

Probabilistic hazard assessment: Application to geomagnetic activity

Gemma S. Richardson*  and Alan W. P. Thomson 

British Geological Survey, The Lyell Centre, Research Avenue South, Edinburgh EH14 4AP, United Kingdom

Received 22 January 2021 / Accepted 6 January 2022

Abstract—Probabilistic Hazard Assessment (PHA) provides an appropriate methodology for assessing space weather hazards and their impact on technology. PHA is widely used in geosciences to determine the probability of exceedance of critical thresholds, caused by one or more hazard sources. PHA has proved useful with limited historical data to estimate the likelihood of specific impacts. PHA has also driven the development of empirical and physical models, or ensembles of models, to replace measured data. Here we aim to highlight the PHA method to the space weather community and provide an example of how it could be used. In terms of space weather impact, the critical hazard thresholds might include the Geomagnetically Induced Current in a specific high voltage power transformer neutral, or the local pipe-to-soil potential in a particular metal pipe. We illustrate PHA in the space weather context by applying it to a twelve-year dataset of Earth-directed solar Coronal Mass Ejections (CME), which we relate to the probability that the global three-hourly geomagnetic activity index Kp exceeds specific thresholds. We call this a “Probabilistic Geomagnetic Hazard Assessment”, or PGHA. This provides a simple but concrete example of the method. We find that the cumulative probability of $Kp > 6-$, $> 7-$, $> 8-$ and $Kp = 9_0$ is 0.359, 0.227, 0.090, 0.011, respectively, following observation of an Earth-directed CME, summed over all CME launch speeds and solar source locations. According to the historical Kp distribution, this represents an order of magnitude increase in the a priori probability of exceeding these thresholds. For the lower Kp thresholds, the results are somewhat distorted by our exclusion of coronal hole high-speed stream effects. The PGHA also reveals useful probabilistic associations between solar source location and subsequent maximum Kp for operational forecasters.

Keywords: geomagnetic hazard / probabalistic hazard assessment / coronal mass ejection

1 Introduction

The impact of severe space weather on technological systems can be estimated by deterministic or probabilistic means. For example, a series of coupled physical or empirical models of processes between the Sun and Earth may provide a deterministic estimate of ground-level magnetic field variations (dB/dt). These variations, in turn, may drive geomagnetically induced currents (GIC) in power grids and other grounded networks. The model (or coupled models) may be driven by solar activity and account for interactions between erupting structures from the Sun and the ambient heliospheric, magnetospheric, and ionospheric environments. However, at the current level of development, the accuracy of such a model may be poor, particularly because of a lack of spatial and temporal detail at the local (e.g. ground) level, where the impacts are felt.

One way of quantifying the precision and uncertainty in deterministic model outputs is ensemble modelling. This can be useful in determining a mean and a range of behaviours commensurate with varied input conditions and uncertain knowledge of variables (or physical models) at key steps in the modelling chain. Other issues for deterministic models include their domains of applicability, including extrapolation to extreme environments, their treatment of missing data and detail (e.g. the interplanetary magnetic field orientation between Sun and Earth) and the lack of supporting measurements as input and/or validation data. Different models may also produce conflicting outputs. Probabilistic techniques are therefore an attractive alternative, or at least complementary method, to the deterministic, physics-driven approach.

Probabilistic hazard assessment (PHA) is common in geosciences, for example, in seismic, volcanic and tsunami impact analysis. In the deterministic approach, geoscientists examine long time series of damaging events to determine a “maximum credible scenario” for some regions. This scenario might be

*Corresponding author: gemk@bgs.ac.uk

described in terms of, in the case of tsunamis, maximum wave height, wave run-up, or inundation area (e.g. Grezio et al., 2017). That is, one expresses the hazard in terms of variables that are directly relevant for society, for example, the specific threat to buildings and populations. However, the lack of measured data for some rare events and important variables has led to the development of physical models of geohazard processes of varying levels of sophistication to help estimate quantities such as the maximum wave height. Combining these models with appropriate weights is then the function of PHA. The major variables of PHA are typically the hazard source type, location, source to site path and any propagation effects. The outputs of PHA include the annual probability of exceedance curves and frequency of exceedance versus strength or amplitude (e.g. of ground motion in seismic studies).

In the space weather context, there are a variety of source mechanisms for the generation of, for example, ground-level impacts. These sources include coronal mass ejections (CME) and high-speed solar wind streams rooted in coronal holes (CH), both of varying “intensity”, in the most general sense. Source-to-site factors include the interaction of eruptive solar magnetic structures with the ambient solar wind, stream interfaces and interactions between the modulated solar wind and the magnetosphere and ionosphere. Across this chain of events, Sun to Earth, nonlinear interactions are common.

In the next section, we outline the PHA method before describing the data sources and processing used to identify Earth-directed Coronal Mass Ejections. We then discuss the results of the PHA analysis before summarising the findings.

2 Method

The general form of PHA is given by (1), in terms of cumulative probabilities P (see e.g. review articles referenced by Gerstenberger et al., 2020; Baker, 2013; or Kijko, 2011, for a seismic interpretation; or Grezio et al., 2017, for a tsunami hazard review).

$$P(\text{IM} > x) = \int_{m_{\min}}^{m_{\max}} \int_0^{r_{\max}} P(\text{IM} > x | m, r) f_M(m) f_R(r) dr dm, \quad (1)$$

where $P(\text{IM} > x | m, r)$ represents the cumulative probability that some “Intensity Measure” (IM) exceeds a given threshold “ x ”, given that a hazard source of magnitude “ m ” occurs, at some location “ r ” distant with respect to the IM impact site. In equation (1), $f_M(m)$ and $f_R(r)$ are probability density functions for the source magnitude and location, assumed independent in (1). In addition, in (1) we integrate over all magnitudes and locations, with the assumption that there is some minimum magnitude, m_{\min} and a most-distant source location, r_{\max} , beyond which the impact on IM is essentially zero. There may also be (or not be) some maximum source magnitude, m_{\max} . In practise, the integral is usually approximated by a discrete sum.

For space weather hazards, we adopt (1), with some approximations as described below, and we assume here that the main variables driving the impact IM are an estimate of solar source magnitude and location. Equation (1) applies to one type of source (e.g. to CMEs) but can be generalised to include a

variety of sources, in the form of a sum of contributing integrals, one for each source type (e.g. CME and CH), each with its own source magnitude and location.

For space weather applications, there is typically limited data from which one can estimate $P(\text{IM} > x | m, r)$ and hence calculate $P(\text{IM} > x)$. This is particularly true where we are more interested in to what extent, for example, a given CME or CH will result in a geomagnetically induced current (GIC) of some magnitude in some transformer, rather than a measure derived from a more generic geomagnetic data product, such as an activity index. For such a specific IM, a model, or ensemble of models, is probably the most realistic way to evaluate the integral (examples in the GIC context might include the NASA “Solar Shield” model (Pulkkinen et al., 2010), or some future development of “WSA-ENLIL” (Pizzo et al., 2011), or a similar model of solar wind propagation that couples to a magnetospheric/ionospheric model, driving ground magnetic field variations or GIC).

