Pick, L., Effenberger, F., Zhelavskaya, I., Korte, M. (2019): A Statistical Classifier for Historical Geomagnetic Storm Drivers Derived Solely From Ground-Based Magnetic Field Measurements. - *Earth and Space Science*, *6*, 10, pp. 2000—2015.

DOI: http://doi.org/10.1029/2019EA000726

Earth and Space Science

RESEARCH ARTICLE
10.1029/2019EA000726

A Statistical Classifier for Historical Geomagnetic Storm Drivers Derived Solely From Ground-Based Magnetic Field Measurements

Leonie Pick[1,2,3], **Frederic Effenberger**[1], **Irina Zhelavskaya**[1,2], **and Monika Korte**[1]

[1]Helmholtz Centre Potsdam, GFZ German Research Center for Geosciences, Potsdam, Germany, [2]University of Potsdam, Faculty of Science, Potsdam, Germany, [3]Jacobs University, Physics and Earth Sciences, Bremen, Germany

Key Points:
- Ground-level historical (1930–2015) geomagnetic storm characteristics are used for an automated classification of the storm drivers
- Based on machine learning 7,546 storm drivers are predicted skillfully and reliably in terms of two classes (ICMEs and C/SIRs)
- The results and underlying methods are fully reproducible and made available via a Python notebook and a conveniently prepared input data set

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Pick,
leonie.pick@gfz-potsdam.de

Citation:
Pick, L., Effenberger, F., Zhelavskaya, I., Korte, M. (2019). A statistical classifier for historical geomagnetic storm drivers derived solely from ground-based magnetic field measurements. *Earth and Space Science*, *6*, 2000–2015, https://doi.org/10.1029/2019EA000726

Received 27 MAY 2019
Accepted 17 SEP 2019
Accepted article online 16 OCT 2019
Published online 28 OCT 2019

Abstract Solar wind observations show that geomagnetic storms are mainly driven by interplanetary coronal mass ejections (ICMEs) and corotating or stream interaction regions (C/SIRs). We present a binary classifier that assigns one of these drivers to 7,546 storms between 1930 and 2015 using ground-based geomagnetic field observations only. The input data consists of the long-term stable Hourly Magnetospheric Currents index alongside the corresponding midlatitude geomagnetic observatory time series. This data set provides comprehensive information on the global storm time magnetic disturbance field, particularly its spatial variability, over eight solar cycles. For the first time, we use this information statistically with regard to an automated storm driver identification. Our supervised classification model significantly outperforms unskilled baseline models (78% accuracy with 26[19]% misidentified interplanetary coronal mass ejections [corotating or stream interaction regions]) and delivers plausible driver occurrences with regard to storm intensity and solar cycle phase. Our results can readily be used to advance related studies fundamental to space weather research, for example, studies connecting galactic cosmic ray modulation and geomagnetic disturbances. They are fully reproducible by means of the underlying open-source software (Pick, 2019, http://doi.org/10.5880/GFZ.2.3.2019.003)

Plain Language Summary The Earth's magnetic field reaches out into space where it constantly interacts with the solar wind, a stream of charged particles from the Sun. Geomagnetic storms occur when discontinuities in the solar wind disturb the geomagnetic field, possibly causing failures of, for example, electricity transmission and satellite communications. In order to mitigate the socioeconomic risk, a better understanding of these processes is necessary. So far, the two main storm drivers have been identified from satellite observations: interplanetary coronal mass ejections and corotating or stream interaction regions. However, quantitative investigations require as many storm events as possible from multiple decades with varying solar conditions. We present an innovative method that classifies the drivers of 7,546 storms since 1930, using ground-based magnetic field measurements only. This effectively increases the sample size, because geomagnetic observatories were operational long before the space era. Our results are directly applicable to other current space weather studies.

