Correspondence to:

D. Mourenas,
didier.mourenas@cea.fr

A Climatology of Long-Duration High 2-MeV Electron Flux Periods in the Outer Radiation Belt

D. Mourenas[1,2] , **O. V. Agapitov[3]** , **A. V. Artemyev[4]** , and **X.-J. Zhang[4]**

[1]CEA, DAM, DIF, Arpajon, France, [2]Laboratoire Matière en Conditions Extrêmes, Paris-Saclay University, CEA, Bruyères-le-Châtel, France, [3]Space Sciences Laboratory, University of California, Berkeley, CA, USA, [4]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, CA, USA

Abstract   Since the advent of the Space Age, the importance of understanding and forecasting relativistic electron fluxes in the Earth's radiation belts has been steadily growing due to the threat that such particles pose to satellite electronics. Here, we provide a model of long-duration periods of high time-integrated 2-MeV electron flux deep inside the outer radiation belt, based on the significant correlation obtained in 2001–2017 between time-integrated electron flux measured by satellites and a measure of the preceding time-integrated homogenized aa_H geomagnetic index. We show that this correlation is likely due to a stronger cumulative chorus wave-driven acceleration of relativistic electrons and a stronger cumulative inward radial diffusion of such electrons during periods of higher time-integrated geomagnetic activity. Return levels of 2-MeV electron flux are provided based on Extreme Value analysis of time-integrated geomagnetic activity over 1868–2017, in rough agreement with estimates based on 20-year data sets of measured flux. A high correlation is also found between our measure of time-integrated geomagnetic activity averaged over each solar cycle and averaged sunspot numbers, potentially paving the way for forecasts of time-integrated relativistic electron flux during future solar cycles based on predictions of solar activity.

1. Introduction

The ever-growing reliance of modern technological society on space-based connections and services in many economic and military sectors (e.g., internet, financial transactions, navigation, various control systems) has increased hazards associated with potential failures of Global Navigation Satellite Systems and other satellites in the aftermath of strong solar wind disturbances (Eastwood et al., 2017; Glauert et al., 2018; Meredith et al., 2015; Riley et al., 2018). Indeed, electronic devices inside satellites are particularly at risk from the total ionizing dose (TID) of radiation, corresponding to direct or indirect ionization of a semiconductor by incident relativistic electrons in the Earth's radiation belts (Cochran et al., 2009; Zheng et al., 2019), which can lead to satellite anomalies and sometimes even total failure (Chen et al., 2021; Ecoffet, 2013; Iucci et al., 2005). It is worth emphasizing that this TID or radiation absorbed dose (in rad or Gray units) is a long-term cumulative effect, causing a gradual degradation of semiconductor performance (Chen et al., 2021; Stassinopoulos & Raymond, 1988; Zheng et al., 2019).

Spacecraft electronic devices are usually protected from TID by an aluminum shielding of ∼100–300 mil thickness (corresponding to 2.5–7.5 mm or ∼0.7–2 g/cm²), which strongly reduces the penetrating electron flux at energies below ∼1–2 MeV (Chen et al., 2021; Stassinopoulos & Raymond, 1988). Therefore, the main TID risk for satellites is due to long-duration periods (typically days to weeks) of high relativistic (>1–2 MeV) electron flux, which correspond to the periods of highest time-integrated electron flux (also called fluence; e.g., see Chen et al., 2021). Such periods represent a much bigger TID threat than the much shorter-lived periods of highest instantaneous (or hourly or daily) electron flux. Relativistic electron fluxes represent a particularly important hazard for spacecraft electronics at altitudes comprised between 20,000 and 40,000 km in the outer radiation belt, where geomagnetic storms or high-speed solar wind streams can trigger prolonged periods of elevated 2-MeV electron flux (Baker et al., 1990; Glauert et al., 2018; Iucci et al., 2005; Li & Hudson, 2019; Mourenas, Artemyev, & Zhang, 2019; Murphy et al., 2018; Ozeke et al., 2020; Thorne et al., 2013). Such long-duration (typically ∼10 days at $L \sim 4.5$), high 2-MeV electron flux periods have been found to be the main contributors to the total yearly time-integrated electron flux during active years (Mourenas, Artemyev, & Zhang, 2019) and, therefore, to the TID in satellites.

Assessing radiation hazards in the heart of the outer radiation belt (around $L \sim 4.5$) is important for Global Positioning System (GPS) satellites, but also for spacecraft on geostationary transfer orbit (GTO) using slow electric orbit raising, which spend months in this region (Glauert et al., 2018). Various models have been developed to forecast MeV electron fluxes up to $\sim$1–2 days ahead based on the past history of geomagnetic indices and/or solar wind parameters, often with additional inputs such as the measured electron flux at the same or lower energy during the preceding days (Boynton, Balikhin, et al., 2016; Chu et al., 2021; Glauert et al., 2021; Pires de Lima et al., 2020). In the present study, we develop a predictive model of long-duration 10-day periods of high time-integrated 2-MeV electron flux based on the past time-integrated geomagnetic activity, similar to our previous models relying on AE and ap indices (Mourenas, Artemyev, & Zhang, 2019), but using now the homogenized aa index, denoted aa_H, continuously available from 1868 to 2017 (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018). The aa and aa_H indices are based on the range of variation of the horizontal component of the geomagnetic field (Mayaud, 1980).

The proposed model is developed in Section 2 based on Van Allen Probes electron flux data in 2012–2017 (Baker et al., 2013) and GPS satellite data in 2003–2005 (Morley et al., 2016), and its performance is tested using various metrics (Zheng et al., 2019). A key advantage of the proposed model is that it can approximately predict both the start of each 10-day period of high 2-MeV electron flux and the total time-integrated electron flux that will be accumulated over these next 10 days. Therefore, it can provide 1-day-ahead to 7-day-ahead warnings of a high TID based on the past history of only one ground-based geomagnetic index, allowing satellites operators to take temporary measures to prevent damage. In contrast, other forecasting models usually predict electron flux at most 1–2 days in advance and often need additional inputs from satellites (e.g., see Boynton, Balikhin, et al., 2016; Chu et al., 2021; Glauert et al., 2021; Pires de Lima et al., 2020). In essence, the proposed model can be viewed as a very simplified short-term (10-day) climatological model of space weather, compared to other forecasting models more similar to meteorological models. However, an inherent drawback of the proposed model is its inability to predict electron flux outside prolonged periods of high time-integrated flux. A brief analysis of the physical phenomena producing these 10-day periods of elevated relativistic electron flux is provided at the end of Section 2.

A complementary strategy for mitigating total radiation dose effects on spacecraft is to develop a long-term climatology of space weather events, allowing satellite designers to harden electronic devices for surviving the strongest events expected during the satellite lifetime (typically 10–20 years). However, the longest data set of 2-MeV electron flux near $L \sim 4.5$, obtained from modern particle sensors on GPS satellites, represents less than 2 solar cycles of relatively homogeneous data (Morley et al., 2016), and other missions provide even shorter data sets (e.g., 6 years for the Van Allen Probes). As an alternative to statistics of satellite data, a physics-based three-dimensional Fokker-Planck code has been used to simulate electron flux variations over 30 years (Glauert et al., 2018). Nevertheless, reliable estimates of maximum event strength require a statistical analysis of longer-duration (>100 years) data sets (Riley et al., 2018).

Accordingly, we analyze in Section 3 the aa_H index available from 1868 to 2017 (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018) to estimate the return levels of periods of high time-integrated geomagnetic activity. Next, building on the significant correlation obtained in Section 2 between time-integrated aa_H and long-duration periods of high 2-MeV electron flux, we provide estimates of the return levels of such long-duration high 2-MeV electron flux periods. These results are compared with previous estimates based on the available data sets of electron flux. Finally, the high correlation found between time-integrated aa_H and sunspot numbers is used to estimate secular variations of time-integrated electron fluxes.

2. Parameterization of High Time-Integrated 2-MeV Electron Flux Periods as a Function of $Int(aa_H)$

2.1. Geomagnetic Activity Data

The original aa index, based on the range of variation of the horizontal component of the geomagnetic field after subtraction of quiet day variation, has been devised to mimic the ap index over a much longer time span starting in 1868 (Mayaud, 1980). It is constructed from K indices determined from 3-hr measurements at two antipodal middle latitude stations in England and Australia (whereas ap and Kp are based on measurements at

13 stations), normalized to geomagnetic latitudes ±50° (Mayaud, 1980). A recent work has shown that middle latitude indices, such as ap and aa, should probably be preferred to the more stochastic auroral indices AE and AL for long-term reconstruction/forecasting (Mourenas et al., 2020). Therefore, in the present study we make use of the 1868–2017 data set of the homogenized aa_H index (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018). This homogenized aa_H index has been constructed by correcting individual aa values for secular changes in geomagnetic field, for variations in measurement station calibrations, and for asymmetries between hemispheres, and the final aa_H index has been scaled to match the similar am index over 2012–2017 (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018).

We consider here strong events of continuously elevated aa_H index. Integrating aa_H during such events, we obtain a measure $Int(aa_H)$ of their cumulative strength, similar to the $Int(ap)$ measure considered in a previous work (Mourenas, Artemyev, & Zhang, 2019), but somewhat different from the $Int(Dst)$ and $Int(AE)$ measures corresponding to long-duration storms and High Intensity Long Duration Continuous AE Activity (HILDCAA) events, respectively (Mourenas, Artemyev, & Zhang, 2019; Mourenas et al., 2018, 2020; Tsurutani et al., 2006). Mourenas, Artemyev, and Zhang (2019) have shown that the 60 10-day periods of highest time-integrated 1.8-MeV electron flux measured in 2013–2017 by the Van Allen Probes (Baker et al., 2013) at adiabatically invariant shell $L^* \sim 4.5$, have occurred just after (∼0–2 days after) significant $Int(ap) > 800$ nT·hr events, calculated with an integration threshold $ap \geq 15$. Mourenas, Artemyev, and Zhang (2019) have further shown that the magnitudes of such electron flux peaks are well correlated with $Int(ap)$, and that this correlation can be used to hindcast GPS 2-MeV electron flux at $L \sim 4.2–4.4$ in 2002–2012, demonstrating its long-term usefulness. Hereafter, we build on this study by Mourenas, Artemyev, and Zhang (2019), but we parameterize the same 10-day periods of highest time-integrated 2-MeV electron flux by $Int(aa_H)$ instead of $Int(ap)$. Since $aa_H \sim 1.7ap$ on average, we use a threshold $Int(aa_H) > 1,400$ nT·hr, equivalent to the threshold $Int(ap) > 800$ nT·hr used by Mourenas, Artemyev, and Zhang (2019), to define strong events. As aa_H sometimes decreases below 20 nT when $ap = 15$, we also use a conservatively low integration threshold $aa_H \geq 18$ nT, to prevent several strong $Int(ap)$ events of $ap \geq 15$ from being split into two weaker $Int(aa_H)$ events.

2.2. Model of High Time-Integrated Electron Flux

In this section, we investigate the impact of significant $Int(aa_H) > 1,400$ nT·hr events on time-integrated electron flux in the heart of the outer radiation belt. We use the daily-averaged omnidirectional 2.1 MeV electron flux (level 2, release 03) measured by the Relativistic Electron-Proton Telescope (REPT) instrument on board the Van Allen Probes (Baker et al., 2013), at magnetic latitudes <25° and at an adiabatically invariant coordinate $L^* \sim 4.5$ calculated using the TS04D (Tsyganenko & Sitnov, 2005) external field model and the International Geomagnetic Reference Field (IGRF) internal field model. This 2.1 MeV electron flux from REPT has been carefully cross-calibrated using simultaneous Magnetic Electron Ion Spectrometer (MagEIS) electron flux data from the Van Allen Probes (Boyd et al., 2019), making it more reliable than the 1.8 MeV electron flux from REPT used in a previous study (Mourenas, Artemyev, & Zhang, 2019).

To include in our analysis additional data from a more active solar cycle, we also analyze daily-averaged omnidirectional 2-MeV electron fluxes measured at $L = 4.2–4.4$ by GPS satellites in 2001–2011 (Morley et al., 2016, 2017). L is determined using the T89 (Tsyganenko, 1989) and IGRF models. GPS satellites have near-circular orbits at 20,200 km altitude, with a period of 12 hr and an inclination of 55°. GPS electron fluxes are provided by the Combined X-ray and Dosimeter (CXD) instrument developed at Los Alamos National Laboratory, in 11 energy channels between 0.14 and 6 MeV, the final fluxes being re-calculated using a sophisticated fitting procedure after subtraction of proton counts (Morley et al., 2016). We use GPS omnidirectional electron flux averaged over $4.2 \leq L \leq 4.4$, because it is measured at low geomagnetic latitudes <25° (as for Van Allen Probes data) where the flux is usually higher (Li et al., 2014; Thorne et al., 2013).

Morley et al. (2016) have noted that GPS 2-MeV electron fluxes were usually ∼2 times smaller than 2.1-MeV electron fluxes measured by the Van Allen Probes at the same L-shells. We further checked that during the eight 10-day periods following $Int(aa_H)$ events in March-July 2013, the GPS 2-MeV electron flux measured at $L \sim 4.2–4.4$ was on average ∼2.5 times smaller than the 2.1 MeV flux simultaneously recorded at $L^* \sim 4.5$ by the Van Allen Probes (with identical average and median values of the 10-day-averaged Van Allen Probes to GPS flux ratio and 75% of these ratios within [1.6, 3.0]). This is probably mainly due to different calibrations of

the detectors (Morley et al., 2016). Indeed, the GPS and Van Allen Probes measurements used here have been performed during quiet to moderately disturbed 10-day periods beginning at the end of $Int(aa_H)$ events of strong geomagnetic activity—that is, such 10-day periods usually start in the late recovery phase of weak to strong storms. At such nearly quiet times, the geomagnetic field usually comes back to a nearly dipolar configuration, such that $L^* \sim 4.5 \sim L$. Therefore, the analyzed GPS and Van Allen Probes data are likely obtained at very similar radial distances from the Earth. The 10-day time-integrated fluxes obtained from GPS satellites at $L \sim 4.2$–4.4 during these particular periods should also be weakly dependent on the exact magnetic field model used. The remarkable coherence of $\sim$1–2 MeV electron fluxes over the region $4.0 < L < 5.5$ (with typical correlation lengths $\Delta L \sim 0.5$) after enhancement events (Pinto, Bortnik, et al., 2020; Walton et al., 2021) further confirms that the analyzed GPS fluxes obtained at $L = 4.2$–4.4 can be used as good proxies for fluxes at slightly higher L-shells.

Accordingly, we select the $\sim$30 strongest (among a cluster, or isolated) $Int(aa_H) > 1,400$ nT·hr events in 2003–2005 with available GPS electron flux and similarly the $\sim$60 strongest $Int(aa_H)$ events in 2012–2017. We multiply the GPS electron flux at $L = 4.2$–4.4 by a constant factor of 2.5 to obtain electron fluxes approximately equivalent to 2.1-MeV electron fluxes measured by the Van Allen Probes at $L^* \sim 4.5$ during these quiet to weakly disturbed periods. To examine the effects of such strong $Int(aa_H)$ events on time-integrated electron flux, we integrate electron flux over fixed 10-day periods. The starting time t_0 of such 10-day periods is fixed at the end of each $Int(aa_H)$ event, or else at its starting time plus 3–4 days when a longer-than-3-day $Int(aa_H)$ event has then already reached $Int(aa_H) > 1,400$ nT·hr before its end (some examples of $Int(aa_H)$ events are provided in Appendix A). Indeed, 2–3 days of continuously elevated geomagnetic activity are usually sufficient to produce significant increases of 2-MeV electron flux at $L^* \sim 4.5$, via a prolonged local acceleration by chorus waves and/or inward radial diffusion by ultra low frequency (ULF) waves (Baker et al., 1994; Horne et al., 2005; Mourenas, Artemyev, & Zhang, 2019; Ozeke et al., 2020; Thorne et al., 2013; Zhao et al., 2019).

