Clustering of Fast Coronal Mass Ejections during Solar Cycles 23 and 24 and the Implications for CME–CME Interactions

Jenny M. Rodríguez Gómez1 ⊙, Tatiana Podladchikova1 ⊙, Astrid Veronig2, Alexander Ruzmaikin3, Joan Feynman4, and Anatoly Petrukovich5

1 Skolkovo Institute of Science and Technology, Territory of Innovation Center “Skolkovo,” Bolshoy Boulevard 30, bld. 1, Moscow 121205, Russia; Jenny.RodriguezGomez@skoltech.ru
2 Institute of Physics & Kanzelhöhe Observatory, University of Graz, Universitätsplatz 5, 8010 Graz, Austria
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
4 Heliophysics, Glendale, CA 91214, USA
5 Space Research Institute, Moscow, Russia

Received 2019 December 31; revised 2020 June 17; accepted 2020 June 17; published 2020 August 11

Abstract

We study the clustering properties of fast coronal mass ejections (CMEs) that occurred during solar cycles 23 and 24. We apply two methods: the Max Spectrum method can detect the predominant clusters, and the declustering threshold time method provides details on the typical clustering properties and timescales. Our analysis shows that during the different phases of solar cycles 23 and 24, CMEs with speeds \( >1000 \text{ km s}^{-1} \) preferentially occur as isolated events and in clusters with, on average, two members. However, clusters with more members appear, particularly during the maximum phases of the solar cycles. Over the total period and in the maximum phases of solar cycles 23 and 24, about 50% are isolated events, 18% (12%) occur in clusters with two (three) members, and another 20% in larger clusters. These findings suggest that \( \tau_c = 28–32 \text{ hr} \), irrespective of the very different occurrence frequencies of CMEs during a solar minimum and maximum. These timescales for extreme events may reflect the characteristic energy build-up time for large flare and CME-prolific active regions. Statistically associating the clustering properties of fast CMEs with the disturbance storm time index at Earth suggests that fast CMEs occurring in clusters tend to produce larger geomagnetic storms than isolated fast CMEs. This may be related to CME–CME interaction producing a more complex and stronger interaction with Earth’s magnetosphere.

Unified Astronomy Thesaurus concepts: Solar coronal mass ejections (310); Solar corona (1483)

1. Introduction

Coronal mass ejections (CMEs) are a manifestation of solar variability and the main source of major space weather events. CMEs eject substantial amounts of mass and magnetic flux from the Sun to the heliosphere and cause disturbances in the interplanetary medium (Gopalswamy et al. 2009). The initiation and impulsive acceleration of CMEs occur on timescales of a few minutes to several hours with a kinetic energy that may exceed \( 10^{32} \text{ erg} \) (Hundhausen et al. 1984; Schwenn 1996; Vourlidas et al. 2000; Hudson et al. 2006; Bein et al. 2011). CME speeds range from \(<100 \text{ to } >2000 \text{ km s}^{-1} \), occasionally reaching up to \( 3500 \text{ km s}^{-1} \) (Yashiro et al. 2004; Gopalswamy et al. 2009; Chen 2011; Webb & Howard 2012). During their propagation, the interplanetary manifestations of the CMEs (ICME) may interact with Earth (and other planets), producing space weather impacts on their environment and technology (Schwenn 2006; Baker et al. 2013; Riley et al. 2018).

The characteristics of extreme solar phenomena and extreme space weather events help us to better understand the dynamics and variability of the Sun as well as the physical mechanisms behind these events (Koskinen et al. 2017; Green et al. 2018). In this article, we will characterize extreme events in the form of fast CMEs. Large flares and fast CMEs predominantly originate from complex active regions that contain large amounts of magnetic flux (Sammis et al. 2000; Murray et al. 2018; Tschernitz et al. 2018; Toriumi & Wang 2019). Observations have shown that active regions tend to occur in clusters. This behavior is related to the magnetic flux emergence of new active regions, which preferentially emerge in the vicinity of old ones (Gaizauskas et al. 1983; Harvey & Zwaan 1993; Ruzmaikin 1998; Ruzmaikin et al. 2011).

Multiple CMEs launched from complex active regions are not rare. Ruzmaikin et al. (2011) showed that fast CMEs in particular tend to occur in clusters. This clustering may lead to interactions of successive CMEs, either already close to the Sun or in interplanetary space. Solar observations reveal that CME–CME interaction occurs in particular for CMEs that are launched in sequence from the same active region. During their propagation from the Sun to Earth, the CME–CME interaction can be related to enhanced particle acceleration and can generate more intense geomagnetic storms than isolated CMEs when arriving at Earth (Farrugia & Berdichevsky 2004; Dumbović et al. 2015; Vennerstrom et al. 2016; Lugaz et al. 2017).

Here we study the temporal distribution of fast CMEs, with the main focus on the statistical interpretation of the occurrence and clustering properties of CMEs, how these change for different solar cycle phases, and also how the clustering properties may affect the geoeffectivity of CMEs. To this aim, we apply two different approaches, the Max Spectrum method and the declustering threshold time method (Ferro & Segers 2003; Stoev et al. 2006; Ruzmaikin et al. 2011). The Max Spectrum method provides two exponents: the power-tail exponent (\( \alpha \)) describing the probability distribution of the speeds of fast CMEs and the extremal index (\( \theta \)) that separates individual
clusters and also provides an estimate of the predominant cluster size. The declustering threshold time is used to identify clusters in time series of CMEs with speeds larger than 1000 km s\(^{-1}\). This method provides information about the cluster size, the mean cluster duration and the mean time between successive fast CMEs.

The paper is structured as follows. In Section 2, we describe the CME data set. In Section 3, we explain the Max Spectrum method and the concept of declustering threshold time. In Section 4, the main results of our study on the clustering of fast CMEs that occurred during solar cycles 23 and 24 are presented. In addition, the same analysis is performed separately for different phases of the solar cycles. In Section 5 we present a statistical approach to relate the CME-clustering properties to their potential geoeffectivity. In Section 6, we summarize our main findings and discuss their implications.