1. Introduction

Understanding cause and effect of solar processes and geomagnetic disturbances is crucial in order to mitigate our socioeconomic vulnerability to space weather (e.g., Oughton et al., 2018). Measurements of the magnetic field disturbances at ground level (δB) lead to the discovery that geomagnetic storms tend to be either weak and periodic, recurring after ≈27 days (one synodic solar rotation), or strong and sporadic (Greaves & Newton, 1929; Maunder, 1904). However, it was not until the beginning of the space age that stream interaction regions (SIRs) and interplanetary coronal mass ejections (ICMEs) were identified as major storm drivers (both reviewed by Kilpua et al., 2017). Their geoeffectiveness depends primarily on a sustained southward directed Z component ($B_{sw,z}$) of the interplanetary magnetic field (IMF, B_{sw}) with regard to the ecliptic pole (Rostoker & Fälthammar, 1967). SIRs are regions of compressed plasma that typically form when tenuous, fast solar wind streams from coronal holes (Krieger et al., 1973) collide with preceding dense, slow streams from the streamer belt (Feldman et al., 1981). If the coronal hole survives long enough, the SIR corotates with the Sun and is called a corotating interaction region (CIR), causing

recurrent geomagnetic disturbances (Snyder et al., 1963; Wilcox & Ness, 1965). On the other hand, ICMEs are the interplanetary manifestations of transient magnetized plasma ejections from the Sun, causing sporadic geomagnetic disturbances (Gosling, 1993; Gosling et al., 1975).

Today, SIRs and ICMEs are identified based on characteristic signatures in solar wind plasma and IMF observations from the Wind, ACE (e.g., Jian et al., 2006a, 2006b), and, since 2015, DISCOVR spacecraft near the Sun-Earth L1 Lagrangian Point. L1 is located at $\approx 1.5 \times 10^6$ km from the Earth toward the Sun. There, the slow-fast stream interfaces of SIRs are marked by a proton density drop and a rise of solar wind speed (v_{sw}), proton temperature (T_p), and plasma beta (β), accompanied by large $|B_{\mathrm{sw}}|$ fluctuations. ICMEs are often associated with a preceding shock and a sheath region between the shock and the ejecta, which can be "magnetic clouds." These are marked by a slow rotation of B_{sw} with enhanced field magnitude and reduced variability, as well as low T_p and β. Due to their ambiguity a combination of these signatures is required for a reliable classification.

The type of geomagnetic disturbance provoked by SIRs and ICMEs differs according to the response system. In this study, we focus on hourly δB measurements from geomagnetic observatories at low and middle latitudes. It is well known that the Dst index (Sugiura, 1964), that is, the symmetric part of the horizontal disturbance field ($\delta B_{\boldsymbol{H,sym}}$), responds more intensely to ICMEs than to SIRs, which is commonly attributed to differences in ring current (RC) intensity (e.g., Borovsky & Denton, 2006). At the same time, it is known that $\delta B_{\boldsymbol{H}}$ exhibits a pronounced asymmetry ($\delta B_{\boldsymbol{H,asy}}$), particularly between dawn and dusk (e.g., Akasofu & Chapman, 1964), which is quantified by the ASY-H index (Iyemori, 1990). This is associated with the partial ring current (PRC) system superposing the symmetric RC on the nightside (e.g., Kamide & Fukushima, 1971). While the contributions of different magnetospheric current systems, that is, cross-tail current, RC, and PRC, to $\delta B_{\boldsymbol{H,asy}}$ are studied extensively (e.g., Dubyagin et al., 2014), $\delta B_{\boldsymbol{H,asy}}$ has not been tested for its statistical C/SIR versus ICME discrimination capability.

To this end, we demonstrate that the long-term record of δB can be exploited beyond the current state of the art. This is done by means of a binary classifier that assigns driver classes $Y \in \{0, 1\}$, 0: CIRs or SIRs, 1: ICMEs, to geomagnetic storms from 1930–2015. Although the vast majority of strong events falls into one of these two categories (cf. Table 1 of Richardson & Cane, 2012), the binary nature of the classifier is a simplification that does not appropriately account for mixed, slow solar wind or unclear drivers. This is particularly relevant for the relatively weak events we incorporate in our analysis (section 2.2). However, the fact that we can predict whether a storm event is clearly ICME driven or not (mostly C/SIRs) from ground-level data alone is already a noteworthy result.