To verify that using such fixed 10-day periods of flux integration is appropriate for characterizing the impact of $Int(aa_H)$ events on space weather, we show in Figure 1a the distribution of the start and end times of the periods of high 2-MeV electron flux measured by the Van Allen Probes or GPS satellites following the selected strong $Int(aa_H) > 1,400$ nT·hr events. Such start and end times are respectively fixed at $\sim$1/2 and $\sim$1/3 of the maximum daily flux reached during each 10-day period. Such start and end times of high flux periods are calculated here with respect to the time t_0 defined above, corresponding to the end of each $Int(aa_H)$ event. We also indicate the full duration (between their start and end times) of these periods of high 2-MeV electron flux. Figure 1a shows that $Int(aa_H) > 1,400$ nT·hr events are followed by prolonged peaks of 2-MeV electron flux, starting on average immediately at the time t_0 and lasting in general 10 ± 4 days. The average start of such flux peaks occurs at $t \simeq -0.15$ days (slightly earlier than t_0), their end at $t \simeq 10.3$ days, for a full duration of $\simeq$10.5 days (with standard errors smaller than $\pm$0.4 days). Accordingly, we henceforth examine time-integrated electron fluxes calculated over these fixed 10-day periods following strong $Int(aa_H)$ events.

The events with $Int(aa_H) > 1,400$ nT·hr are very good precursors/predictors of prolonged peaks of 2-MeV electron flux reaching 10-day time-integrated fluxes larger than $2 \cdot 10^{11}$ e/cm²/sr/MeV, since they allow to recover nearly all (92%) of these flux peaks in 2012–2017. This is due to the fast decrease of the average time-integrated flux as $Int(aa_H)$ decreases below 2000 nT·hr in Figure 1b. Therefore, decreasing the $Int(aa_H)$ threshold below 1,400 nT·hr would not significantly improve the modeling of these flux peaks. Conversely, increasing this threshold to 2,000 nT·hr would result in missing $\sim$30% of these events (see Figure 1b). This indicates that the 1,400 nT·hr threshold is appropriate.

Figure 1b shows that all $Int(aa_H) > 1,400$ nT·hr events are followed by 10-day periods of high time-integrated 2-MeV electron flux. There is a significant Pearson linear correlation $R = 0.64$ ($R^2 = 0.41$) between $Int(aa_H)$ and 10-day-integrated $\sim$2-MeV electron fluxes $F(t)$ at $L^* \sim 4.5$ measured by the Van Allen Probes or inferred from measurements by GPS satellites. The best least-squares fit to the 10-day-integrated 2-MeV electron flux is given by

$$\int F(t)dt \simeq (0.4283 \ln [Int\,(aa_H)] - 2.963) \cdot 10^{12} \text{ e/cm}^2/\text{sr/MeV}. \tag{1}$$

Nearly 81% (70%) of the measured values remain within a factor 2.5 (2) of the best fit in Equation 1 and $\sim$98% (95%) remain below 2.5 (2) times the best fit level. The Spearman Rank Order Correlation Coefficient (ROCC) between the obtained best least-squares fit and the measured time-integrated flux is ROCC = 0.67. The z-score
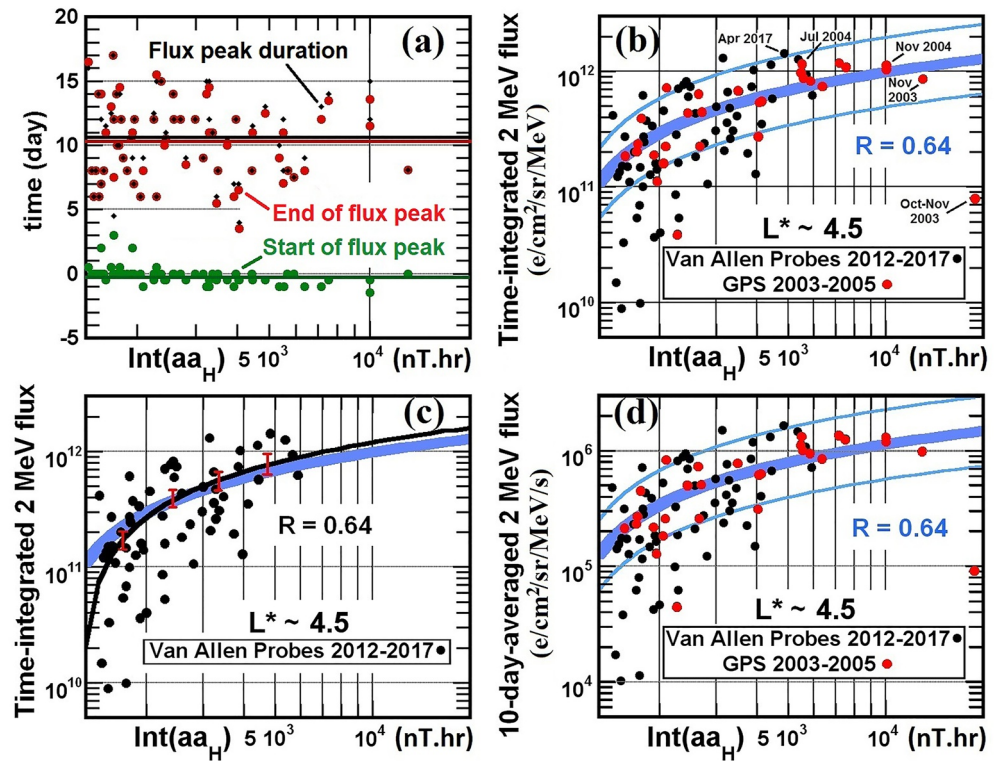
Figure 1. (a) Distribution of starting time (green), end time (red), and duration (black) of periods of high time-integrated 2-MeV electron flux measured by the Van Allen Probes near $L^* \sim 4.5$ or by GPS satellites at $L = 4.2$–4.4 following $Int(aa_H) > 1,400$ nT·hr events. Time 0 corresponds to the end of an $Int(aa_H)$ event, or to its starting time plus 3–4 days if it is longer than 3 days and has already reached 1,400 nT·hr. Average values are shown by horizontal lines of same colors. (b) Time-integrated 2-MeV electron flux measured by the Van Allen Probes at $L^* \sim 4.5$ in 2012–2017 (black circles) and by GPS satellites at $L = 4.2$–4.4 (rescaled by a factor 2.5) in 2003–2005 (red circles) during 10-day periods immediately following $Int(aa_H) > 1,400$ nT·hr events, as a function of $Int(aa_H)$. The best least-squares fit is shown by a thick blue curve. A factor 2 range around this best fit is delimited by thin blue curves. The dates of several events are indicated. (c) Same as (b) but keeping only time-integrated electron fluxes measured by the Van Allen Probes in 2012–2017 (black circles). The corresponding best fit is shown in black (with standard error bars in red), together with the best fit from panel (b) in blue. (d) Same as (b) but showing the corresponding average electron flux over the same 10-day periods.

of the ROCC is $z > 4$, that is, the Fisher transformation of the ROCC is more than four standard deviations away (corresponding to a p-value < 0.0001) from the null hypothesis of a statistical independence of time-integrated flux and $Int(aa_H)$ in Figure 1b (e.g., Fieller et al., 1957). These results suggest that the dependence of time-integrated flux on $Int(aa_H)$ is both real and statistically significant.

Figure 1c shows that the best fit to all Van Allen Probes and GPS data (in blue) given by Equation 1 is very close (within standard error bars) to the best fit to Van Allen Probes data alone (in black, corresponding to $\int F(t) dt = (0.575 \ln[Int(aa_H)] - 4.10)10^{12}$ e/cm^2/sr/MeV and $R = 0.64$). The difference between these two fits remains smaller than 20% between 1800 and 20,000 nT·hr, despite the absence of very strong $Int(aa_H) > 6,000$ nT·hr events during the 2012–2017 interval, characterized by a weaker geomagnetic activity than in 2003–2005. This demonstrates that the fit to 2003–2005 and 2012–2017 data given in Equation 1 remains reliable even over a subset of the total time interval, where only data from the same satellite are used. Figures 1b and 1c also show that during these two very different solar cycles, an $Int(aa_H)$ event of a given strength produces on average the same 10-day-integrated electron flux. Therefore, the parameterization by $Int(aa_H)$ obtained in Figure 1b can be used with a reasonable confidence to estimate time-integrated relativistic electron fluxes during other solar cycles.

Figure 1d finally shows that strong $Int(aa_H)$ events produce both high 10-day fluences (as seen in Figure 1b) and high time-averaged 2-MeV electron fluxes, often exceeding 10^6 e/cm^2/sr/MeV/s. After rescaling, the same best least-squares fit as in Figure 1b indeed provides a best fit to the time-averaged 2-MeV flux $\langle F \rangle$, as

$$\langle F \rangle \simeq (0.49566 \ln [Int (aa_H)] - 3.429) \cdot 10^6 \text{ e/cm}^2/\text{sr/MeV/s}. \tag{2}$$

It is worth emphasizing the logarithmic increase of the high time-averaged relativistic electron fluxes found in Figure 1b, especially for $Int(aa_H) > 4,000$ nT·hr. This may be due to the fact that the strongest $Int(aa_H)$ events often result from the impact on the magnetosphere of a long succession of solar wind structures (Tsurutani et al., 2006). Such a succession of solar wind disturbances generally leads to an initial dropout of relativistic electron flux, followed by a strong increase due to local chorus-driven electron acceleration and/or inward radial diffusion of electrons by ULF waves (Mourenas, Artemyev, & Zhang, 2019; Murphy et al., 2018; Ozeke et al., 2020; Thorne et al., 2013; Tsurutani et al., 2006). However, it can ultimately lead to a second dropout of electron flux in the few days following the $Int(aa_H)$ event, thereby efficiently limiting the maximum 10-day-integrated flux. Many events of comparatively smaller time-integrated flux in Figure 1b are probably partly due to such dropouts, produced by magnetopause shadowing, outward radial diffusion, or wave-induced electron precipitation (Boynton et al., 2017; Mourenas et al., 2016; Olifer et al., 2018; Pinto, Zhang, et al., 2020; Shprits et al., 2006; Su et al., 2016) in the days following the $Int(aa_H)$ period. It is exactly what happened just after the huge $Int(aa_H) = 18,800$ nT·hr Halloween superstorm of October 2003, when a subsequent solar flare impacted the magnetosphere on 4 November and led to a fast dropout, abruptly ending a period of enhanced electron flux only a few days after its start (Mourenas, Artemyev, & Zhang, 2019), resulting in an unexpectedly low 10-day-integrated flux level (indicated in Figure 1b).

2.3. Quantification of Model Performance

To take into account the presence of successive $Int(aa_H) > 1,400$ nT·hr events occurring in clusters during very active periods, we slightly refine our forecast model by simply assuming that the time-averaged flux at a given time is equal to the maximum flux given by the best fit in Equation 2 based on the strongest $Int(aa_H) > 1,400$ nT·hr event ending during the preceding 10 days. Comparisons between the modeled (or predicted) and measured 2-MeV electron fluxes are displayed in Figure 2 during various years. Henceforth we use daily-averaged 2-MeV electron fluxes from GPS satellites at $L \simeq 4.2$–4.4 in 2001–2011 (multiplied by a factor 2.5 as before) and from the Van Allen Probes near $L* \sim 4.5$ in 2012–2017. Qualitatively, Figure 2 shows that strong $Int(aa_H) > 1,400$ nT·hr events are good predictors (and precursors) of long-duration high 2-MeV electron flux peaks, which correspond to periods of high time-integrated fluxes. A similarly good qualitative agreement between model and data is found for the in-sample years 2003, 2015, and 2017 used to build the model, and for the out-of-sample years 2002 and 2010 (near solar maximum and solar minimum, respectively), demonstrating the predictive capacity of the model. Nearly all strong $Int(aa_H) > 1,400$ nT·hr events are indeed followed by periods of high time-integrated 2-MeV electron flux, while only a few of the highest long-duration flux peaks are not predicted.

Several points are worth emphasizing concerning the training and validation procedures used for our forecast model. First, this forecast model mainly aims at predicting the 10-day periods of highest time-integrated electron flux in a distant future, presumably during a different solar cycle. To have a sufficient amount of meaningful data in the training set used for building the model, the years 2003–2005, containing some of the highest relativistic electron fluxes recorded in the past 20 years, have been selected in Section 2.2, together with 2012–2017 during which Van Allen Probes data is available. Second, to provide a fair test of the forecasting ability of a model in a distant future, the validation data set should be as independent as possible of the training data set (e.g., Wang et al., 2020). The existence of significant temporal correlations in the *aa* index time series (similar to aa_H), corresponding to the 27-day recurrence period of solar activity and its multiples up to 108 days (Lockwood et al., 2019), and the well-known seasonal and yearly recurrences of geomagnetic storm patterns (Chapman, McIntosh, et al., 2020; Hathaway, 2015; Owens et al., 2021), suggest to use well-separated groups of years for the validation data set and the training data set, to exclude possible correlations between these two data sets. This led us to select for the validation data set the out-of-sample years 2001–2002 and 2006–2011, which are sufficiently distant from the in-sample years 2003–2005 and 2012–2017 of the training set. This way, moreover, a large majority ($\sim$2/3) of the validation data set belongs to solar cycle 23, whereas a large majority of the training data set belongs to solar cycle 24. This should ensure that our validation data set is as independent as possible of our training data set, and representative of a distant period belonging to a different solar cycle. This validation data set will represent a hard test for the forecast model, but it should better show its skill and possible limitations related to the evolution of solar activity over successive cycles. Alternatively, we could have selected training and
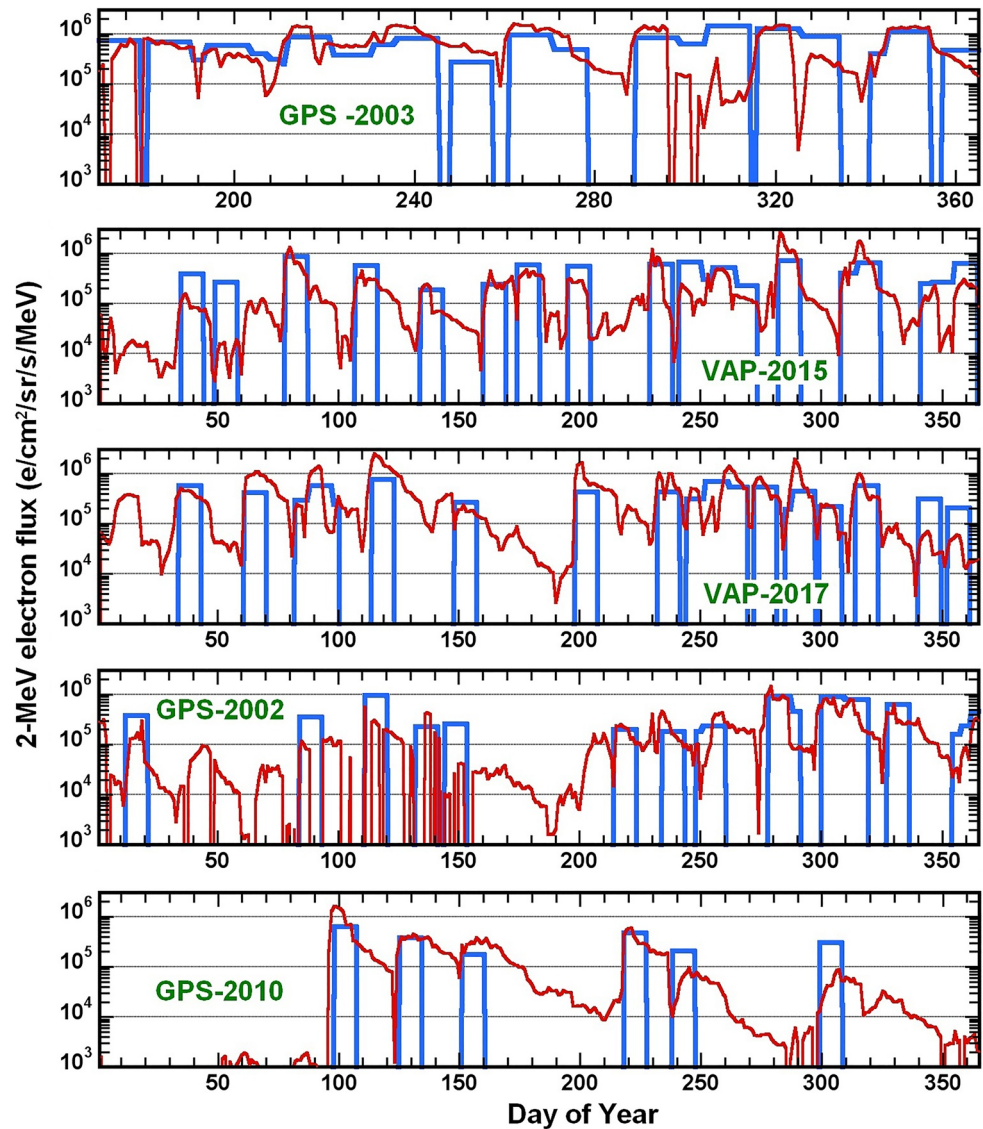
Figure 2. Daily-averaged 2-MeV electron flux (in e/cm²/sr/s/MeV) measured by the Van Allen Probes at $L^* \simeq 4.5$ in 2015 and 2017 and by Global Positioning System satellites at $L = 4.2 - 4.4$ (rescaled by a factor 2.5) in 2002, 2003, 2010 (red), together with daily-averaged 2-MeV electron flux predicted by the model in Equation 2 after strong $Int(aa_H) > 1,400$ nT·hr events (blue). Note that 2002 and 2010 are out-of-sample years.

