Characterizing extreme geomagnetic storms using Extreme Value Analysis: a discussion on the representativeness of short datasets

Guillerme Bernoux, Vincent Maget

AGU ADVANCING EARTH AND SPACE SCIENCE

Space Weather

RESEARCH ARTICLE
10.1029/2020SW002450

Key Points:

- Ca is a new time-integrated geomagnetic index based on aa, defined to characterize geoeffectiveness from the radiation belts point of view
- Dedicated extreme value analysis methods are applied to a 150-year-long data set of geomagnetic events based on Ca index
- The use of the same method on data sets based on periods covered by just one space mission can give different results for two main reasons

Correspondence to:
G. Bernoux,
guillerme.bernoux@onera.fr

Citation:
Bernoux, G., & Maget, V. (2020). Characterizing extreme geomagnetic storms using extreme value analysis: A discussion on the representativeness of short data sets. *Space Weather*, *18*, e2020SW002450. https://doi.org/10.1029/2020SW002450

Received 13 JAN 2020
Accepted 20 MAY 2020
Accepted article online 23 MAY 2020

Characterizing Extreme Geomagnetic Storms Using Extreme Value Analysis: A Discussion on the Representativeness of Short Data Sets

G. Bernoux[1] **and V. Maget**[1]

[1]ONERA/DPHY, Université de Toulouse, Toulouse, France

Abstract One of the main goals when studying Space Weather is to characterize extreme events occurrences and related characteristics. To do so, dedicated statistical methods from the so-called extreme value analysis (EVA) field have been developed. In this study we used Ca index, derived from aa, in order to characterize geoeffectiveness from the radiation belts point of view with a 150-year-long data set. The analysis performed in this study thus focuses on this newsworthy index to provide clues on the reliability of EVA methods. The first main result we present here is that the 1-in-10-, 1-in-50-, and 1-in-100-year events, respectively, match Ca values of 100.39, 131.39, and 142.84 nT. Consequently, the only 1-in-100 event observed during the Space Era would be the "Halloween Storm" in 2003 that reached a Ca value of 147.6 nT. The second main result highlighted in this work is that performing the same analysis with shorter subsets (20 years long) can give significantly different results for two reasons. The first reason is that some short time periods do not display the same distribution of events as the full period. The second reason is that the choice of the correct threshold (when using a Peaks Over Threshold approach) is made difficult with a short data set and leads to inaccurate results. This is a strong result as for accurate estimation of the induced effects of extreme events in radiation belts, we may only rely on short flux data sets from one or another mission (mostly shorter than 20 years).

1. Introduction

The Sun-Earth connection is a complex system, which is composed of multiple interactions transferring energy from their source locations to the Earth's magnetosphere. This transfer of energy is very difficult to quantify. Many authors have tried to identify key parameters (Akasofu, 1981; Alves et al., 2006; Borovsky & Denton, 2006; Tenfjord & Østgaard, 2013), but the mechanisms remain so complex that a simple law cannot fully describe the effects of the solar activity on the different parts of the magnetosphere. In particular, these effects become even more complex in one of the most internal regions of the magnetosphere, which are the radiation belts.

These interactions are strong and may vary (in terms of frequency of occurrence and intensity) independently. Therefore, the observed effects on the vicinity of the Earth also vary greatly and it can be difficult to accurately determine the origin and the extent of an event. The solar wind is the main medium of energy transfer from the Sun to the Earth. Strong dynamical fluctuations of its characteristics are mostly induced by active regions and coronal holes located at the Sun chromosphere from which plasmas are accelerated and propelled in the interplanetary medium. They are the so-called coronal mass ejections (CMEs) and high-speed streams (HSS). These large structures of solar wind may impact the Earth's magnetosphere, consequently inducing direct effects on the near Earth's space (at satellites altitudes) and fluctuations in the Earth's magnetic field strength as measured on ground, from which geomagnetic activity indices are defined (Menvielle et al., 2011). To illustrate the complexity of the Sun-Earth connection, Reeves et al. (2011) showed that the relationship between the solar wind speed and the 1.8- to 3.5-MeV electron fluxes at geosynchronous orbit is hardly described by a linear correlation but is instead considerably more complex and shows that high-velocity solar wind does not always drive high electron fluxes the way it could be expected to.

The core of the Earth's magnetosphere is filled by radiation belts containing trapped energetic particles (mostly high-energy protons and electrons from a few keV up to hundreds of MeV for protons and few MeV for electrons) due to its intense magnetic field. Since the advent of the Space Era, the study and

knowledge of these belts has become of prime importance, since their energetic population has dynamic variations in accordance with the current state of the magnetosphere, occasionally having dramatic impact on inflight space hardware (Bourdarie & Maget, 2012; Horne et al., 2013).

The geomagnetic activity and the magnetospheric currents can be described with many different indices that can be based on different data sets. Today, the most commonly used, Kp index, is built from experimental ground-based data since 1932 by 12 stations located in near subauroral zones. It is, more precisely, the average of the geomagnetic disturbances on a logarithmic scale between 0 and 9. It provides the planetary level of disturbance caused by the horizontal component of the geomagnetic field (Bartels et al., 1939). Other indices, such as Cp, C9 are basically based on the same database, but the mathematical formula varies as well as the stations used (Mayaud, 1980). In parallel, solar indices (directly linked to the Sun's surface activity), such as the sunspot number (SSN) and F10.7 (measure of the solar radio flux per unit frequency at a wavelength of 10.7 cm), have also been defined to monitor and anticipate the effects of solar activity. Taking into account indices from the same kind will usually show a strong correlation between them. However, the correlation between indices representing the activity of the magnetosphere and those related to the Sun activity show a rather weak correlation, hence constituting a first obstacle for forecasting the activity in the magnetosphere. Indeed, the quantity of energy transferred from the solar wind to the inner part of the magnetosphere rarely exceeds 5% to 10% (Sauvaud, 2002).

The disturbances induced on the magnetosphere vary daily and even hourly. Because of the intense solar activity, magnetic storms can occur and induce brutal and intense variations in the terrestrial magnetosphere. For example, during typical geomagnetic storms, as observed from commonly used satellites' orbits, changes in the particles flux levels may increase by 2 to 3 orders of magnitude in a few hours only (Reeves et al., 2003). Induced effects on satellites equipment can consequently be critical. Extreme events are events that can reach very high levels of disturbances. The most famous one is the “Carrington event”: from 28 August to 4 September 1859, extremely bright auroras were observed even in near-equatorial zones, high magnetic disturbances were denoted and the intense solar activity was remarked. The related Dst index (another measure of disturbances) has been evaluated to −900 nT (Siscoe et al., 2006). A geomagnetic storm is considered as strong when Dst index reaches −200 nT. Similar intense events happened since then, especially during the Space Era, having harmful impact on both ground and orbital hardware, particularly inflight satellites flying through radiation belts. Among these storms, one can mention the “Halloween Storm” of October 2003, the storm of July 2004 or the storm of April 2010 that caused electrostatic discharges on Galaxy-15 satellite (Odenwald, 2015). While the Space Era is now barely in its seventh decade, questions arise: Have we already encountered 1-in-10-, 1-in-50-, and 1-in-100-year events during this time period, for which in situ measurements are available? How can we define extreme geomagnetic events?

This research intends to analyze extreme events by using statistical methods adapted to the study of extreme values for Space Weather purposes and determine up to which confidence level we can trust them. These extreme value analysis (EVA) methods are described as solutions to estimate typical return levels such as 1-in-10 years or 1-in-100 years on limited time ranged data sets. Definition and characterization of what is an extreme event is required to drive these statistical methods (even standard and straightforward ones like direct statistical analysis). This definition has (1) to be based on historical data and (2) to focus on the human and industrial needs perspective. The aim is here to quantify and qualify the occurrence and return levels of extreme events and to be able to describe them statistically. A discussion on the suitability of applying such methods to geomagnetic data and particle fluxes information is the following objective.

In this paper we base our EVA on 150 years of geomagnetic data in order to assess the 1-in-N-year events. To achieve this and to avoid any bias and clustering, we propose to rely on a new physics-based definition of a radiation belt event to drive our statistical analysis. Section 2 introduces the geomagnetic indices and parameters used in this study. In section 3 we perform a first basic analysis and suggest a definition of a geomagnetic event that may suit our EVA. Then elements of extreme value theory are given in section 4, and the EVA is detailed in section 5. A further discussion on the confidence levels one can have on performing such methods with shorter geomagnetic data is led in section 6. In section 7 the reader will find the conclusions.

2. How to Characterize Geomagnetic Storms Intensity From the Radiation Belts Point of View

2.1. A Dedicated Time-Integrated Geomagnetic Index

Following Rochel et al. (2016), we based our study on the aa index and its related time-integrated geomagnetic storm index Ca (see below for details). Ca has been designed to quantify the geoeffectiveness of solar wind structures impacting the magnetosphere from the radiation belts point of view. We will compare this integrated index to the typical reaction time of the radiation belts in the next section.

The aa index is a 3-hr K-based index that was first introduced by (Mayaud, 1971), with its first values for the 1868–1968 period being published in (Mayaud, 1972). The aa index is computed with data from two subauroral opposite observatories, the first one being in the United Kingdom and the second one in Australia. The point in using aa was to have an index that could go back to 1868. Despite its wide data range (from 1868 onward), this index is considered as very homogenous and stands as a simple and easy mean of monitoring the planetary geomagnetic activity (Mayaud, 1980). The aa index ranges from 0 onward and its unit is the nanotesla (nT). The higher the aa value is, the stronger the disturbance of the magnetosphere would be. The aa index is made available on the ISGI—International Service of Geomagnetic Indices website. The user is referred to Mayaud (1980) and Menvielle et al. (2011) for more details.

The relaxation characteristic time in the radiation belts for high-energy electrons (order of magnitude of 100 keV and above) after a strong magnetospheric disturbance is of the order of 4 days (Meredith et al., 2006; Rochel et al., 2016). Using this statement, it is possible to convolute it with the magnetic activity (represented by the index aa) to define a new index named Ca:

$$Ca(t) = \frac{1}{\tau}\int_0^{\infty} aa(t-t')\exp\left(-\frac{t'}{\tau}\right)\mathrm{d}t'$$

where aa is the magnetic index and τ the relaxation characteristic time (Rochel et al., 2016).

Ca appears to be a very convenient index for our study. Indeed, since Ca is based on aa, data are available from 1868 onward, which is the longest available data set for any index. For the purpose of EVA, having a very wide data range is significant. In this paper we use 150 years of data from 1868 to 2017, which offers an interesting opportunity to gauge evolved and complex EVA methods against usual statistics in determining typical return levels from such a long data set.

Defining such an index is also interesting because it can take into account both multiple close events and energy accumulation, which could explain some extreme events (Benacquista et al., 2017; Lugaz et al., 2015). Besides, it has been shown in Rochel et al. (2016) and more detailed in the next paragraph that Ca is well correlated to energetic particle fluxes (the highest correlations being found at $L^* \approx 4$). Hence using Ca as the main index in our study is adequate in order to perform an EVA on the planetary geomagnetic activity and in order to establish a link between this analysis and the dynamics of relativistic electron fluxes in Earth's radiation belts.