Covering 85 years, the event detection method should consider the ≈11-year solar cycle (SC) variation of geomagnetic activity. A Dst-based selection is inappropriate, because the index suffers from a baseline error that causes an underestimation of SC variability (e.g., Lühr & Maus, 2010; Temerin & Li, 2015). Therefore, we use a different δB-based index, which is morphologically similar to Dst but represents the low-frequency variations more robustly: the Hourly Magnetospheric Currents (HMC) index (Pick et al., 2019).

Our results advance the understanding of how SIR and ICME structures are typically imprinted at ground level and how their occurrence varies over eight SCs. They can contribute to other studies linking long-term geomagnetic activity measurements to solar wind evolution (e.g., Martini et al., 2015; Mursula et al., 2004) but can also assist studies in related areas, for example, the long-term correlation between galactic cosmic ray modulation and geomagnetic activity (e.g., Dumbović et al., 2012).

2. Data

2.1. Geomagnetic Data and HMC Index

We use the HMC index (H_t, $t = 1, \ldots, N_h$) and the underlying processed geomagnetic observatory data, both available from Pick and Korte (2018). The bases are the time series of hourly vector magnetic field measurements, $B_{\boldsymbol{it}}$, taken by $i = 1, \ldots, 28$ geomagnetic observatories (Figure 1a, International Association of Geomagnetism and Aeronomy codes from http://www.bgs.ac.uk/iaga/vobs/) and distributed by the World Data Centre for Geomagnetism, Edinburgh (http://www.wdc.bgs.ac.uk/data.html). Throughout 1930–2015, they cover geomagnetic latitudes 52° (SED) $\leq \theta'_{M,it} \leq$ 9° (ABG) on the Northern and −43° (GNA) $\leq \theta'_{M,it} \leq$ −12° (VSS) on the Southern Hemisphere (for a definition of geomagnetic coordinates see Laundal & Richmond, 2017). As described in detail by Pick et al. (2019), a quality check of the