2. Data Set

We use data from the Large Angle and Spectrometric Coronagraph\(^\text{5}\) (LASCO; Brueckner et al. 1995) on board the Solar and Heliospheric Observatory (SOHO) satellite. This catalog contains all the CMEs detected by the LASCO coronagraphs C2 and C3, which cover a combined field of view (FOV) from 2.1 to 32\(R_\odot\). The catalog provides several attributes to characterize the CMEs: date and time of appearance in the coronagraph FOV, angular width, speed from the linear fit to the height–time measurements, speed from the quadratic fit at the last height of measurement, and speed from the quadratic fit at 20\(R_\odot\) (Yashiro et al. 2004; Gopalswamy et al. 2009). We use the speed from the quadratic fit at the time/distance of the first data point, which gives the speed closer to the initiation of the CME eruption in the low corona before other interactions occur.

We follow the approach outlined in Ruzmaikin et al. (2011) and build the time series from the hourly spaced time series of CME speeds. The hours with no CME occurrence are assigned a zero speed. In the few cases where more than one CME occurred within one hour, the highest CME speed is chosen. We use the entire LASCO data available from 1996 January to 2018 March (resulting in a total set of 25,895 CMEs) almost completely covering the last two solar cycles. We note that our data also include the recently occurring strongest events of cycle 24, namely the X9.3 and X8.2 flare/CME events from 2017 September 6 and 10 (e.g., Yang et al. 2017; Guo et al. 2018; Liu et al. 2018; Mitra et al. 2018; Seaton & Darnel 2018; Veronig et al. 2018; Romano et al. 2019).

Figure 1 shows the CME speeds and their number together with the sunspot number from 1996 to 2018. The monthly mean sunspot number was obtained from WDC-SILSO, Royal Observatory of Belgium (top panel). The second and third panels show the CME speed and the monthly means of the daily number of CMEs from the CDAW LASCO Catalog. The monthly mean of the daily number of CMEs with speeds \(v \geq 1000\text{ km s}^{-1}\) are plotted in the bottom panel. These curves were smoothed over 13 months (red lines). The vertical dotted lines mark three 4 yr intervals centered at the maximum and minimum phases of solar cycles 23 and 24, for which we study the CME-clustering properties separately.

\(^{5}\) https://cdaw.gsfc.nasa.gov/CME_list/

Figure 2 shows the probability distribution function (PDF) for the solar cycles 23 and 24 and a close-up of the distribution for speeds \(v \geq 1000\text{ km s}^{-1}\). These PDFs of the CME speeds reveal a non-Gaussian heavy-tailed distribution (Yurchyshyn et al. 2005; Ruzmaikin et al. 2011). This means that fast CMEs occur with a much higher probability than expected from a normal or exponential distribution. The distribution of CME speeds shown on top has a mode of 195\text{ km s}^{-1} and a mean speed of 357 \(\pm\) 282\text{ km s}^{-1}. There are 899 CMEs (3.5% of the overall sample) with speeds exceeding 1000\text{ km s}^{-1}, 2847 CMEs (11%) with speeds \(>700\text{ km s}^{-1}\), and 63 CMEs (0.2%) achieve speeds \(>2000\text{ km s}^{-1}\).

The dominant interplanetary phenomena causing intense magnetic storms (\(Dst < -100\text{ nT}\)) are ICMEs. The statistical dependence of disturbance storm time \((Dst)\) minima during storms were widely explored (Kane & Echer 2007; Echer et al. 2008, 2013; Podladchikova & Petrukovich 2012; Podladchikova et al. 2018). The main properties that determine the geoeffectivity of an ICME impacting Earth’s magnetosphere are its arrival speed and the strength of the \(B_t\) component of the interplanetary magnetic field (Gosling et al. 1991; Gonzalez et al. 1999; Borovsky & Denton 2006). While currently we still have no proper handle to derive estimates of \(B_t\) from solar observations (e.g., Green et al. 2018; Vourlidas et al. 2019), the ICME speed impacting at Earth (though evolution in interplanetary space due to the drag force exerted by the ambient solar wind) is related to the CME speed that we can derive from the coronagraph observations near the Sun (e.g., Vršnak et al. 2013). Thus, in order to characterize extreme events, we focus our analysis here on fast CMEs in particular, defined as those with speeds \(v > 1000\text{ km s}^{-1}\), which comprise about 3.5% of the overall sample.

3. Methods

Here we describe two statistical methods to study and characterize the clustering properties of the fast CMEs. These are the Max Spectrum and the declustering threshold time methods.

3.1. Max Spectrum Method

The Max Spectrum method is based on the block maxima technique (Stoev et al. 2006). In general, a real valued random variable \(X\) with a cumulative distribution function \(F(x) = P\{X \leq x\}, x \in R\) is said to have a right heavy tail if

\[ P\{X > x\} = 1 - F(x) = L(x)x^{-\alpha}, \text{ as } x \rightarrow \infty \]

for some \(\alpha > 0\), where \(L(x) > 0\) is a slowly varying function. The tail exponent \(\alpha > 0\) controls the rate of decay of \(F\) and hence characterizes its tail behavior. If we consider the case where the slowly varying function \(L\) is trivial, when

\[ P\{X > x\} = 1 - F(x) \sim \sigma_0 x^{-\alpha}, \text{ as } x \rightarrow \infty \]

with \(\sigma_0\) and where \(\sim\) means that the ratio of the left-hand side to the right-hand side in Equation (1) tends to 1, as \(x \rightarrow \infty\). We assume that the \(X(i)\)’s are almost surely positive \((F(0) = 0)\) (Stoev et al. 2006).

In the application of this method, we use the hourly time series of CME speeds created, without using a CME speed threshold. In our case the variable \(X(i)\) corresponds to the CME speed \(v(i)\). This method starts with taking averages of data maxima in time intervals (blocks) with a fixed size. The block
size is then progressively increased. Here we consider the time series of total length \( N \) for the CME speed \( v(i) \), where \( 1 \leq i \leq N \) and \( j \) introduce the time-interval scale index \( j = 1, 2, 3, \ldots, \log_2(N) \). To form nonoverlapping time blocks of length \( 2^j \), at each fixed scale \( j \) we calculate the maximum CME speed within each block

\[
D(j, k) = \max_{1 \leq i < 2^j} v(2(k - 1) + i) \quad k = 1, 2, \ldots, b_j + 1
\]

where \( b_j = \left\lceil \frac{N}{2^j} \right\rceil \) is the number of blocks (of length \( 2^j \)) at each scale \( j \) and \( k \) defines the location of the block on the time axis.