As an aside, we note that the likely role of space weather modellers, in terms of PHA, is the development of better models that define the integrand $P(\text{IM} > x | m, r)$ for “useful” IM. The role of space weather hazard assessors is then to combine these model outputs with appropriate weights, as in equation (1).

In the application of equation (1) and the construction of $P(\text{IM} > x | m, r)$, there are both aleatory and epistemic uncertainties with which one has to contend (e.g. Grezio et al., 2017). Aleatory uncertainties refer to essentially random factors, including measurement uncertainty, that remain even where models are improved. An example of such uncertainty might be the local time at the power transformer of interest at the instant an unforeseen CME impacts the magnetosphere. Epistemic uncertainties refer to those uncertainties that can be minimised in time through improvements in modelling capability. An example here might be knowledge of the prior state of the solar wind at the instant of CME eruption, the way that the Interplanetary CME evolves and propagates through the heliosphere, or whether the magnetosphere is already in a disturbed state at the expected time of the CME arrival.

In the following sections, we demonstrate the PHA method empirically. Ideally, we would relate CME observation to, for example, a specific transformer-level GIC, but that is beyond our scope and data. Instead, we provide a simple example of CME observations in relation to the Kp index (Matzka et al., 2021a: the Kp index is expressed on a scale of thirds from 0o, 0+, 1–, 1o, 1+, 2–, 2o, and so on, up to 9o), to illustrate the general method. Specifically, we use observations of Earth-directed CMEs and their plane-of-sky speeds as a proxy for “ m ”, and their solar source latitude and longitude to represent “ r ”. We relate these to the subsequent maximum 3-hour Kp , representing IM, in any 24-hour period in the 5 days post-CME launch. We refer to this as a Probabilistic Geomagnetic Hazard Assessment (PGHA).

In our analysis, we do not take into account, for example, the interplanetary magnetic field orientation at Earth, which determines the geo-effectiveness of any given CME, or indeed any other physical process occurring between observation of an Earth-directed CME that affects subsequent Kp levels. Such epistemic uncertainties could be reduced in future work but are beyond the scope here.

To use equation (1), we also make some simplifications and approximations, reflecting the nature of the data we have

available for analysis. Firstly, the CME speed, “ m ”, and the source location, “ r ”, might not be independent, though we assume that they are, as a reasonable assumption. Also, the reported sky-plane speed, which we use, is not the same as true CME radial velocity and depends on solar longitude (we comment further on the significance of this point in Section 5). We also do not take into account measurement or observational uncertainties, which will feed through to error estimates on the $P(\text{IM} > x)$.

We approximate (1) as follows by integrating over r :

$$P(\text{IM} > x) = \int P(\text{IM} > x | r) f_R(r) dr, \quad (2)$$

where we have assumed

$$P(\text{IM} > x | r) = \int P(\text{IM} > x | m, r) f_M(m) dm. \quad (3)$$

This calculation can be cross-checked by integrating over m , as follows:

$$P(\text{IM} > x) = \int P(\text{IM} > x | m) f_M(m) dm, \quad (4)$$

assuming

$$P(\text{IM} > x | m) = \int P(\text{IM} > x | m, r) f_R(r) dr. \quad (5)$$

Equation (2) (assuming Eq. (3) holds) is referred to as “Method 1”, and equation (4) (assuming Eq. (5) holds) is “Method 2”.

In the next section, we describe the data sources that have been used and then build the components of equations (2) and (4) in Section 4, by means of Table 1 and Figures 3–5. For future reference, the various elements of equations (2) and (4) are given by the following:

- $f_R(r)$: Number distribution of helio-latitude/longitude data, as provided in each bin of Figure 4, divided by the total number of Earth-directed CME in that dataset (i.e. divided by 255);
- $f_M(m)$: Number distribution of speed data, as provided in each speed bin of Figure 3, divided by the total number of Earth-directed CME in that dataset (i.e. 261);
- $P(\text{IM} > x | r)$: Probability of Kp over each Kp threshold, for a given helio-latitude/longitude bin in Figure 5, divided by 100 (%), i.e. to obtain the occurrence probability irrespective of speed;
- $P(\text{IM} > x | m)$: Probability of Kp over each Kp threshold, for a given speed bin, in Table 1, divided by 100 (%), i.e. the occurrence probability irrespective of location.

3 CME data sources and processing

For this study, we examined the following resources to identify Earth-directed Coronal Mass Ejections (EDCME) and consequent geomagnetic activity for the period 1998–2009. These resources provide data that is either identical, or similar to, the sort of real-time information that would have been available to operational forecasters at the time of each CME.

Table 1. Percentage likelihood of Earth-directed CME (total of 261) causing a geomagnetic storm, based on 500 km/s bins of LASCO speed estimates.

V km/s	$Kp > 6-$	$Kp > 7-$	$Kp > 8-$	$Kp > 9o$
<500	38	22	4	
500–1000	35	18	7	
1000–1500	29	23	8	
1500–2000	44	24	16	4
2000–2500	58	50	33	17
2500–3000	75	50	0	

1. Images from the Large Angle and Spectrometric Coronagraph Experiment (LASCO)¹ instrument on board the Solar and Heliospheric Observatory (SOHO)². LASCO provides coronagraph images of the solar corona from 1.1 to 32 solar radii, allowing coronal mass ejections to be observed and studied.
2. NOAA Space Weather Prediction Centre (SWPC) reports, forecasts and summaries of space weather³. These provide a discussion of solar activity, forecasts of solar and geomagnetic activity, events lists, information about solar active regions and more.
3. The Advanced Composition Explorer (ACE) list of solar wind shocks⁴, observed by either the MAG or SWEPAM instruments on-board the ACE spacecraft.
4. The “Richardson and Cane” list (Richardson & Cane, 2010) details Interplanetary Coronal Mass Ejections (ICME), identified in the near-Earth solar wind, and relates these back to CME at the Sun where possible.
5. GOES X-ray event observations⁵ identify solar flare time, location, size and the associated solar active region.
6. The Heliophysics Integrated Observatory (HELIO)⁶ Context Service. HELIO allows coronagraphs, solar images, GOES X-ray curves etc., to be presented together, making it easier to correlate events.
7. The SOHO LASCO CME catalogue (Yashiro et al., 2004), which provides a description of all CME manually identified in the LASCO coronagraphs since 1996. This includes the time of first appearance, speed estimates, the central position angle and apparent width of the CME, as well as estimates of CME mass and kinetic energy.