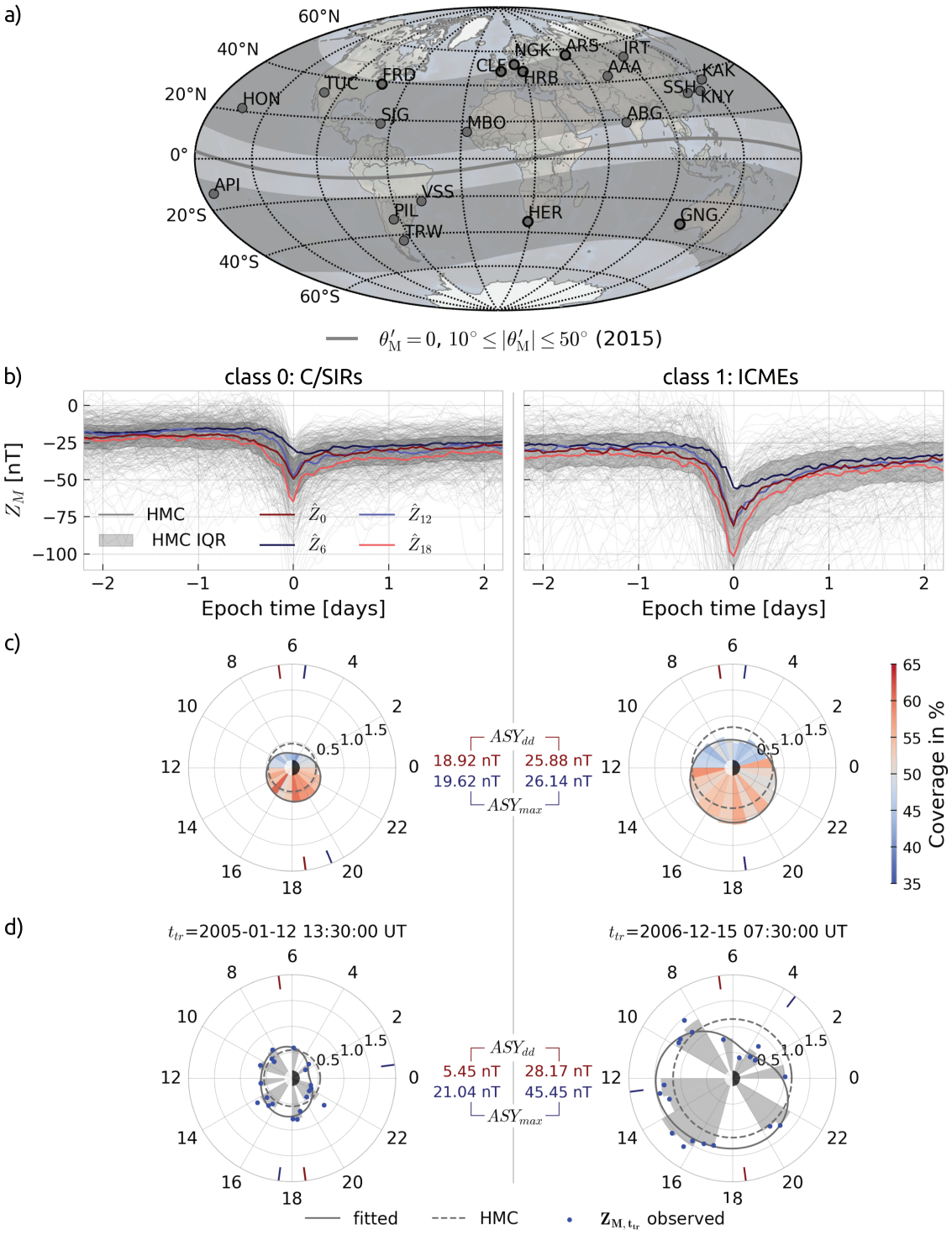

Figure 1. Input data. (a) HMC observatories with respect to the geomagnetic equator in 2015 from IGRF-12 (gray line; Thébault et al., 2015). Thick circles mark stations with predecessors; gray bands mark (10°, 50°) latitude ranges. (b) HMC (gray) and MLT-resolved disturbances ($\hat{\boldsymbol{Z}}$, colored) of training events superposed at peak times $\boldsymbol{t_{tr}}$. (c) Polar plots showing $-\boldsymbol{Z}/100$ (nT; radius) averaged over $\boldsymbol{t_{tr}}$ in dependence on MLT (angle) with a second-order Fourier fit to the data (gray solid) and MLT-symmetric HMC (dashed). The color bar indicates the sampling degree per MLT bin. Red and blue dashes mark the MLTs for ASY_{dd} and ASY_{max}. (d) Same as (c) for two events with particular observatory data distributions (dots). HMC = Hourly Magnetospheric Currents index; MLT = magnetic local time; IQR = interquartile range; C/SIRs = corotating or stream interaction regions; ICMEs = interplanetary coronal mass ejections.

observatory time series led to a modified data set. Core, ionospheric (both time varying), and crustal field (static) contributions were calculated using third-party geomagnetic field models and subtracted from the data matrix. The residuals were transformed into the geomagnetic frame, $\delta B_{\boldsymbol{M},it}$, and fit hour-by-hour to the gradient of a scalar potential, V, which was modeled by spherical harmonics of degree one (dipole field). With regard to the Earth's surface, internal and external contributions to V were separated resulting in three Gauß coefficients describing the core field and one, the HMC index, describing the external magnetic field directed along the dipole axis. Similar to Dst, HMC is expected to measure the diamagnetic effect of the magnetospheric currents, especially the RC. Throughout the study, we use a 3-hr running mean with weights $[0.25, 0.5, 0.25]$ of the original HMC. This prevents a possible impairment of the autocorrelation of HMC values on the time scale of geomagnetic storms due to, for example, sporadic observatory dropouts and thus supports the identification of storm peaks.