validation data sets more uniformly distributed over 2001–2017, by selecting for example, odd and even years. But this would have introduced some correlations between nearby years from the two sets, and the validation set would not anymore have been mainly selected during a different solar cycle, making it less representative of an independent period located in a distant future.

Hereafter, several quantitative measures of the model performance are provided. To quantify the performance of the model on all days of each year, we first provide in Table 1 the True Skill Statistics (TSS), Heidke Skill Score (HSS), probability of detection (POD), probability of false detection (POFD), and false alarm ratio (FAR) (Woodcock, 1976; Zheng et al., 2019) of the $Int(aa_H)$ model for the prediction of days of high daily-averaged 2-MeV electron flux higher than $3.5 \cdot 10^5$ e/cm²/sr/MeV/s at $L^* \sim 4.5$. Such high daily fluxes have been observed only $\sim$15% of the time in 2001–2017. The POD (also known as hit rate) is equal to the fraction of actual events (here, days of high flux) correctly predicted, and the POFD is the fraction of non-event days that were incorrectly forecast as events (Zheng et al., 2019). The FAR is the fraction of predicted events that turn out to be non-events. The HSS is an equitable measure of categorical performance, based on the fraction of correctly

Table 1
Skill Scores and Accuracy Metrics of the Model Given by Equation 2 Together With a Threshold Int(aa_H) > 1,400 nT·hr for Predicting Individual Days of Average 2-MeV Electron Flux ⟨F⟩ > 3.5 · 10⁵ e/cm²/sr/MeV/s, and of the Model Given by Equation 1 Together With Int(aa_H) > 3,650 nT·hr for Predicting 10-Day Periods With ∫F dt > 8 · 10¹¹ e/cm²/sr/MeV, Near L ~ 4.5 in 2001–2017*

	2003–2005 and 2012–2017 (in-sample)	2001–2002 and 2006–2011 (out-of-sample)	2001–2017 (ALL)
Skill and accuracy of model (2) with $Int(aa_H) > 1{,}400$ nT·hr for predicting days of $\langle F\rangle > 3.5 \cdot 10^5$ e/cm²/sr/MeV/s			
TSS	0.60	0.60	0.62
HSS	0.58	0.43	0.55
POD	0.73	0.68	0.72
POFD	0.13	0.08	0.10
FAR	0.35	0.64	0.43
MEF	1.6	3.1	2.0
SSPB	−30%	+210%	+43%
Skill and accuracy of model (1) with $Int(aa_H) > 3{,}650$ nT·hr for predicting 10-day periods with $\int F\,dt > 8 \cdot 10^{11}$ e/cm²/sr/MeV			
TSS	0.78	0.96	0.81
HSS	0.61	0.42	0.56
POD	0.84	1.0	0.86
POFD	0.06	0.04	0.05
FAR	0.48	0.7	0.55
MEF	1.4	2.1	1.6
SSPB	−15%	+110%	−10%

Note. The predicted 10-day periods begin immediately after these $Int(aa_H)$ events, or at most 3–4 days after the start of longer-than-3-day $Int(aa_H)$ events when $Int(aa_H)$ has already reached 1,400 or 3,650 nT·hr.

predicted events after elimination of correct predictions arising purely from random chance (Heidke, 1926; Hogan & Mason, 2011). A null or negative HSS indicates no prediction skill, while a perfect model has a HSS of 1. The TSS (also known as Peirce Skill Score or Hanssen and Kuipers discriminant) is an equitable measure of categorical forecast performance similar to the HSS, based on the fraction of correctly predicted events after elimination of correct predictions arising from random chance, written as TSS = POD − POFD (Peirce, 1884; Woodcock, 1976). The TSS has the advantage of being unbiased with respect to event/non-event sample ratio, allowing more accurate comparisons between skill scores for different samples than the HSS (Woodcock, 1976). A forecast model is deemed excellent for TSS $\in$ (0.8, 1.0), good for TSS $\in$ (0.6, 0.8), fair for TSS $\in$ (0.4, 0.6), whereas TSS <0.2 indicates no predictive ability (Landis & Koch, 1977). However, the TSS may become less suitable than the HSS in the case of rare events with occurrences smaller than 1% (Doswell III et al., 1990).

To quantify the accuracy of the model during days belonging to the predicted 10-day periods of high flux following $Int(aa_H) > 1{,}400$ nT·hr events, we use two additional metrics. The Median Error Factor (MEF) between predicted and measured values is derived from the Median Symmetric Accuracy (MSA) introduced by Morley et al. (2018), under the form MEF = 1 + MSA/100 = $\exp(M(|\ln(Q_i)|))$, where M denotes the median and Q_i the ratio of modeled to measured values (Glauert et al., 2018; Morley et al., 2018). Half of the predictions remain within a factor of MEF of the data. The MEF and MSA metrics are especially appropriate for electron flux data spanning several orders of magnitude, and robust to the presence of outliers or bad data (Morley et al., 2018; Zheng et al., 2019). In Table 1, the MEF is first calculated for daily time-averaged 2-MeV electron fluxes during the 10-day periods predicted by the $Int(aa_H)$ model using Equation 2. For this evaluation, the predicted continuous 10-day periods of high time-integrated flux are therefore split into 10 separate days. This procedure has the advantage of increasing the number of points, but at the expense of somewhat underestimating the true skill of the model in predicting long-duration periods of high time-integrated flux, since the accuracy of such predictions only requires a good performance on the $n < 10$ days of highest flux within each 10-day period. The symmetric signed percentage bias (SSPB), defined as SSPB = 100 $Sgn(M(\ln(Q_i)))(\exp(|M(\ln(Q_i))|) − 1)$ (with Sgn the sign function), gives a robust and unbiased measure of the mean percentage error of the model (Morley et al., 2018).

An intrinsic limitation of the present model is that it cannot forecast low to moderate electron fluxes occurring outside of the predicted 10-day periods of high flux. But this should not be considered a problem in practice, since this type of model is mainly designed to provide advance warning of periods at risk of high time-integrated flux and TID.

The high probability of detection POD(days) $\simeq 0.68 - 0.72$ found in Table 1 shows that a large majority (~70%) of the days of high 2-MeV electron flux are correctly predicted, during both in-sample years (used to derive the model) and out-of-sample years. Note also that part of the non-predicted days of high flux simply belong to periods of high flux lasting longer than the assumed 10 days. The POFD is much smaller than the POD, with POFD(days) $\simeq 0.1$. This gives a True Skill Statistics TSS(days) $\simeq 0.62$, indicating a good forecast efficiency (Landis & Koch, 1977) for both the in-sample and out-of-sample data sets. The HSS(days) $\simeq 0.43 - 0.55$ is similarly elevated. It is worth emphasizing that the high TSS and HSS skill scores of the model correspond to days of high flux that are predicted, on average, 5 days in advance (between 0 and 10 days ahead).

The Median Error Factor is MEF(days) $\simeq 2$ for the full data set, showing a good accuracy for electron fluxes that vary by orders of magnitude (similar to the accuracy of much more sophisticated models; e.g., see Glauert et al., 2018). The MEF is larger (~3) during out-of-sample years than during in-sample years (~1.6). These different MEF values could stem from the different origins of the data at $L^* \sim 4.5$, since 63% of the in-sample data directly come from Van Allen Probe measurements at $L^* \sim 4.5$, whereas 100% of the out-of-sample data are inferred from measurements by GPS satellites at $L = 4.2$–4.4, slightly closer to the Earth, which introduces some additional uncertainty. Alternatively, these different MEF values could be due to the fact that 75% of the out-of-sample years (i.e., 2006–2011) take place around solar minimum, whereas all in-sample years belong to solar maximum or the declining phase of a solar cycle. The periods of solar maximum and declining phase are known to be more active in terms of geomagnetic storms and substorms (driven by coronal mass ejections or high-speed solar wind streams) than during solar minimum (Richardson et al., 2000; Tsubouchi & Omura, 2007; Tsurutani et al., 2006). The weakest storms present during solar minimum are expected to produce enhancements of 2-MeV electron flux and phase space density (PSD) at higher L than stronger storms (Tverskaya et al., 2003; Zhao & Li, 2013). The subsequent enhancement of 2-MeV electron flux at $L^* \sim 4.5$ may then occur after a significant delay, because these electrons need to be diffused radially inward to $L^* \sim 4.5$ by ULF waves from their initial PSD peak at $L^* \simeq 5$–6 (Li et al., 2014). This could explain the higher MEF, FAR and SSPB and the slightly lower HSS during out-of-sample years. But we checked that only a minor fraction of the days of false alarm (erroneously predicted) during out-of-sample years indeed correspond to prolonged peaks of 2-MeV electron flux starting sensibly later than their predicted start (like in September and November 2010 in Figure 2), and that most solar minimum flux peaks are as well predicted (like in April-August 2010 in Figure 2) as solar maximum flux peaks in 2013–2017. Actually, the majority of the days of false alarm (erroneously predicted) belong to 2001 and the first half of 2002, near solar maximum. This particular period corresponds to more frequent large geomagnetic storms (reaching $Dst < -130$ nT) than in 2003–2004 or 2013–2015 (near the next solar maximum). Such large storms, which are mostly caused by coronal mass ejections, are usually less efficient than weaker storms caused by corotating interaction regions and high-speed solar wind streams in producing prolonged peaks of 2-MeV electron flux at $L > 4$ in the outer belt (Miyoshi & Kataoka, 2011; Mourenas, Artemyev, & Zhang, 2019; Spasojevic, 2014). Nevertheless, it is worth noting that the SSPB of the model over the full data set is limited, with SSPB ~+43%, indicating daily predicted fluxes only slightly higher than measured fluxes.

The results in Figure 1b also suggest a model for predicting the 10-day periods of highest time-integrated 2-MeV electron flux $\int F\, dt > 8 \cdot 10^{11}$ e/cm²/sr/MeV near $L^* \sim 4.5$, given by Equation 1 together with a threshold $Int(aa_H) > 3{,}650$ nT·hr. Such 10-day periods of high time-integrated measured electron flux have been observed only ~4% of the time in 2001–2017, all of them after $Int(aa_H) > 1{,}900$ nT·hr events (and 63% after $Int(aa_H) > 5{,}000$ nT·hr events). Such 10-day periods are the most dangerous time intervals for satellites in terms of Total Ionizing Dose. Table 1 shows that both the TSS and HSS of this forecast model are high, with TSS = 0.81 and HSS = 0.56 (and POD = 0.86) over the full 2001–2017 data set, indicating its good to excellent efficiency. The TSS is even higher (0.96) during out-of-sample years, although the HSS is reduced to 0.42. This is simply due to the ~4 times more rare occurrences of such 10-day periods of high time-integrated measured flux during out-of-sample years: although this leads to a higher FAR, it is important to note that all real events are correctly predicted (POD = 1) during out-of-sample years, despite the fact that they are observed during solar minimum. Moreover, during both out-of-sample and in-sample years, roughly 65% of the false alarms still

correspond to 10-day periods of high time-integrated flux $\int F\,dt > 2.66 \cdot 10^{11}$ e/cm²/sr/MeV that represent an important risk. The moderate values of MEF $\simeq 1.6$ and 2.1 and SSPB $\simeq -10\%$ and $+110\%$ over the full data set and over out-of-sample years, respectively, show that the accuracy of this model is significantly higher than for predicting individual days of high flux. Increasing the threshold to $Int(aa_H) > 5{,}000$ nT·hr in the model still gives high TSS $\simeq$ HSS $\simeq 0.6$ over 2001–2017, but lower POD $\simeq 0.63$ and FAR $\simeq 0.4$. Despite the intrinsic limitations of these models, we underline that most of the days of high time-integrated flux inside the predicted 10-day periods are predicted more than 2–3 days in advance, based on the past history of only one ground-based geomagnetic index—without any needed input from satellites, contrary to many other forecast models (e.g., Boynton, Balikhin, et al., 2016; Chu et al., 2021; Glauert et al., 2021; Pires de Lima et al., 2020).

2.4. Physical Insights

How to explain the formation and duration of the long 10-day periods of high time-integrated relativistic electron flux analyzed in Section 2.1? Let us first examine their formation at $L^* \sim 4.5$. Whistler-mode chorus wave intensity is correlated with geomagnetic activity (Agapitov et al., 2015, 2018; Li et al., 2009; Meredith et al., 2003). Therefore, disturbed periods of high $Int(aa_H)$ can lead to a local quasi-linear energization of part of the abundant 100–300 keV electrons up to ~ 2 MeV via cyclotron resonance with intense ($B_w^2 \sim 100^2$ pT²) chorus waves, producing elevated fluxes of 2-MeV electrons over time scales of ~ 1–3 days in the low density region near $L^* \sim 4.5$ when $AE \sim 400$–600 nT (O. Agapitov et al., 2019; Horne et al., 2005; Thorne et al., 2013; Summers et al., 1998), roughly corresponding to an average $aa_H \sim 50$–100 nT (Rostoker, 1991). Including non-linear interactions with a population of more intense, but relatively short and phase-decorrelated, chorus wave packets should produce an only moderately faster electron acceleration over 1–3 days (Artemyev et al., 2021; Zhang et al., 2020). ULF wave intensity is also correlated with geomagnetic activity (Ozeke et al., 2014). Accordingly, a prolonged inward radial diffusion of electrons by ULF waves during disturbed $Int(aa_H)$ periods can also explain the formation of elevated 2-MeV electron fluxes near $L^* \sim 4.5$ over time scales of 1–3 days for $Kp \sim 4$–5 (Ozeke et al., 2014, 2020), roughly corresponding to an average $aa_H \sim 50$–100 nT. Chorus and ULF wave driven electron acceleration can occur separately, simultaneously, or in close succession (Li et al., 2014).