In order for the reader to have a more precise idea of the typical behavior of Ca during corotating interaction region (CIR)-induced and CME-induced storms, Figures 1 and 2 show the dynamics of the geomagnetic indices Kp, aa, and Ca with two solar wind parameters (plasma bulk speed and flow pressure), and the fluxes of electrons in the radiation belts for three ranges of energies ($0.309 < E < 0.379$ MeV, $1.539 < E < 1.995$ MeV, and $3.299 < E < 3.969$ MeV) measured by the instrument MagEIS aboard the Radiation Belt Storm Probe A (RBSP A) spacecraft (Blake et al., 2013) between 15 April and 15 May 2017 (Figure 1) and between 28 September and 28 October 2013 (Figure 2). Kp index and the solar wind parameters are obtained from the hourly OMNIWeb database (King & Papitashvili, 2005). The CIR-induced event (Figure 1) is particularly characterized by the arrival of a high-speed solar wind and the filling of the radiation belts at all energy ranges, in particular with very high energy electrons (3.299–3.969 MeV). The (triple) CME-induced event (Figure 2) is characterized by a spontaneous high increase of the flow pressure along with a moderate increase in the solar wind speed, while the radiation belts are filled mostly with electrons in the energy range $0.309 < E < 0.379$ MeV, no increase being observed in the fluxes of very high energy electrons (3.299–3.969 MeV). Whereas the CIR-induced event is the one that fills the most the radiation belts at all energy ranges and demonstrates the highest energy accumulation, the geomagnetic indices aa and Kp show higher peaks for the CME-induced event. One can see however that Ca increases and

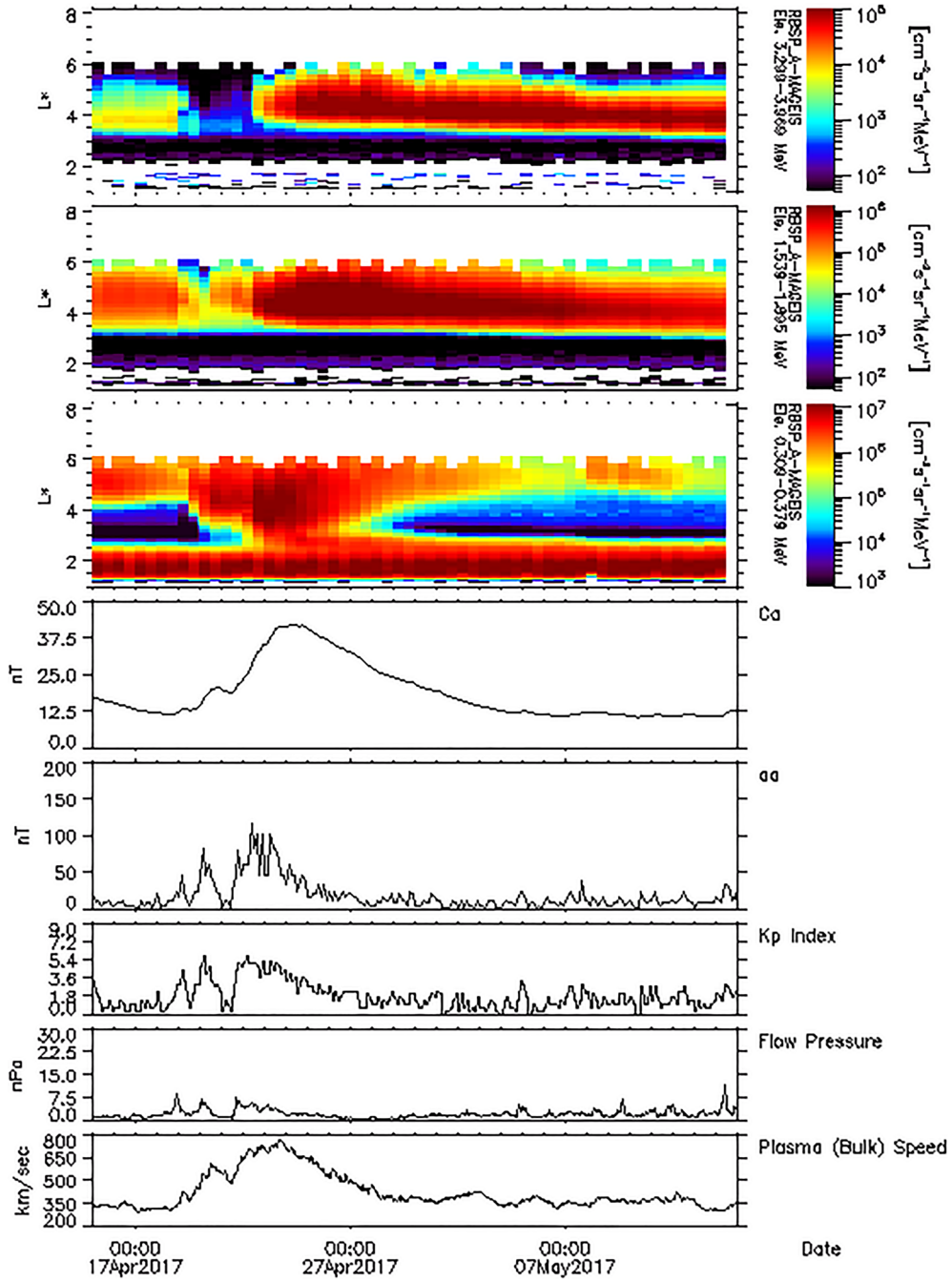

Figure 1. Evolution of the geomagnetic indices Kp, aa, and Ca with two solar wind parameters (plasma bulk speed and flow pressure) and the fluxes of electrons in the radiation belts for three ranges of energies ($0.309 < E < 0.379$ MeV, $1.539 < E < 1.995$ MeV, and $3.299 < E < 3.969$ MeV) measured by the instrument MagEIS aboard the Radiation Belt Storm Probe A (RBSP A) spacecraft between 15 April and 15 May 2017 period, which displayed a CIR-induced storm.

decreases more slowly than aa, but it reaches higher values (above 37.5 nT) in the first case than in the second one (below 25 nT), showing a better consistency with the geoeffectiveness from the radiation belts point of view than aa and Kp indices.

2.2. Geoeffectiveness From the Radiation Belts Point of View

As mentioned in the previous paragraph, Ca index has been defined to characterize geoeffectiveness from radiation belts perspectives. The time convolution, performed in the computation of Ca index, allows

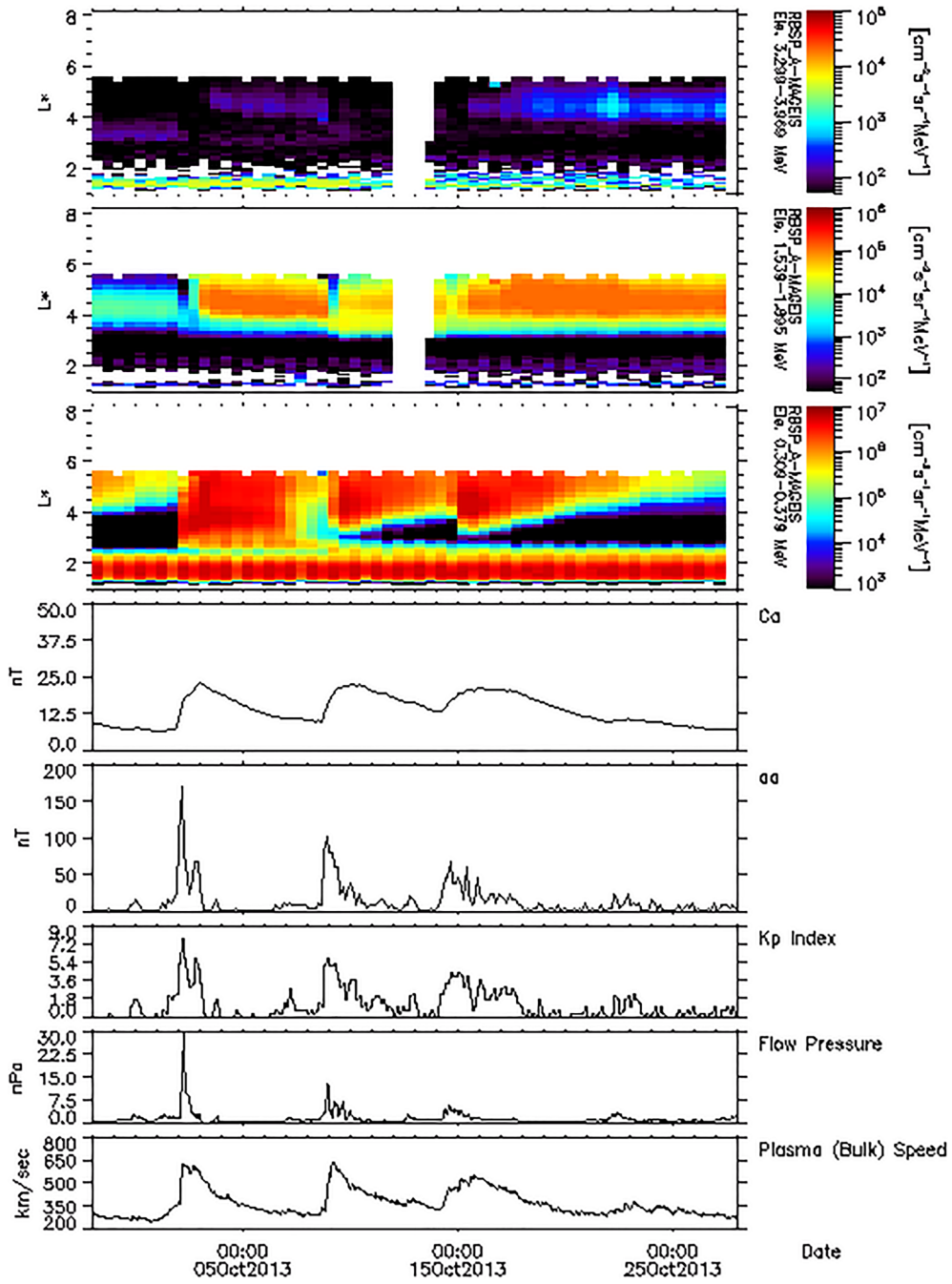

Figure 2. Evolution of the geomagnetic indices Kp, aa, and Ca with two solar wind parameters (plasma bulk speed and flow pressure) and the fluxes of electrons in the radiation belts for three ranges of energies (0.309 < E < 0.379 MeV, 1.539 < E < 1.995 MeV, and 3.299 < E < 3.969 MeV) measured by the instrument MagEIS aboard the Radiation Belt Storm Probe A (RBSP A) spacecraft between 28 September and 28 October 2013 period, which displayed a CME-induced storm.

taking into account the induced and not immediate reaction of the radiation belts system when an intense disturbance impacts the magnetosphere. Wave-particle-induced energization is in particular well correlated with this index. Thus, the global energy transfer from the solar wind-magnetosphere-radiation belts coupled system is at first order well quantified. Figure 3 presents the direct correlation of aa and Ca index against >100-keV electrons fluxes as observed by NOAA-POES15 satellite, between July 1998 and December 2015. This clearly shows the improvement in the correlation coefficient when using Ca index. In

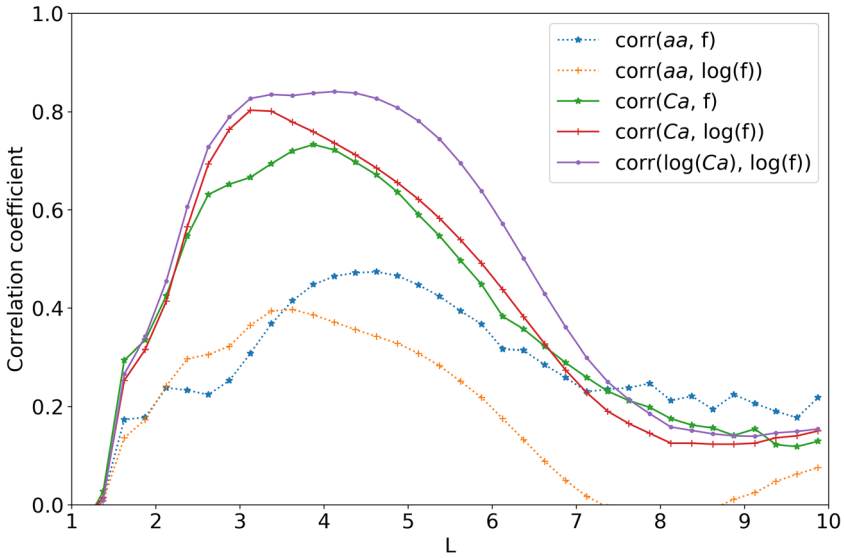

Figure 3. Correlation coefficients as a function of L parameter computed between E > 100-keV electrons fluxes (f) measured by NOAA-POES15-MEPED instrument and aa on one side and Ca on the other side. Different combinations have been tested. The time period used to compute these correlations extends from July 1998 to December 2015.

particular, this improvement is nicely observed at any L values inside the radiation belts domain. To go further, we plotted in Figure 4 the correlation coefficients between Ca, and the convolved flux Cf (using the same formula as Ca index for construction) for different energies. We see that the correlation coefficient is even closer to one, highlighting the benefit of Ca to quantify the global energy transfer to the radiation belts.