The blocks of scale \( j \) are naturally contained in the blocks of scale \( j + 1 \). Now, we average the binary logarithms of the block maxima \( D(j, k) \) over all blocks at a fixed scale \( j \), i.e.,

\[
Y(j) = \frac{1}{b_j} \sum_{k=1}^{b_j} \log_2 D(j, k). \tag{3}
\]

The function \( Y(j) \) is called the max spectrum of the data. Stoever et al. (2006) established an important result: the max spectrum for a time series with sufficiently large \( j \) scales can be expressed as

\[
Y(j) \approx \frac{j}{\alpha} + C, \tag{4}
\]

where \( C \) is a constant and \( \alpha > 0 \). The tail of the data distribution follows a power law with exponent \(-\alpha\). The exponent \( \alpha \) is called the power-tail exponent, which allows us...
to characterize what kind of distribution is related to the time series. In general, under the generalized extreme value (GEV) theory, some distributions are obtained depending on location and a scale parameter. One of them is called the Fréchet distribution. In general, this distribution shows a right-side tail that decays like a power law (Leadbetter et al. 1983; Hsing 1988; McNeil et al. 2005). Ruzmaikin et al. (2011) have shown that fast solar CMEs can be described as a Fréchet distribution. In this paper, we are interested in the Fréchet distribution and parameters describing the clustering of the fast CMEs that occurred during solar cycles 23 and 24.

Equation (4) is valid for statistically independent events. If we have dependent data with the same distribution function (Stoev et al. 2006), then Equation (4) can be transformed into

$$Y(j) \approx j^{-1/\alpha} + C + \log_2(\theta)/\alpha,$$

where $C$ is a constant and $\theta$ is called “the extremal index” which takes values in the interval $[0, 1]$ (Leadbetter et al. 1983). Values of $\theta$ close to 0 indicate a strong dependence and the possibility to form clusters, while values close to 1 show a weak dependence indicating individual independent events. Note that this index characterizes only the dependence of the extremal values in the time series data (Hamidieh et al. 2010).

Equations (4) and (5) suggest a method of estimating $\alpha$ and $\theta$ (Stoev et al. 2006; Hamidieh et al. 2010). The power-tail exponent $\alpha$ is obtained on the self-similar range of the max spectrum $Y(j)$. This range can be related to the self-similar cascade process in turbulence (Frisch 1995; Ruzmaikin et al. 2011). Similarly, here we check the self-similar interval to obtain the slope of the line fitted and obtain the power-tail exponent $\alpha$. The inverse exponent $1/\alpha$ is obtained as a slope of the line fitted to the max spectrum of the data in the self-similar range (Stoev et al. 2006; Ruzmaikin et al. 2011). The process to select the self-similar range is fundamental for obtaining the power-tail exponent and it influences the extremal index $\theta$ and the cluster number. In general, different intervals were checked on the self-similar range to obtain the slope and the power-tail exponent $\alpha$. The best fitted line and its corresponding correlation coefficient guides the choice of this interval. The max spectrum $Y(j)$ and the power-tail exponent $\alpha$ are key parameters in the estimation of the extremal index $\theta$.

The extremal index $\theta$ defines the number of independent clusters and provides an estimate of the cluster size given as $(1/\theta)$ (Leadbetter et al. 1983). To calculate the extremal index $\theta$, the original data is first randomly permuted. The new data series $(Y^{(\text{p})})$ is obtained in the interval $1 \leq t \leq N$, and has the same distribution function as the original data. The randomization destroys the dependence structure of the data, resulting in an approximately independent sample (Hamidieh et al. 2010). The new max spectrum $Y(j)^{\text{p}}$ is related to the data series randomly permuted, and satisfies Equation (4), which means that the tail of the new data distribution follows a power law with exponent $-\alpha$. Since the max spectrum of the original data $Y(j)$ satisfies Equation (5) with the same constant $C$, the difference between the two spectra yields an estimate of the extremal index of

$$\theta = 2^{-\alpha(Y(j)^{\text{p}} - Y(j))},$$

where $\alpha$ is the fitting parameter in the power-tail exponent obtained from the max spectrum $Y(j)$. Then, we calculate the differences of $Y(j)^{\text{p}} - Y(j)$ and compute the mean for positive differences to obtain an estimate of the extremal index $\theta$ at each scale $j$. This procedure is repeated 100 times and 100 $\theta(j)$ values are calculated to produce a sequence of $\theta(j)$ box plots for each scale $j$. We obtain an empirical 95% confidence interval, based on the 0.025th and 0.975th empirical quantiles of $\theta(j)$ from the histogram of $\theta$ values. In practice, Hamidieh et al. (2010) recommend selecting the middle range of scales for $\theta$ estimation. At large scales $j$ (larger block sizes) the bias is lower and the number of block maxima is reduced. At lower scales $j$ (smaller block sizes) the bias grows.

We developed a code to calculate the max spectrum and extremal index $\theta$. The performance of these estimations is examined carefully. We compared our results with the code7 from Stoev et al. (2006) and the results presented in Ruzmaikin et al. (2011) for the years 1999–2006. Additionally, we examined our code using the Max-AutoRegressive model of order one (Max-AR(1)) to obtain the extremal index $\theta$. We use the Max-AR(1) series as $y = \max_{ar}(b, a, \text{iter}, N)$, with the length of time series $N = 2^{15}$, number of iteration $\text{iter} = 1$, and the parameters $a = 0.7$, $b = 1.5$. Our results are in agreement with the results based on the code from Stoev et al. (2006).

3.2. Declustering Threshold Time $\tau_c$

In addition, to derive a more detailed characterization of the cluster properties and how they change over different phases of a solar cycle, we use the declustering threshold time method. The main idea is to identify clusters in the time series of CMEs with speeds $v \geq 1000$ km s$^{-1}$. The mean time interval between CMEs within a cluster depends on the speed threshold $v$. In our case, we choose CME speeds $v \geq 1000$ km s$^{-1}$ to characterize extreme events. The extremal index provides an estimate of the number of clusters $(m \times \theta)$, where $m$ is the number of extreme events within a given time interval, e.g., $m$ CMEs with speeds exceeding a threshold $v$ occur during this interval. These CMEs are on average grouped into a cluster. The declustering threshold time $\tau_c$ concept is useful to group CMEs into clusters. Consider the time intervals $\tau_i$ between consecutive fast CMEs. If the time interval between two fast CMEs is less than $\tau_c$, then these CMEs can be grouped into a cluster (Smith 1989; Ferro & Segers 2003; Beirlant et al. 2004; Ruzmaikin et al. 2011).