To investigate these potential sources of 2-MeV electrons, we use analytical estimates of the MLT-averaged and bounce-averaged quasi-linear energy diffusion rate $D_{EE}(AE)$ of 1-MeV electrons by chorus waves at $L = 4.5$ (Mourenas et al., 2014), provided by O. Agapitov et al. (2019) based on Van Allen Probes statistics of simultaneously measured local chorus wave frequency, amplitude, and plasma density. We show such diffusion rates $D_{EE}(AE)$ in Figure 3 during four time intervals corresponding to two moderate and two strong $Int(aa_H)$ events. The analytical electric field radial diffusion rate $D_{LL}(Kp)$ of electrons by ULF waves at $L = 4.5$ from Ozeke et al. (2014) is also shown, together with the time-integrated D_{EE} and D_{LL} during these events, denoted as $Int(D_{EE})$ and $Int(D_{LL})$.

The two moderate $Int(aa_H) \simeq 1{,}600$–1,950 nT·hr events took place on 6–7 March and 12–13 April 2016, in the wake of strong (min(Dst) ~ -100 nT) and moderate (min(Dst) ~ -60 nT) geomagnetic storms. These two moderate $Int(aa_H)$ events of continuously high $aa_H > 18$ nT last only ~ 1.5 days, mostly during the recovery phase of the first geomagnetic storm, corresponding to increases by factors of ~ 5–20 of the 2-MeV electron flux from its initial level (see Figure 3). However, on 14 April 2016, just after a first $Int(aa_H) \simeq 1{,}600$ nT·hr event, a second, much weaker $Int(aa_H) = 950$ nT·hr event occurred during a second moderate storm, corresponding to a second, weaker increase of 2-MeV flux by a factor ~ 2–3. The lowest flux reached during storm main phase is likely related to both the Dst effect of adiabatic electron motion (Kim & Chan, 1997) and magnetopause shadowing loss (Shprits et al., 2006; Turner et al., 2012). The final 2-MeV electron flux is $\sim(2.5 - 5.5) \times 10^5$ e/cm²/sr/MeV/s after these two moderate events.

The two strong $Int(aa_H) \simeq 3{,}900$ and 5,700 nT·hr events took place on 8–10 May, and 24–27 October 2016. They correspond to a strong (min(Dst) ~ -95 nT) storm and a moderate (min(Dst) ~ -65 nT) storm, respectively. These $Int(aa_H)$ events start during storm main phase and continue for ~ 2 days during the recovery phase. They correspond to similar enhancements (by factors of ~ 6–15) of the 2-MeV electron flux from its initial level as during the two moderate $Int(aa_H)$ events. However, the final 2-MeV electron flux is $\sim(2.4 - 2.9) \times 10^6$ e/cm²/sr/MeV/s, roughly ~ 5–10 times higher than after the two moderate $Int(aa_H)$ events.
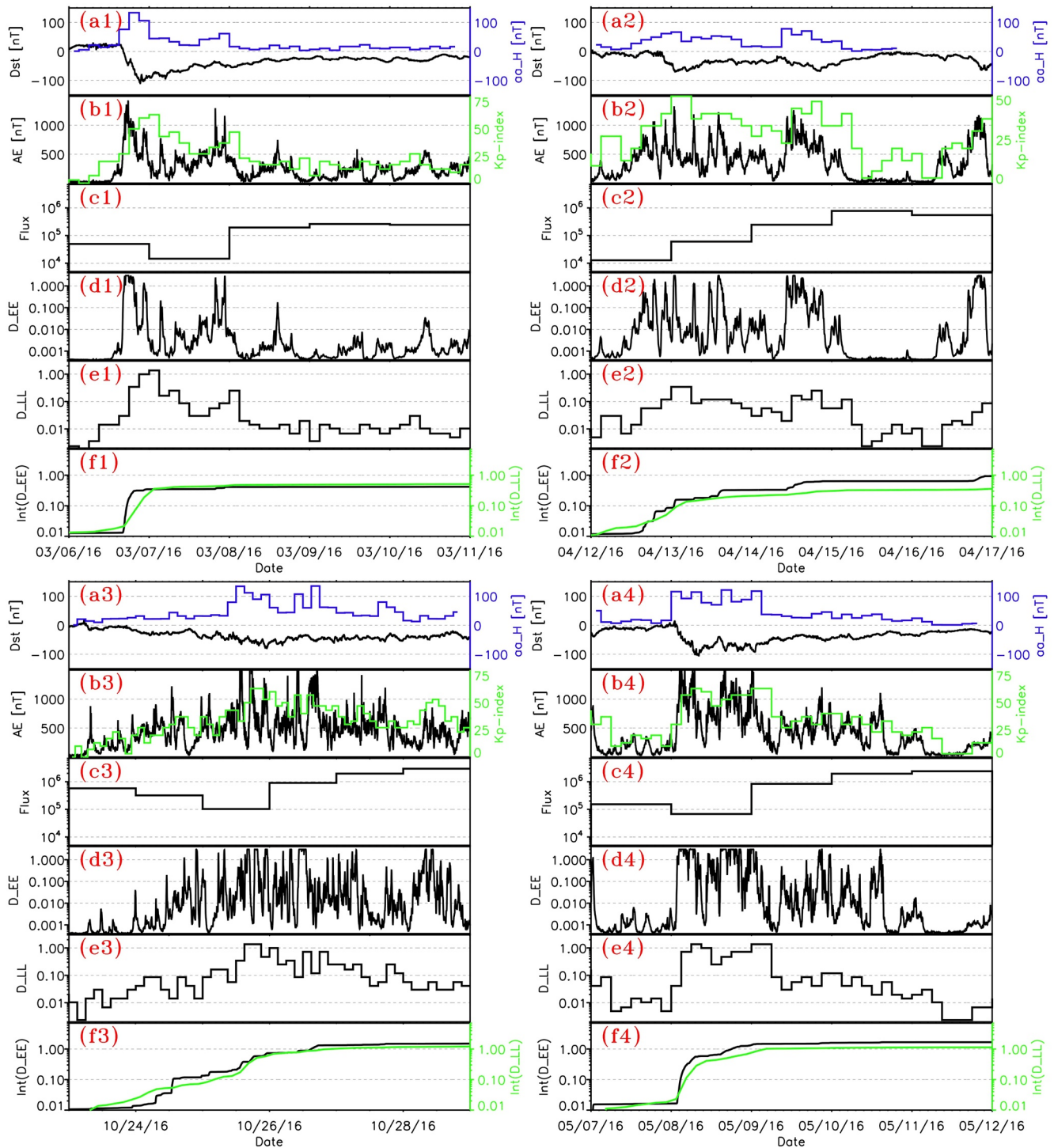
Figure 3. (a1–a4) aa_H (blue) and Dst (black) indices during four events in 2016 (top: moderate $Int(aa_H)$ events, bottom: strong $Int(aa_H)$ events). (b1–b4) AE (black) and $Kp \times 10$ (green). (c1–c4) Daily-averaged 2-MeV electron flux (in e/cm²/sr/s/MeV) measured by the Van Allen Probes at $L^* \simeq 4.5$. (d1–d4) Chorus wave-driven quasi-linear energy diffusion rate D_{EE} at $L = 4.5$. (e1–e4) ULF wave-driven radial diffusion rate D_{LL} at $L = 4.5$. (f1–f4) Time-integrated $Int(D_{EE})$ (black) and $Int(D_{LL})$ (green) during these intervals.

Therefore, the final 2-MeV electron flux appears to depend on the strength of these $Int(aa_H)$ events, as in statistical results displayed in Figure 1. Since the initial 2-MeV flux varies wildly from event to event, it suggests that the factor of increase of the 2-MeV flux from its initial level is not the appropriate parameter to classify these events. Instead, it is important to note that the final 2-MeV flux is usually deconnected from

its initial level by a steep dropout of electron flux occurring during storm main phase (or slightly earlier) due to magnetopause shadowing (Murphy et al., 2018; Turner et al., 2012, 2014). This is analogous to a 'hard reset' of the outer radiation belt (Turner et al., 2012, 2014). In such a case, the final 2-MeV flux level should be determined only by the time-integrated strength of the chorus-driven energization rate D_{EE} and of the radial diffusion rate D_{LL} of electrons, and by the flux level of 100–300 keV seed electrons injected from the plasma sheet that are accelerated to 2-MeV (Horne et al., 2005; Ozeke et al., 2020). But during storm main phase and recovery phase, the PSD of 100–300 keV injected electrons at $L^* \sim 4$–6 remains nearly identical (within a factor of $\sim$2) from event to event in 2012–2016 statistics (Murphy et al., 2018). Consequently, the final 2-MeV electron flux should be mostly determined by the time-integrated D_{EE} and D_{LL}. This conjecture is supported by the results in Figure 3, which show that the time-integrated $Int(D_{EE})$ and $Int(D_{LL})$ indeed reach higher levels during stronger $Int(aa_H)$ events and correspond to a higher final 2-MeV electron flux. Electron energization related to $Int(D_{EE})$ and $Int(D_{LL})$ apparently starts within a few hours of the flux dropout, as soon as injected 100–300 keV electrons can be accelerated without being immediately lost (Murphy et al., 2018; Turner et al., 2012). The higher cumulative rates of chorus wave-driven energization, $Int(D_{EE})$, and inward radial diffusion, $Int(D_{LL})$, of electrons during the stronger $Int(aa_H)$ events likely explain the formation of much higher fluxes of 2-MeV electrons.

Finally, the average $\sim$10-day duration of the prolonged periods of high 2-MeV electron flux still needs to be explained. The dynamic evolution of the background plasma density N_e during and after a storm likely plays an important role in both the formation and duration of such 10-day peaks of flux. Since the chorus wave-driven electron acceleration rate D_{EE} varies like $1/N_e^{3/2}$, the reduction of plasma density due to plasmasphere erosion during the initial phase of a storm can strongly increase 2-MeV electron flux near $L = 4.5$, just outside the plasmapause, rapidly forming such peaks of flux in one or a few days (Agapitov et al., 2019; Horne et al., 2005; Summers et al., 1998). After the storm, plasmasphere refilling takes place over a few more days, until the plasmapause usually reaches $L > 4.5$ again (Goldstein et al., 2014; O'Brien & Moldwin, 2003). Most of the 10 days of high flux occur during the following period, characterized by a weak geomagnetic activity, when the $L \sim 4.5$ region should be either just above or inside the plasmasphere. But, both outside and inside the plasmasphere, the lifetimes of 2-MeV electrons due to chorus or hiss wave-driven precipitation into the atmosphere are statistically longer than 10 days during such moderately active periods (Agapitov et al., 2020; Aryan et al., 2020; Mourenas et al., 2017), that is, insufficiently short to explain the ending of a flux peak in less than $\sim$10 days.

However, if electromagnetic ion cyclotron (EMIC) waves are simultaneously present at $L \sim 4.5$ inside high-density regions where they can resonantly interact with not-too-high energy electrons (Summers et al., 1998), either within the refilled plasmasphere or inside a remnant plasmaspheric plume, a combined pitch-angle scattering of 2-MeV electrons by both whistler-mode and EMIC waves (even at different local times) can reduce the lifetimes of 2-MeV electrons to less than 10 days (Li et al., 2007; Mourenas et al., 2016; Pinto et al., 2019; Zhang et al., 2017). This has been confirmed by electron lifetime measurements around $L = 4.5$ during weakly disturbed periods with $Kp \sim 1$–2 and $\langle AE \rangle < 200$ nT (Mourenas et al., 2017, 2021).

An alternative explanation for the relatively short duration of these 10-day periods of high flux could be the presence of a dropout of 2-MeV electron flux caused by magnetopause shadowing and electron outward radial diffusion by ULF waves toward the last closed drift shell (Boynton et al., 2017; Olifer et al., 2018; Ozeke et al., 2020; Pinto, Zhang, et al., 2020; Shprits et al., 2006). Although such dropouts statistically occur only after a median waiting time of $\sim$20 days at $L \sim 4.2$ (Boynton et al., 2017), the presence of a plasmaspheric plume during storm recovery (Goldstein et al., 2014) can facilitate the dropout by allowing an easier propagation of intense ULF waves to low $L \sim 4$ (Degeling et al., 2018). The shorter duration of these flux peaks at higher L (Mourenas, Artemyev, & Zhang, 2019) is also consistent with the more frequent occurrence of dropouts at higher L (Boynton, Mourenas, & Balikhin, 2016; Boynton et al., 2017). Therefore, a dropout due to magnetopause shadowing or strong electron precipitation by combined EMIC and whistler-mode waves, favored by the appearance of high density plasmaspheric regions at $L \geq 4.5$ during the late recovery phase of a storm, can probably account for the average $\sim$10-day duration of the analyzed periods of high time-integrated 2-MeV electron flux.
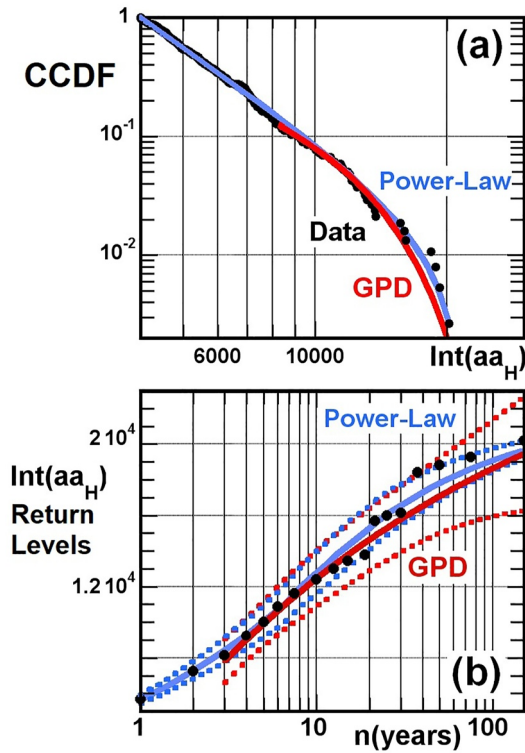
Figure 4. (a) Complementary to the Cumulative Distribution Function of $Int(aa_H)$ events (black circles), with corresponding best ML power-law (with upper-cutoff) fit (blue) and best ML Generalized Pareto Distribution (GPD) fit (red). (b) Return levels of observed $Int(aa_H)$ events (black circles) as a function of the period considered (in number of years), with GPD fit (red) and ML power-law fit (blue). Dotted lines show 95% confidence intervals (for clarity, not all data points are shown at $Int(aa_H) < 13,750$ nT·hr, all of them remaining within 95% confidence intervals).