These simple correlation estimations are very interesting and show the ability of such an index to better classify events as a function of the quantity of energy transferred to the radiation belts. Of course, as any simple index, it cannot totally quantify this complex process. Some limitations are of course observable. In particular, as energy increases, the correlation coefficient tends to decrease, as it can be observed for >900-keV convolved flux in Figure 2. This is mainly due to the fact that the characteristic time of 4 days used for the convolution has been chosen to best fit the observations for any L values and energies. As a consequence,

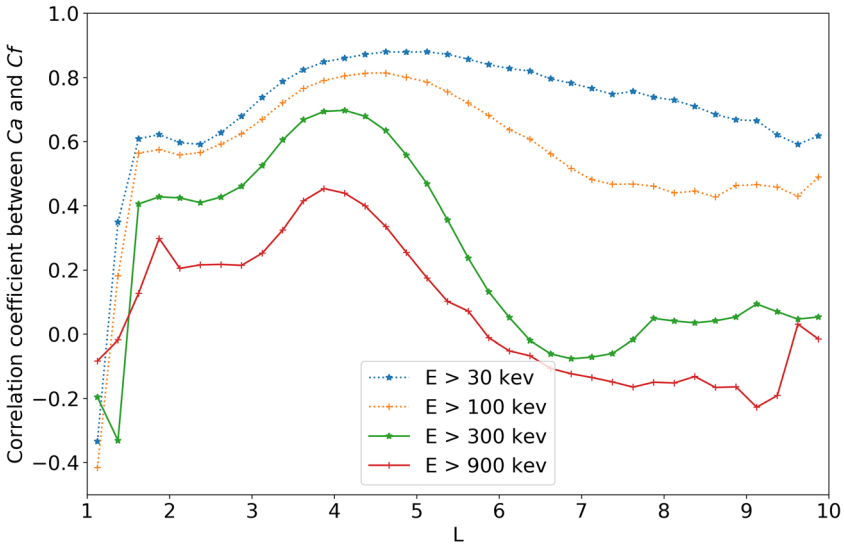

Figure 4. Correlation coefficients as a function of L parameter computed between convolved electrons fluxes (Cf, as for Ca) measured by NOAA-POES15-MEPED instrument for different energies and Ca index. The time period used to compute these correlations extends from July 1998 to December 2015.

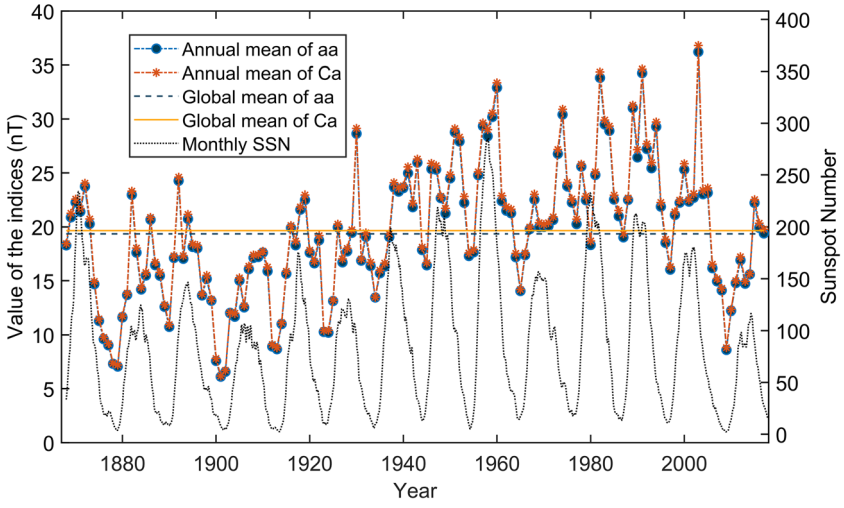

Figure 5. Evolution of the annual means of aa and Ca between 1868 and 2017 along with their global means over the full period. The 13-month smoothed monthly total sunspot number is also plotted for a better visualization of the solar cycle modulation.

on both extremes this Ca index will less correlate to observations. However, this can be easily adapted according to which energy and/or effects analysis is focused on. Finally, it can be seen that at GEO orbit, close to L equal 6, the correlation tends to decrease for any energies. This highlights the differences in the radiation belts dynamics between this particular orbit, close to the outer boundary of the radiation belts, and other smaller L values, deep inside the heart of the radiation belts. The predominant processes are not the same, and the dynamics is not driven in the same way. As a consequence, we will see in the following that our conclusions may differ significantly from (Meredith et al., 2015), which focused only for E > 2 MeV on the GEO orbit.

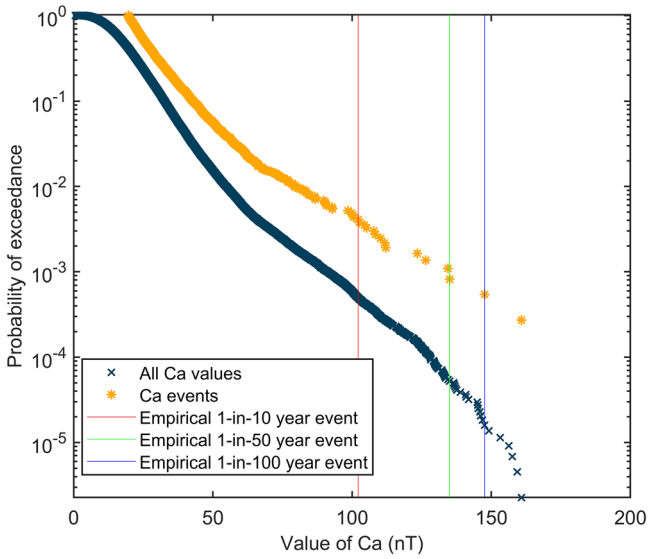

Figure 6. Probability of exceedance plot of (a) all the Ca values between 1868 and 2017 (as dark blue crosses) and (b) the geomagnetic events following their maximum Ca value between 1868 and 2017 (as yellow asterisks). The red (resp. green and blue) line represents the value of the empirical 1-in-10-year (resp. 1-in-50- and 1-in-100-year) event.

3. Data Analysis and Event Definition

3.1. A First Statistical Approach

Definitive aa values are available from 1868 to 2017. It is calculated on a 3-hr basis, and hence, 438,296 values are available for this period. The global average of aa during the whole period is 19.36 nT, and the global one of related Ca is 19.66 nT. However, the annual average of Ca may vary greatly from year to year. The year with the highest annual mean is 2003 with an annual average of 36.81 nT, and the year with the lowest average is 1901 with 6.18 nT. The evolution of the annual averages of aa and Ca is given in Figure 5. This figure also displays the evolution of 13-month smoothed monthly total SSN between 1868 and 2017, which is representative of the solar cycle modulation. One can note in this figure that the solar cycle modulation and the weakness of the last 10 years are also observable in the curves of the annual averages of aa and Ca. In particular, year 2009 has a quite low annual average Ca value (8.80 nT).

Extreme events can be defined as periods when the geomagnetic activity is very intense. Therefore, extreme events correspond to high values of Ca. Such events are hence displayed in the tail of the distributions of Ca. The cumulative distribution of the Ca values is plotted (as a probability of exceedance curve) in Figure 6. With this long data set, it is possible to directly infer on the intensity of extreme disturbances periods. Looking at the tail of the Ca distribution, we can note that Ca seems to converge and not exceed a value of about 165 nT. This first guess would denote the existence of a threshold, which could not be overpassed.

Table 1
List of the 15 Most Extreme Events Since 1868 With Their Dates, Intensities, and Empirical Return Period

Date	Ca (nT)	Empirical return period (years)
20/11/1882	160.9	150
31/10/2003	147.6	75
16/05/1921	135.1	50
14/03/1989	134.5	37.5
19/09/1941	126.4	30
28/03/1946	123.4	25
31/03/1940	112.1	21.4
16/11/1960	111.6	18.8
18/10/1872	110.3	16.7
26/01/1938	108.3	15.0
02/04/1960	107.8	13.6
18/07/1959	105.2	12.5
10/11/2004	104.5	11.5
23/09/1946	102.6	10.7
07/10/1960	102.2	10

Note. Dates are formatted as DD/MM/YYYY.

However, this is a first assumption only made on instantaneous Ca data. To well analyze extreme time periods, we have to first define what an event is and then carefully analyze its distribution. Thanks to this long-duration data set, we will then be able to determine the typical return levels and quantify the intensity of extreme disturbances based on usual statistics. Besides, these long-duration data sets are a perfect test bed to analyze the efficiency of EVA methods and get the most of them, as we are to present further on in this study.

3.2. Event Definition

It is necessary in our study to define a geomagnetic event from a mathematical point of view. A simple way to define an event is to consider that an event starts when the value of Ca exceeds a given threshold and the event ends when the value of Ca falls below this threshold. Then, the characteristic value of the event is chosen as the event maximum. However, the choice of such a threshold is not straightforward.

A first guess would be to take the global mean of Ca. However, this appears not to be always adequate. For example, using this method for the strongly disturbed year of 2003 would show an extremely long event (longer than half the year), as only 6.68% of Ca values were below 19.66 nT. Thus, we decided to use an annual threshold also depending on the annual mean, which could resolve at least partially this kind of issue without introducing a bias in our statistics. Indeed, as we only focus on extreme values, this adaptive threshold (as long as it does not reach extreme values itself) will not modify the tail of the distribution.

For the years that have annual Ca mean higher than the global Ca mean, the annual threshold is defined as the average between the global and the annual Ca mean values. For the other years, the threshold is forced to be equal to the global mean, in order to avoid taking into account very small values of Ca that cannot physically be considered as events. In case there is an event overlapping over 2 years, the value of the threshold is updated only after the end of this event. This new choice leads to a threshold of 28.23 nT in 2003 (24.97% of the values stay below the threshold) and a much better accuracy in terms of events selection. However, this definition is poorer from the physics point of view, since the same event could be or could not be taken into account according to its year of occurrence. We justify our choice by the fact that, doing it this way, we can stabilize our statistical pedestal, without modifying its high values tail. It prevents us also from taking into account unrealistically long events as it was the case before in 2003. The forced minimum value of the threshold finally prevents us from introducing bias in the pedestal of the statistical distribution, which may lead to inaccurate fitting of the high values tail.