Our storm driver classification (section 3) exploits not only the symmetric disturbance field, measured by HMC, but also the asymmetric one. We calculate the field residuals as before, but we now remove our coestimated core field (see above). Consistent with the definition of HMC, we transform the residuals into a Cartesian system, $\alpha \boldsymbol{B}_{\boldsymbol{M},it} = \left(X_M, Y_M, Z_M\right)_{it}$ and consider only the Z_M components, which are aligned with the dipole axis. This data set, $Z_{M,it}$, corresponds to different distributions in magnetic local time (MLT), which change throughout universal time (UT) according to the availability of observatory measurements. We follow the definition $\mathrm{MLT} = \left(\phi_M - \phi_{M,S}\right)/15 + 12$, where ϕ_M and $\phi_{M,S}$ are the geomagnetic longitudes of the observatory and the subsolar point, respectively. The largest gap in MLT ranges between 15 (March 1936) and 3 hr (January 1949) with a mean of ≈4.3 hr. We average $Z_{M,it}$ over 1-hr MLT bins, centered on half hours, and get $Z_{jt}, j = 0, \ldots, 23$ per UT hour. In recognition of the prominent dawn-dusk asymmetry in the magnetic disturbance field (e.g., Newell & Gjerloev, 2012), and the ASY-H index, we define

$$ASY_{\mathrm{dd},t} = \frac{Z_{6t} - Z_{18t}}{2}\ [\mathrm{nT}], \quad ASY_{\mathrm{max},t} = \frac{\max\left(Z_{jt}\right) - \min\left(Z_{jt}\right)}{2}\ [\mathrm{nT}] \tag{1}$$

as measures of the asymmetry size. To mitigate the problem of empty MLT bins, Z is again averaged over six 1-hr bins, giving a total of four new bins: $\hat{Z}_{jt}, j = 0, 6, 12, 18$, centered on 00:30 (midnight), 06:30 (dawn), 12:30 (noon), and 18:30 (dusk). Figure 1b shows HMC and $\hat{\boldsymbol{Z}}$ for selected "training events" (section 2.2) driven by either C/SIRs or ICMEs, superposed at the event peaks. Panel (c) shows $\boldsymbol{Z}$ averaged over all peak hours, while two examples of individual events, for which some MLT bins are not sampled by observatory measurements, are given in panel (d). Several previously discovered storm characteristics are evident, namely, that (1) ICME-driven storms trigger a stronger response, that is, larger absolute HMC peak values (panel b; see section 1), (2) ASY_{dd} and ASY_{max} are larger for stronger than for weaker storms (panels (b) and (c); see also Love & Gannon, 2009; Siscoe et al., 2012; Yakovchouk et al., 2012), and (3) departures from the typical behavior can be large at any time (panel d).

2.2. Training and Target Events

Fundamental to our supervised classification method (section 3) is the definition of a training set (d_{tr}), from which the model parameters are "learned." This data set is defined for geomagnetic storm events $t_{\mathrm{tr},i}$, $i = 1, \ldots, N_{\mathrm{tr}}$ with known input data and driver class labels as output ($\boldsymbol{Y}_{\boldsymbol{tr}}$). The events are referred to by the UT hours at which HMC reaches local minima, that is, the event peak times. The zero entries of $\boldsymbol{Y}_{\boldsymbol{tr}}$ ($i_{\mathrm{tr0},j}, j = 1, \ldots, N_{\mathrm{tr0}}$) refer to C/SIR-driven events, while the remaining ones ($i_{\mathrm{tr1},j}, j = 1, \ldots, N_{\mathrm{tr1}}$) have class label one and refer to ICME-driven events.