3. Distribution of $Int(aa_H)$ Events in 1868–2017

3.1. Distribution of Extreme $Int(aa_H)$ Events

In the homogenized aa index, called aa_H, individual aa data points have been modified using time- and station-dependent scale factors to correct aa for secular changes (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018). However, some information has been lost through the initial logarithmic discretization of the aa index and the existence of an upper limit $K = 9$ for the related K index (Chapman, Horne, & Watkins, 2020; Mayaud, 1980). Consequently, the maximum value of aa_H reached during an event does not quantify well the real extremum of activity, and Extreme Value Theory (Coles, 2001; Tsubouchi & Omura, 2007) is not directly applicable to individual aa_H values (Chapman, Horne, & Watkins, 2020). Nevertheless, Chapman, Horne, and Watkins (2020) have shown that yearlong averages of the largest 0.5% of aa_H values (corresponding to averages over 14–15 aa_H points) can still be used as reliable estimates of extreme activity. In a similar way, we use here as an estimate of extreme activity an integrated parameter, $Int(aa_H)$, equal to the sum of all aa_H values (corresponding to ~10–30 aa_H points) recorded during an event where aa_H remains continuously above 18 nT. This summation over time, as long as aa_H remains above a low threshold, should partly correct for both the discretization of the aa_H index and the upper limit on K and aa_H. Indeed, only a small number (at most one or two in general) of the aa_H values reach extremely high levels $aa_H > 450$ nT (or non-homogenized $aa > 450$ nT) during a given $Int(aa_H)$ event. This means that more than ~90% of the summed aa_H values during an $Int(aa_H)$ event correspond to a K index smaller than 8.2 (Mayaud, 1980) and should not be affected by the saturation $K \leq 9$ of the corresponding K index. Accordingly, the potential error on the integrated quantity $Int(aa_H)$ should remain small. In addition, $Int(aa_H)$ clearly has no pre-determined upper limit, since in principle it may increase indefinitely with the duration of integration of aa_H (i.e., with the duration of the event). This justifies the applicability of Extreme Value Theory for examining the probability of rare extreme $Int(aa_H)$ events (Coles, 2001).

In analogy to the central limit theorem, it has been shown that the exceedances over a threshold in a sample of N independent extreme events tend to follow a Generalized Pareto Distribution (GPD) for sufficiently high N and threshold values (Coles, 2001). A reliable Extreme Value Theory method therefore consists in fitting the tail of the distribution of exceedances of independent (by construction) $Int(aa_H)$ events over a well-chosen and sufficiently high threshold $\min[Int(aa_H)]$ by a GPD of the form

$$P_{GPD}(\xi, \sigma) = \frac{1}{\sigma}\left(1 + \frac{\xi}{\sigma} \cdot (Int(aa_H) - \min[Int(aa_H)])\right)^{-1-1/\xi} \tag{3}$$

for its probability distribution (Coles, 2001).

The appropriate threshold $\min[Int(aa_H)]$ for a reliable GPD fit corresponds to a sufficiently high range of $\min[Int(aa_H)]$, where a linear relationship should exist between mean exceedance $\langle Int(aa_H) - \min[Int(aa_H)]\rangle$ and $\min[Int(aa_H)]$, as well as nearly stable (constant) estimates of GPD parameters ξ and $\sigma^* = \sigma - \xi \cdot \min[Int(aa_H)]$ obtained via the Maximum Likelihood (ML) method (Coles, 2001; Love et al., 2015; Tsubouchi & Omura, 2007). The optimal (ξ, σ) usually correspond to the lowest appropriate threshold $\min[Int(aa_H)]$, because more events are taken into account (Coles, 2001). Here, it gives $\min[Int(aa_H)] = 8,400$ nT·hr (higher $\min[Int(aa_H)]$ give similar ξ and σ but larger uncertainties, see details in Appendix B), $\xi \simeq -0.213 \pm 0.33$, and $\sigma \simeq 4,260 \pm 1,880$, where minimum and maximum parameter values correspond to 95% confidence intervals calculated via the delta method (Coles, 2001). Figure 4a shows the Complementary to the Cumulative Distribution Function (CCDF) of extreme $Int(aa_H) > 4,260$ nT·hr events. The CCDF, also called "tail distribution", gives the fraction of events with

higher $Int(aa_H)$ than a given $Int(aa_H)$ value in abscissa. The GPD fit (red curve) is fairly close to the data (black points) in Figure 4a, with a maximum Kolmogorov-Smirnov distance $D \simeq 0.053$ between the fitted CCDF and the data (Clauset et al., 2009), corresponding to a p-value = 0.88. Accordingly, the hypothesis that extreme $Int(aa_H)$ events actually have such a GPD distribution cannot be confidently rejected at the $p \leq 0.05$ level, and this GPD fit appears plausible (Coles, 2001). In addition, ξ remains negative over a 90% confidence interval, indicating the likely presence of an upper limit $\max[Int(aa_H)] \approx 20{,}000-28{,}000$ nT·hr (a previous analysis of the $Int(aa)$ data set has yielded similar results, see Mourenas, Artemyev, Zhang, et al., 2019).

An alternative ML fit to the CCDF of $Int(aa_H)$ is searched in the form of a power-law yearly probability distribution with upper-cutoff $P_v[Int(aa_H)] = C \cdot \mathrm{H}(\max[Int(aa_H)] - Int(aa_H))/Int(aa_H)^\alpha$ (with H the Heaviside function and an upper-cutoff $\max[Int(aa_H)] \simeq 21{,}000$ nT·hr), as done before for time-integrated Dst events (Mourenas et al., 2018). For a threshold $Int(aa_H) > 4{,}000$ nT·hr (using lower thresholds yielded less good fits), we obtain $\alpha = 3.583$ and $C = 1.32 \cdot 10^{10}$, with a small maximum Kolmogorov-Smirnov distance $D = 0.043$ between fit and data, corresponding to a p-value of 0.5 for 376 data points. Accordingly, the hypothesis that extreme $Int(aa_H)$ events have such a power-law distribution with upper-cutoff cannot be confidently rejected at the $p \leq 0.05$ level (Coles, 2001). This ML power-law fit performs well in Figure 4a (see blue curve). The root-mean-squared error (RMSE) between the power-law fit and the CCDF of $Int(aa_H) > 4{,}000$ nT·hr is small, RMSE = 0.016, with a squared Pearson correlation coefficient $R^2 = 0.998$, indicative of a very good fit. We also checked that this ML power-law fit is only weakly sensitive to the choice of the upper-limit $\max[Int(aa_H)]$, since varying it by 5% changes CCDF values by less than 1% on average for the 10 highest $Int(aa_H)$ values, and much less at lower $Int(aa_H)$. The delta method applied to the ML estimate of α (Coles, 2001) gives a 95% confidence interval $3.26 < \alpha < 3.9$ (other methods give similar values, see Clauset et al., 2009). Therefore, this ML power-law fit appears plausible, and it remains accurate over a much wider $Int(aa_H)$ domain than the GPD fit. Interestingly, the peak aa_H magnitude of $aa_H > 40$ nT storms has also an approximately power-law distribution (Haines et al., 2019).

Physically, a power-law distribution of the most extreme $Int(aa_H)$ events could result from protracted periods of strong solar wind driving that compel the magnetosphere-ionosphere system to assume a particular self-organized critical configuration in nearly stable non-equilibrium (Aschwanden et al., 2016; Valdivia et al., 2013). A saturation process progressively more efficient from $Int(aa_H) \approx 6{,}000$ nT·hr to $Int(aa_H) = 20{,}000$ nT·hr could also limit the decrease of the probability to get very strong events as $Int(aa_H)$ increases toward its upper limit $\max[Int(aa_H)]$, potentially leading to an approximate power-law shape (Mourenas et al., 2018; Mourenas, Artemyev, Zhang, et al., 2019; Zhang & Du, 2010). Such a saturation could stem from various physical mechanisms, such as enhanced ring current dropout due to magnetopause shadowing and outward radial diffusion (Boynton et al., 2017; Turner et al., 2014), ionospheric feedback (Toledo-Redondo et al., 2021), or saturation of the solar wind-magnetosphere coupling during stronger events (Kivelson & Ridley, 2008; Lopez et al., 2010; Siscoe et al., 2002).

3.2. Return Levels of Extreme $Int(aa_H)$ Events

For risk assessment, it is useful to consider the n-year return level $Int(aa_H)_n$ of an extreme event—the expected $Int(aa_H)$ level exceeded once every n years. For extreme $Int(aa_H) \geq \min[Int(aa_H)] = 8{,}400$ nT·hr events in 1868–2017 having a GPD, it can be estimated as

$$Int(aa_H)_n \simeq \min(Int(aa_H)) + \frac{\sigma}{\xi}\left[\left(\frac{44n}{150}\right)^\xi - 1\right], \tag{4}$$

with $\xi = -0.213$ and $\sigma = 4{,}260$ (Coles, 2001). For the power-law distribution ML fit with upper cutoff, it is given by

$$Int(aa_H)_n = \left(\frac{150}{376n}\left[\min(Int(aa_H))^\beta - \max(Int(aa_H))^\beta\right] + \max(Int(aa_H))^\beta\right)^{1/\beta}, \tag{5}$$

with $\beta = 1 - \alpha = -2.583$ for $\alpha = 3.583$, a threshold $\min[Int(aa_H)] = 4{,}000$ nT·hr, and an assumed upper cutoff at $\max[Int(aa_H)] = 21{,}000$ nT·hr in rough agreement with GPD estimates (the negative $\xi = -0.213$ indicating the likely presence of an upper limit).

Figure 4b shows that the GPD and power-law with upper cutoff distribution fits in Equations 4 and 5 yield very similar return levels, increasing with the number n of years until they reach a similar upper limit $\max[Int(aa_H)] \simeq 20,000$ nT·hr. The corresponding estimated return levels are close to the observed return levels over 1,868–2017. All the observed return levels are found within the 95% confidence intervals of the GPD and power-law fits, except for one or two events (among $\max(n) = 150$ in total) located slightly outside them. The highest return level (for a 95% confidence interval) of the GPD fit remains smaller than ~22,500 nT·hr over 150 years. The upper limit estimate $\max[Int(aa_H)] \simeq (\min[Int(aa_H)] - \sigma/\xi) \approx 28,000$ nT·hr provided by Extreme Value Theory is ~40% higher than the 4 largest events observed in 1868–2017, which reached $Int(aa_H) = 18,400$, 18,800, 19,300, 20,200 nT·hr in April 1994, October 2003, May 1921 (Love et al., 2019), and November 1882 (Love, 2018), respectively. These four events can be considered as typical 1 in 70 ± 40 years to 1 in 200 ± 130 years events. They agree well with the power-law distribution fit with upper-cutoff in Figure 4b. All these results suggest that $Int(aa_H)$ events larger than 21,000–22,500 nT·hr are unlikely to be observed in the next 50–150 years without an important change in the solar wind behavior. Let us caution, however, that the above estimates of return levels are based on the $Int(aa_H)$ distribution recorded in 1868–2017. Any forecast on this basis must further assume that this 150-year distribution from the past will remain representative of future 10-year to 150-year distributions.

Unfortunately, no reliable estimate is available for the Carrington superstorm of September 1859, which has been estimated to have reached a slightly higher peak disturbance level in terms of Dst than the May 1921 superstorm (Cliver & Dietrich, 2013). Rough estimates (due to important data gaps) of the aa index during the Carrington storm available from the Helsinki magnetic observatory suggest a maximum daily-averaged value only ~15% higher than during four more recent storms, such as the November 1960 and March 1989 events (Nevanlinna, 2006). The estimated time-integrated Dst was also probably weaker during this event than during more recent events, due to a fast recovery (Mourenas et al., 2018). Accordingly, the Carrington 1859 superstorm might not have exceeded the maximum $Int(aa_H)$ levels reached in 1868–2017.

3.3. Return Levels of High Time-Integrated 2-MeV Electron Flux Periods

Figure 4b shows that the ML power-law fit from Equation 5 gives $Int(aa_H)_n$ return levels very close to (albeit slightly higher than) GPD fit values. In addition, the 10-day time-integrated 2-MeV electron flux increases only weakly (logarithmically) with $Int(aa_H)$ above 5,000 nT·hr in Figure 1b. Therefore, we can safely use the ML power-law fit to estimate the return levels of time-integrated 2-MeV electron flux near $L^* \sim 4.5$. Upper and lower bounds corresponding to a ~70% confidence interval are obtained by combining the maximum 95% confidence intervals of GPD and ML power-law fits with the 70% confidence interval of the best fit in Figure 1b. Such estimated return levels of 10-day time-integrated 2-MeV electron flux at $L^* \sim 4.5$ are displayed in Figure 5a. They are only weakly increasing with the number n of considered years (by a factor ~1.7 from 1 to 100 years). 100-year return levels of 10-day-integrated 2-MeV electron flux reach $1.25 \cdot 10^{12}$ e/cm²/sr/MeV at $L^* \sim 4.5$, with an upper bound estimated as $\sim 2.6 \cdot 10^{12}$ e/cm²/sr/MeV.

Based on a 2002–2016 statistics from INTEGRAL IREM calibrated to MagEIS fluxes from the Van Allen Probes, Meredith et al. (2017) have also provided the return levels of daily 2-MeV electron fluxes at $L^* \sim 4.25 - 4.75$, making use of Extreme Value Theory to extrapolate over a much longer 100-year interval. Since the 2.1 MeV electron flux from REPT is on average ~1.8 times larger than the corresponding MagEIS flux (Morley et al., 2016), and since such fluxes generally remain near their peak level for ~7–10 days at $L^* = 4.5$, we simply multiply by a factor 1.8 the daily flux return levels inferred from INTEGRAL IREM data and assume that they last 10 days to compare them with our results over 10-day periods. Figure 5a shows that such return levels inferred directly from INTEGRAL IREM statistics are in good agreement with our estimated return levels at $L^* \sim 4.5$. A 30-year Fokker-Planck simulation utilizing measured electron fluxes from GOES as boundary condition (Glauert et al., 2018) also gives a similar maximum 10-day-integrated 2-MeV electron flux at $L^* = 4.6$ over 30 years.