Moreover, the fact that we are using the index Ca, which is an integrated version of aa (as seen above) gives us a smoother index that does not display the short-term variability of aa. That is why we define the end of the event as the time when the Ca value falls below the threshold without other condition (e.g., over the duration for which the Ca value must remain below this threshold). Finally, each event is characterized by the maximum value of Ca during it, thus, preventing bias to be introduced by the sliding-year threshold criteria.

Using this definition, it is possible to list the most extreme events since 1868. This list is given in Table 1. As we can see, with this classification, the most intense event is November 1882, followed by Halloween Storm. March 1989 is the fourth and, for example, July 2004, a triple CME-induced event, is not in the top 10. This classification indeed fosters extreme events, which impact the whole radiation belts region and are sufficiently long to induce long-term changes in the electrons radiation belts, especially in the $3 < L^* < 5$ region. The particular case of the triple CME of July 2004 is difficult to classify. For instance, when considering only the $E > 2$ MeV electrons at geostationary, (Meredith et al., 2015) showed that it can be considered as a 1-in-50-year event. This may indicate a limitation in our methodology as our study relies on the aa index, which is, although the longest ground-based time series we have, an index that is possibly not well adapted to quantify the disturbances at each orbit, in particular the geostationary one. However, the Ca index was

defined to promote the events allowing intensification of the electrons flux inside the radiation belts, that is, when there was time for energization due to wave-particle interaction. Our classification is therefore complementary from both Meredith et al. (2015) as we focus on the global intensification of the electron radiation belts region and also from Meredith et al. (2016), a paper that includes an analysis of POES E > 30, E > 100, and E > 300-keV electrons, since we focus on an independent ground-based index aiming at covering the intensification of a large range of the energy band of trapped electrons. The cumulative distribution of the geomagnetic events (defined using the yearly sliding threshold methodology) ranked according to their maximum Ca value is also given in Figure 6. The tail of this distribution clearly shows the extreme events classification. Compared to the probability of exceedance curve of all Ca values, one can note that the maximal value of Ca is conserved, that is, 160.9 nT. However, the classification in terms of events (and one point by event) allows deskewing the distribution, which thus becomes usable. In addition, one can note that using this definition of the events allows for a declustering of our data set since we keep only the highest Ca value for a given event and discard all the intermediary values that were observed in the ascending and descending phase of the event, that may be very high themselves. With such a long data set, we are directly able to estimate the typical return levels as highlighted in Figure 6 and the return periods given in Table 1. With this empirical methodology, we could now assert that November 1882 is a 1-in-150-year event, Halloween Storm a 1-in-75 one, March 1989 a 1-in-37.5 one, and July 2004 only a 1-in-6.25-year event. However, this direct statistic cannot be performed on a data set shorter than at least the return level we want to estimate, and it is not possible to infer whether a saturation of extreme events exists or not. We show in the next sections that EVA methods are designed to answer such questions, and we will give insights on their accuracy and limitations.

4. EVA Methodology

In order to perform an efficient analysis of the extreme geomagnetic events, we use a set of specific methods developed for the analysis of extreme values. EVA is widely used in many different research fields such as hydrology (Tramblay et al., 2013), climate study (Cooley, 2009), or finance (Embrechts et al., 1997). Some Space Weather studies apply extreme value theory in order to analyze energetic electron fluxes in radiation belts (Koons, 2001) or in order to model geomagnetic activity, either from the magnetic indices point of view (Siscoe, 1976) or directly from raw ground-based magnetic data (Thomson et al., 2011). We adopt in this study the magnetic indices point of view for the reasons detailed in sections 1 and 2.

Coles (2001) details EVA methods and highlights two principal techniques. The first method is based on a blocks maxima approach and fits the data set to a generalized extreme value (GEV) distribution, as done in Koons (2001). This approach is adapted to situations where blocks of data can be clearly defined or when only maximum values for a fixed period are available.

The second method fits a generalized Pareto distribution (GPD) to points above a predefined threshold (Pickands, 1975). It appears that this Peaks Over Threshold (POT) approach is more adequate to our study since it can take into account multiple extreme events that occurred the same year, and in return, the model does not take into account the periods when no geomagnetic storm occurred. For example, Meredith et al. (2015) used this method to discuss on extreme events at GEO altitude, using NOAA-GOES electron flux data. Let Z be a random variable following a GPD. Let us denote G as the GPD. For a value z above a high threshold u, the GPD can be written as follows:

$$G(z) = P(Z > z \mid z > u) = 1 - \left[1 + K\left(\frac{z-u}{S}\right)\right]^{-\frac{1}{K}}$$

where K and S are the shape and scale parameters, respectively. According to Coles (2001), the shape parameter is physically meaningful: If $K < 0$, the distribution has an upper limit, whereas if $K \geq 0$ the distribution does not have an upper limit. The first step to use the POT method is to estimate accurately these parameters. To do so, the maximum of likelihood method (Piera-Martinez, 2008) is used and confidence intervals are also provided using the deviance statistic (Nelder & Wedderburn, 1972).

To drive and ensure their accurate estimation, we followed Coles (2001) who suggested to use different plots described below. Let k be the number of exceedances above the threshold u. The discrimination plots are as follows:

1. The "Probability Plot" that consists in the points $\left\{\left(G(z_i), \frac{i}{k+1}\right), i = 1, ..., k\right\}$, the z_i being the data values above the threshold ordered in ascending order. If the model is adequate, then the points should be close to the identity map.
2. The "Quantile Plot" that consists in the points $\left\{\left(G^{-1}\left(\frac{i}{k+1}\right), z_i\right), i = 1, ..., k\right\}$, where $G^{-1}$is the inverse function of the GPD and the z_i are the data values above the threshold ordered in ascending order. If the model is adequate, then the points should be close to the identity map.
3. The "Density plot," which compares the density estimate with the normalized histogram of the data.
4. The "Return level plot," which plots the estimated return levels against the return periods and displays the empirical estimates of the return levels. This plot also gives information of interest since we are interested in estimating the 1 in 10-, 1 in 50-, and 1 in 100-year events. The return level z_N associated with a return period N (in years) is given by the formula:

$$z_N = u + \frac{S}{K}\left[\left(Nn_y\zeta_u\right)^K - 1\right], \text{ if } K \neq 0$$

$$z_N = u + S * \log\left(Nn_y\zeta_u\right), \text{ if } K = 0$$

where n_y is the number of observations per year and ζ_u is the probability that an individual observation exceeds the threshold u. Still according to Coles (2001), ζ_u has a natural estimator: $\widehat{\zeta}_u = \frac{k}{n}$, where $\widehat{\zeta}_u$ is the estimate of ζ_u and n is the total number of observations.

The last step to tune the POT method is to choose the right threshold for our specific problem. If the threshold is too high, there will not be enough data and the variance will be too high. If the threshold is too low, there will be too much bias (Roth et al., 2015). Mean Residual Life (MRL) plots and Hill plots were used in this study to drive this estimation. The MRL plot consists in the points

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(z_i - u)\right), u < z_{max}\right\}$$

where z_{max} is the highest empirical value, u is a threshold and n_u is the corresponding number of empirical values above the threshold.

The Hill plot consists in two plots. The first one displays the points $\{(u, K'_u), \ u < z_{max}\}$, where K'_u is the estimated shape parameter with a given threshold u. The second plot consists in the points $\left\{\left(u, \widehat{S}'_u\right), \ u < z_{max}\right\}$, where $\widehat{S}'_u$ is the estimated scale parameter that has been modified following $\widehat{S}'_u = S'_u - K'_u * u$ in order to eliminate the scale's parameter linear dependence in u.

A valid choice of threshold is found when

1. the MRL plot becomes approximatively linear,

and

2. the Hill plot becomes approximatively constant.

However, the interpretation of these plots can sometimes be rather complex (see Coles, 2001; Piera-Martinez, 2008, for details) and we took great care in analyzing our choices. In order to use the POT approach, the data sample should ideally be statistically independent and stationary. Since extreme values tend to cluster during a single event (Charras-Garrido & Lezaud, 2013), declustering is often necessary. The fact that we have already separated our data into physical events as explained in section 3 ensures this declustering, which is furthermore physics based, and not only mathematically based. Even though underlying solar cycles does not allow our data to meet the stationarity condition, no detrending method was used in this paper.

5. Extreme Events Characterization Using the 150 Years of Data

We have first applied the EVA to all the geomagnetic events observed between 1868 and 2017. During this 150-year period, 438,296 values of Ca were available. The number of events observed is 3,672. The

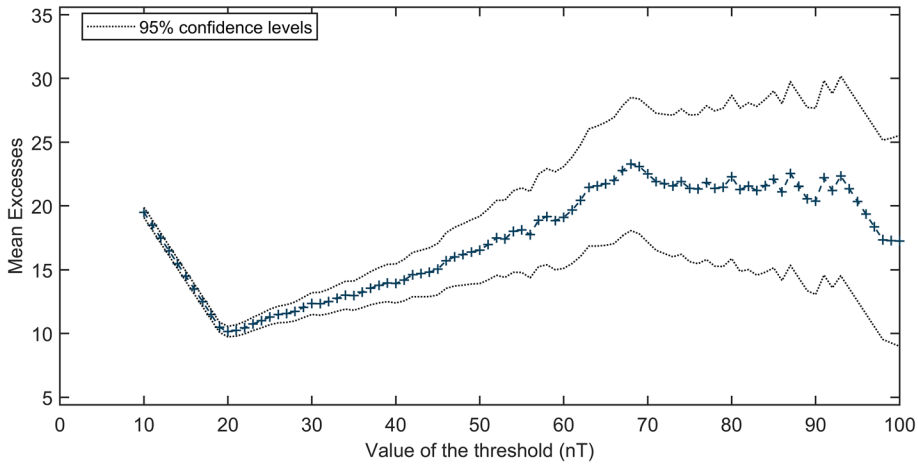

Figure 7. Mean Residual Life plot for geomagnetic events data displayed with the 95% confidence interval.

selected POT threshold, estimated with the methods mentioned in section 4, is $u = 67$ nT. The MRL plot and the Hill plot are given in Figures 7 and 8, respectively. This situation illustrates the fact that the threshold selection can be rather difficult, since there is no clear evidence of linearity in the MRL plot, and the Hill plot is not perfectly constant for $u > 67$ nT. However, 67 nT is already a high value for Ca, and choosing this value as the POT threshold should ensure us that only extreme events are being taken into account. Sixty-two events are above the threshold (it corresponds to the 99.986th percentile of the whole data set).