For the 14,320 CME in the LASCO CME catalogue, for 1998–2009 inclusive, we identified those CME that were Earth-directed and which, if any, solar events were associated with them (i.e. flare or filament). “Earth-directed” here means that in the SWPC reports available at the time, at least some component of the CME or associated shock front was expected to impact the magnetosphere, even if the CME was travelling at an angle to the Sun–Earth line. Reports of CME that had no earthward component or that are not recorded at all in the

¹ <http://lasco-www.nrl.navy.mil/>

² <http://sohowww.nascom.nasa.gov/>

³ <http://www.swpc.noaa.gov/ftpdir/warehouse/>

⁴ http://www.ssg.sr.unh.edu/mag/ace/ACElists/obs_list.html

⁵ <ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/>

⁶ <http://www.helio-vo.eu/>

SWPC reports were, in general, counted as non-Earth-directed. However, there are a few known cases where a CME was classified (by SWPC) as not having an Earth-directed component but then did have a shock arrival at Earth. These examples are excluded from the study and results here, but we estimate these examples amount to only about 7% of the known Earth-directed events. Therefore, we note that our results are partly based on the observations and decisions of forecasters at one forecast centre (SWPC). Observations and forecasts made by other centres might yield different results and a useful future study would be to cross-compare, where such data are available.

From a combination of the SWPC reports, GOES X-ray lists and the HELIO context service, we identified any flares and filaments associated with CME. To be considered “associated”, in this work, the flare or filament needed to have occurred within 4 h preceding, or the hour after the time of first appearance in the LASCO catalogue, and from a region on the solar disk close to the apparent eruption site of the CME. (We extended the time window to an hour after the CME first appearance to ensure we did not miss any associated flares, as flare onset can occur several tens of minutes after a CME (Harrison, 1995)). For flares, the location was taken from the GOES X-ray list where available; otherwise, the location is given as the location of the active region from which the flare originated. For filament eruptions, the location is generally given as the centre of the filament. Where a flare and filament occur close in time and location and appear to be related, they were both considered to be associated with the CME. There are also a few examples where the source choice was ambiguous and a unique location could not be attributed to the CME.

To identify CME arrivals at Earth, we used a combination of the Richardson and Cane (2010) ICME list, the ACE solar wind shock list, the 3-hour Kp index and the SWPC summaries between 1 and 5 days after a CME, looking for shocks and storm periods. CME, which did not appear to strike the Earth (i.e. no shock signature was identified and no apparent increase in activity was detected within 5 days), were assumed to have missed the Earth.

All events were associated with the maximum (three-hourly) Kp index observed in the 24 h following CME arrival. Kp is used to indicate the general level of geomagnetic activity. As discussed earlier, in the context of PGHA and those impacts of greatest concern to specific end-users, there may be other measures of technology-specific impact that could be constructed. However, for this “test” example, we grouped storms, in terms of Kp , according to the NOAA Space Weather Scale for Geomagnetic storms (G-scales)⁷. Here a G2 storm is defined as being in the range $6- \leq Kp \leq 6+$, G3 is in the range $7- \leq Kp \leq 7+$, G4 is $8- \leq Kp \leq 9-$, and G5 is $Kp = 9o$. In the following, the CME are discussed as having resulted in storms reaching a given value or higher, for example, greater than $Kp = 6-$ (G2) includes *all* storms with Kp 6- to 9o inclusive. We choose this approach, as we believe it is of value in forecasting to provide statistics about reaching a storm of a certain minimum value. However, we acknowledge that predicting a specific classification, for example, precisely G4 could also be a sensible alternative approach. However, being so category-specific will also reduce the sample sizes of big events and hence the statistical robustness of results.

Separately we compiled a list of every day for which the maximum Kp in that day reached at least 6- to provide a means of checking that no significant storms had been missed. The source of each Kp event was then identified either as one of the CME listed or as a result of CH activity. A few storms that are attributed to CME that do not appear in our list, and there are also a few instances when there is no obvious cause identified. This is partly due to data gaps in the LASCO catalogue, which is mostly a problem in 1998–1999, but it is also known that around 19% of CME are not observable in LASCO (Wang et al., 2011).

Therefore, we found that there were 315 EDCME between January 1998 and December 2009 identified by this process, of which 142 (45%) caused geomagnetic storms ($\geq G2$), and 48 (15%) caused severe geomagnetic storms where Kp reached at least 8- (G4).

Up to this point, if multiple CME resulted in one storm, they were counted as being individually geo-effective, but that is clearly potentially misleading. Stating that a CME with a certain speed and from a particular solar latitude/longitude caused a storm of a given size is not entirely true if the storm actually resulted from the interaction of several CME. Some authors (e.g. Wang et al., 2002) deal with this by attributing all effects to the first CME, but that may also be misleading if that first CME is overtaken by a second or third faster CME. Therefore, we use the following prescription to get around this issue. If a CME gives rise to a clear max in Kp before an obvious second CME arrival then that maximum Kp value is attributed to the first CME. Later geomagnetic activity is then considered to result from the CME in combination, and these combined CME are excluded from the analysis. If the arrivals are not clearly separated, are close together (<12 h), or the CME are thought to have combined in transit, the CME are also counted as “combined” and are excluded. This results in 47 CME being excluded, approximately 15% of the Earth-directed total. $Kp = 9o$ events are still counted separately through this definition. For example, although the severe storm in October 2003 comprised two events in two days, the shock arrivals were far enough apart (by more than 24 h), that it is reasonable to claim both CME resulted in Kp of 9o, although arguably the magnetic field was already disturbed ahead of the second CME arrival.

There are 22 events (47 CME in total) resulting from CME in combination (two or more arriving close together or combining in the heliosphere), so it is difficult to draw any robust conclusions about specific parameters. However, we do find that 100% of these events caused at least $Kp = 5-$ (G1), 95.5% led to G2 storms, and 50% led to G4 storms ($Kp > 8-$). We note that these percentages are significantly larger than for storms resulting from a single CME (c.f. 45% at G2 and 15% at G4 reported above). We also note that Gopalswamy et al. (2007) found the most intense storms occur when there are successive CME, and Schwenn et al. (2005) noted that the effects of CME that interact and merge are highly unpredictable.

There were a few occasions when CME were not originally thought to be Earth-directed, but did then have an arrival at Earth. We identified 20 such CME, 80% of which caused minor geomagnetic activity (<G2); of the rest, three led to G2, one caused G3, and one CME caused G4. These are excluded here, in the spirit of analysing those EDCME that would have been likely to have been identified in a “real-time” forecast situation.

