCO). *Solar Phys.* **162**, 357. DOI. ADS.

Byrne, J.P., Maloney, S.A., McAteer, R.T.J., Refojo, J.M., Gallagher, P.T.: 2010, Propagation of an Earth-directed coronal mass ejection in three dimensions. *Nature Communications* **1**, 74. DOI. ADS.

Byrne, J.P., Morgan, H., Habbal, S.R., Gallagher, P.T.: 2012, Automatic Detection and Tracking of Coronal Mass Ejections. II. Multiscale Filtering of Coronagraph Images. *Astrophys. J.* **752**, 145. DOI. ADS.

Campi, C., Benvenuto, F.: 2017, Feature selection for flarecast. *private communication.*

Conlon, P.A., Gallagher, P.T., McAteer, R.T.J., Ireland, J., Young, C.A., Kestener, P., Hewett, R.J., Maguire, K.: 2008, Multifractal properties of evolving active regions. *Solar Phys.* **248**(2), 297. DOI. https://doi.org/10.1007/s11207-007-9074-7.

Davies, J.A., Harrison, R.A., Perry, C.H., Möstl, C., Lugaz, N., Rollett, T., Davis, C.J., Crothers, S.R., Temmer, M., Eyles, C.J., Savani, N.P.: 2012, A Self-similar Expansion Model for Use in Solar Wind Transient Propagation Studies. *Astrophys. J.* **750**, 23. DOI. ADS.

Davies, J.A., Perry, C.H., Trines, R.M.G.M., Harrison, R.A., Lugaz, N., Möstl, C., Liu, Y.D., Steed, K.: 2013, Establishing a Stereoscopic Technique for Determining the Kinematic Properties of Solar Wind Transients based on a Generalized Self-similarly Expanding Circular Geometry. *Astrophys. J.* **777**, 167. DOI. ADS.

Deng, N., Xu, Y., Yang, G., Cao, W., Liu, C., Rimmele, T.R., Wang, H., Denker, C.: 2006, Multiwavelength study of flow fields in flaring super active region noaa10486. *Astrophys. J.* **644**(2), 1278. http://stacks.iop.org/0004-637X/644/i=2/a=1278.

Drake, J.F.: 1971, Characteristics of Soft Solar X-Ray Bursts. *Solar Phys.* **16**, 152. DOI. ADS.

Eastwood, J.P., Biffis, E., Hapgood, M.A., Green, L., Bisi, M.M., Bentley, R.D., Wicks, R., McKinnell, L.-A., Gibbs, M., Burnett, C.: 2017, The economic impact of space weather: Where do we stand? *Risk Analysis* **37**(2), 206. DOI. http://dx.doi.org/10.1111/risa.12765.

Eyles, C.J., Harrison, R.A., Davis, C.J., Waltham, N.R., Shaughnessy, B.M., Mapson-Menard, H.C.A., Bewsher, D., Crothers, S.R., Davies, J.A., Simnett, G.M., Howard, R.A., Moses, J.D., Newmark, J.S., Socker, D.G., Halain, J.-P., Defise, J.-M., Mazy, E., Rochus, P.: 2009, The Heliospheric Imagers Onboard the STEREO Mission. *Solar Phys.* **254**, 387. DOI. ADS.

Falconer, D.A., Moore, R.L., Gary, G.A.: 2008, Magnetogram Measures of Total Nonpotentiality for Prediction of Solar Coronal Mass Ejections from Active Regions of Any Degree of Magnetic Complexity. *Astrophys. J.* **689**, 1433. DOI. ADS.

Falconer, D.A., Moore, R.L., Barghouty, A.F., Khazanov, I.: 2012, Prior Flaring as a Complement to Free Magnetic Energy for Forecasting Solar Eruptions. *Astrophys. J.* **757**, 32. DOI. ADS.

Georgoulis, M.: 2013, Toward an efficient prediction of solar flares: Which parameters, and how? *Entropy* **15**(11), 5022. DOI. http://dx.doi.org/10.3390/e15115022.