To determine the declustering threshold time $\tau_c$, we use the PDF of time intervals $\tau_i$ between consecutive fast CMEs and the GEV distribution. $\tau_c$ was defined as the maximum $(\sigma)$ value of the GEV distribution of time intervals $(\tau_i)$ between consecutive fast CMEs. Figure 3 shows the PDF of time intervals $\tau_i$ between consecutive fast CMEs $(v \geq 1000$ km s$^{-1})$ and the GEV distribution (red line) at each period of the solar cycle. The declustering threshold time during the whole interval (solar cycles 23 and 24) is $\tau_c \leq 28.0 \pm 1.4$ hr (Figure 3(a)). Applying the method separately to the different phases of the solar cycle, we find $\tau_c \leq 28.2 \pm 2.0$ hr for the maximum of cycle 23, $\tau_c \leq 32.0 \pm 3.6$ hr for the maximum of cycle 24, and $\tau_c \leq 32.5 \pm 17.5$ hr for the minimum phase between cycles 23 and 24 (Figures 3(b)–(d)).

7 https://sites.lsa.umich.edu/sstoev/software/
4. Clustering of the Observed Fast CMEs

Here we apply the methodology described in Section 3 to fast CMEs listed in the LASCO CDAW catalog. The max spectrum $Y(j)$ within the self-similar range is used to obtain the power-tail exponent ($\alpha$). In the max spectrum plots, the log$_2$ units along the y-axis are converted to km s$^{-1}$, using $2^{Y(j)}$ and the scales $j$ on the x-axis are converted into time units using $2^j$. See the top panels in Figures 4–7, which show the results for different periods during solar cycles 23 and 24. We obtained a set of box plots with $\theta$ values (middle panels in Figures 4–7). The central mark in the box plots is the median, the box edges are the 10th and 90th percentiles, and the whiskers extend to the most extreme data points. Additionally, we built a histogram with the $\theta$ values and we calculate an empirical 95% confidence interval, based on 0.025th and 0.975th empirical quantiles in each histogram (vertical dotted lines in the bottom panels in Figures 4–7). This procedure allows us to obtain an interval of the extremal index $\theta$ values as well as to obtain an estimate of the predominant cluster size $(1/\theta)$.

In Section 4.1 we present the analysis applied to the whole time period under study, i.e., from 1996 January to 2018 March. However, the CME occurrence rate and speeds vary over the solar cycle (see Figure 1). To take this variability into account, in Sections 4.2–4.4 we apply the same method for different subperiods to study the variations of the CME clustering over different solar cycle phases. In particular, we select three periods each covering 4 yr, representative of the maximum phase of cycles 23 and 24 as well as the minimum between cycles 23 and 24, as marked in Figure 1.

4.1. Full Period 1996–2018

The full interval covers the time range from 1996 January to 2018 March, i.e., it covers almost entirely cycles 23 and 24. During this period the length of the hourly CME speed time series is $N = 193,944$ and the number of scales is $j = \log_2(N)$. Thus, we have $j = 17$ available scales to apply the Max Spectrum method. Our best fit to the slope gives evidence that the cumulative distribution function of the CME speeds has a Fréchet-type power-law tail, with a power-law exponent.

Figure 3. Determination of the declustering threshold time $\tau_c$ at each phase of the solar cycle. PDF and GEV distribution (red lines) of time intervals $\tau_i$ between consecutive fast CMEs. (a) Solar cycles 23 and 24, derived declustering time $\tau_c = 28.0$ hr. (b) Maximum of solar cycle 23, $\tau_c = 28.2$ hr. (c) Maximum of solar cycle 24, $\tau_c = 32.0$ hr. (d) Solar minimum $\tau_c = 32.5$ hr.
\(\alpha = 3.5\) (Figure 4 (top)). Box plots of the extremal index (\(\theta\)) in the scales related to the self-similar range are from 644–1753 km s\(^{-1}\) (Figure 4 (middle)). The 0.025th and 0.975th empirical quantiles of the histogram of the \(\theta\) box plots (Figure 4, bottom) allow us to obtain an estimate of the extremal index \(\theta\), which shows values from 0.36–0.66 within the 95% confidence interval, with a mean of \(\theta = 0.43\). The corresponding cluster size is \(1/\theta \approx 2–3\), which means that in the whole time period CMEs with speeds higher than 644 km s\(^{-1}\) occur preferentially in groups of two or three.

When selecting CMEs with speeds \(v \geq 1000\) km s\(^{-1}\), we obtain \(m = 913\) extreme events, with \(\theta \approx 0.36–0.66\) and the estimated number of clusters being \(\theta \times m \approx 328–603\), with a declustering threshold time \(\tau_c = 28.0\) hr as derived from the maximum of the GEV fit (Figure 3). Table 1 summarizes the derived information on the cluster size, mean cluster duration, mean time between successive CMEs, and an estimate of the probabilities that a cluster of the corresponding size is recorded, using the declustering threshold time (\(\tau_c\)) method. The cluster duration was calculated as the time difference \(\Delta t\) between the end and the start of the cluster, with the start time of the cluster being defined as the first appearance of the first CME of the
Table 1

CME Clusters in Solar Cycles 23 and 24, with Speeds Exceeding 1000 km s⁻¹ and τc ≤ 28.0 ± 1.4 hr

| Cluster Size | Number of Clusters | Number of CMEs in Clusters (%) | Mean Cluster Duration (hr) | Mean Time between Successive CMEs (hr) | Recording Probabilities |
|--------------|---------------------|--------------------------------|---------------------------|----------------------------------------|-------------------------|
| 1            | 449                 | 449(49.2)                      | ...                       | ...                                    | 0.742                   |
| 2            | 84                  | 168(18.4)                      | 13.5(1.0)                 | 13.4(0.3)                              | 0.140                   |
| 3            | 38                  | 114(12.5)                      | 28.4(2.0)                 | 13.7(8.3)                              | 0.063                   |
| 4            | 16                  | 64(7.0)                        | 39.2(5.0)                 | 12.7(8.0)                              | 0.026                   |
| 5            | 5                   | 25(2.7)                        | 60.4(10.1)                | 13.1(7.2)                              | 0.008                   |
| 6            | 5                   | 30(3.3)                        | 59.4(11.4)                | 10.2(7.7)                              | 0.008                   |
| 7            | 5                   | 35(3.8)                        | 68.2(14.0)                | 12.7(8.2)                              | 0.008                   |
| 8            | 1                   | 8(0.9)                         | 81.0                      | 10.5(7.2)                              | 0.002                   |
| 10           | 2                   | 20(2.2)                        | 133.7(27.1)               | 13.5(7.4)                              | 0.003                   |
| Total        | 605                 | 913(100)                       | ...                       | ...                                    | 1                       |