⁷ <http://www.swpc.noaa.gov/NOAA scales/>

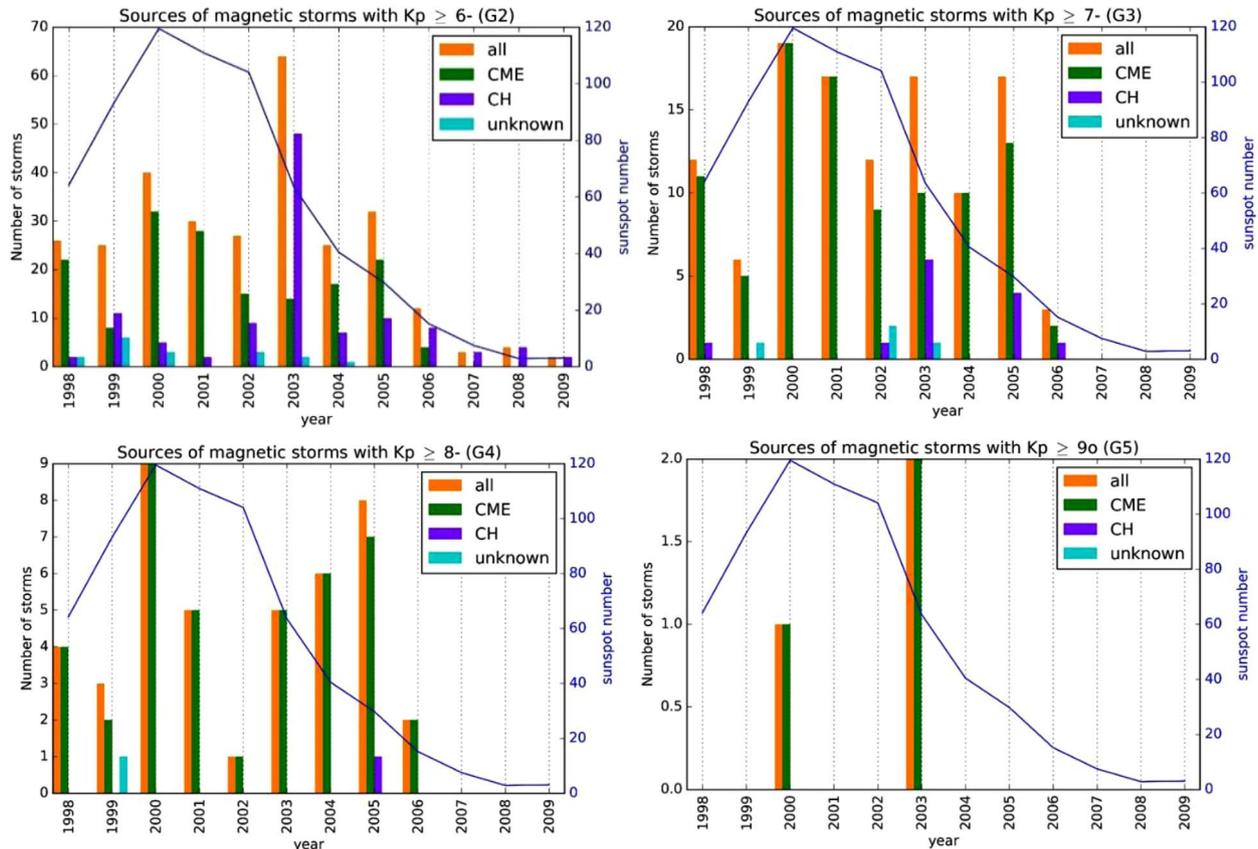


Fig. 1. Number of storms per year that led from $Kp \geq 6-$ to 9o (top left to bottom right), between 1998 and 2009 inclusive (total of 290), separated by source (with total number in orange). Also shown is the average yearly sunspot number (blue line) to provide an indication of solar cycle dependence.

4 Analysis and results

We examined all occasions in the dataset where Kp reached 6– or higher and, where possible, identified the source of each magnetic disturbance. This allowed confirmation of CME that had been missed and identified where geomagnetic storms resulted from other solar or heliospheric activity. We identified 290 storms between 1998 and 2009 inclusive, 163 of which were related to CME, 111 to coronal holes (CH) and 16 for which the source could not be identified. Figure 1 shows the number of storms between 1998 and 2009 for each level of Kp , identified by their solar origin, if known.

In general, CH caused lower levels of activity than CME, with only one event at G4 resulting from a CH. CME also caused more storms at G3 and above for all years; at G2 there were two years when coronal holes caused more storms than CME, particularly in 2003, during the descending phase of the solar cycle. The number of storms appears to increase towards the maximum in the solar cycle (~2000), but there is no clear reduction in the number of storms during the descending phase, until we reach a solar minimum (2007–2009).

As described in Section 1, PHA can include all solar, heliospheric and magnetospheric source mechanisms, but for simplicity in the following, we concentrate solely on those storms that result from Earth-directed CME. Figure 1

demonstrates, at least for G3 and higher, that this is a reasonable first approximation.

The solar source location is a major factor in whether a CME is Earth-directed. We consider the source location to be the same as the location of any associated flare/filament site. Some studies have suggested this may not be the best way to identify the location (e.g. Wang et al., 2002; Moon et al., 2009), as this can differ from the apparent origin of the CME in LASCO images. However, in the absence of a better method, we use the associated flare/filament site, particularly as this is most readily available close to real-time. Any differences between the apparent CME origin and the flare/filament site may be caused by the CME not propagating radially with respect to the source location (Plunkett et al., 2001), or deflected by interaction with the ambient corona (Cremades & Bothmer, 2004; Wang et al., 2011), or because of interaction with existing solar wind structures far from the Sun. Of the 269 Earth-directed CME in this study, we identified source locations for 255 CMEs.

The location of Earth-directed CME, binned by helio-longitude and helio-latitude, are shown in Figure 2. Most Earth-directed CME originate between $\pm 15^\circ$ longitude (positive being westward), with more CME considered to be Earth-directed in the western hemisphere (141) than in the east (110), and four CME, which were located directly on the central meridian. Skirgiello (2005) also found a persistent predominance