Georgoulis, M.K.: 2010, Pre-eruption magnetic configurations in the active-region solar photosphere. **6**(S273), 495. DOI.

Georgoulis, M.K., Rust, D.M.: 2007, Quantitative Forecasting of Major Solar Flares. *Astrophys. J. Lett.* **661**, L109. DOI. ADS.

Georgoulis, M.K., Titov, V.S., Miki, Z.: 2012, Non-neutralized electric current patterns in solar active regions: Origin of the shear-generating lorentz force. *Astrophys. J.* **761**(1), 61. http://stacks.iop.org/0004-637X/761/i=1/a=61.

Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., Howard, R.: 2009, The SOHO/LASCO CME Catalog. *Earth Moon and Planets* **104**, 295. DOI. ADS.

Guerra, J.A., Pulkkinen, A., Uritsky, V.M., Yashiro, S.: 2015, Spatio-Temporal Scaling of Turbulent Photospheric Line-of-Sight Magnetic Field in Active Region NOAA 11158. *Solar Phys.* **290**, 335. DOI. ADS.

Guo, J., Zhang, H.Q., Chumak, O.V.: 2007, Magnetic properties of flare-CME productive active regions and CME speed. *Astron. Astrophys.* **462**, 1121. DOI. ADS.

Harra, L.K., Schrijver, C.J., Janvier, M., Toriumi, S., Hudson, H., Matthews, S., Woods, M.M., Hara, H., Guedel, M., Kowalski, A., Osten, R., Kusano, K., Lueftinger, T.: 2016, The Characteristics of Solar X-Class Flares and CMEs: A Paradigm for Stellar Superflares and Eruptions? *Solar Phys.* **291**, 1761. DOI. ADS.

Harrison, R.A.: 1995, The nature of solar flares associated with coronal mass ejection. *Astron. Astrophys.* **304**, 585. ADS.

Harrison, R.A., Davies, J.A., Biesecker, D., Gibbs, M.: 2017, The application of heliospheric imaging to space weather operations: Lessons learned from published studies. *Space Weather* **15**(8), 985. 2017SW001633. DOI. http://dx.doi.org/10.1002/2017SW001633.

Haynes, A.L., Parnell, C.E.: 2007, A trilinear method for finding null points in a three-dimensional vector space. *Physics of Plasmas* **14**(8), 082107. DOI. http://dx.doi.org/10.1063/1.2756751.

Hewett, R.J., Gallagher, P.T., McAteer, R.T.J., Young, C.A., Ireland, J., Conlon, P.A., Maguire, K.: 2008, Multiscale Analysis of Active Region Evolution. *Solar Phys.* **248**, 311. DOI. ADS.

Higgins, P.A.: 2012, Sunspot group evolution and the global magnetic field of the Sun. *Thesis, Trinity College Dublin.*

Higgins, P.A., Gallagher, P.T., McAteer, R.T.J., Bloomfield, D.S.: 2011, Solar magnetic feature detection and tracking for space weather monitoring. *Advances in Space Research* **47**, 2105. DOI. ADS.

Jing, J., Yurchyshyn, V.B., Yang, G., Xu, Y., Wang, H.: 2004, On the Relation between Filament Eruptions, Flares, and Coronal Mass Ejections. *Astrophys. J.* **614**, 1054. DOI. ADS.

Kaiser, M.L., Kucera, T.A., Davila, J.M., St. Cyr, O.C., Guhathakurta, M., Christian, E.: 2008, The STEREO Mission: An Introduction. *Space Sci. Rev.* **136**, 5. DOI. ADS.

Korsós, M.B., Baranyi, T., Ludmny, A.: 2014, Pre-flare dynamics of sunspot groups. *Astrophys. J.* **789**(2), 107. http://stacks.iop.org/0004-637X/789/i=2/a=107.