In order to compile the training set, we collected a reference set from published catalogues (d_{re}) with $N_{\mathrm{re}} = 868$ already classified events, dated between 1995 and 2015. Of these, 571 are C/SIR and 297 ICME driven, giving a class ratio of $\gamma_{\mathrm{re}} = N_{re0}/N_{re1} \approx 1.9$ in favor of the C/SIR drivers. Specifically, 745 events ($N_{re0} = 522$, $N_{re1} = 223$) are taken from Jian et al. (2006a, 2006b, 2011) during 1995–2009, 77 additional events ($N_{re0} = 29, N_{re1} = 48$) from Turner et al. (2009) during 1995–2004 (classification following Richardson et al., 2000) and 46 events ($N_{re0} = 20$, $N_{re1} = 26$) from Shen et al. (2017) during 2013–2015. The underlying classification methods are based on a manual identification of characteristic signatures in plasma and IMF

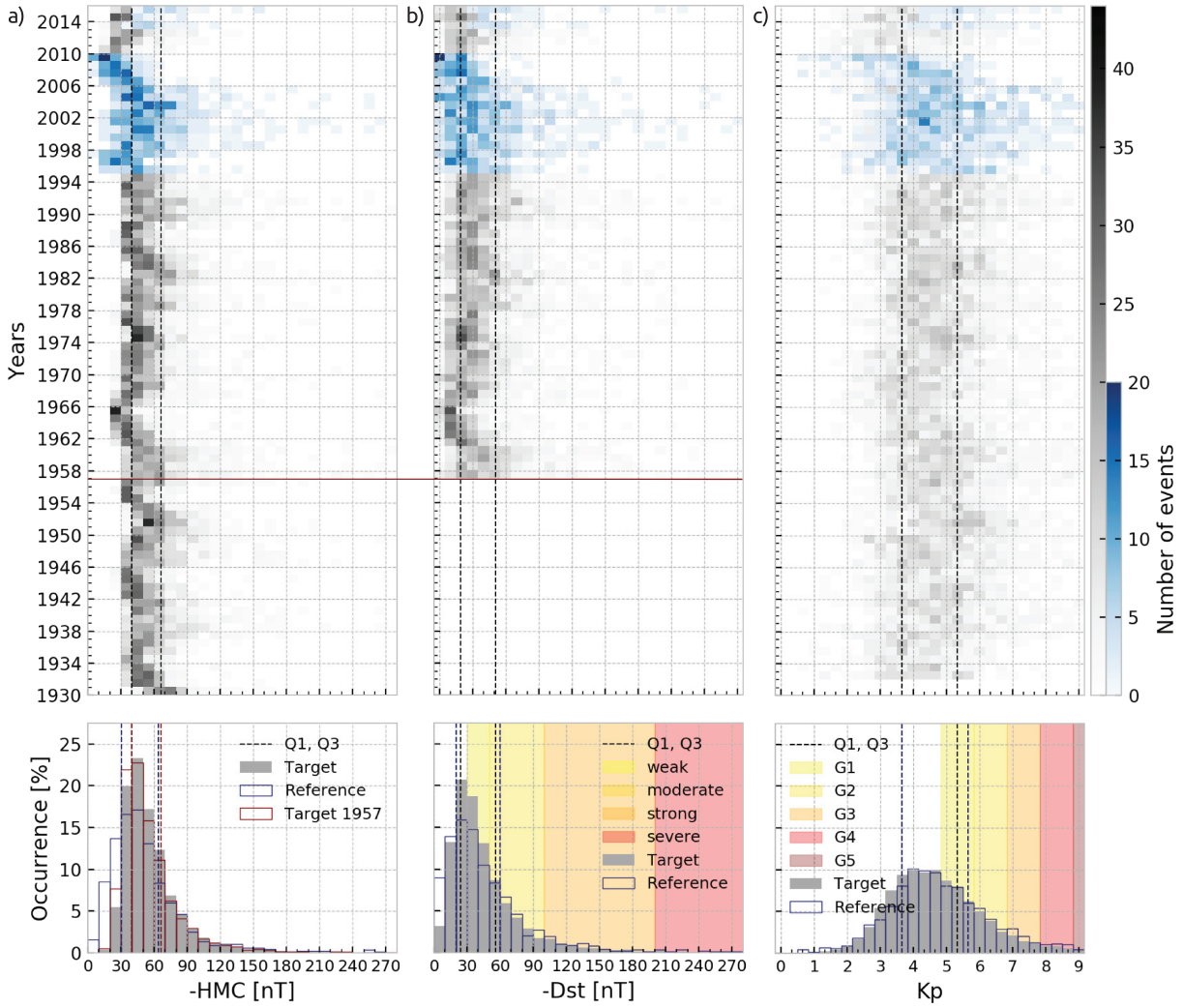

Figure 2. Event histograms. Distribution of HMC (a), Dst (b), and Kp (c) for 7,547 target (gray) and 868 reference (blue) event peaks. Top panels show the number of events per year and index bin. Bottom panels show the occurrence per index bin, summed over all years. The red bar plot in (a) refers only to years since 1957 (red horizontal line). For Kp, the largest value in 26-hr intervals centered on the HMC peaks is considered ($\boldsymbol{t_{ta}}$). Dashed lines mark the 25th (Q1) and 75th (Q3) percentiles. Colored bars indicate commonly used Dst- and Kp-based storm intensity scales. HMC = Hourly Magnetospheric Currents index.

observations from ACE and Wind as outlined in section 1.

Once the model is trained, it can predict the driver classes of storm events from the target set (d_{ta}) between 1930 and 2015. The target event peak hours are identified with regard to HMC (H) in three steps, which are visualized for clarity in Appendix A.

1. All hours, t, for which $H_t < Hl_t$ are marked:

$$H_t < Hl_t = P_n\left(\frac{H_{11y,t}}{\min\left(H_{11y,t}\right)}\right)^{p_{\mathrm{sc}}} \tag{2}$$