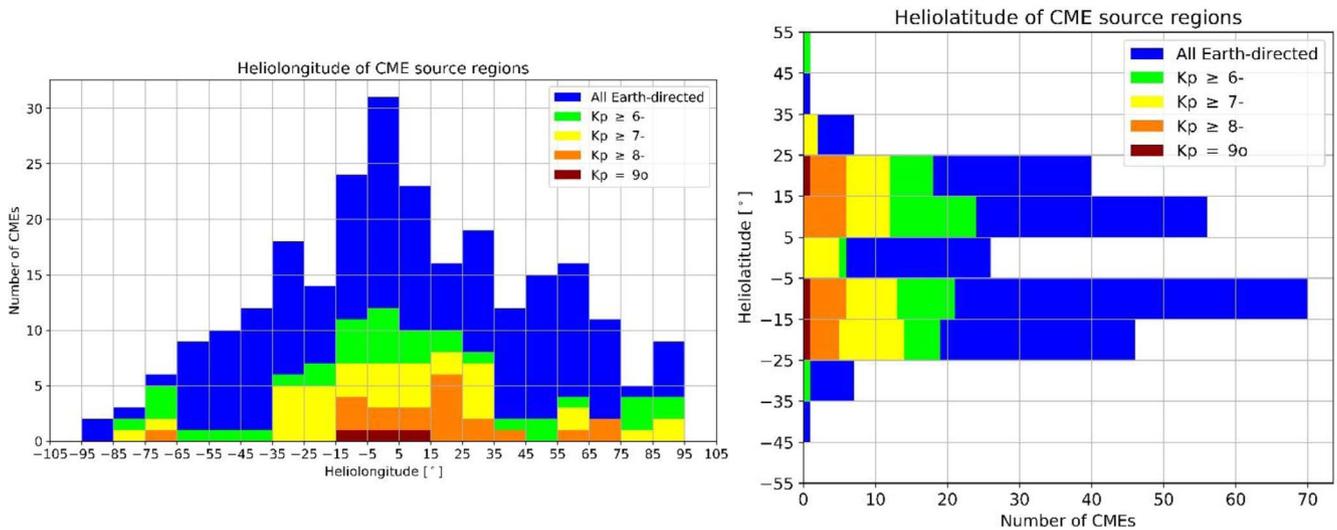


Fig. 2. Count of Earth-directed CME whose source location is known (255 in total), binned by helio-longitude (left) and helio-latitude (right) of any associated flare or filament (blue), together with the count per bin of events that caused storms of $Kp \geq 6-$ (green) through to $Kp = 9o$ (red).

of CME in the western hemisphere, which may be due to some instrumental or observational bias, perhaps caused by the geometry of asymmetrically shaped CME.

CME in the western hemisphere (particularly helio-longitudes of 5° – 25° , though also to larger longitudes) are also more likely to cause storms than other bins, consistent with previous studies (e.g. Kim et al., 2005; Zhang et al., 2007). However, the largest storms ($Kp = 9o$) follow from CME from the central 30° of helio-longitude, consistent with Gopalswamy et al. (2007), who found that the overall strength of geomagnetic storms decreased as the solar source location changes from the centre to the limb. It is also worth noting that CME occurring on either Eastern or Western limbs ($>65^{\circ}$ helio-longitude), which are big enough to have an Earth-directed component also have an increased chance of causing storms. Shen et al. (2014) noted that limb CME could become geo-effective if they have a large angular width.

In helio-latitude, the Earth-directed CME were distributed approximately symmetrically about the equator (119 north, 134 south and two on the equator). The ± 5 – 15° bins have the most Earth-directed CME: 70 in the south and 56 in the north. Wang et al. (2011) identified a peak between 15° and 30° latitude, with very few CME originating equatorward of 15° latitude. This discrepancy between the two studies is likely to be due to the method of defining the source location. Wang et al. (2011) defined the source as the centre of the eruption feature identified in the SOHO/EIT images, then converted the apparent coordinates to heliographic coordinates. All the Earth-directed CME were within $\pm 55^{\circ}$, and all but three were within $\pm 35^{\circ}$ helio-latitude.

From the LASCO catalogue, we also extracted sky-plane speed information for 261 CME, 97 (36%) of which caused storms $\geq G2$ and 23 (9%) caused storms $\geq G4$. The LASCO speeds have a range of 179–3387 km/s, with an average (median) of 996 km/s (911 km/s), which is consistent with Gopalswamy et al. (2007), who found an average of 933 km/s for on-disk halo CME, and 1421 km/s for limb halo CME in

between 1996 and 2005. Zhang et al. (2007) presented a range of around 60–2800 km/s with a mean (median) of 945 km/s (875 km/s) for CME during 1996–2005. The differences in the speed range are in part due to the selection criteria and time span investigated and in part due to Zhang et al. (2007) only including those CME deemed to have caused storms. When we only include those CME that caused storms ($Kp \geq 6-$), we find a range in speeds of 206–2657 km/s and an increase (decrease) in the mean (median) to 1052 km/s (885 km/s).

Figure 3 shows the number of Earth-directed CME in 500 km/s speed bins, regardless of source helio-latitude and helio-longitude, and the count in each bin that caused storms of specific magnitudes. Previous studies show a peak in the distribution of all CME speeds at around 350 km/s (e.g. Yurchyshyn et al., 2005; Mittal et al., 2009) and average CME speeds of around 450 km/s (e.g. Gopalswamy, 2004; Mittal et al., 2009), but here we find relatively few Earth-directed CME with speeds less than 500 km/s. We suspect that the number of Earth-directed CME in this bin may, therefore, be underestimated due to observational uncertainties or biases.

Most Earth-directed CME have speeds of 500–1000 km/s (103 events), and CME in this bin also caused the largest number of geomagnetic storms (36 with $Kp \geq 6-$). The percentage of CME that caused a geomagnetic storm increased with speed (Table 1), up to 75% for CME with speeds in the range 2500–3000 km/s. However, the likely statistical significance of events with the highest speeds is low due to the small number (only 3 events) of Earth-directed CME with a speed greater than 2500 km/s. We note that $Kp = 9o$ events only occur for speeds greater than 1500 km/s, which agrees with Srivastava & Venkatkrishnan (2004), who found CME with speeds greater than 1500 km/s cause “super-intense” storms (based on $Dst < -200nT$).

The source helio-latitude and helio-longitude of CME can also be used to provide information about the likelihood of storms of different magnitudes. Figure 4 shows the number (255 total) of Earth-directed CME in each latitude/longitude

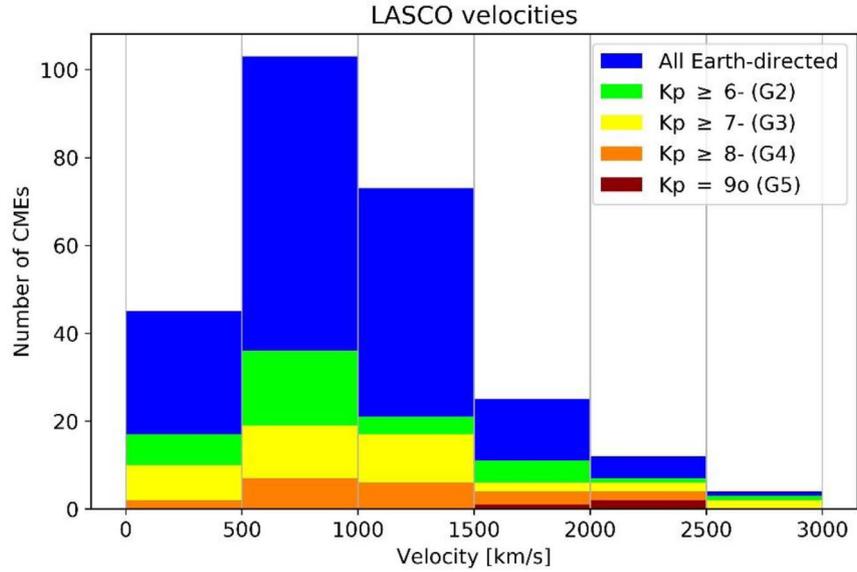


Fig. 3. The total number (261) of Earth-directed CME identified during the study (blue), and the numbers of these CME that resulted in geomagnetic storms (green to red), for Kp of 6– to 9o (G2–G5), binned by LASCO solar wind speed.