Künzel, H.: 1965, Zur Klassifikation von Sonnenfleckengruppen. *Astronomische Nachrichten* **288**, 177. ADS.

Kusano, K., Maeshiro, T., Yokoyama, T., Sakurai, T.: 2002, Measurement of magnetic helicity injection and free energy loading into the solar corona. *Astrophys. J.* **577**(1), 501. http://stacks.iop.org/0004-637X/577/i=1/a=501.

Lara, A.: 2008, The Source Region of Coronal Mass Ejections. *Astrophys. J.* **688**, 647. DOI. ADS.

Lee, K., Moon, Y.-J., Lee, J.-Y.: 2015, Forecast of a Daily Halo CME Occurrence Probability Depending on Class and Area Change of the Associated Sunspot. *Solar Phys.* **290**, 1661. DOI. ADS.

Liu, Y.: 2008, Magnetic field overlying solar eruption regions and kink and torus instabilities. *Astrophys. J. Lett.* **679**(2), L151. http://stacks.iop.org/1538-4357/679/i=2/a=L151.

Magdalenić, J., Marqué, C., Krupar, V., Mierla, M., Zhukov, A.N., Rodriguez, L., Maksimović, M., Cecconi, B.: 2014, Tracking the CME-driven Shock Wave on 2012 March 5 and Radio Triangulation of Associated Radio Emission. *Astrophys. J.* **791**, 115. DOI. ADS.

Mason, J.P., Hoeksema, J.T.: 2010, Testing Automated Solar Flare Forecasting with 13 Years of Michelson Doppler Imager Magnetograms. *Astrophys. J.* **723**, 634. DOI. ADS.

McCauley, P.I., Su, Y.N., Schanche, N., Evans, K.E., Su, C., McKillop, S., Reeves, K.K.: 2015, Prominence and Filament Eruptions Observed by the Solar Dynamics Observatory: Statistical Properties, Kinematics, and Online Catalog. *Solar Phys.* **290**, 1703. DOI. ADS.

McIntosh, P.S.: 1990, The classification of sunspot groups. *Solar Phys.* **125**, 251. DOI. ADS.

Moestl, C., Isavnin, A., Boakes, P.D., Kilpua, E.K.J., Davies, J.A., Harrison, R.A., Barnes, D., Krupar, V., Eastwood, J.P., Good, S.W., Forsyth, R.J., Bothmer, V., Reiss, M.A., Amerstorfer, T., Winslow, R.M., Anderson, B.J., Philpott, L.C., Rodriguez, L., Rouillard, A.P., Gallagher, P., Nieves-Chinchilla, T., Zhang, T.L.: 2017, Modeling observations of solar coronal mass ejections with heliospheric imagers verified with the heliophysics system observatory. *Space Weather* **15**(7), 955. 2017SW001614. DOI. http://dx.doi.org/10.1002/2017SW001614.

Moon, Y.-J., Choe, G.S., Wang, H., Park, Y.D., Gopalswamy, N., Yang, G., Yashiro, S.: 2002, A Statistical Study of Two Classes of Coronal Mass Ejections. *Astrophys. J.* **581**, 694. DOI. ADS.

Morgan, H., Byrne, J.P., Habbal, S.R.: 2012, Automatically Detecting and Tracking Coronal Mass Ejections. I. Separation of Dynamic and Quiescent Components in Coronagraph Images. *Astrophys. J.* **752**, 144. DOI. ADS.

Murray, S.A., Bingham, S., Sharpe, M., Jackson, D.R.: 2017a, Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather* **15**, 577. DOI. ADS.

Murray, S.A., Zucca, P., Carley, E., Gallagher, P.: 2017b, HELCATS LOWCAT. *figshare*. DOI.

Pant, V., Willems, S., Rodriguez, L., Mierla, M., Banerjee, D., Davies, J.A.: 2016, Automated Detection of Coronal Mass Ejections in STEREO Heliospheric Imager Data. *Astrophys. J.* **833**, 80. DOI. ADS.