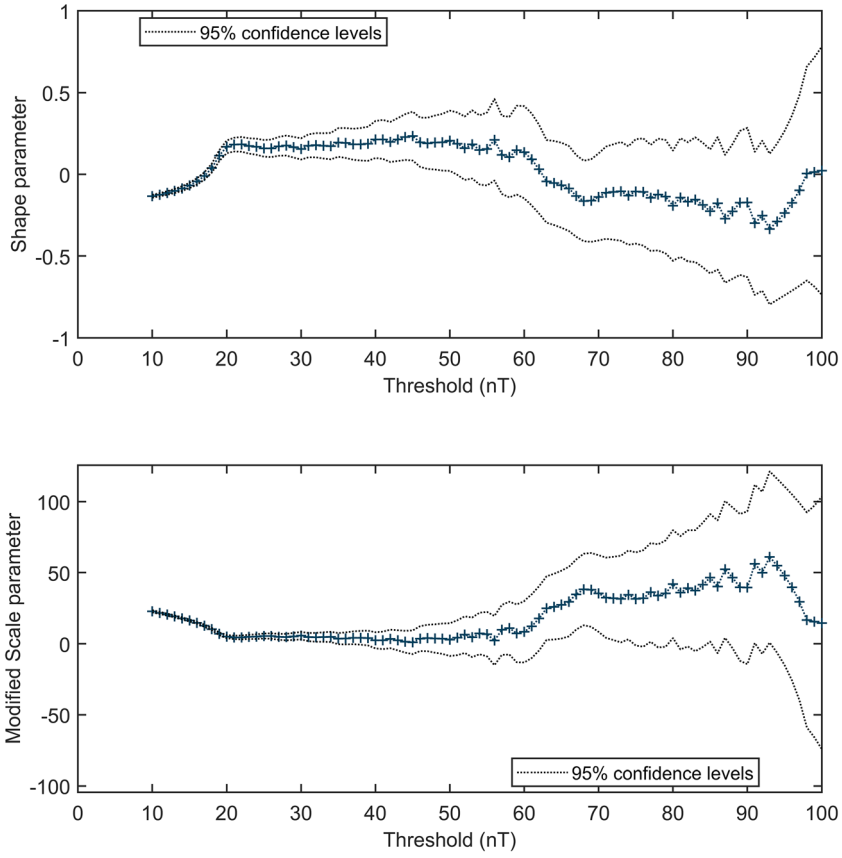

Figure 8. Hill plot for the geomagnetic events data displayed with the 95% confidence interval (computed with the standard deviation-based method).

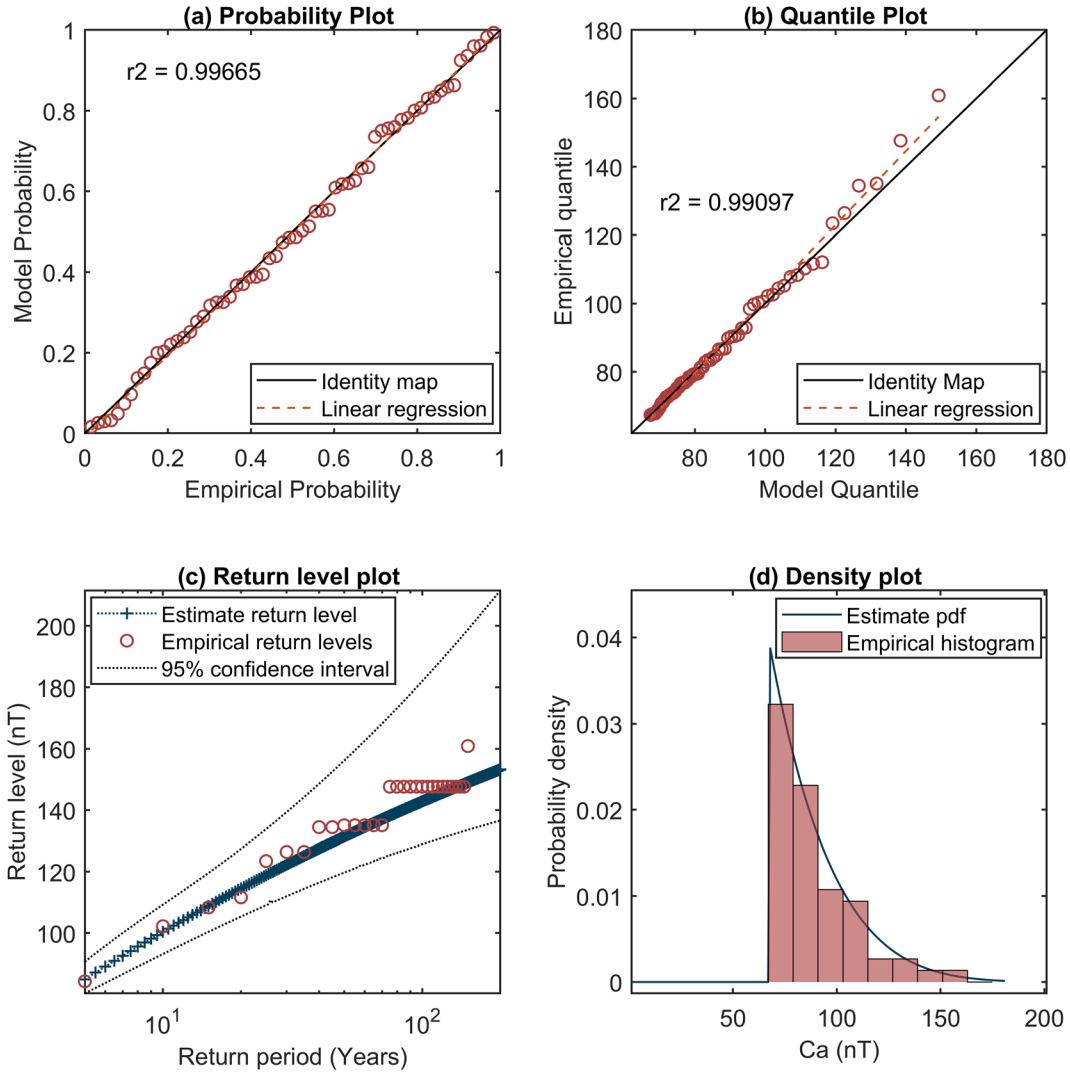

Figure 9. Diagnostic plots for extreme value analysis on the 150-year long geomagnetic events data set. (a) Probability plot, (b) Quantile plot, (c) Return level plot, and (d) Density plot.

The maximum likelihood estimate of the shape parameter K is −0.13, and its 95% confidence interval is [−0.33, 0.18]. The maximum likelihood estimate of the scale parameter S is 25.80, and its 95% confidence interval is [17.9, 36.3]. The negative value of K indicates that the distribution has an upper limit denoted z_{inf} and given by

$$z_{inf} = u - \frac{S}{K}.$$

Here, z_{inf} equals 261.9 nT, which indicates that, considering the most likely values of K and S, an event could not exceed a Ca value of 261.9 nT. This shows that, as expected from the physics point of view, the existence of an upper bound is most likely, even if this value is still very far from what has been observed over the last 150 years. However, taking into account the 95% confidence intervals of K and S, we obtain a 95% confidence interval of z_{inf} estimates spreading from 168.1 nT to infinity (when K equals positive values). Hence, we cannot state with a 95% confidence level that the distribution of Ca events has an upper limit. The highest confidence level that gives us an interval containing only negative values for the shape parameter is 64%, with a confidence interval ranging from −0.23 to −0.01.The diagnostic plots are given in Figure 9. The Probability Plot (Figure 9a) and Quantile plot (Figure 9b) are very close to linear (with both coefficients of determination

Table 2
Return Level Estimates With the POT Approach for the 1868–2017 Yearlong Period

Return period	Level estimate	95% confidence interval	Empirical estimate
10 years	100.39 nT	[93.1, 109.2] nT	102.2 nT
50 years	131.39 nT	[119.7, 156.1] nT	135.0 nT
100 years	142.84 nT	[128.9, 182.0] nT	147.6 nT
200 years	153.28 nT	[136.6, 211.6] nT	≥160.9 nT

higher than 0.99) and to the unit diagonal and hence do provide some confidence that the POT approach is relevant in this situation. The Return level plot (Figure 9c) provides a satisfactory representation of the empirical estimates. The probability density estimate in (Figure 9d) seems to fit well the histogram of the data. Therefore, the diagnostic plots tend to confirm that the generalized Pareto model is adequate to our EVA when using the whole available data set.

Using this model, it becomes possible to determine estimates of the 1-in-10-, 1-in-50-, 1-in-100-, and 1-in-200-year events. The results, along with their 95% confidence intervals are given in Table 2.

This first analysis gives rather accurate estimates for all the 1-in-N-year events. The fact that we used a very long range of homogenous data allows for a high confidence in these results. We can compare these results with the straightforward statistical analysis conducted in previous section. These values have also been reported in Table 2 and show very good agreement, thus validating the POT approach. According to these results, the "Halloween Storm" of 31 October 2003 would be close to a 1-in-130-year event. The storm of 14 March 1989 that caused many damages in Quebec would be close to a 1-in-50-year event. The storm of 10 November 2004 would have not only been a 1-in-10-year event but also the last 1-in-10-year event to have occurred from late 2017. This shows that since the advent of the Space Era we have already witnessed extreme events of the highest intensity, but that these events are not equally distributed among the years.

Finally, as the data set covers 150 years, the 1-in-100-year return level should be better estimated using the POT approach than the straightforward analysis as the data set duration becomes comparable to the return level looking for (at best one or two events of this strength are observed during this time period). This shows the benefits of such an evolved method compared to standard ones. Nevertheless, the question remains the same: What is the real confidence we can have in this method?

6. Influence of the Time Period and Time Range of Data Used on the EVA Results

6.1. EVA on Shorter Subsets

The heterogeneity discussed in the previous section can be explained by the discrepancies in terms of geomagnetic activity that can be observed within and between two different solar cycles. This constitutes a first indicator that all time periods are not equally representative regarding extreme geomagnetic events.

For instance, the highest event that has been observed between 2010 and 2017 reached a maximum Ca level of 53.63 nT on 19 March 2015. According to our analysis, this barely matches a 1-in-1.5-year event. Hence, this period does not even contain a single event that could be considered as extreme. Performing an extreme event analysis on only such a short and nonrepresentative period would consist in fitting a distribution to irrelevant data, since the tail of the distribution of the events would be truncated when compared to the full-range data set or another subset that contains extreme events.

The problem here is that most space missions that collect measurements of electron and proton fluxes in the radiation belts are usually not designed to last a very long time (usually between 2 and 10 years) and they effectively last at most two decades. Hence, data collected from a single mission (e.g., the Van Allen Probes, launched on 30 August 2012) are only representative of the time range covered by the mission, and in the given example, it is a nonrepresentative time period from the extreme events point of view, as we have just exposed.

To illustrate this fact with figures, we performed an EVA on data from the 2010–2017 subset of Ca-based events as if we were "blind" regarding the rest of the full data set. The first difficulty arises for the choice of the threshold. Here, the interpretation using the MRL plot (given in Figure 10) is not straightforward. We chose to use a threshold $u = 35$ nT, which gives 21 events above the threshold to fit the distribution. The results are that the estimated 1-in-10-year event is 52.08 nT and the 1-in-100-year event is 54.04 nT. It appears immediately that these results are very far from those obtained with the full data set. In addition, the diagnostic plots in Figure 11 are very poor, which indicates that the results from this analysis should not be considered as definitive values.

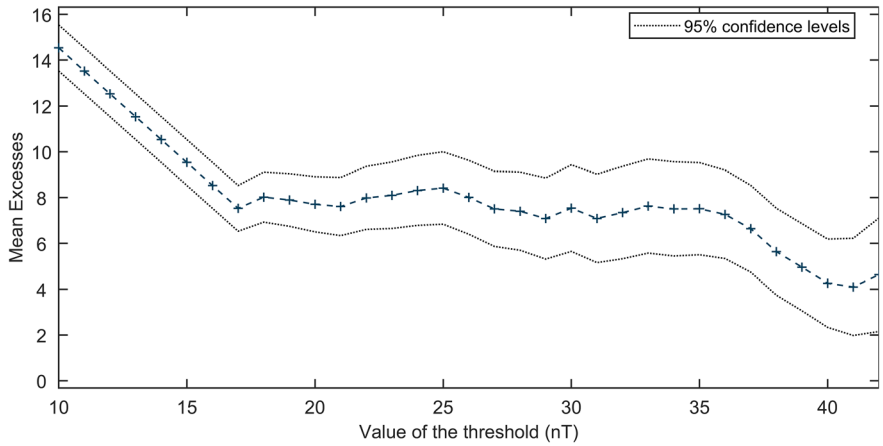

Figure 10. Mean Residual Life plot for geomagnetic events between 2010 and 2017 alone displayed with the 95% confidence interval.