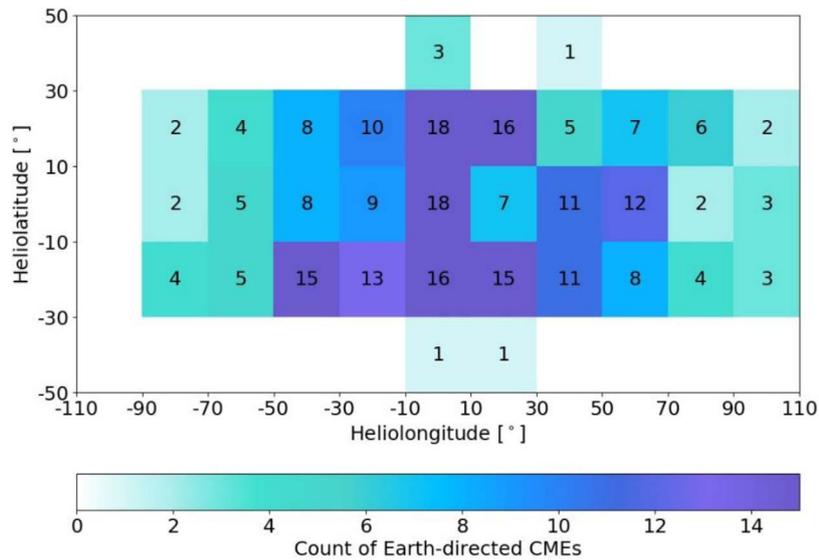


Fig. 4. Total count of Earth-directed CME (255) in January 1998–December 2009, binned by source region latitude and longitude.

bin (note the bin sizes are increased compared to earlier figures to provide a higher count in each bin). Near the limbs and poleward of 30° helio-latitude, there are at most only six CME in any of these bins, so subsequent analysis may be unlikely to be statistically robust.

Figure 5 then shows the percentage of CME in each bin that caused storms with $Kp \geq 6-$ (Fig. 5a), up to $Kp = 9o$ (Fig. 5d). The small sample (3 events) of $Kp = 9o$ storms only occurs from CME located at central helio-longitudes and $\pm 10-30^\circ$ helio-latitude. The highest percentage of G4 storms follow from CME in the bins $10^\circ-30^\circ$ west and at equatorial latitudes. Note that within each bin, we do not distinguish CME speed, that is, all speeds contribute within each bin (c.f. Fig. 3, where we show storms in terms of CME speed bins, irrespective of source location on the Sun).

5 A probabilistic geomagnetic hazard assessment for Kp and EDCME

If we identify the source magnitude, “ m ”, in equation (1), with the EDCME speed, V , and the EDCME source location bin with the parameter “ r ”, then the “Intensity Measure” (IM) is Kp , and the thresholds for Kp are from 6–, up to 9o. We know the speed distribution (i.e. the normalised form of Fig. 3) and the position distribution (normalised form of Fig. 4) for the CME source location, so we can then compute the distribution functions $f(m = V)$ and $f(r = R)$, respectively. To recap the comments at the end of Section 2, the probability distributions $P(\text{IM} > x | r)$ and $P(\text{IM} > x | m)$ can then be obtained by dividing the percentages given in Figures 5 and

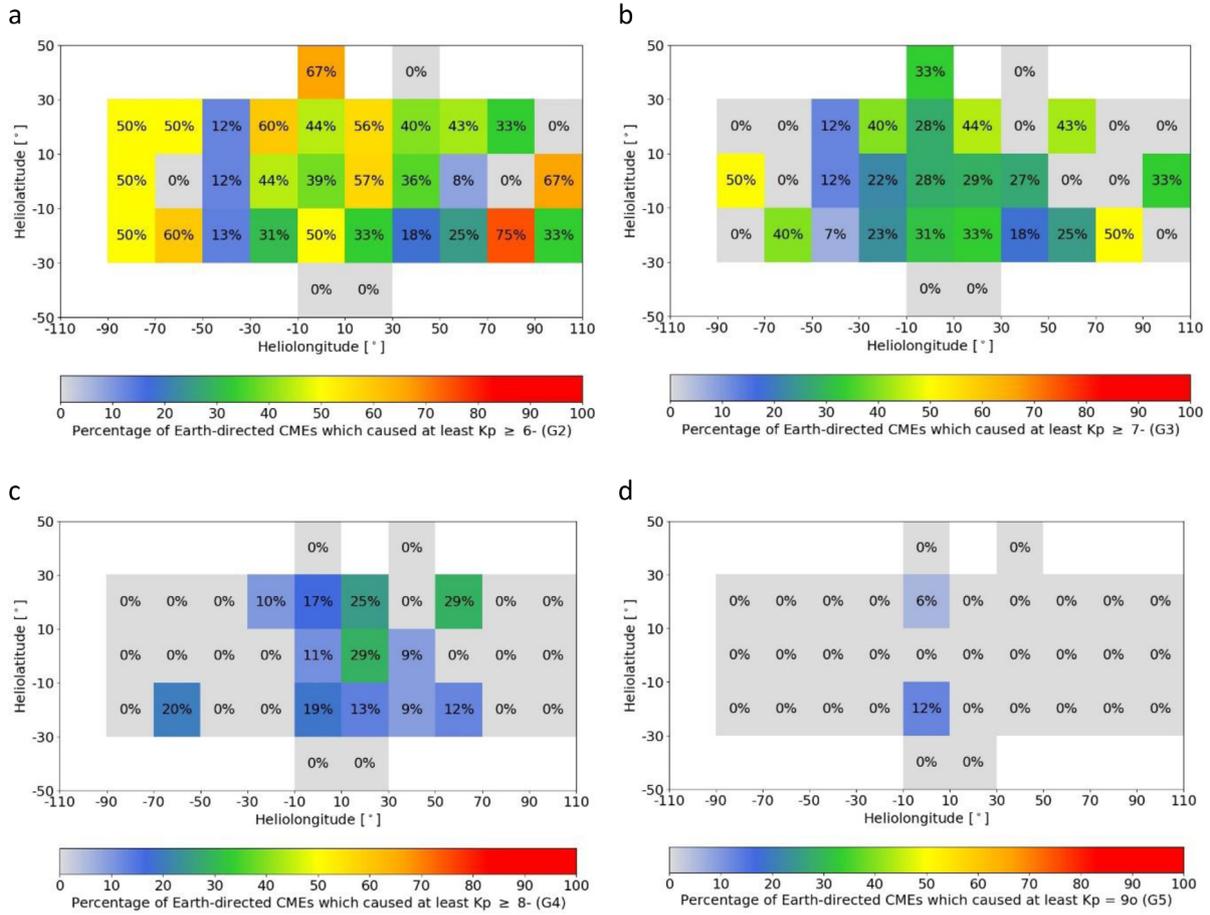


Fig. 5. Percentages of Earth-directed CME (total of 255) in January 1998–December 2009, irrespective of speed (= summed over all speeds), binned by source region latitude and longitude, which caused storms with $Kp \geq 6-$ (a), $Kp \geq 7-$ (b), $Kp \geq 8-$ (c) and $Kp = 9o$ (d).