Park, S.-H., Chae, J., Wang, H.: 2010, Productivity of solar flares and magnetic helicity injection in active regions. *Astrophys. J.* **718**(1), 43. http://stacks.iop.org/0004-637X/718/i=1/a=43.

Park, S.-H., Cho, K.-S., Bong, S.-C., Kumar, P., Chae, J., Liu, R., Wang, H.: 2012, The occurrence and speed of cmes related to two characteristic evolution patterns of helicity injection in their solar source regions. *Astrophys. J.* **750**(1), 48. http://stacks.iop.org/0004-637X/750/i=1/a=48.

Park, S.-H., Cho, K.-S., Bong, S.-C., Kumar, P., Chae, J., Liu, R., Wang, H.: 2012, The Occurrence and Speed of CMEs Related to Two Characteristic Evolution Patterns of Helicity Injection in Their Solar Source Regions. *Astrophys. J.* **750**, 48. DOI. ADS.

Plotnikov, I., Rouillard, A.P., Davies, J.A., Bothmer, V., Eastwood, J.P., Gallagher, P., Harrison, R.A., Kilpua, E., Möstl, C., Perry, C.H., Rodriguez, L., Lavraud, B., Génot, V., Pinto, R.F., Sanchez-Diaz, E.: 2016, Long-Term Tracking of Corotating Density Structures Using Heliospheric Imaging. *Solar Phys.* **291**, 1853. DOI. ADS.

Pontin, D.I., Priest, E.R., Galsgaard, K.: 2013, On the nature of reconnection at a solar coronal null point above a separatrix dome. *Astrophys. J.* **774**(2), 154. http://stacks.iop.org/0004-637X/774/i=2/a=154.

Robbrecht, E., Berghmans, D.: 2004, Automated recognition of coronal mass ejections (CMEs) in near-real-time data. *Astron. Astrophys.* **425**, 1097. DOI. ADS.

Rouillard, A.P., Lavraud, B., Genot, V., Bouchemit, M., Dufourg, N., Plotnikov, I., Pinto, R.F., Sanchez-Diaz, E., Lavarra, M., Penou, M., Jacquey, C., Andre, N., Caussarieu, S., Toniutti, J.-P., Popescu, D., Buchlin, E., Caminade, S., Alingery, P., Davies, J.A., Odstrcil, D., Mays, L.: 2017, A propagation tool to connect remote-sensing observations with in-situ measurements of heliospheric structures. *ArXiv e-prints*. ADS.

Sammis, I., Tang, F., Zirin, H.: 2000, The Dependence of Large Flare Occurrence on the Magnetic Structure of Sunspots. *Astrophys. J.* **540**, 583. DOI. ADS.

Scherrer, P.H., Bogart, R.S., Bush, R.I., Hoeksema, J.T., Kosovichev, A.G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T.D., Title, A., Wolfson, C.J., Zayer, I., MDI Engineering Team: 1995, The Solar Oscillations Investigation - Michelson Doppler Imager. *Solar Phys.* **162**, 129. DOI. ADS.

Scherrer, P.H., Schou, J., Bush, R.I., Kosovichev, A.G., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Zhao, J., Title, A.M., Schrijver, C.J., Tarbell, T.D., Tomczyk, S.: 2012, The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 207. DOI. ADS.

Scherrer, P.H., Schou, J., Bush, R.I., Kosovichev, A.G., Bogart, R.S., Hoeksema, J.T.: 2012, The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo). *Solar Phys.* **275**, 207. DOI. http://dx.doi.org/10.1007/s11207-011-9834-2.

Schrijver, C.J.: 2007a, A Characteristic Magnetic Field Pattern Associated with All Major Solar Flares and Its Use in Flare Forecasting. *Astrophys. J. Lett.* **655**, L117. DOI. ADS.