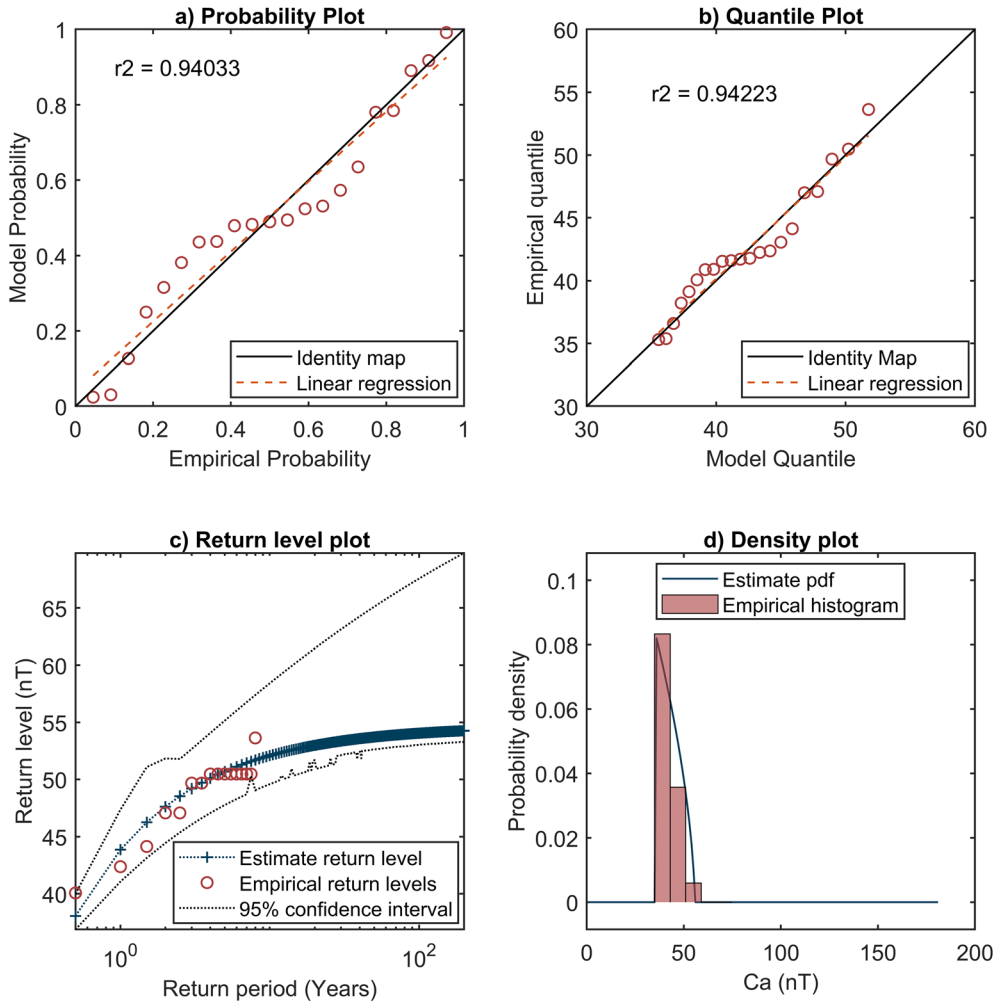

Figure 11. Diagnostic plots for extreme value analysis on the 2010–2017 (included) geomagnetic events data. (a) Probability plot, (b) Quantile plot, (c) Return level plot, and (d) Density plot.

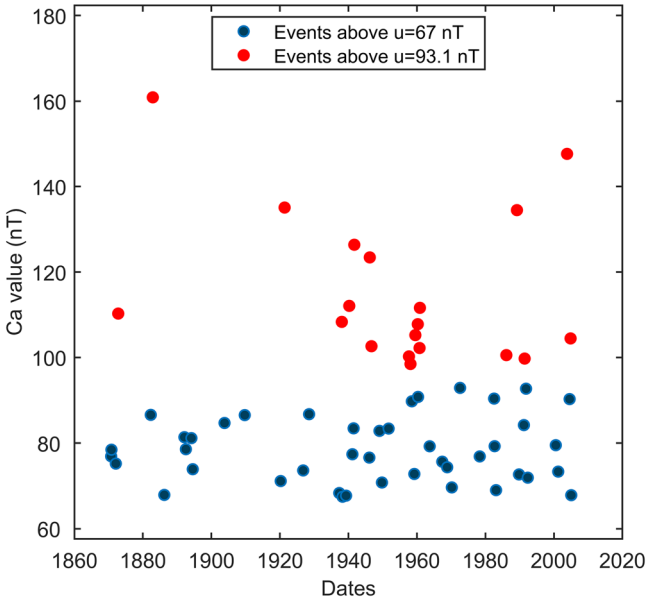

Figure 12. Events that displayed a maximum Ca index stronger than 67 nT (in blue) and 93 nT (in red) over time.

Conversely, other periods contain more numerous extreme events than the average. As an example, nine intense events (events above the threshold $u = 67$ nT) were observed between 1957 and 1964, which equals an average of 1.125 event per year, while for the whole period (1868–2017) this average is 0.41 event per year. Whereas this could be considered good news if one were to study the extreme events using data including this period, this could also introduce a bias if data from this period *alone* were to be used in such a study, since the distribution of the events in this sole period does not seem to match the overall distribution. To illustrate the non-uniformity of the distribution of the events, Figure 12 shows all the events stronger than 67 nT between 1868 and 2017 and highlight those whose maximum Ca value reached at least 93 nT (which is the 95% confidence lower boundary to the estimated 1-in-10-year event) between 1868 and 2017. In this figure it becomes clear that some periods (e.g., around 1960) contained many very intense events, whereas other periods (e.g., circa 1900–1910) contained very few.

Therefore, performing an accurate EVA on shorter data sets appears to be a challenge, not only because a shorter data set contains less numerous values as a longer data set but also because all subsets are not equally representative of the full range of the Ca-based events that could be observed. That is why in the next subpart we attempt to assess a first estimation of the representativeness of shorter data sets.

6.2. Estimating the Representativeness of Short Data Sets

As mentioned above, the first benefit of EVA methods is to be able to accurately estimate return levels for time scales larger than the duration of the used data set itself. This is also one of the mains reasons that in the following we assert the representativeness of short data sets using EVA return levels and not empirical ones. The point is that for Space Weather investigation, it is important to be able to quantify such extreme events, from the flux point of view. However, missions measuring radiation flux never exceed 20 years of continuous data, nor do they cover the whole radiation belts region. For the aforementioned reasons and in order to be able to draw a comparison with studies such as Meredith et al. (2015), which perform extreme value analyses on relativistic electron fluxes and thus only use less than 20 years of data, we here perform multiple EVAs using a 20-year-long sliding window included in the full 150-year-long data set (e.g., 1868–1887 and 1869–1888). We try to determine if there exists any 20-year period that is representative of the 1868–2017 full period in terms of extreme geomagnetic events. Before doing so, let us simply note that, as shown in Figure 4, the linear correlation between Ca and Cf for energies $E > 900$ keV at $L = 6.6$ is close to 0, and therefore, no conclusion on the validity of the results of Meredith et al. (2015) can be directly drawn from the results using Ca.

We must first give a more precise definition of a representative period. A representative period is a period that displays a distribution of Ca-based events very similar to the distribution of the events during the whole 150-year period. Therefore, performing an EVA on a representative period should lead to the fitting of a similar Pareto distribution as the one fitted with the full data set. From this assertion we extracted two quantifiable criteria to determine whether a 20-year period is representative or not:

1. When using the same threshold as the one used with the full period, the maximum-likelihood estimates of the shape parameter, the scale parameter (and hence of the 1-in-10-year and the 1-in-100-year events) must fall within the 95% confidence interval obtained with the full data set. When using a different threshold, the shape and scale parameters are necessarily changed. In this case, the criterion becomes that all the estimated 1-in-N-year event must fall within their 95% confidence interval obtained with the full data set.
2. The Pareto distribution must have been fitted with more than six points (i.e., there must be more than six events above the threshold u).

The first criterion seems straightforward, since if a 20-year period is representative, then the results of the EVA should be very close to the results obtained in section 5. The second criterion means that a

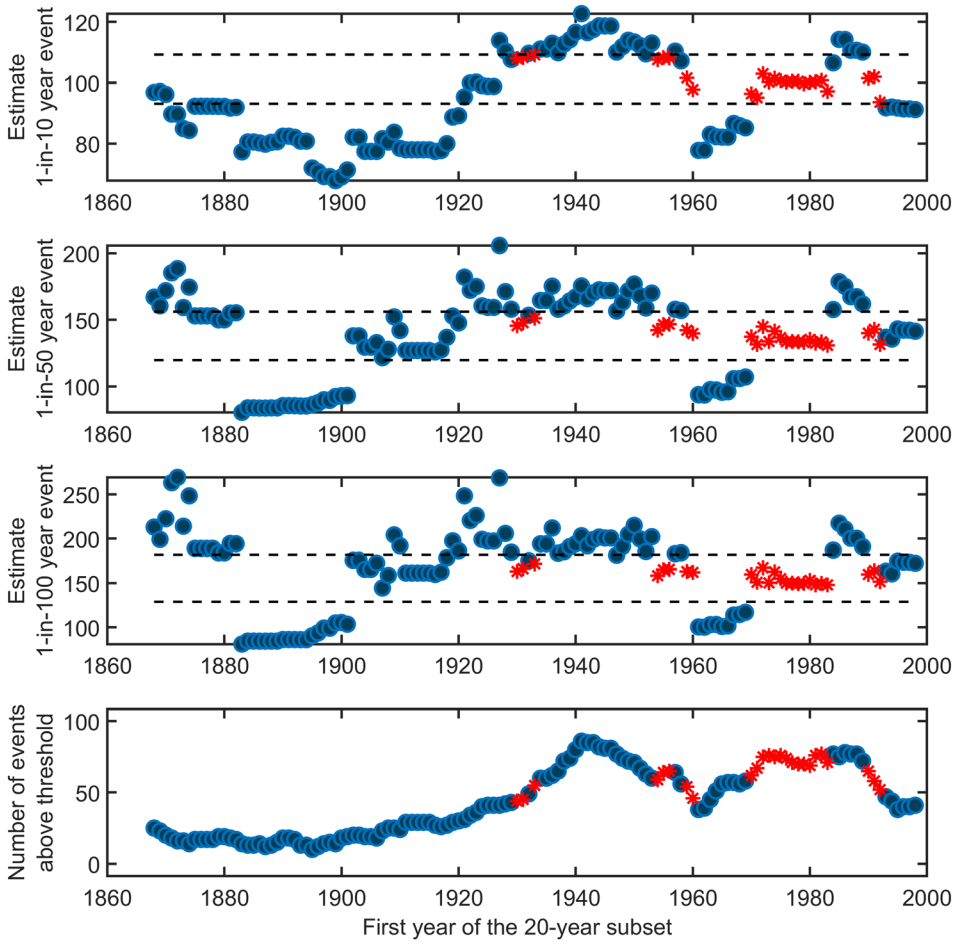

Figure 13. Estimation of the 1-in-10-, 1-in-50-, and 1-in-100-year events for all 20-year-long subsets comprised between 1868 and 2017 using a threshold $u = 45$ nT. The 95% confidence intervals obtained with the full 150-year-long data set and a threshold set at 67 nT are given in dashed lines. The number of events above the threshold is given in the last subplot. The representative subsets obtained with this method are displayed as a red asterisk.

distribution fitted on only a few data points is not as reliable as a fit using multiple data points. The threshold of six points means that there must be at least 0.3 event-above-the-threshold per year during the 20-year window (there is 0.41 event above the threshold $u = 67$ nT per year during the full 1868–2017 period). One could argue that six points is a very low number to fit a distribution. This uncertainty should translate in very large confidence intervals after doing the EVA.