Table 2. Computed probability of Kp exceeding each threshold, post-observation of an EDCME, via the PGHA method via Method 1 and 2 (2nd and 3rd row respectively); the a priori probability of Kp above each threshold (4th row), and probability scaling factors, i.e. “EDCME-observed/a priori” (5th and 6th row). Since statistical uncertainty is not quantified, data have been rounded to three figures.

Kp threshold	$Kp > 6-$ (G2)	$Kp > 7-$ (G3)	$Kp > 8-$ (G4)	$Kp = 9o$ (G5)
Method 1: PGHA computed probability for storm from a CME, after EDCME observation (A1)	0.359	0.227	0.0903	0.0113
Method 2: PGHA computed probability for storm from a CME, after EDCME observation (A2)	0.360	0.230	0.0881	0.0115
A priori probability for storm from all sources, from Kp distribution (B)	0.0661	0.0258	0.00980	0.00068
Check: Ratio (A1/B)	5.43	8.82	9.21	16.6
Check: Ratio (A2/B)	5.45	8.92	8.99	16.9

Table 1, respectively, by 100%. From these we can compute $P(IM = Kp > \text{threshold})$, by means of equations (2) and (4).

In Table 2, we show the results of the analysis. The second row of Table 2 describes the computed post-EDCME probability, i.e. $P(Kp > \text{threshold})$, using Method 1. The third row then cross-checks this via Method 2. The fourth row provides the a priori probability of exceeding each threshold from all solar sources, based on the known historical Kp distribution and which is therefore independent of solar observations. This a priori probability is calculated as the number of days where

Kp exceeded each threshold, divided by the total number of days in the dataset (1998–2009), using Kp collected from GFZ (Matzka et al., 2021b). In rows five and six, we give a check on the scaling between the “post-EDCME-observed” and “a priori” cumulative probabilities for each method.

We note that the probability of Kp exceeding each threshold increases by around an order of magnitude, compared to the a priori probability, after observation of an EDCME (i.e. rows five and six). The difference between rows 2 and 3 also indicates the impact that some of the assumptions and

simplifications have made, resulting in a variation of 5% ($Kp > 6-$) to 15% ($Kp = 9o$) in event probabilities.

Finally, we have used the sky-plane speeds, from height-time measurements, as these are closest to what is available in real-time. However, these measurements are subject to projection effects, with a dependence on the longitude of the solar source (Gopalswamy et al., 2000). Indeed, Gopalswamy et al. (2010) found that correcting the sky-plane speeds to use space speeds led to an increase in the median and mean speeds of around 17% for a set of halo CMEs, with CMEs from source longitudes $<45^\circ$ the most affected. This will change the underlying EDCME speed distribution (Fig. 3), which should be included in a future analysis.

6 Summary and concluding remarks

We introduced PHA as a methodology for evaluating the probability of exceeding space weather hazard thresholds for which there may be limited measured data. We then chose to illustrate the method empirically by use of measured CME and Kp . This has had the useful consequences of a) allowing calculation of the “raw” $P(Kp > \text{threshold})$ probabilities through two implementations of equation (1) and, b) determining the extent to which the “raw” or a priori likelihoods increase post-EDCME observation (by a factor of ~ 10).

The results are, of course, dependent on the assumptions made in selecting the solar and magnetic data sets, projection effects, errors arising from missing data, measurement errors and the exclusion of coronal hole impacts from the analysis. The latter is likely to be particularly significant at $Kp < 7-$ (i.e. Fig. 1a) since, of the 290 storm events in the sample, 163 were due to CME and 111 to CH. One might therefore expect that a more complete PGHA estimate at G2 (0.359) could increase by a factor of order $\sim 290/163$, or 1.8 if we were to include CH impacts. This would increase the “post-EDCME”/“a priori” storm ratio at G2 and above to around 9.7, more in line with results for the higher Kp thresholds, for which CH activity proves much less significant.

The results are also dependent on computational assumptions we have made, such as the CME speed and location being independent, as discussed in Sections 2 and 5. We have also approximated how we calculate equation (1) through equations (2) or (4), and this is something that could be possibly done differently. Our results also do not factor in the interplanetary magnetic field direction at Earth, or other consequences of physical processes from EDCME observation, through to Kp impact. Finally, more sophisticated statistical analysis of the difference between Method 1 and 2, or consequence of solar measurement uncertainty, other than already offered here, we leave to future work.

However, to conclude, we emphasise that the general point of this paper is to introduce PHA as a general framework to determine hazard probabilities, return times etc., for more novel measures of space weather impact. The full realisation of PHA in space weather science will rely on the development of Sun to Earth model(s) for $P(IM > x | m, r)$, which comprehensively capture the physics of the solar source and the nonlinear interplanetary-magnetospheric-ionospheric medium, and how this affects particular technologies, according to different, impact measures, IM.

Acknowledgements. This paper is published with the permission of the Executive Director, BGS (NERC), and we thank colleagues for their comments and suggestions. We would also like to thank the two referees for taking the time to provide in-depth reviews, which have greatly improved the paper. We acknowledge support for this work from project “SWIGS” (Space Weather Impact on Ground-based Systems), under UK Natural Environment Research Council grant number NE/P017231/1. The figures in this paper were all made using matplotlib v2.0 (Hunter, 2007). The editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