It is well known that space weather approximately follows the solar cycle, with more frequent geomagnetic storms at solar maximum than solar minimum (Hathaway, 2015), higher 2–6 MeV electron fluxes in the outer radiation belt during the declining phase than near solar minimum (Baker et al., 2004). The most extreme daily-averaged aa_H events are also more frequent during large solar cycles than small cycles (Owens et al., 2021). The average geomagnetic activity aa index is indeed correlated with sunspot number, albeit with significant scatter (Feynman, 1982). In Figure 5b, we find a high correlation ($R^2 = 0.87$) between the sum of $Int(aa_H) > 1,400$ nT·hr
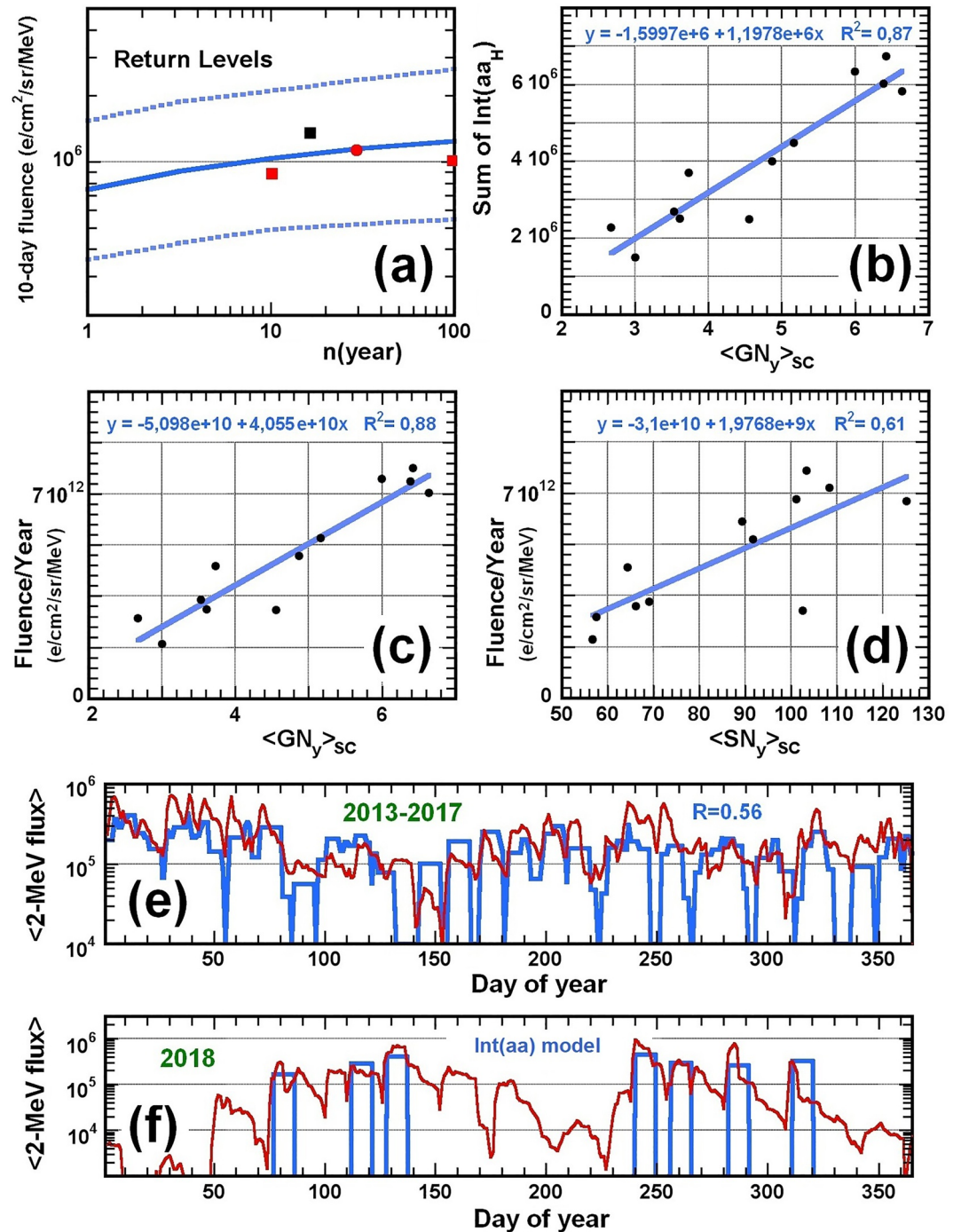
Figure 5. (a) Return levels of 10-day-integrated 2-MeV electron flux (here in units of 10^6 e/cm^2/sr/MeV) at $L^* \sim 4.5$ (solid blue) based on ML power-law fit in Equation 5, after a period of n years. Upper and lower bounds to these return levels are shown (dashed lines), as well as estimates based on a 14-year statistics from INTEGRAL IREM (Meredith et al., 2017) at $L^* = 4.25 - 4.75$ (red squares), the result (red circle) of a 30-year Fokker-Planck simulation at $L^* = 4.6$ (Glauert et al., 2018), and the maximum level in 2001–2017 from Van Allen Probes and GPS satellite data (black square). (b) Variation of the sum of $Int(aa_H) > 1,400$ nT·hr events as a function of the average yearly active-day sunspot Group Number $\langle GN_y \rangle_{SC}$ during solar cycles in 1868–1995, with best fit. (c) Yearly time-integrated 2-MeV electron flux (i.e., fluence) over 10-day periods following $Int(aa_H) > 1,400$ nT·hr events at $L^* \sim 4.5$ based on the model, as a function of $\langle GN_y \rangle_{SC}$, with best fit. (d) Same as (c) but as a function of solar cycle-averaged yearly International Sunspot Number $\langle SN_y \rangle_{SC}$ in 1868–2008. (e) Average daily 2-MeV electron flux at $L^* \sim 4.5$ from the $Int(aa_H)$ model (blue) versus daily 2-MeV electron flux measured by the Van Allen Probes in 2013–2017 (red), as a function of day of year. (f) Same as (e) for 2018, with model flux calculated from Equation 2 using $Int(aa)$ instead of $Int(aa_H)$.

values during a solar cycle and the corresponding average yearly active-day sunspot Group Number GN_y (Usoskin et al., 2016). This high correlation can be used to estimate relativistic electron fluence during past or future solar cycles, based on past (Clette et al., 2014; Usoskin et al., 2016) or predicted (Petrovay, 2010) sunspot numbers. Figure 5c shows that the yearly time-integrated 2-MeV electron flux at $L^* \sim 4.5$ estimated using the $Int(aa_H)$ model in Equation 1 varies almost linearly with the average GN_y from solar cycle to solar cycle, with a similarly high correlation coefficient ($R^2 = 0.88$). Figure 5d shows a similar, albeit smaller, correlation ($R^2 = 0.61$) with the average yearly International Sunspot Number SN_y (Clette et al., 2014). Such results suggest a potential new way of estimating 2-MeV electron fluence over the long term. However, comparisons with measurements would need to be performed over at least two different full solar cycles to validate this method. Moreover, a background of quiet-day relativistic electron fluence should be added to the predicted active-day fluence, although this quiet-day fluence represents less than 25%–35% of the total fluence during years of moderate to high geomagnetic activity (Mourenas, Artemyev, & Zhang, 2019).

In the same vein, we show in Figure 5e the average daily 2-MeV electron flux at $L^* \sim 4.5$ predicted by the $Int(aa_H)$ model and the daily 2-MeV electron flux measured by the Van Allen Probes in 2013–2017, as a function of day of year. There is a significant correlation ($R = 0.56$) between modeled and measured average daily fluxes, demonstrating the model's capacity for providing good estimates of statistical seasonal and weekly variations of flux and time-integrated electron flux.

Since the aa_H index is not (yet) available after 31 December 2017, it is important to check whether useful forecasts can still be provided after that date based on the proposed model, by simply using the non-homogenized aa index (as readily available as Kp) instead of aa_H in Equation 2. Such a substitution is a priori justified, because differences between individual aa and aa_H values are small (<5–15%) and vary in sign during an event (Lockwood, Chambodut, Barnard, Owens, & Clarke, 2018; Lockwood, Chambodut, Barnard, Owens, Clarke, & Mendel, 2018), leading to even smaller differences between the time-integrated aa_H and aa levels analyzed in the present study during strong events of continuously high $aa_H \geq 18$ nT (or $aa \geq 18$ nT). Figure 5f shows the daily 2-MeV electron flux at $L^* \sim 4.5$ measured by the Van Allen Probes in 2018 (an out-of-sample year), compared to the modeled flux during the 10-day periods that follow $Int(aa) > 1,400$ nT·hr events. The agreement between the $Int(aa)$ model and measurements in 2018 remains similarly good as the agreement between the $Int(aa_H)$ model and measurements in 2017 (an in-sample year) in Figure 2, with all the highest flux peaks exceeding $3 \cdot 10^5$ e/cm²/sr/MeV/s well reproduced by the model. This confirms that both the aa_H index and the aa index can be used to provide useful predictions of time-integrated 2-MeV electron flux.

4. Conclusions

We developed a predictive model of long-duration periods of high time-integrated 2-MeV electron flux near $L^* \sim 4.5$ deep inside the outer radiation belt, based on the significant correlation obtained between time-integrated electron flux measured by the Van Allen Probes and GPS satellites in 2001–2017 and a peak-over-threshold measure (denoted $Int(aa_H)$) of the preceding time-integrated homogenized aa_H geomagnetic index. An analysis of four different events shows that this correlation is likely due to a stronger cumulative chorus wave-driven acceleration of relativistic electrons and a stronger cumulative inward radial diffusion of such electrons by ULF waves during periods of higher time-integrated geomagnetic activity. The predictive ability of the model has been assessed during both individual days and continuous 10-day periods, using various skill scores and accuracy metrics, attesting its good efficiency in 2001–2017.

A key point of the present model is that the days of high 2-MeV electron flux (and time-integrated flux) are predicted, on average, 5 days in advance (between 0 and 10 days ahead). This suggests that this simple but efficient model could be used as a complement to other, more sophisticated forecast models (e.g., Glauert et al., 2021; Pires de Lima et al., 2020) for providing a far-ahead warning of dangerous long-duration periods of particularly elevated time-integrated relativistic electron flux that should be examined in more detail. In addition, the present model only relies on the past history of a unique aa_H ground-based geomagnetic index. We showed that the aa index can be used instead of aa_H in the model, thanks to small differences between time-integrated $Int(aa_H)$ and $Int(aa)$ parameters. Although this model has been developed for 2-MeV electron flux because it represents an important risk for spacecraft electronics, electron flux energy spectra are often coherent over ~1.5–3 MeV during prolonged periods of high flux, suggesting that the model could be applied over this whole energy range, using

typical energy spectral shapes observed during such periods of high flux. At lower and higher energy, however, the presence of electron injections and the different time scales of electron acceleration would require to develop other specific models.

Return levels of 2-MeV electron flux have been provided based on Extreme Value analysis of time-integrated geomagnetic activity over 1868–2017. Let us caution, however, that such forecasts need to assume that the analyzed 150-year distribution of aa_H is representative of future distributions and that the same correlations between peaks of $Int(aa_H)$ and peaks of time-integrated 2-MeV electron flux will remain valid in the future. Notwithstanding these limitations, the provided maximum return levels of 10-day time-integrated 2-MeV electron flux roughly agree with previous independent studies based on statistical analyses of <20-year data sets of measured electron fluxes. The maximum 10-day-integrated flux of 2-MeV electrons at $L^* \sim 4.5$ is estimated as $\simeq 1.25 \cdot 10^{12}$ e/cm^2/sr/MeV for a 1 in 100 years event. Extreme Value theory suggests that $Int(aa_H)$ events probably have an upper limit, which could be related to various physical mechanisms of adaptation of the magnetosphere to strong and prolonged solar wind impacts. Steep dropouts of electron flux due to magnetopause shadowing should also limit time-integrated 2-MeV electron flux between $L \sim 4.2$ and geostationary orbit, with much more frequent dropouts occurring at higher L (Boynton et al., 2017; Boynton, Mourenas, & Balikhin, 2016).

Finally, we found a high correlation between the $Int(aa_H)$ measure of time-integrated geomagnetic activity averaged over each solar cycle and the similarly averaged International Sunspot Number and active-day sunspot Group Number. This suggests that forecasts of time-integrated relativistic electron flux during future solar cycles might be obtained based on predictions of future solar activity.

Appendix A: Examples of $Int(aa_H)$ Events

Two typical examples of $Int(aa_H) > 1,400$ nT·hr events are displayed in Figure A1 between 12 March and 10 May 2015. The top panel shows the variation of the aa_H index with time, while the middle panel shows the level of $Int(aa_H)$, integrated over time as long as aa_H remains continuously higher than 18 nT. The bottom panel shows the corresponding daily 2-MeV electron flux measured by the Van Allen Probes at $L^* = 4.5$. Blue rectangles denote the 10-day periods of high flux considered in this study, which start at the end of each $Int(aa_H) > 1,400$ nT·hr event. Black arrows indicate the start and end of actual flux peaks, determined at 1/2 and 1/3 of the maximum flux.
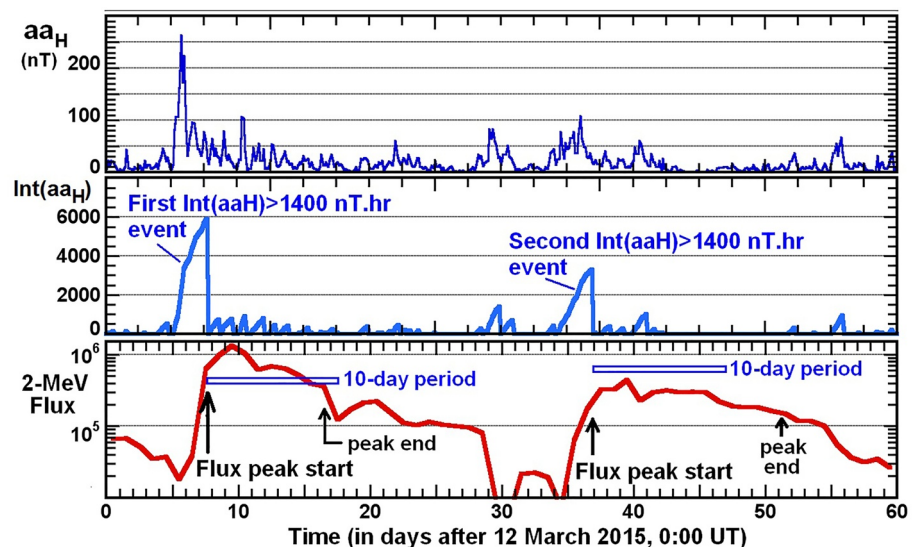

Figure A1. Two $Int(aa_H) > 1,400$ nT·hr events in March–May 2015. (top) aa_H (in nT) index as a function of time. (Middle) $Int(aa_H)$ (in nT·hr), calculated by summing hourly aa_H values as long as aa_H remains continuously ≥18 nT. (Bottom) daily 2-MeV electron flux (in e/cm^2/s/sr/MeV) measured by the Van Allen Probes at $L^* \sim 4.5$. The start and end of flux peaks are indicated by black arrows. The considered 10-day periods of high time-integrated flux starting immediately after each $Int(aa_H) > 1,400$ nT·hr event are marked by blue rectangles.

Appendix B: Determination of GPD Fit Parameters

Here, we provide a brief description of the derivation of appropriate scale and shape parameters of the GPD distribution fit shown in Figure 4. The appropriate threshold $\min[Int(aa_H)]$ for a reliable GPD fit must correspond to (a) a sufficiently high $\min[Int(aa_H)]$ in the distribution tail (usually within the 60 upper points), (b) a linear relationship between mean exceedance $\langle Int(aa_H) - \min[Int(aa_H)] \rangle$ and $\min[Int(aa_H)]$ over some range of $\min[Int(aa_H)]$, and (c) nearly stable (constant) estimates of GPD parameters ξ and $\sigma^* = \sigma - \xi \cdot \min[Int(aa_H)]$ obtained via the Maximum Likelihood (ML) method over the same $\min[Int(aa_H)]$ range (Coles, 2001; Tsubouchi & Omura, 2007). In Figures B1a and B1b, the mean exceedance (black points) shows some evidence of linearity over the same domain $\min[Int(aa_H)] \simeq 8{,}400\text{--}10{,}150$ nT·hr (corresponding to the 44 upper data points) where ξ and σ^* are simultaneously nearly constant in Figures B1a and B1b.
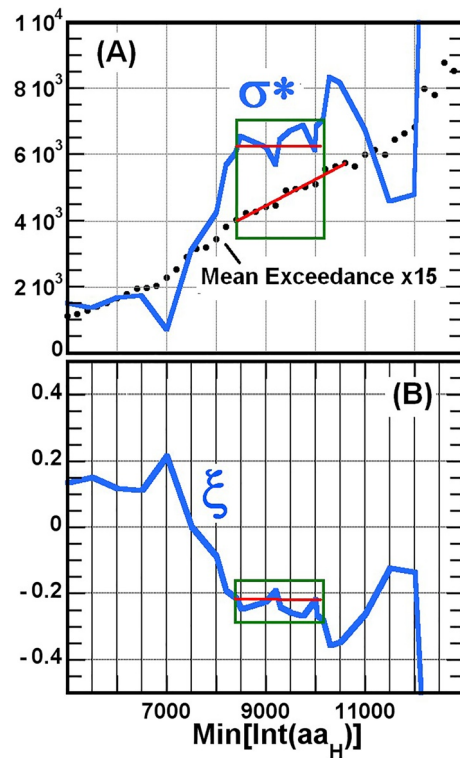

Figure B1. (a) Mean exceedance $\langle Int(aa_H) - \min[Int(aa_H)] \rangle$ (black points) and σ^* (blue) as a function of the threshold $\min[Int(aa_H)]$ in 1868–2017. (b) ξ GPD parameter (blue) as a function of the threshold $\min[Int(aa_H)]$ in 1868–2017. Green boxes show the domain of simultaneously nearly constant σ^* and ξ and linearly increasing mean exceedance.