Note. The first column gives the size of clusters. The second column lists the number of clusters of a certain size. The third column gives the total number of CMEs in these clusters and their percentages. The fourth column lists the mean duration of the cluster with the standard error in parentheses. The fifth column lists the mean time between successive CMEs in each cluster with the standard deviation in parentheses. The sixth column provides an estimate of the probabilities that a cluster of the corresponding size is recorded, which is the number of clusters of specific size divided by the total number of clusters.

To study the cluster behavior during the solar maximum of solar cycle 23, we select the data for the time range from 2006 March to 2010 March. The length of this time series is 35,698 and j = 15 scales are available in this interval. The max spectrum is self-similar in a range from 1000 km s⁻¹ and the power-tail exponent is α = 2.8 (Figure 6). The extremal index θ shows values from 0.64–0.96, with a mean value of 0.77. The predominant cluster sizes have values from 1 to 2. We obtain m = 225 extreme events in this time interval. Using the extremal index values, the estimated cluster number θ × m ≈ 144–216. The declustering threshold time is τc = 3.2 ± 0.6 hr (Figure 3(c)).

4.3. Maximum of Solar Cycle 24

To characterize the clustering properties of fast CMEs during the maximum phase of cycle 24, we select the data set from 2011 June to 2015 June. The length of the time series is 35,688 and j = 15 scales are available in this interval. The max spectrum is self-similar in a range from 1000 km s⁻¹ and the power-tail exponent is α = 2.8 (Figure 5). The extremal index θ shows values from 0.64–0.96, with a mean value of 0.77. The predominant cluster sizes have values from 1 to 2. We obtain m = 225 extreme events in this time interval. Using the extremal index values, the estimated cluster number θ × m ≈ 144–216. The declustering threshold time is τc = 3.2 ± 0.6 hr (Figure 3(c)).

Table 3 summarizes the CME-clustering properties during the maximum of solar cycle 24. Fast CMEs occur preferentially as individual events (52.4%) and in clusters with two members (22.2%). However, clusters with three (13.3%) and four (9.0%) members show a considerable percentage during the maximum of solar cycle 24. The probability of recording an isolated fast CME is 0.742. The probability of recording a cluster of two or three fast CMEs within the declustering threshold time τc = 28 hr is 0.142 and 0.077, respectively. The probability of larger clusters, i.e., ≥4 members, is 0.060.

4.4. Solar Minimum

We selected the interval from 2006 March to 2010 March to investigate the cluster behavior during a solar minimum. The length of this time series is 35,712, i.e., there are j = 15 scales in this time period. The Max Spectrum method shows a self-similar range from 1000 km s⁻¹ and a power-tail exponent α = 2.7 (Figure 7(a)). The extremal index θ shows values from about 0.56–0.79 (Figures 7(b)–(c)) with a mean value of 0.71, and the cluster size is 1/θ ≈ 1–2. In this period, we find m = 21 extreme events with CME speeds v ≥ 1000 km s⁻¹. Using the extremal index values, we estimate the number of
clusters as $\theta \times m \approx 10–16$. In this interval, the declustering threshold time is $\tau_c \leq 32.5 \pm 17.5$ hr (Figure 3(d)). During this minimum period, fast CMEs occur preferentially as isolated events (61.9%). In this phase, only two CME clusters occurred, one with two members and interestingly also a large one with six members. The probability of recording an isolated fast CME is 0.866. The probability of recording a cluster of two or more fast CMEs within the declustering threshold time $\tau_c \leq 32$ hr is 0.067.

Table 2

| Cluster Size | Number of Clusters | Number of CMEs in Clusters (%) | Mean Cluster Duration (hr) | Mean Time between Successive CMEs (hr) | Recording Probabilities |
|--------------|--------------------|-------------------------------|---------------------------|----------------------------------------|-------------------------|
| 1            | 178                | 178(47.2)                     | ...                       | ...                                    | 0.721                   |
| 2            | 35                 | 70(18.6)                      | 12.6(1.4)                 | 14.0(9.7)                              | 0.142                   |
| 3            | 19                 | 57(15.1)                      | 30.0(2.8)                 | 14.3(8.8)                              | 0.077                   |
| 4            | 8                  | 32(8.5)                       | 39.1(6.4)                 | 13.0(8.2)                              | 0.032                   |
| 5            | 3                  | 15(4.0)                       | 71.6(18.0)                | 13.9(6.4)                              | 0.012                   |
| 6            | 3                  | 18(4.8)                       | 67.2(20.3)                | 10.0(7.7)                              | 0.012                   |
| 7            | 1                  | 7(1.8)                        | 106.9                     | 17.7(8.9)                              | 0.004                   |
| Total        | 247                | 377(100)                      | ...                       | ...                                    | 1                       |

Note. For a detailed description of the columns, see Table 1.

Table 3

| Cluster Size | Number of Clusters | Number of CMEs in Clusters (%) | Mean Cluster Duration (hr) | Mean Time between Successive CMEs (hr) | Recording Probabilities |
|--------------|--------------------|-------------------------------|---------------------------|----------------------------------------|-------------------------|
| 1            | 118                | 118(52.4)                     | ...                       | ...                                    | 0.742                   |
| 2            | 25                 | 50(22.2)                      | 12.6(2.0)                 | 13.2(10.5)                             | 0.158                   |
| 3            | 10                 | 30(13.3)                      | 22.0(4.2)                 | 14.2(10.9)                             | 0.063                   |
| 4            | 5                  | 20(9.0)                       | 31.0(8.6)                 | 10.8(8.8)                              | 0.031                   |
| 7            | 1                  | 7(3.1)                        | 8.6                       | 13.4(9.4)                              | 0.006                   |
| Total        | 159                | 225(100)                      | ...                       | ...                                    | 1                       |

Note. For a detailed description of the columns, see Table 1.