   P_n is the nth HMC percentile during 1930–2015 and $H_{\mathbf{11y}}$ is a low-pass-filtered HMC with a cutoff period of 11 years. With $p_{\mathrm{sc}} = 1$, the term in brackets acts as a scaling factor in dependence of the SC, ranging between 0.39 (October 2009) and 1 (May 1951). Small factors correspond to solar minimum years when disturbed times have smaller absolute HMC values compared to solar maximum years.
2. The local HMC minima of consecutively marked hours (Step 1) are identified by changes in the sign of $\mathrm{d}H_t/\mathrm{d}t$. Should two successive local minima be separated by less than Δt hours only the deeper one is kept, so that

$$t_{i+1} - t_i \geq \Delta t \tag{3}$$

3. Only those local minima hours (Step 2) associated with a HMC drop of at least Hs nT are finally selected as target event peak hours t_{ta}:

$$|H_t - H_{5d,t}| \geq Hs \tag{4}$$

Here, H_{5d} is the 5-day low-pass-filtered HMC and a measure of the disturbance background level.

We define the training set as the intersection between the selected target events and the reference events, that is, $d_{tr} = d_{re} \cap d_{ta}$. The parameters of the selection scheme, $P_{29} = -37.6$ nT, $\Delta t = 27$ hours, and $Hs = 7$ nT, are chosen so that d_{ta} is as small while d_{tr} is as large as possible, that is, at least 60% of d_{re}. This guarantees that the target event selection scheme recovers most of the reference events (Figure 2) and that the training events are geoeffective at low and middle latitudes; that is, they produce a storm-like geomagnetic disturbance with regard to the prevailing activity level on both annual (equation (2)) and daily (equation (4)) time scales (see also Figure 1b). Given the variability in the events' recovery times, there is no perfect choice of Δt (equation (3)). We justify our choice of 27 hr by the fact that it is not chosen subjectively, but optimized with regard to the reference set.