The optimal (ξ, σ) parameters generally correspond to the lowest appropriate threshold $\min[Int(aa_H)]$, because more events are then taken into account, which decreases uncertainties (Coles, 2001). Here, it gives $\min[Int(aa_H)] = 8{,}400$ nT·hr, corresponding to a shape parameter $\xi \simeq -0.213 \pm 0.33$ and a scale parameter $\sigma \simeq 4{,}260 \pm 1{,}880$ (a higher $\min[Int(aa_H)] = 9{,}300$ nT·hr gives similar $\xi = -0.24 \pm 0.398$ and $\sigma \simeq 4{,}230 \pm 2160$ values but with larger uncertainties), where minimum and maximum parameter values correspond to 95% confidence intervals calculated via the delta method (Coles, 2001). This GPD fit is shown in Figure 4a.

Data Availability Statement

Van Allen Probes REPT electron flux data (REL03 L2) is available from NASA at https://cdaweb.gsfc.nasa.gov/cgi-bin/eval1.cgi, and LANL CXD data of GPS electron flux (2017 release) is available from NOAA at https://www.ngdc.noaa.gov/stp/space-weather/satellite-data/satellite-systems/gps/. The aa_H index can be retrieved at

https://www.swsc-journal.org/articles/swsc/olm/2018/01/swsc180022/swsc180022-2-olm.txt. The *aa* index is calculated by ISGI collaborating institutes from data collected at magnetic observatories and available at http://isgi.unistra.fr. The OMNI data of *AE*, *Kp*, and *Dst* are available from the GSFC/SPDF OMNIWeb interface at https://omniweb.gsfc.nasa.gov. The active-day Sunspot Group Number and International Sunspot Number V2.0 are available from SILSO (Royal Observatory of Belgium, Brussels) at http://sidc.be/silso/.

References

Agapitov, O., Mourenas, D., Artemyev, A., Claudepierre, S. G., Hospodarsky, G., & Bonnell, J. W. (2020). Lifetimes of relativistic electrons as determined from plasmaspheric hiss scattering rates statistics: Effects of ω_{pe}/Ω_{ce} and wave frequency dependence on geomagnetic activity. *Geophysical Research Letters*, *47*(13), e2020GL088052. https://doi.org/10.1029/2020GL088052

Agapitov, O., Mourenas, D., Artemyev, A., Hospodarsky, G., & Bonnell, J. W. (2019). Time scales for electron quasi-linear diffusion by lower-band chorus waves: The effects of ω_{pe}/Ω_{ce} dependence on geomagnetic activity. *Geophysical Research Letters*, *46*(12), 6178–6187. https://doi.org/10.1029/2019GL083446

Agapitov, O. V., Artemyev, A. V., Mourenas, D., Mozer, F. S., & Krasnoselskikh, V. (2015). Empirical model of lower band chorus wave distribution in the outer radiation belt. *Journal of Geophysical Research*, *120*(12), 10. https://doi.org/10.1002/2015JA021829

Agapitov, O. V., Mourenas, D., Artemyev, A. V., Mozer, F. S., Hospodarsky, G., Bonnell, J., & Krasnoselskikh, V. (2018). Synthetic empirical chorus wave model from combined Van Allen Probes and Cluster statistics. *Journal of Geophysical Research*, *123*(1), 297–314. https://doi.org/10.1002/2017JA024843

Artemyev, A. V., Neishtadt, A. I., Vasiliev, A. A., & Mourenas, D. (2021). On a transitional regime of electron resonant interaction with whistler-mode waves in inhomogeneous space plasma. *Physical Review E*, *104*(5), 055203. https://doi.org/10.1103/PhysRevE.104.055203

Aryan, H., Agapitov, O. V., Artemyev, A., Mourenas, D., Balikhin, M. A., Boynton, R., & Bortnik, J. (2020). Outer radiation belt electron lifetime model based on combined Van Allen Probes and Cluster VLF measurements. *Journal of Geophysical Research*, *125*(8), e28018. https://doi.org/10.1029/2020JA028018

Aschwanden, M. J., Crosby, N., Dimitropoulou, M., Georgoulis, M. L., Hergarten, S., McAteer, J., et al. (2016). 25 years of self-organized criticality—Solar and astrophysics. *Space Science Reviews*, *198*(1–4), 47–166. https://doi.org/10.1007/s11214-014-0054-6

Baker, D. N., Blake, J. B., Callis, L. B., Cummings, J. R., Hovestadt, D., Kanekal, S., et al. (1994). Relativistic electron acceleration and decay time scales in the inner and outer radiation belts: SAMPEX. *Geophysical Research Letters*, *21*(6), 409–412. https://doi.org/10.1029/93GL03532

Baker, D. N., Kanekal, S. G., Hoxie, V. C., Batiste, S., Bolton, M., Li, X., et al. (2013). The Relativistic Electron-Proton Telescope (REPT) instrument on board the Radiation Belt Storm Probes (RBSP) spacecraft: Characterization of Earth's radiation belt high-energy particle populations. *Space Science Reviews*, *179*(1–4), 337–381. https://doi.org/10.1007/s11214-012-9950-9

Baker, D. N., Kanekal, S. G., Li, X., Monk, S. P., Goldstein, J., & Burch, J. L. (2004). An extreme distortion of the Van Allen belt arising from the 'Hallowe'en' solar storm in 2003. *Nature*, *432*(7019), 878–881. https://doi.org/10.1038/nature03116

Baker, D. N., McPherron, R. L., Cayton, T. E., & Klebesadel, R. W. (1990). Linear prediction filter analysis of relativistic electron properties at 6.6 RE. *Journal of Geophysical Research*, *95*(A9), 15133–15140. https://doi.org/10.1029/JA095iA09p15133

Boyd, A. J., Reeves, G. D., Spence, H. E., Funsten, H. O., Larsen, B. A., Skoug, R. M., et al. (2019). RBSP-ECT combined spin-averaged electron flux data product. *Journal of Geophysical Research*, *124*(11), 9124–9136. https://doi.org/10.1029/2019JA026733

Boynton, R. J., Balikhin, M. A., Sibeck, D. G., Walker, S. N., Billings, S. A., & Ganushkina, N. (2016). Electron flux models for different energies at geostationary orbit. *Space Weather*, *14*(10), 840–860. https://doi.org/10.1002/2016SW001506

Boynton, R. J., Mourenas, D., & Balikhin, M. A. (2016). Electron flux dropouts at Geostationary Earth Orbit: Occurrences, magnitudes, and main driving factors. *Journal of Geophysical Research*, *121*(9), 8448–8461. https://doi.org/10.1002/2016JA022916

Boynton, R. J., Mourenas, D., & Balikhin, M. A. (2017). Electron flux dropouts at $L \sim 4.2$ from global positioning system satellites: Occurrences, magnitudes, and main driving factors. *Journal of Geophysical Research*, *122*, 11. https://doi.org/10.1002/2017JA024523

Chapman, S. C., Horne, R. B., & Watkins, N. W. (2020). Using the index over the last 14 solar cycles to characterize extreme geomagnetic activity. *Geophysical Research Letters*, *47*(3), e2019GL086524. https://doi.org/10.1029/2019GL086524

Chapman, S. C., McIntosh, S. W., Leamon, R. J., & Watkins, N. W. (2020). Quantifying the solar cycle modulation of extreme space weather. *Geophysical Research Letters*, *47*(11), e2020GL087795. https://doi.org/10.1029/2020GL087795

Chen, Y., Carver, M. R., Morley, S. K., & Hoover, A. S. (2021). Determining ionizing doses in medium Earth orbits using long-term GPS particle measurements. In *2021 IEEE Aerospace Conference (50100)* (pp. 1–21). https://doi.org/10.1109/AERO50100.2021.9438516

Chu, X., Ma, D., Bortnik, J., Tobiska, W. K., Cruz, A., Bouwer, S. D., et al. (2021). Relativistic electron model in the outer radiation belt using a neural network approach. *Space Weather*, *19*(12), e2021SW002808. https://doi.org/10.1029/2021SW002808

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Powerlaw distributions in empirical data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111

Clette, F., Svalgaard, L., Vaquero, J. M., & Cliver, E. W. (2014). Revisiting the sunspot number. A 400-year perspective on the solar cycle. *Space Science Reviews*, *186*(1–4), 35–103. https://doi.org/10.1007/s11214-014-0074-2

Cliver, E. W., & Dietrich, W. F. (2013). The 1859 space weather event revisited: Limits of extreme activity. *Journal of Space Weather and Space Climate*, *3*, A31. https://doi.org/10.1051/swsc/2013053

Cochran, Y., Buchner, S. P., Chen, D., Kim, H. S., LaBel, K. A., Oldham, T. R., et al. (2009). Total ionizing dose and displacement damage compendium of candidate spacecraft electronics for NASA. In *Proceedings of 2009 IEEE Radiation Effects Data Workshop, Quebec City, Canada* (pp. 25–31). https://doi.org/10.1109/REDW.2009.5336318

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

Degeling, A. W., Rae, I. J., Watt, C. E. J., Shi, Q. Q., Rankin, R., & Zong, Q.-G. (2018). Control of ULF wave accessibility to the inner magnetosphere by the convection of plasma density. *Journal of Geophysical Research*, *123*(2), 1086–1099. https://doi.org/10.1002/2017JA024874

Doswell, C. A., III, Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, *5*(4), 576–585. https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2

Eastwood, J. P., Biffis, E., Hapgood, M. A., Green, L., Bisi, M. M., Bentley, R. D., et al. (2017). The economic impact of space weather: Where do we stand? *Risk Analysis*, *37*(2), 206–218. https://doi.org/10.1111/risa.12765

Ecoffet, R. (2013). Overview of in-orbit radiation induced spacecraft anomalies. *IEEE Transactions on Nuclear Science*, *60*(3), 1791–1815. https://doi.org/10.1109/TNS.2013.2262002

Feynman, J. (1982). Geomagnetic and solar wind cycles, 1900–1975. *Journal of Geophysical Research*, *87*(A8), 6153–6162. https://doi.org/10.1029/JA087iA08p06153

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for rank correlation coefficients. *I. Biometrika*, *44*(3/4), 470–481. https://doi.org/10.2307/2332878

Glauert, S. A., Horne, R. B., & Kirsch, P. (2021). Evaluation of SaRIF high-energy electron reconstructions and forecasts. *Space Weather*, *19*(12), e2021SW002822. https://doi.org/10.1029/2021SW002822

Glauert, S. A., Horne, R. B., & Meredith, N. P. (2018). A 30 year simulation of the outer electron radiation belt. *Journal of Geophysical Research*, *16*(10), 1498–1522. https://doi.org/10.1029/2018SW001981

Goldstein, J., de Pascuale, S., Kletzing, C., Kurth, W., Genestreti, K. J., Skoug, R. M., et al. (2014). Simulation of Van Allen Probes plasmapause encounters. *Journal of Geophysical Research*, *119*(9), 7464–7484. https://doi.org/10.1002/2014JA020252

Haines, C., Owens, M. J., Barnard, L., Lockwood, M., & Ruffenach, A. (2019). The variation of geomagnetic storm duration with intensity. *Solar Physics*, *294*(11), 154. https://doi.org/10.1007/s11207-019-1546-z

Hathaway, D. H. (2015). The solar cycle. *Living Reviews in Solar Physics*, *12*, 1–87. https://doi.org/10.1007/lrsp-2015-4

Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler Stockholm*, *8*(4), 301–349. https://doi.org/10.1080/20014422.1926.11881138

Hogan, R. J., & Mason, I. B. (2011). Deterministic forecasts of binary events. In I. T. Jolliffe & D. B Stephenson (Eds.), *Forecast verification: A practitioner's guide in atmospheric Science* (pp. 31–60). John Wiley. https://doi.org/10.1002/9781119960003.ch3

Horne, R. B., Thorne, R. M., Glauert, S. A., Albert, J. M., Meredith, N. P., & Anderson, R. R. (2005). Timescale for radiation belt electron acceleration by whistler mode chorus waves. *Journal of Geophysical Research*, *110*(A3), 3225. https://doi.org/10.1029/2004JA010811

Iucci, N., Levitin, A. E., Belov, A. V., Eroshenko, E. A., Ptitsyna, N. G., Villoresi, G., et al. (2005). Space weather conditions and spacecraft anomalies in different orbits. *Space Weather*, *3*(1), 1001. https://doi.org/10.1029/2003SW000056

Kim, H.-J., & Chan, A. A. (1997). Fully adiabatic changes in storm time relativistic electron fluxes. *Journal of Geophysical Research*, *102*(A10), 22107–22116. https://doi.org/10.1029/97JA01814

Kivelson, M. G., & Ridley, A. J. (2008). Saturation of the polar cap potential: Inference from Alfvén wing arguments. *Journal of Geophysical Research*, *113*(A5), A05214. https://doi.org/10.1029/2007JA012302

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Li, W., & Hudson, M. K. (2019). Earth's Van Allen radiation belts: From discovery to the Van Allen Probes era. *Journal of Geophysical Research*, *124*(11), 8319–8351. https://doi.org/10.1029/2018JA025940

Li, W., Shprits, Y. Y., & Thorne, R. M. (2007). Dynamic evolution of energetic outer zone electrons due to wave-particle interactions during storms. *Journal of Geophysical Research*, *112*(A10), 10220. https://doi.org/10.1029/2007JA012368

Li, W., Thorne, R. M., Angelopoulos, V., Bortnik, J., Cully, C. M., Ni, B., et al. (2009). Global distribution of whistler-mode chorus waves observed on the THEMIS spacecraft. *Geophysical Research Letters*, *36*(9), 9104. https://doi.org/10.1029/2009GL037595

Li, W., Thorne, R. M., Ma, Q., Ni, B., Bortnik, J., Baker, D. N., et al. (2014). Radiation belt electron acceleration by chorus waves during the 17 March 2013 storm. *Journal of Geophysical Research*, *119*(6), 4681–4693. https://doi.org/10.1002/2014JA019945

Lockwood, M., Bentley, S. N., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., et al. (2019). The development of a space climatology: 3. Models of the evolution of distributions of space weather variables with timescale. *Space Weather*, *17*(1), 180–209. https://doi.org/10.1029/2018SW002017

Lockwood, M., Chambodut, A., Barnard, L. A., Owens, M. J., & Clarke, E. (2018). A homogeneous aa index: 1. Secular variation. *Journal of Space Weather and Space Climate*, *8*, A58. https://doi.org/10.1051/swsc/2018044

Lockwood, M., Chambodut, A., Barnard, L. A., Owens, M. J., Clarke, E., & Mendel, V. (2018). A homogeneous aa index: 2. Hemispheric asymmetries and the equinoctial variation. *Journal of Space Weather and Space Climate*, *8*, A53. https://doi.org/10.1051/swsc/2018038

Lopez, R. E., Bruntz, R., Mitchell, E. J., Wiltberger, M., Lyon, J. G., & Merkin, V. G. (2010). Role of magnetosheath force balance in regulating the dayside reconnection potential. *Journal of Geophysical Research*, *115*(A12), A12216. https://doi.org/10.1029/2009JA014597

Love, J. J. (2018). The electric storm of November 1882. *Space Weather*, *16*(1), 37–46. https://doi.org/10.1002/2017SW001795

Love, J. J., Hayakawa, H., & Cliver, E. W. (2019). Intensity and impact of the New York Railroad superstorm of May 1921. *Space Weather*, *17*(8), 1281–1292. https://doi.org/10.1029/2019SW002250

Love, J. J., Rigler, E. J., Pulkkinen, A., & Riley, P. (2015). On the lognormality of historical magnetic storm intensity statistics: Implications for extreme-event probabilities. *Geophysical Research Letters*, *42*(16), 6544–6553. https://doi.org/10.1002/2015GL064842