Figure 8. Distribution of cluster sizes (left), the number of CMEs in a cluster (middle), and the cluster duration (right) for fast CMEs with speeds exceeding 1000 km s$^{-1}$. From top to bottom: solar cycles 23 and 24 (whole interval), maximum of solar cycle 23, maximum of solar cycle 24, and solar minimum between cycles 23 and 24.
Table 4
CME Clusters during the Solar Minimum between Cycles 23 and 24, with Speeds Exceeding 1000 km s\(^{-1}\) and \(\tau_i \leq 32.5 \pm 17.5\) hr

| Cluster Size | Number of Clusters | Number of CMEs in Clusters (%) | Mean Cluster Duration (hr) | Mean Time between Successive CMEs (hr) | Recording Probabilities |
|--------------|--------------------|-------------------------------|---------------------------|----------------------------------------|------------------------|
| 1            | 13                 | 13(61.9)                      | ...                       | ...                                    | 0.866                  |
| 2            | 1                  | 2(9.5)                        | 16.5(8.3)                 | 16.5(11.7)                             | 0.067                  |
| 6            | 1                  | 1(28.6)                       | 72.4                      | 14.5(8.1)                              | 0.067                  |
| Total        | 15                 | 21(100)                       | ...                       | ...                                    | 1                      |

Note. For a detailed description of the columns, see Table 1.

Table 5
Cluster Information during Solar Cycles 23 and 24

| Interval       | Power-tail Exponent \(\alpha\) | Extremal Index \(\theta\) | Cluster Size \((1/\theta)\) | \(\tau_i\) (hr) | Isolated Events (%) | Cluster (2 Members) (%) | Cluster (3 Members) (%) | Cluster \((\geq 4)\) Members (%) |
|----------------|---------------------------------|--------------------------|-----------------------------|-----------------|---------------------|------------------------|------------------------|---------------------------|
| Full period    | 3.4                             | 0.36-0.66                | 2-3                         | 28.0            | 49.2                | 18.4                   | 12.5                   | 19.9                      |
| Max SC 23      | 2.8                             | 0.49-0.91                | 1-2                         | 28.3            | 47.2                | 18.6                   | 15.1                   | 19.1                      |
| Max SC 24      | 2.8                             | 0.64-0.96                | 1-2                         | 32.0            | 52.4                | 22.2                   | 13.3                   | 12.1                      |
| Minimum        | 2.7                             | 0.56-0.79                | 1-2                         | 32.5            | 61.9                | 9.5                    | ...                    | 28.6                      |

Note. Power-tail exponent \(\alpha\), extremal index \(\theta\) and cluster size from the Max Spectrum method as well as percentage of isolated events and clusters with two, three, and four or more members found using the declustering threshold time \((\tau_i)\) method.

4.5. Summary of CME Cluster Behavior and Illustrative Examples

Figure 8 shows the distribution of cluster sizes, the number of CMEs in clusters, and the cluster duration of fast CMEs \((v \geq 1000\) km s\(^{-1}\)) for the different periods studied (Tables 1–4). A summary of the main cluster properties of the different periods is given in Table 5. In all periods, we find that the predominant occurrence is as isolated events. The fraction of isolated events is about 50% during the overall period as well as during the maximum phases of the solar cycles, whereas it is as high as 62% during the solar minimum. During the full period and the maximum phases, a significant fraction of clusters with two and three members also occur, with percentages of 22% and 15%, respectively. Clusters with \(\geq 4\) members cover 20% in total. In contrast, in the solar minimum, only two clusters occur, and all other CMEs are isolated events. This is in agreement with the results from the Max Spectrum method. However, larger clusters with more than four members (up to 10) also exist, and include about 20% of all CMEs in total. The mean durations of the clusters during the whole period are 13.5 hr for clusters consisting of two CMEs, 28.4 hr for clusters of three, 39.2 hr for clusters of four, and maybe as long as 133.7 hr for the largest cluster consisting of 10 CMEs.

For the declustering times derived from the maximum of the GEV fit to the distributions of the time differences between successive fast CMEs, we find values in the range \(\tau_i = 28–32\) hr for the different phases of the solar cycle. This is an interesting result. Although the occurrence of CMEs is much reduced during solar minimum periods than in solar maximum periods, this does not affect the basic clustering timescales. The CME declustering times are very similar during the different phases of the solar cycle.

In the following we show for illustration some examples of the clusters we identified, using white-light coronagraph images from LASCO C2. Figure 9 shows a cluster with two members and a time difference between the successive CMEs \(\tau_i = 16.8\) hr which occurred on 2017 September 9 at 23:12:12 and 2017 September 10 at 16:00:07 (Figure 9). During 2017 September 9–10, AR 12673 produced a cluster with two fast CMEs, the first one had a speed of \(v = 1148\) km s\(^{-1}\) and is followed after about 17 hr by another very fast CME with \(v = 3703\) km s\(^{-1}\). Note that AR 12673 was the source of the two largest flare/CME events of solar cycle 24, i.e., the X9.3 flare on 2017 September 6 and the X8.2 flare on 2017 September 10, which is associated with the second CME of the cluster described here (e.g., Veronig et al. 2018). The CME–CME interaction of the two fast CMEs of this cluster and their space weather effects are studied in detail in Guo et al. (2018).

Figure 10 shows a cluster with three members that occurred during the decreasing phase of cycle 23, namely on 2005 August 22 at 02:30:05, 2005 August 22 at 18:06:05, and 2005 August 23 at 15:06:05 with speeds from 1291–2150 km s\(^{-1}\) and a mean time difference between the successive CMEs of \(\tau_i = 18.7\) hr. Figure 11 shows one widely studied case of homologous CMEs that occurred in the time period 2000 November 23–25 (Nitta & Hudson 2001; Lugaz et al. 2017), which is related to a cluster detected with six members. This cluster starts on 2000 November 23 at 21:30:08 and lasts until 2000 November 25 at 01:31:58 with the speeds of the CMEs in the cluster ranging from 1002–2528 km s\(^{-1}\) and a mean time between successive CMEs of \(\tau_i = 5.6\) hr.