The first difficulty arises from the choice of the threshold. Indeed, since all 20-year-long periods are not equal in terms of geomagnetic activity, some periods do not contain many, if not any, very intense events. By plotting the MRL and Hill plots for many subsets it appears that a suitable value for the threshold is always comprised between 35 and 50 nT. Hence, we decide to use a fixed threshold $u = 45$ nT for all EVA analyses performed here.

The results of the multiple EVAs are presented in Figure 13, which shows the estimated 1-in-10-year, 1-in-50-year, and the 1-in-100-year events along with the number of events above the threshold for all 20-year subsets. The subsets whose results match all the criteria of representativeness are highlighted as a red asterisk. We have not displayed all the 1-in-N-year estimated events for clarity and evident lack of place, but all the subsets for which the estimated 1-in-10-, 1-in-50-, and 1-in-100-year events matched the criteria also had well estimated 1-in-N-year event for N between 3 and 200. According to these results, 25 subsets (among which 19 subsets started during the Space Era) are as representative as the whole 150-year-long data set in terms of extreme events. However, to correctly analyze these results, one shall keep in mind that the

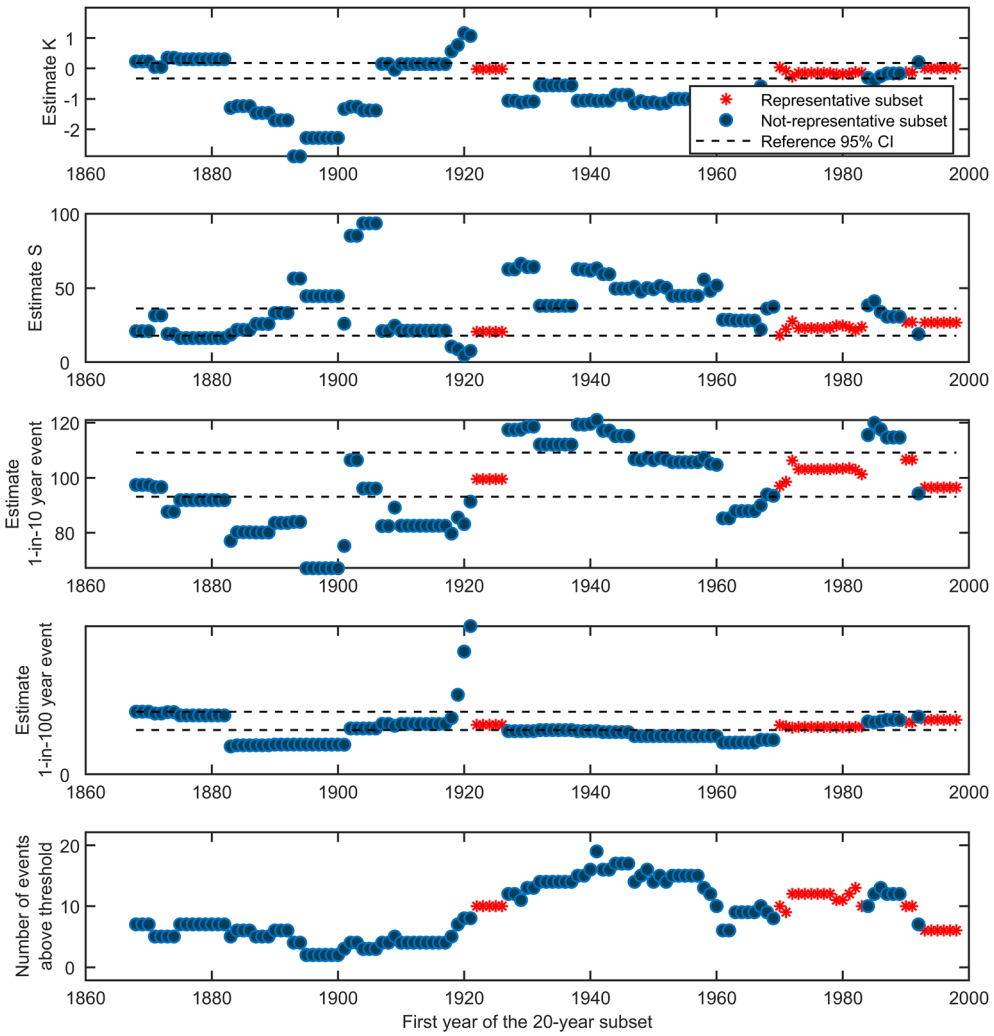

Figure 14. Estimation of the Shape and Scale parameters as well as the 1-in-10- and 1-in-100-year events for all 20-year-long subsets comprised between 1868 and 2017 using a threshold $u = 67$ nT. The 95% confidence intervals obtained with the full 150-year-long data set and a threshold set at 67 nT are given in dashed lines. The number of events above the threshold is given in the last subplot. The representative subsets obtained with this method are displayed as a red asterisk.

threshold chosen here for the POT analyses (45 nT) is lower than the one chosen previously for the whole data set (67 nT). Yet the threshold also conveys a specific mathematical and physical meaning. Indeed, mathematically this threshold quantifies the value above which the tail of the distribution of the Ca events displays a different behavior as the rest of the distribution. Physically, it also gives a value above which an event could be considered as extreme. Hence, if we want to be able to draw a good comparison we should always use the same threshold.

That is why we perform again multiple EVAs on the same multiple 20-year-long sliding window, but this time with the threshold $u = 67$ nT. The results of this experiment are given in Figure 14. This time it appears that 27 subsets can be considered representative of the whole data set, among which 22 subsets started during the Space Era. Nine subsets that previously seemed to give us good results do not match all the criteria anymore (e.g., the subsets starting in 1930 and 1960). Conversely, 11 subsets now match all the criteria, including all the subsets that started after 1993 included.

The 1930–1949 subset that previously gave us good results is not as satisfying anymore. These results can be considered with a higher trust than the results obtained with the threshold set at 45 nT since we are now using a threshold that has a true physical and mathematical meaning, even though it is not a threshold that we would have used if we only had knowledge of the given 20-year-long subsets. In particular, the period

encompassing the years 2003 and 2004 are good candidates. We can consequently have good characterizations of extreme events using observed fluxes during such periods. However, if we refine our sliding window again and, for example, keep only last solar cycle observations such as the valuable Van Allen Probes ones, we see that the time period is not representative for such analysis.

The main lesson learnt here is that there are 20-year-long subsets that are representative of the full data set given the right threshold. The main obstacle to performing an accurate EVA on such a short but representative data set is that, quite often, the analyst will probably not chose the right threshold using the usual methods for threshold selection, which will give misleading estimates such as the ones we obtained with eleven subsets when the threshold was set to a lower value (45 nT).

7. Conclusions and Perspectives

We have performed an EVA on a 150-year-long data set using an index derived from the aa index, which is considered to be homogenous in time. Therefore, the following results can be trusted with a high confidence level. Prior to the EVA itself we had to define what a geomagnetic event is. Our definition is probably not perfect but accurate for computing the EVA. By using the POT method with the threshold carefully set to 67 nT, it appears that the estimated 1-in-10-, 1-in-50-, and 1-in-100-year events, respectively, reached a Ca value of 100.39, 131.39, and 142.84 nT.

Moreover, our results show that we cannot assert with a 95% confidence level that the distribution of Ca events has an upper limit, even if the estimated (most likely) shape parameter suggests so. Finding either there is an upper limit to the particle fluxes in the radiation belts is an important topic in Space Weather research. For instance, in Meredith et al. (2017) the authors conduct an EVA on relativistic electron fluxes (for $0.69 < E < 2.05$ MeV) in the Earth's outer radiation belt. In particular, they find that there is most probably an upper bound to the flux of relativistic electrons in the region $4 < L^* < 6$ but that further out it is not possible to reach a final conclusion since the shape parameter is close to 0, with confidence boundaries overlapping positive values. In this context, let us simply note that even if Ca is, by construction and as we have shown, a good indicator of the general state of disturbance of the radiation belts, we cannot consider Ca as a direct proxy for the particle fluxes saturation since Ca contains an integrated part of temporal information. Thus, the existence of an upper limit on Ca is a different problem from the existence of upper limits on particle fluxes in radiation belts. That being said, our study shows that even with a data set overlapping an extended time period such as ours, the problem of the existence of an upper limit to a distribution of an index representing extreme geomagnetic storms is not easily solvable, which echoes one of the conclusions in (Meredith et al., 2017).

Indeed, while Ca index is accurate for monitoring the global geomagnetic activity on a planetary scale, it may not always be representative of the local situation for a given altitude or orbit. Moreover, Ca value, as defined here with a 4 days characteristic time, will not be perfectly correlated with the dynamics of electrons at any energy. For such purposes, other data sets and measures have to be used, which leads to a great problem regarding EVA: Such data sets are usually much shorter in time. Using shorter (20-year-long) data sets may lead to misleading results when conducting an EVA for two main reasons. The first one is that all periods do not always contain a similar distribution of events. The second reason is that, even for 20-year-long time periods that are quite as representative as the full 150-year-long data set, it is not always straightforward to choose the right threshold for performing a POT analysis, and choosing a different threshold may give altered results. This is why in this study we give time periods that could be more representative than others, based on first order Ca index and our definition of events.

It is interesting to compare the results of this study to, for example, Meredith et al. (2015, 2016), bearing in mind the precautions outlined in section 6.2 above. In Meredith et al. (2015), an EVA is performed on electron fluxes with energy greater than 2 MeV measured by GOES fleet from 1995 to mid-2014. It is found that the distribution of extreme events does not have an upper bound in this particular study. We find the same result when using the data covering the 1998–2017 period with the most likely scale parameter and a threshold set at 45 nT, but not when using a threshold set at 67 nT (which we have shown to be more accurate), neither with the whole 150-year-long data set. Moreover, according to the conclusions of Meredith et al. (2015), the strong magnetic storm of July 2004 is considered as a 1-in-50-year event, while it is barely a 1-in-10-year event in our study. Conversely, the storm of October 2003 that is seen as a 1-in-130-year event

in our study when using the whole data set is not even a 1-in-10-year event when considering GOES > 2-MeV electron fluxes. On the other hand, the conclusions form our study echo some of the results from Meredith et al. (2016), where it is found in particular that the fluxes of E > 30-keV electrons as observed by the POES satellites from 1998 to 2014 tend to a limiting value. This highlights once again the complexity of the Earth's magnetospheric system and that Ca index is a global proxy for the whole energy band and region of the electron radiation belt.