Mayaud, P. N. (1980). Derivation, meaning, and use of geomagnetic indices. In *Washington DC American Geophysical Union Geophysical Monograph Series* (Vol. 22, p.607). https://doi.org/10.1029/GM022

Meredith, N. P., Cain, M., Horne, R. B., Thorne, R. M., Summers, D., & Anderson, R. R. (2003). Evidence for chorus-driven electron acceleration to relativistic energies from a survey of geomagnetically disturbed periods. *Journal of Geophysical Research*, *108*(A6), 1248. https://doi.org/10.1029/2002JA009764

Meredith, N. P., Horne, R. B., Isles, J. D., & Rodriguez, J. V. (2015). Extreme relativistic electron fluxes at geosynchronous orbit: Analysis of GOES 2 MeV electrons. *Space Weather*, *13*(3), 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*(7), 917–933. https://doi.org/10.1002/2017SW001651

Miyoshi, Y., & Kataoka, R. (2011). Solar cycle variations of outer radiation belt and its relationship to solar wind structure dependences. *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(1), 77–87. https://doi.org/10.1016/j.jastp.2010.09.031

Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, *16*(1), 68–88. https://doi.org/10.1002/2017SW001669

Morley, S. K., Sullivan, J. P., Carver, M. R., Kippen, R. M., Friedel, R. H. W., Reeves, G. D., & Henderson, M. G. (2017). Energetic particle data from the global positioning system constellation. *Space Weather*, *15*(2), 283–289. https://doi.org/10.1002/2017SW001604

Morley, S. K., Sullivan, J. P., Henderson, M. G., Blake, J. B., & Baker, D. N. (2016). The Global Positioning System constellation as a space weather monitor: Comparison of electron measurements with Van Allen Probes data. *Space Weather*, *14*(2), 76–92. https://doi.org/10.1002/2015SW001339

Mourenas, D., Artemyev, A. V., Agapitov, O. V., & Krasnoselskikh, V. (2014). Consequences of geomagnetic activity on energization and loss of radiation belt electrons by oblique chorus waves. *Journal of Geophysical Research*, *119*(4), 2775–2796. https://doi.org/10.1002/2013JA019674

Mourenas, D., Artemyev, A. V., Ma, Q., Agapitov, O. V., & Li, W. (2016). Fast dropouts of multi-MeV electrons due to combined effects of EMIC and whistler mode waves. *Geophysical Research Letters*, *43*(9), 4155–4163. https://doi.org/10.1002/2016GL068921

Mourenas, D., Artemyev, A. V., & Zhang, X.-J. (2018). Statistics of extreme time-integrated geomagnetic activity. *Geophysical Research Letters*, *45*(2), 502–510. https://doi.org/10.1002/2017GL076828

Mourenas, D., Artemyev, A. V., & Zhang, X.-J. (2019). Impact of significant time-integrated geomagnetic activity on 2-MeV electron flux. *Journal of Geophysical Research: Space Physics*, *124*(6), 4445–4461. https://doi.org/10.1029/2019JA026659

Mourenas, D., Artemyev, A. V., & Zhang, X.-J. (2020). Dynamical properties of peak and time-integrated geomagnetic events inferred from sample Entropy. *Journal of Geophysical Research: Space Physics*, *125*(2), e27599. https://doi.org/10.1029/2019JA027599

Mourenas, D., Artemyev, A. V., Zhang, X.-J., & Angelopoulos, V. (2019). Extreme time-integrated geomagnetic activity: Ap index statistics. *ArXiv*. arXiv:1901.09823.

Mourenas, D., Artemyev, A. V., Zhang, X.-J., Angelopoulos, V., Tsai, E., & Wilkins, C. (2021). Electron lifetimes and diffusion rates inferred from ELFIN measurements at low altitude: First results. *Journal of Geophysical Research: Space Physics*, *126*(11), e2021JA029757. https://doi.org/10.1029/2021JA029757

Mourenas, D., Ma, Q., Artemyev, A. V., & Li, W. (2017). Scaling laws for the inner structure of the radiation belts. *Geophysical Research Letters*, *44*(7), 3009–3018. https://doi.org/10.1002/2017GL072987

Murphy, K. R., Watt, C. E. J., Mann, I. R., Rae, I. J., Sibeck, D. G., Boyd, A. J., et al. (2018). The global statistical response of the outer radiation belt during geomagnetic storms. *Geophysical Research Letters*, *45*(9), 3783–3792. https://doi.org/10.1002/2017GL076674

Nevanlinna, H. (2006). A study on the great geomagnetic storm of 1859: Comparisons with other storms in the 19th century. *Advances in Space Research*, *38*(2), 180–187. https://doi.org/10.1016/j.asr.2005.07.076

O'Brien, T. P., & Moldwin, M. B. (2003). Empirical plasmapause models from magnetic indices. *Geophysical Research Letters*, *30*(4), 1152. https://doi.org/10.1029/2002GL016007

Olifer, L., Mann, I. R., Morley, S., Ozeke, L. G., & Choi, S. (2018). On the role of last closed drift shell dynamics in driving fast losses and Van Allen radiation belt extinction. *Journal of Geophysical Research*, *123*(5), 3692–3703. https://doi.org/10.1029/2018JA025190

Owens, M. J., Lockwood, M., Barnard, L. A., Scott, C. J., Haines, C., & Macneil, A. (2021). Extreme space-weather events and the solar cycle. *Solar Physics*, *296*(5), 82. https://doi.org/10.1007/s11207-021-01831-3

Ozeke, L. G., Mann, I. R., Dufresne, S., Olifer, L., Morley, S., Claudepierre, S., et al. (2020). Rapid outer radiation belt flux dropouts and fast acceleration during the March 2015 and 2013 storms: The role of ULF wave transport from a dynamic outer boundary. *Journal of Geophysical Research*, *125*(2), e2019JA027179. https://doi.org/10.1029/2019JA027179

Ozeke, L. G., Mann, I. R., Murphy, K. R., Jonathan Rae, I., & Milling, D. K. (2014). Analytic expressions for ULF wave radiation belt radial diffusion coefficients. *Journal of Geophysical Research*, *119*(3), 1587–1605. https://doi.org/10.1002/2013JA019204

Peirce, C. S. (1884). The numerical measure of the success of predictions. *Science*, *4*(93), 453–454. https://doi.org/10.1126/science.ns-4.93.453.b

Petrovay, K. (2010). Solar cycle prediction. *Living Reviews in Solar Physics*, *7*, 6. https://doi.org/10.12942/lrsp-2010-6

Pinto, V. A., Bortnik, J., Moya, P. S., Lyons, L. R., Sibeck, D. G., Kanekal, S. G., et al. (2020). Radial response of outer radiation belt relativistic electrons during enhancement events at geostationary orbit. *Journal of Geophysical Research*, *125*(5), e27660. https://doi.org/10.1029/2019JA027660

Pinto, V. A., Mourenas, D., Bortnik, J., Zhang, X.-J., Artemyev, A. V., Moya, P. S., & Lyons, L. R. (2019). Decay of ultra-relativistic remnant belt electrons through scattering by plasmaspheric hiss. *Journal of Geophysical Research*, *124*(7), 5222–5233. https://doi.org/10.1029/2019JA026509

Pinto, V. A., Zhang, X.-J., Mourenas, D., Bortnik, J., Artemyev, A. V., Lyons, L. R., & Moya, P. S. (2020). On the confinement of Ultrarelativistic electron remnant belts to low L-shells. *Journal of Geophysical Research*, *125*(3), e2019JA027469. https://doi.org/10.1029/2019JA027469

Pires de Lima, R., Chen, Y., & Lin, Y. (2020). Forecasting megaelectron-volt electrons inside Earth's outer radiation belt: PreMevE 2.0 based on supervised machine learning algorithms. *Space Weather*, *18*(2), e2019SW002399. https://doi.org/10.1029/2019SW002399

Richardson, I. G., Cliver, E. W., & Cane, H. V. (2000). Sources of geomagnetic activity over the solar cycle: Relative importance of coronal mass ejections, high-speed streams, and slow solar wind. *Journal of Geophysical Research*, *105*(A8), 18203–18213. https://doi.org/10.1029/1999JA000400

Riley, P., Baker, D., Liu, Y. D., Verronen, P., Singer, H., & Gudel, M. (2018). Extreme space weather events: From cradle to Grave. *Space Science Reviews*, *214*(1), 21. https://doi.org/10.1007/s11214-017-0456-3

Rostoker, G. (1991). A quantitative relationship between AE and Kp. *Journal of Geophysical Research*, *96*(A4), 5853–5857. https://doi.org/10.1029/90JA02752

Shprits, Y. Y., Thorne, R. M., Friedel, R., Reeves, G. D., Fennell, J., Baker, D. N., & Kanekal, S. G. (2006). Outward radial diffusion driven by losses at magnetopause. *Journal of Geophysical Research*, *111*(A11), 11214. https://doi.org/10.1029/2006JA011657

Siscoe, G. L., Crooker, N. U., & Siebert, K. D. (2002). Transpolar potential saturation: Roles of the region 1 current system and solar wind ram pressure. *Journal of Geophysical Research*, *107*(A10), 1321. https://doi.org/10.1029/2001JA009176

Spasojevic, M. (2014). Statistical analysis of ground-based chorus observations during geomagnetic storms. *Journal of Geophysical Research*, *119*(10), 8299–8317. https://doi.org/10.1002/2014JA019975

Stassinopoulos, E. G., & Raymond, J. P. (1988). The space radiation environment for electronics. *Proceedings of the IEEE*, *76*(11), 1423–1442. https://doi.org/10.1109/5.90113

Su, Z., Gao, Z., Zhu, H., Li, W., Zheng, H., Wang, Y., et al. (2016). Nonstorm time dropout of radiation belt electron fluxes on 24 September 2013. *Journal of Geophysical Research*, *121*(7), 6400–6416. https://doi.org/10.1002/2016JA022546

Summers, D., Thorne, R. M., & Xiao, F. (1998). Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. *Journal of Geophysical Research*, *103*(A9), 20487–20500. https://doi.org/10.1029/98JA01740

Thorne, R. M., Li, W., Ni, B., Ma, Q., Bortnik, J., Chen, L., et al. (2013). Rapid local acceleration of relativistic radiation-belt electrons by magnetospheric chorus. *Nature*, *504*(7480), 411–414. https://doi.org/10.1038/nature12889

Toledo-Redondo, S., André, M., Aunai, N., Chappell, C. R., Dargent, J., Fuselier, S. A., et al. (2021). Impacts of ionospheric ions on magnetic reconnection and Earth's magnetosphere dynamics. *Reviews of Geophysics*, *59*(3), e2020RG000707. https://doi.org/10.1029/2020RG000707

Tsubouchi, K., & Omura, Y. (2007). Long-term occurrence probabilities of intense geomagnetic storm events. *Space Weather*, *5*(12), S12003. https://doi.org/10.1029/2007SW000329

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Guarnieri, F. L., Gopalswamy, N., Grande, M., et al. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research*, *111*(A7), A07S01. https://doi.org/10.1029/2005JA011273

Tsyganenko, N. A. (1989). A magnetospheric magnetic field model with a warped tail current sheet. *Planatary Space Science*, *37*(1), 5–20. https://doi.org/10.1016/0032-0633(89)90066-4

Tsyganenko, N. A., & Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *Journal of Geophysical Research*, *110*(A3), A03208. https://doi.org/10.1029/2004JA010798

Turner, D. L., Angelopoulos, V., Morley, S. K., Henderson, M. G., Reeves, G. D., Li, W., et al. (2014). On the cause and extent of outer radiation belt losses during the 30 September 2012 dropout event. *Journal of Geophysical Research: Space Physics*, *119*(3), 1530–1540. https://doi.org/10.1002/2013JA019446

Turner, D. L., Shprits, Y., Hartinger, M., & Angelopoulos, V. (2012). Explaining sudden losses of outer radiation belt electrons during geomagnetic storms. *Nature Physics*, *8*(3), 208–212. https://doi.org/10.1038/nphys2185

Tverskaya, L. V., Pavlov, N. N., Blake, J. B., Selesnick, R. S., & Fennell, J. F. (2003). Predicting the L-position of the storm-injected relativistic electron belt. *Advances in Space Research*, *31*(4), 1039–1044. https://doi.org/10.1016/S0273-1177(02)00785-8

Usoskin, I. G., Kovaltsov, G. A., Lockwood, M., Mursula, K., Owens, M., & Solanki, S. K. (2016). A new calibrated sunspot group series since 1749: Statistics of active day fractions. *Solar Physics*, *291*(9–10), 2685–2708. https://doi.org/10.1007/s11207-015-0838-1

Valdivia, J. A., Rogan, J., Munoz, V., Toledo, B., & Stepanova, M. (2013). The magnetosphere as a complex system. *Advances in Space Research*, *51*(10), 1934–1941. https://doi.org/10.1016/j.asr.2012.04.004

Walton, S. D., Forsyth, C., Rae, I. J., Watt, C. E. J., Thompson, R. L., Horne, R. B., et al. (2021). Cross-L* coherence of the outer radiation belt during storms and the role of the plasmapause. *Journal of Geophysical Research*, *126*(10), e29308. https://doi.org/10.1029/2021JA029308

Wang, C., Rosen, I. G., Tsurutani, B. T., Verkhoglyadova, O. P., Meng, X., & Mannucci, A. J. (2020). Medium-range forecasting of solar wind: A case study of building regression model with space weather forecast testbed (SWFT). *Space Weather*, *18*(12), e2019SW002433. https://doi.org/10.1029/2019SW002433

Woodcock, F. (1976). The evaluation of yes/no forecasts for scientific and administrative purposes. *Monthly Weather Review*, *104*(10), 1209–1214. https://doi.org/10.1175/1520-0493(1976)104(1209:TEOYFF)2.0.CO;2

Zhang, L., & Du, Z. (2010). On the reflected Geometric brownian motion with two barriers. *Intelligent Information Management*, *2*(04), 295–298. https://doi.org/10.4236/iim.2010.23034

Zhang, X. J., Agapitov, O., Artemyev, A. V., Mourenas, D., Angelopoulos, V., Kurth, W. S., et al. (2020). Phase decoherence within intense chorus wave packets constrains the efficiency of nonlinear resonant electron acceleration. *Geophysical Research Letters*, *47*(20), e89807. https://doi.org/10.1029/2020GL089807

Zhang, X.-J., Mourenas, D., Artemyev, A. V., Angelopoulos, V., & Thorne, R. M. (2017). Contemporaneous EMIC and whistler mode waves: Observations and consequences for MeV electron loss. *Geophysical Research Letters*, *44*(16), 8113–8121. https://doi.org/10.1002/2017GL073886

Zhao, H., Baker, D. N., Li, X., Malaspina, D. M., Jaynes, A. N., & Kanekal, S. G. (2019). On the acceleration mechanism of ultrarelativistic electrons in the center of the outer radiation belt: A statistical study. *Journal of Geophysical Research*, *124*(11), 8590–8599. https://doi.org/10.1029/2019JA027111

Zhao, H., & Li, X. (2013). Inward shift of outer radiation belt electrons as a function of Dst index and the influence of the solar wind on electron injections into the slot region. *Journal of Geophysical Research*, *118*(2), 756–764. https://doi.org/10.1029/2012JA018179

Zheng, D. J., Ganushkina, N. Y., Jiggens, P., Jun, I., Meier, M., Minow, J. I., et al. (2019). Space radiation and plasma effects on satellites and aviation: Quantities and metrics for tracking performance of space weather environment models. *Space Weather*, *17*(10), 1384–1403. https://doi.org/10.1029/2018SW002042