Schrijver, C.J.: 2007b, A Characteristic Magnetic Field Pattern Associated with All Major Solar Flares and Its Use in Flare Forecasting. *Astrophys. J. Lett.* **655**, L117. DOI. ADS.

Shibata, K., Magara, T.: 2011, Solar Flares: Magnetohydrodynamic Processes. *Living Rev. Solar Phys.* **8**, 6. ADS.

Singh, Y.P., Badruddin: 2006, Statistical considerations in superposed epoch analysis and its applications in space research. *J. Atmos. Solar-Terr. Phys.* **68**, 803. DOI. ADS.

Tiwari, S.K., Falconer, D.A., Moore, R.L., Venkatakrishnan, P., Winebarger, A.R., Khazanov, I.G.: 2015, Near-Sun speed of CMEs and the magnetic nonpotentiality of their source active regions. *Geophys. Res. Lett.* **42**, 5702. DOI. ADS.

Venkatakrishnan, P., Ravindra, B.: 2003, Relationship between CME velocity and active region magnetic energy. *Geophys. Res. Lett.* **30**, 2181. DOI. ADS.

Vourlidas, A., Howard, R.A.: 2006, The Proper Treatment of Coronal Mass Ejection Brightness: A New Methodology and Implications for Observations. *Astrophys. J.* **642**, 1216. DOI. ADS.

Wang, S., Liu, C., Deng, N., Wang, H.: 2014, Sudden photospheric motion and sunspot rotation associated with the x2.2 flare on 2011 february 15. *Astrophys. J. Lett.* **782**(2), L31. http://stacks.iop.org/2041-8205/782/i=2/a=L31.

Wang, Y., Chen, C., Gui, B., Shen, C., Ye, P., Wang, S.: 2011, Statistical study of coronal mass ejection source locations: Understanding CMEs viewed in coronagraphs. *J. Geophys. Res.(Space Physics)* **116**, A04104. DOI. ADS.

Webb, D.F., Howard, T.A.: 2012, Coronal Mass Ejections: Observations. *Living Reviews in Solar Physics* **9**, 3. DOI. ADS.

Yang, G., Xu, Y., Cao, W., Wang, H., Denker, C., Rimmele, T.R.: 2004, Photospheric shear flows along the magnetic neutral line of active region 10486 prior to an x10 flare. *Astrophys. J. Lett.* **617**(2), L151. http://stacks.iop.org/1538-4357/617/i=2/a=L151.

Yashiro, S., Gopalswamy, N., Akiyama, S., Michalek, G., Howard, R.A.: 2005, Visibility of coronal mass ejections as a function of flare location and intensity. *J. Geophys. Res.(Space Physics)* **110**, A12S05. DOI. ADS.

Yashiro, S., Michalek, G., Akiyama, S., Gopalswamy, N., Howard, R.A.: 2008, Spatial Relationship between Solar Flares and Coronal Mass Ejections. *Astrophys. J.* **673**, 1174. DOI. ADS.

Youssef, M.: 2012, On the relation between the CMEs and the solar flares. *NRIAG Journal of Astronomy and Geophysics* **1**, 172. DOI. ADS.

Yurchyshyn, V., Yashiro, S., Abramenko, V., Wang, H., Gopalswamy, N.: 2005, Statistical Distributions of Speeds of Coronal Mass Ejections. *Astrophys. J.* **619**, 599. DOI. ADS.

Zheng, Y.: 2013, Improving cme forecasting capability: An urgent need. *Space Weather* **11**(11), 641. 2013SW001004. DOI. http://dx.doi.org/10.1002/2013SW001004.

Zuccarello, F.P., Seaton, D.B., Mierla, M., Poedts, S., Rachmeler, L.A., Romano, P., Zuccarello, F.: 2014, Observational evidence of torus instability as trigger mechanism for coronal mass ejections: The 2011 august 4 filament eruption. *Astrophys. J. Lett.* **785**(2), 88. http://stacks.iop.org/0004-637X/785/i=2/a=88.