5. Geoeffectiveness of CME Clusters

In this section, we study the relationship between clusters of fast CMEs and their potential geoeffectiveness by evaluating the geomagnetic Dst index using the available data from the World Data Center for Geomagnetism, Kyoto, from 1998 April to 2014 December. The main idea is to statistically check
whether fast CMEs that occur in clusters are more geoeffective than fast CMEs that occur in isolation. For CMEs with speeds of 1000 km s\(^{-1}\) in the SOHO/LASCO FOV (\(\sim 2-30 R_\odot\)), a large spread of travel times to 1 au from \(\sim 20\) to \(\sim 80\) hr was found in observations and in drag-based modeling (Schwenn et al. 2005; Vršnak et al. 2013). Geomagnetic storms depend on the arriving ICME speed as well as on the strength and structure of the interplanetary magnetic field (Gonzalez &
Tsurutani 1987; Wilson 1987; Russell 2000). In interplanetary space, the magnetic driving forces are usually assumed to have ceased, and the MHD drag force due to the interaction between the solar wind and ICMEs to be important, which would tend to accelerate slow CMEs (i.e., slower than the ambient solar wind) and to decelerate fast CMEs (Cargill 2004; Temmer et al. 2011; Vršnak et al. 2013; Temmer & Nitta 2015). However, there are also other effects in interplanetary space that are relevant to consider, in particular preceding and interacting CMEs/ICMEs, which also have a strong effect on the propagation behavior (Burlaga 1995; Farrugia & Berdichevsky 2004; Farrugia et al. 2011; Temmer & Nitta 2015; Scolini et al. 2020). Further, cases of very fast CMEs also exist, which show only a little deceleration in interplanetary space (Zastenker et al. 1976; Berdichevsky et al. 2002; Vandas et al. 2009; Temmer et al. 2011; Russell et al. 2013; Winslow et al. 2015). During the maximum of solar cycle 24, it was better appreciated that fast CMEs can occur in quick succession (Lugaz et al. 2012; Möstl et al. 2012; Gopalswamy et al. 2013; Liu et al. 2014a, 2014b; Temmer et al. 2014). This close succession is described through the declustering threshold time $\tau_c$. As a result, the possibility of ICMEs interacting in the inner heliosphere significantly increases. ICME–ICME interactions are important because they affect their interplanetary propagation and evolution (Luhmann et al. 2020 and references therein). Therefore, we used a statistical description of the clustering of fast CMEs to evaluate their potential geoeffectiveness.

Based on these findings, we defined the corresponding potential geoeffective period as the CME start + 1 day to cluster end + 4 days (assuming Sun–Earth travel times of fast ICMEs from 1 to 4 days). For isolated events, we defined the potential geoeffective period as the start time of the cluster + 1 day to the start time of the CME + 4 days. We calculate for each of these periods the total sum of the hourly Dst values normalized by cluster size as well as the negative Dst peak (Dst minimum).

We note that this is a rough and statistical approach. Obviously, it is not the case for all fast CMEs, which we study here, that the corresponding interplanetary manifestation of ICMEs will reach the Earth (the sample also includes back-sided events). Also, it is clear that at times where we have a high occurrence rate of fast CMEs, there will on average be a higher geomagnetic activity as, e.g., evidenced by the Dst index. Thus we look at two specific quantities, with the following hypothesis behind them. We calculate the total hourly Dst summed over the time interval where the corresponding fast CMEs of a cluster might be reaching Earth, but divide it by the number of CMEs in the cluster. This gives us a statistical description of the geoeffectiveness per CME, and to check whether this is different for isolated events than for CMEs occurring in clusters. Second, we also check the minimum Dst value in the given potential “geoeffective interval.” Assuming that CMEs that occur in clusters that might be merging in interplanetary space, they can arrive as one merged CME that causes one bigger storm (see, e.g., the review by Lugaz et al. 2017).

Figure 12 shows the potential geoeffectiveness for isolated CMEs and for CMEs in clusters of 2 or $\geq$3 during the time range 1998 April to 2014 December. The first column shows the distribution of the total hourly sum of Dst/cluster size and the second column the minimum values of Dst index in the geoeffectiveness period. For the total Dst/cluster size, we find some change in the bulk of the distribution toward higher values from isolated CMEs to clusters of two. However, the largest numbers are associated with isolated events. This is different when we look into the distribution of the minimum Dst values, which show a change to larger negative peak values for CMEs in clusters than in isolated events. This is also reflected in a change of the mean values of the distribution $(-51.7\,\text{nT}$ for isolated events, $-67.3\,\text{nT}$ for clusters of two members, and $-88.7\,\text{nT}$ for clusters of $\geq$3). Additionally, we calculate the fraction of minimum values of Dst $\leq -100\,\text{nT}$ for isolated events, and clusters with two and $\geq$3 members. For isolated events the fraction corresponds to 10%, for clusters with two members it is 18%, and for clusters with more than two members it corresponds to 26%. These findings provide indications that the geoeffectiveness per CME is higher in CMEs that occur in clusters than in isolated CMEs.

### 6. Summary and Discussion

Two methods were applied to obtain a statistical description of the occurrence and clustering properties of fast CMEs ($v \geq 1000\,\text{km}\,\text{s}^{-1}$) during solar cycles 23 and 24. These are the Max Spectrum and the declustering threshold time methods. The analysis was performed for the whole interval 1996–2018 that covers almost two full solar cycles, as well as separately for 4 yr subperiods centered at the maximum of cycles 23 and 24 as well as on the minimum in between. The main results we found are the following.

1. In all phases of the solar cycle, we find that isolated events make the largest fraction of the temporal distribution of fast CMEs. However, there are distinct differences between the maximum and minimum solar cycle phases. In the maximum phases of solar cycles 23 (24), about 47% (52%) are isolated events, whereas it is 62% in the minimum period between cycles 23 and 24. During the total period studied, about 50% of CMEs occur as isolated events, 18% (12%) occur in clusters of 2 (3), and another 20% in larger clusters $\geq$4.

2. From the Max Spectrum method, we find in all cases that the speeds of fast CMEs show a Fréchet-type distribution, following a power law with a power-tail exponent $\alpha \approx 3.0$. During the maximum of solar cycles 23 and 24, the power-tail exponent $\alpha$ has values from 2.7–2.8. However, when we checked the whole period, we obtain a value $\alpha = 3.5$. This difference is most probably related to the decreasing and rising phases of each solar cycle.