These differences could be explained in different ways. The fact that the distribution does not have an upper bound when considering GOES electron fluxes might be only a consequence of the short data set used in that study, since we observe the same consequence in our one when shortening the data set and using a lower threshold. The difference of the return levels is most probably mainly a consequence of the data set used in each study. In (Meredith et al., 2015), the data concerns only electrons with an energy greater than 2 MeV and the data are made at the geostationary orbit. Therefore, the data are not representative of the whole situation at each altitude, and in this case, the values of Ca present a lower correlation with the GOES data. Indeed, the geostationary orbit experienced strong and rapid disturbances during storms, in the contrary of other more intern orbits. As for example, we can mention, especially for the 2-MeV fluxes, that during magnetic storms dropouts are very effective on that orbit, which induce rapid and strong losses above a few hundreds of keV (see Herrera et al., 2016, for more details). Such dropouts are driven by both solar wind parameters and geomagnetic activity. Therefore, Ca index may be less representative of what is occurring at geostationary orbit.

In this first work, we have demonstrated the interest of long-term data sets for extreme geomagnetic events analysis, in particular in using dedicated but complex statistical methods. Most of all, we show the limitations and caveats to take care about in such analysis. Finally, an idea to go further would be to take into account solar cycle modulation in such an analysis, as well as discriminating the type of event, especially between CIRs and CMEs. The existence of a solar cycle breaks the stationarity assumption necessary for performing EVA. However, each cycle is different from the others, which makes it difficult to include a trend in our model. But taking a data set of 150 years allows us to take into account almost 14 different cycles. Among these cycles, some were very intense in the late twentieth century, other were rather weak in the late nineteenth century and the early twentieth century, as well as the current one. Therefore, using the whole data set is more representative than using a restricted data set that barely represents two cycles.

Another idea that would constitute future works would be to further quantify the accuracy of a restricted data set (in particular a given mission) to estimate 1-in-N-year events for other given types of particle flux or expected effects. This would help in addressing Space Weather recommendations (Millan et al., 2019) and in particular, the issue of assessing a worst-case scenario, which is complex with a single index or data set.

Data Availability Statement

The authors are thankful to the NOAA-POES and NOAA-GOES for online data access available on the CDAweb (at http://cdaweb.gsfc.nasa.gov/). The results presented in this paper rely on geomagnetic indices calculated and made available by ISGI Collaborating Institutes from data collected at magnetic observatories. We thank the involved national institutes, the INTERMAGNET network and ISGI (isgi.unistra.fr). The OMNI data were obtained from the GSFC/SPDF OMNIWeb interface (at https://omniweb.gsfc.nasa.gov). Sunspot data from the World Data Center SILSO, Royal Observatory of Belgium, Brussels. The authors are thankful to Bern Blake, Joe Fennell, Seth Claudepierre, Drew Tuner and all the involved members of the MagEIS instrument team for the RBSP-ECT data, which are publicly available at the website (http://www.RBSP-ect.lanl.gov/).

Acknowledgments

G. Bernoux is thankful for funding from Région Occitanie and ONERA, under Grant Agreements 19008721/ALDOCT and 30196. V. Maget is thankful for funding from CNES and ONERA, under Grant Agreements 4500054922/DIA094 and 24003.

References

Akasofu, S.-I. (1981). Energy coupling between the solar wind and the magnetosphere. *Space Science Reviews*, *28*(2), 121–190. https://doi.org/10.1007/BF00218810

Alves, M. V., Echer, E., & Gonzalez, W. D. (2006). Geoeffectiveness of corotating interaction regions as measured by Dst index. *Journal of Geophysical Research*, *111*, A07S05. https://doi.org/10.1029/2005JA011379

Bartels, J., Heck, N. H., & Johnston, H. F. (1939). The three-hour-range index measuring geomagnetic activity. *Terrestrial Magnetism and Atmospheric Electricity*, *44*(4), 411–454. https://doi.org/10.1029/TE044i004p00411

Benacquista, R., Rochel, S., & Rolland, G. (2017). Understanding the variability of magnetic storms caused by ICMEs. *Annales Geophysicae*, *35*, 147–159. https://doi.org/10.5194/angeo-35-147-2017

Blake, J. B., Carranza, P. A., Claudepierre, S. G., Clemmons, J. H., Crain, W. R., Dotan, Y., et al. (2013). The Magnetic Electron Ion Spectrometer (MagEIS) instruments aboard the Radiation Belt Storm Probes (RBSP) spacecraft. *Space Science Reviews*, *179*, 383–421. https://doi.org/10.1007/s11214-013-9991-8

Borovsky, J. E., & Denton, M. H. (2006). Differences between CME-driven storms and CIR-driven storms. *Journal of Geophysical Research*, *111*, A07S08. https://doi.org/10.1029/2005JA011447

Bourdarie, S. A., & Maget, V. F. (2012). Electron radiation belt data assimilation with an ensemble Kalman filter relying on the Salammbô code. *Annales Geophysicae*, *30*, 929–943. https://doi.org/10.5194/angeo-30-929-2012

Charras-Garrido, M., & Lezaud, P. (2013). Extreme value analysis: An introduction. *Journal de La Société Française de Statistique*, *154*, 66.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer London. https://doi.org/10.1007/978-1-4471-3675-0

Cooley, D. (2009). Extreme value analysis and the study of climate change. *Climatic Change*, *97*, 77–83. https://doi.org/10.1007/s10584-009-9627-x

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extermal events—For insurance and finance*, (1st ed.). Berlin Heidelberg: Springer-Verlag.

Herrera, D., Maget, V. F., & Sicard-Piet, A. (2016). Characterizing magnetopause shadowing effects in the outer electron radiation belt during geomagnetic storms. *Journal of Geophysical Research: Space Physics*, *121*, 9517–9530. https://doi.org/10.1002/2016JA022825

Horne, R. B., Glauert, S. A., Meredith, N. P., Boscher, D., Maget, V., Heynderickx, D., & Pitchford, D. (2013). Space weather impacts on satellites and forecasting the Earth's electron radiation belts with SPACECAST. *Space Weather*, *11*, 169–186. https://doi.org/10.1002/swe.20023

King, J. H., & Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *Journal of Geophysical Research*, *110*, A02104. https://doi.org/10.1029/2004JA010649

Koons, H. C. (2001). Statistical analysis of extreme values in space science. *Journal of Geophysical Research*, *106*(A6), 10,915–10,921. https://doi.org/10.1029/2000JA000234

Lugaz, N., Farrugia, C. J., Huang, C.-L., & Spence, H. E. (2015). Extreme geomagnetic disturbances due to shocks within CMEs. *Geophysical Research Letters*, *42*, 4694–4701. https://doi.org/10.1002/2015GL064530

Mayaud, P.-N. (1972). The aa indices: A 100-year series characterizing the magnetic activity. *Journal of Geophysical Research*, *77*(34), 6870–6874. https://doi.org/10.1029/JA077i034p06870

Mayaud, P.-N. (1980). *Derivation, meaning, and use of geomagnetic indices*, *Geophysical Monograph Series*, (Vol. 22). Washington DC: American Geophysical Union. https://doi.org/10.1029/GM022

Mayaud, P.-N. (1971). Une mesure planétaire d'activité magnétique basée sur deux observatoires antipodaux. *Annales Geophysicae*, *27*, 67–70.

Menvielle, M., Iyemori, T., Marchaudon, A., & Nosé, M. (2011). Geomagnetic indices. In M. Mandea, & M. Korte (Eds.), *Geomagnetic observations and models* (pp. 183–228). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9858-0_8

Meredith, N. P., Horne, R. B., Glauert, S. A., Thorne, R. M., Summers, D., Albert, J. M., & Anderson, R. R. (2006). Energetic outer zone electron loss timescales during low geomagnetic activity. *Journal of Geophysical Research*, *111*, A05212. https://doi.org/10.1029/2005JA011516

Meredith, N. P., Horne, R. B., Isles, J. D., & Green, J. C. (2016). Extreme energetic electron fluxes in low Earth orbit: Analysis of POES E > 30, *E* > 100, and E > 300 keV electrons. *Space Weather*, *14*, 136–150. https://doi.org/10.1002/2015SW001348

Meredith, N. P., Horne, R. B., Isles, J. D., & Rodriguez, J. V. (2015). Extreme relativistic electron fluxes at geosynchronous orbit: Analysis of GOES E > 2 MeV electrons. *Space Weather*, *13*, 170–184. https://doi.org/10.1002/2014SW001143

Meredith, N. P., Horne, R. B., Sandberg, I., Papadimitriou, C., & Evans, H. D. R. (2017). Extreme relativistic electron fluxes in the Earth's outer radiation belt: Analysis of INTEGRAL IREM data. *Space Weather*, *15*, 917–933. https://doi.org/10.1002/2017SW001651

Millan, R. M., von Steiger, R., Ariel, M., Bartalev, S., Borgeaud, M., Campagnola, S., et al. (2019). Small satellites for space science: A COSPAR scientific roadmap. *Advances in Space Research*, *64*, 1466–1517. https://doi.org/10.1016/j.asr.2019.07.035

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, *135*(3), 370. https://doi.org/10.2307/2344614

Odenwald, S. (2015). *Solar storms: 2000years of human calamity*. San Bernardino, CA: CreateSpasce Independent Publishing Platform.

Pickands, J. III (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Piera-Martinez, M. (2008). Modélisation des comportements extrêmes en ingénierie (phdthesis). Université Paris Sud - Paris XI. Retrieved from https://tel.archives-ouvertes.fr/tel-00338076/document

Reeves, G. D., McAdams, K. L., Friedel, R. H. W., & O'Brien, T. P. (2003). Acceleration and loss of relativistic electrons during geomagnetic storms. *Geophysical Research Letters*, *30*(10), 1529. https://doi.org/10.1029/2002GL016513

Reeves, G. D., Morley, S. K., Friedel, R. H. W., Henderson, M. G., Cayton, T. E., Cunningham, G., et al. (2011). On the relationship between relativistic electron flux and solar wind velocity: Paulikas and Blake revisited. *Journal of Geophysical Research*, *116*, A02213. https://doi.org/10.1029/2010JA015735

Rochel, S., Boscher, D., Benacquista, R., & Roussel, J. F. (2016). A radiation belt disturbance study from the space weather point of view. *Acta Astronautica*, *128*, 650–656. https://doi.org/10.1016/j.actaastro.2016.07.012

Roth, M. G., Jongbloed, G., & Buishand, T. (2015). Threshold selection for regional peaks-over-threshold data. *Journal of Applied Statistics*, *43*, 1291–1309. https://doi.org/10.1080/02664763.2015.1100589

Sauvaud, J.-A. (2002). Earth magnetosphere. In *Space environment: Prevention of risks related to spacecraft charging* (pp. 81–106). Toulouse: Cépaduès.

Siscoe, G. L. (1976). On the statistics of the largest geomagnetic storms per solar cycle. *Journal of Geophysical Research*, *81*(25), 4782–4784. https://doi.org/10.1029/JA081i025p04782

Siscoe, G. L., Crooker, N. U., & Clauer, C. R. (2006). Dst of the Carrington storm of 1859. *Advances in Space Research*, *38*(2), 173–179. https://doi.org/10.1016/j.asr.2005.02.102

Tenfjord, P., & Østgaard, N. (2013). Energy transfer and flow in the solar wind-magnetosphere-ionosphere system: A new coupling function. *Journal of Geophysical Research: Space Physics*, *118*, 5659–5672. https://doi.org/10.1002/jgra.50545

Thomson, A. W. P., Dawson, E. B., & Reay, S. J. (2011). Quantifying extreme behavior in geomagnetic activity: Extreme geomagnetic storms. *Space Weather*, *9*, S10001. https://doi.org/10.1029/2011SW000696

Tramblay, Y., Neppel, L., Carreau, J., & Najib, K. (2013). Non-stationary frequency analysis of heavy rainfall events in southern France. *Hydrological Sciences Journal*, *58*, 280–294. https://doi.org/10.1080/02626667.2012.754988