- Baker JW. 2013. *Probabilistic seismic hazard analysis*. White Paper Version 2.0.1 [Online]. Available at: [http://web.stanford.edu/~bakerjw/Publications/Baker_\(2013\)_Intro_to_PSHA_v2.pdf](http://web.stanford.edu/~bakerjw/Publications/Baker_(2013)_Intro_to_PSHA_v2.pdf).
- Cremades H, Bothmer V. 2004. On the three-dimensional configuration of coronal mass ejections. *A&A* **422**: 307–322. <https://doi.org/10.1051/0004-6361:20035776>.
- Gerstenberger MC, Marzocchi W, Allen T, Pagani M, Adams J, et al. 2020. Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Rev Geophys* **58**: e2019RG000653. <https://doi.org/10.1029/2019RG000653>.
- Gopalswamy N. 2004. A global picture of CMEs in the inner heliosphere. In: *The sun and the heliosphere as an integrated system*. Poletto G, Suess ST (Eds.), Kluwer, Boston. pp. 201–251.
- Gopalswamy N, Kaiser ML, Thompson BJ, Burlaga LF, Szabo A, Lara A, Vourlidas A, Yashiro S, Bougeret J-L. 2000. Radio-rich solar eruptive events. *Geophys Res Lett* **27**: 1427–1430. <https://doi.org/10.1029/1999gl003665>.
- Gopalswamy N, Yashiro S, Akiyama S. 2007. Geoeffectiveness of halo coronal mass ejections. *J Geophys Res (Space Phys)* **112**: A6. <https://doi.org/10.1029/2006JA012149>.
- Gopalswamy N, Yashiro S, Michelak G, Xie H, Mäkelä P, Vourlidas A. 2010. A catalog of halo coronal mass ejections from SOHO. *Sun Geosphere* **5**: 7–16.
- Grezio A, Babeyko A, Baptista MA, Behrens J, Costa A, et al. 2017. Probabilistic tsunami hazard analysis: Multiple sources and global applications. *Rev Geophys* **55**: 1158–1198. <https://doi.org/10.1002/2017RG000579>.
- Harrison RA. 1995. The nature of solar flares associated with coronal mass ejection. *A&A* **585**–594. <https://ui.adsabs.harvard.edu/abs/1995A%26A...304.585H>.
- Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**: 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Kijko A. 2011. Introduction to probabilistic seismic hazard analysis. In: *Encyclopaedia of solid earth geophysics*, Gupta H, (Ed.), Springer, Dordrech. pp. 1107–1120. https://doi.org/10.1007/978-90-481-8702-7_10
- Kim R-S, Cho K-S, Moon Y-J, Kim Y-H, Yi Y, Dryer M, Bong S-C, Park Y-D. 2005. Forecast evaluation of the coronal mass ejection (CME) geoeffectiveness using halo CMEs from 1997 to 2003. *J Geophys Res (Space Phys)* **110**: A11. <https://doi.org/10.1029/2005JA011218>.
- Matzka J, Stolle C, Yamazaki Y, Bronkalla O, Morschhauser A. 2021a. The geomagnetic Kp index and derived indices of geomagnetic activity. *Space Weather* **19**: e2020SW002641. <https://doi.org/10.1029/2020SW002641>.
- Matzka J, Bronkalla O, Tornow K, Elger K, Stolle C. 2021b. *Geomagnetic Kp index V. 1.0*. GFZ Data Services. <https://doi.org/10.5880/Kp.0001>.

- Mittal N, Sharma J, Tomar V, Narain U. 2009. On distribution of CMEs speed in solar cycle 23. *Planet Space Sci* **57**: 53–57. <https://doi.org/10.1016/j.pss.2008.10.013>.
- Moon YJ, Kim RS, Cho KS. 2009. Geometrical implication of the CME earthward direction parameter and its comparison with cone model parameters. *J Korean Astron Soc* **42**: 27–32. <https://doi.org/10.5303/JKAS.2009.42.2.027>.
- Pizzo V, Millward G, Parsons A, Biesecker D, Hill S, Odstreil D. 2011. Wang-Sheeley-Arge-Enlil cone model transitions to operations. *Space Weather* **9**: S0–3004. <https://doi.org/10.1029/2011SW000663>.
- Plunkett SP, Thompson BJ, St. Cyr OC, Howard RA. 2001. Solar source regions of coronal mass ejections and their geomagnetic effects. *J Atmos Sol-Terr Phys* **63**: 389–402. [https://doi.org/10.1016/S1364-6826\(00\)00166-8](https://doi.org/10.1016/S1364-6826(00)00166-8).
- Pulkkinen A, Hesse M, Habib S, Van der Zel L, Damsky B, Policelli F, Fugate D, Jacobs W, Creamer E. 2010. Solar shield: Forecasting and mitigating space weather effects on high-voltage power transmission systems. *Nat Hazards* **53**: 333–345. <https://doi.org/10.1007/s11069-009-9432-x>.
- Richardson IG, Cane HV. 2010. Near-Earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties. *Sol Phys* **264**: 189–237. <https://doi.org/10.1007/s11207-010-9568-6>.
- Schwenn R, Dal Lago A, Huttunen E, Gonzalez WD. 2005. The association of coronal mass ejections with their effects near the Earth. *Ann Geophys* **27**: 1033–1059. <https://doi.org/10.5194/angeo-23-1033-2005>.
- Shen C, Wang Y, Pan Z, Miao B, Ye P, Wang S. 2014. Full-halo coronal mass ejections: Arrival at the Earth. *J Geophys Res (Space Phys)* **119**: 5107–5116. <https://doi.org/10.1002/2014JA020001>.
- Skirgiello M. 2005. The east-west asymmetry in Coronal Mass Ejections: evidence for active longitudes. *Ann Geophys* **23**: 3139–3147. <https://doi.org/10.5194/angeo-23-3139-2005>.
- Srivastava N, Venkatakrisnan P. 2004. Solar and interplanetary sources of major geomagnetic storms during 1996–2002. *J Geophys Res (Space Phys)* **109**: A10. <https://doi.org/10.1029/2003JA010175>.
- 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 Phys)* **116**: A4. <https://doi.org/10.1029/2010JA016101>.
- Wang YM, Ye PZ, Wang S, Zhou GP, Wang JX. 2002. A statistical study on the geoeffectiveness of Earth-directed coronal mass ejections from March 1997 to December 2000. *J Geophys Res (Space Phys)* **107**: SSH 2-1–SSH 2-9. <https://doi.org/10.1029/2002JA009244>.
- Yashiro S, Gopalswamy N, Michalek G, Cyr OCS, Plunkett SP, Rich NB, Howard RA. 2004. A catalog of white light coronal mass ejections observed by the SOHO spacecraft. *J Geophys Res (Space Phys)* **109**: A7. <https://doi.org/10.1029/2003JA010282>.
- Yurchyshyn V, Yashiro S, Abramenko V, Wang H, Gopalswamy N. 2005. Statistical distributions of speeds of coronal mass ejections. *Astrophys J* **619**: 599–603. <https://doi.org/10.1086/426129>.
- Zhang J, Richardson IG, Webb DF, Gopalswamy N, Huttunen E, Kasper JC, Nitta NV, Poomvises W, Thompson BJ, Wu C-C, Yashiro S, Zhukov AN. 2007. Solar and interplanetary sources of major geomagnetic storms ($Dst \leq -100$ nT) during 1996–2005. *J Geophys Res (Space Phys)* **112**: A10. <https://doi.org/10.1029/2007JA012321>.

Cite this article as: Richardson GS & Thomson AWP, 2022. Probabilistic hazard assessment: Application to geomagnetic activity. *J. Space Weather Space Clim.* **12**, 4. <https://doi.org/10.1051/swsc/2022001>.