3. The Max Spectrum method provides an estimate of the extremal index ($\theta$), which gives information about the cluster size. During the whole period covering solar cycles 23 and 24, the extremal index $\theta$ has values from 0.3–0.6. This suggests an average cluster size from two to three. These findings are in agreement with the results obtained in Ruzmaika et al. (2011) for the period from 1999 January to 2006 December.

4. The declustering threshold time method depends on the speed threshold and is purely empirical. Using the time series of fast CMEs, we define a threshold $v \geq 1000\,\text{km}\,\text{s}^{-1}$ to characterize extreme events. The declustering threshold time method confirms the results obtained from the Max Spectrum method, i.e., that fast
CMEs show a tendency to occur in clusters. However, while the Max Spectrum method has the capability to detect the predominant clusters, the declustering threshold time method allows us to obtain more detailed information on the clustering properties, i.e., how the CMEs are distributed over clusters of different sizes and what are the typical timescales of the clustering.

5. Through the declustering threshold time method, we obtained an estimate of the typical timescales ($\tau_c$) between successive fast CMEs. For the entire period and during the maximum of solar cycle 23 the time between successive fast CMEs is $\tau_c \leq 28$ hr, while for the maximum of solar cycle 24 and the solar minimum $\tau_c \leq 32$ hr. It is interesting to note that the declustering times obtained are very similar in all phases with $\tau_c$ in the range $28-32$ hr, although the occurrence rate of fast CMEs is very different in different phases of the solar cycle. These findings suggest that the $\tau_c$ values obtained for fast CMEs may be representative of the characteristic energy build-up time of active regions between the release of successive large events.

6. The mean duration of clusters with two members is 13 hr, for clusters with three members it is between 22–30 hr and for clusters with four members it is between 31–39 hr. The largest clusters identified, i.e., with 10 members, reach durations up to 134 hr.

7. During the full interval studied as well as during the maximum phases of solar cycle 23 and 24, the probabilities that a cluster of the corresponding size is recorded can give us clues about the clustering properties of fast CMEs and their impact in a space weather context. For the overall period studied, we find that the probability of recording a cluster of one (isolated) fast CME is 0.742. The probability of recording a cluster consisting of 2, 3, or $\geq 4$ fast CMEs within the declustering threshold time $\tau_c \leq 28$ hr is 0.140, 0.063, and 0.055, respectively. These probabilities describe the occurrence of a fast CME followed by one, two, or more fast CMEs within the declustering time.

8. The potential geoeffectivity in isolated events and clusters statistically quantified by the total hourly sum of Dst normalized by cluster size shows some change in the bulk of the distribution toward higher values from isolated CMEs to clusters of two. However, the largest values are associated with isolated events. On the other hand, the distribution of the Dst minima values show a distinct change (the mean values change from $-51.7$ nT for isolated events, $-67.3$ nT for clusters of two members, and $-88.7$ nT for clusters $\geq 3$). Also, we find that the fraction of associated large geomagnetic storms as quantified by minimum values of Dst $< -100$ nT is increasing with cluster size: it is 10% for isolated events, 18% for clusters of two, and 26% for clusters of $\geq 3$. These findings indicate that clustering of fast CMEs is not necessarily making the overall geoeffectivity higher during the given period compared to the same number of CMEs occurring in isolation, but that statistically fast CMEs that occur in close successions in clusters have a tendency to produce larger storms than isolated events. This could be due to the interaction of the CMEs in interplanetary space and their arrival as one complex entity at Earth that causes larger geoeffectivity (Lugaz et al. 2017).

Our results of typical declustering timescales of fast CMEs ($v \geq 1000$ km s$^{-1}$) in the range of $\tau_c = 28–32$ hr are in basic agreement with the definition of quasi-homologous CMEs as successive CMEs originating from the same AR with a separation by $\sim 15–18$ hr (Wang et al. 2013; Lugaz et al. 2017). The relevant timescales in fast CME occurrence is described by $\tau_c$. These values are relevant for CME interaction, for magnetosphere preconditioning, as well as for comparison.
with relaxation timescales/duration of geomagnetic storms. The interactions in the heliosphere play an important role in the production of solar energetic particles (SEPs) and strong geomagnetic effects, e.g., statistical studies showed that the presence of a previous fast CMEs within 12 hr increases the probability that this second fast CME contributes to SEP production (Gopalswamy et al. 2002; Berdichevsky et al. 2003; Yashiro et al. 2004; Farrugia et al. 2006; Lugaz et al. 2017). These findings emphasize the crucial importance of ICME–ICME interactions for space weather (Liu et al. 2014a, 2015).

The fact that the geoeffectivity per CME is higher when the fast CMEs occur in clusters than when they occur as isolated events may be related to different aspects:

(a) That CMEs in clusters have a higher probability to interact, and that interacting CMEs have a tendency to be more geoeffective (Liu et al. 2014a; Riley & Love 2017).
(b) That the occurrence of multiple CMEs along the Sun–Earth line has preconditions in interplanetary space, e.g., reducing the interplanetary density along the paths of a following fast CME, which reduces the drag force exerted on it (e.g., Liu et al. 2014a; Temmer & Nitta 2015).
(c) That the subsequent disturbance of the Earth’s magnetosphere within short times may lead to differently strong effects, in the form of pre-conditioning of the magnetosphere under repeated strong energy inputs by the arrival of fast CMEs. These large perturbations induced by consecutive CMEs (clusters) in the coupled magnetosphere–ionosphere system causes a higher geoeffectiveness. The strongly varying field and plasma density in the sheath region preceding the ICME, the fast solar wind speed, as well as the interplanetary shock itself are all effective drivers of geomagnetic activity (Farrugia et al. 1997; Pulkkinen 2007; Vennerstrom et al. 2012).

An extreme space weather event caused by preconditioning of the upstream solar wind by an earlier CME, in-transit interaction between subsequent fast CMEs in close succession, as well as their typical timescales, can give clues as to what the main ingredients are for the most extreme space weather events and how to obtain a better forecast of these combined conditions.

We thank the geomagnetic observatories (Kakioka [JMA], Honolulu and San Juan [USGS], Hermanus [RSA]), INTERMAGNET, and many others for their cooperation to make the final Dst index available. We appreciate the use of the CME catalog generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. This research has received financial support from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 824135 (SOLARNET). J.M.R. acknowledges the funding obtained within the SOLARNET Mobility Programme 2019.