The procedure gives $N_{ta} = 7,547$ target events, of which $N_{tr} = 538$ overlap with the reference events ($N_{tr0} = 342$ C/SIRs, $N_{tr1} = 196$ ICMEs). The distribution of target events is skewed toward smaller absolute HMC values (median at -50.4 nT). It is similar to the corresponding Dst distribution, but notably different from the relatively symmetric Kp index (Bartels et al., 1939) distribution (Figure 2). This is expected given the different latitudinal sensitivity ranges of HMC and Dst on the one hand and Kp on the other hand. The comparison shows that the SC variability is indeed more pronounced for HMC than for Dst. We note that some of the target events are not considered to be geomagnetic storms according to a frequently used minimum-Dst categorization proposed by Loewe and Prölss (1997, ≈25%) or the Kp-based National Oceanic and Atmospheric Administration G-scale (https://www.swpc.noaa.gov/noaa-scales-explanation; >50%). However, since our target events produce storm-like geomagnetic disturbances and the majority (≈75%) meets the above Dst definition (Figure 2b), we refer to our target events as storms.

Although ≈38% of the reference events are not taken into account, the training class ratio $\gamma_{tr} \approx 1.7$ is comparable to that of the reference events ($\gamma_{re} \approx 1.9$). We have thoroughly tested the sensitivity of the classification result to the parameters of the event selection method (supporting information Text S1) and conclude that the presented selection of training and target events is optimal given the available reference events.

In summary, we have $N_{tr} = 538$ training storm events with known inputs and class labels. The inputs are HMC and the MLT-resolved disturbance field ($\boldsymbol{Z}, \hat{\boldsymbol{Z}}$). Furthermore, we have $N_{ta} = 7,547$ target events with known inputs, but unknown class labels (supporting information Figure S1).

3. Binary Logistic Regression Model

3.1. Feature Definition

The core of this work is the identification of driver-characteristic features from the input data. Among the considered possibilities, 11 features are listed according to their relative importance (section 3.3) in Table 1. Their derivation consists of a physically motivated initial feature definition (Column 1), followed by a refinement aimed at maximizing the separation of the standardized feature distributions for C/SIR and ICME-driven training events (Figure 3). Standardization is a common preprocessing step in machine learning and means that the features are centered, that is, the means are removed, and scaled, that is, through a division by the standard deviation. The further the class medians are separated from one another and the narrower their interquartile ranges (75th–25th percentiles; Figure 3, colored bars) are, the better the feature is. For comparison, we synthetically generated an ideal feature ($k = 12$) by randomly drawing samples from two normal distributions centered on 0.25 (C/SIRs) and 0.75 (ICMEs), each with a standard deviation of 0.01 (Figure 3, bottom right).

Features $k = 1, 4, 5$ describe the well-known RC enhancement, SC phase, and the recurrence pattern associated with the storms (e.g., Borovsky & Denton, 2006). As expected, ICMEs are statistically more intense and occur closer to solar maximum (Figures 1b and 3). The deviation of the data-based features from the ideal case is particularly striking for the "Recurrence" feature, for which the medians coincide at 0.5. Nevertheless, the feature still has skill by contributing the information that 75% of C/SIRs recur at least once, while this is true for just 25% of ICMEs (Figure 3, $k = 5$). The latter poses a problem that is discussed in section 5. Of particular interest are the innovative, high-priority features $k = 2, 3$, which are designed to

Table 1
Feature Definitions

ID (k)	Description	Formal definition	$\tilde{\boldsymbol{X}}_{\boldsymbol{tr}}[\boldsymbol{i}_{\boldsymbol{tr0}}, k]$, $\tilde{\boldsymbol{X}}_{\boldsymbol{tr}}[\boldsymbol{i}_{\boldsymbol{tr1}}, k]$	α_k
1	Event intensity.	$\Delta \boldsymbol{H} = \boldsymbol{H}_{\boldsymbol{t_{ta}}} - \boldsymbol{H}_{\boldsymbol{1y},\boldsymbol{t_{ta}}}$ (nT).	0.08, 0.15	0.63
2	Correlation of disturbances at dawn and dusk.	$r\left(\hat{\boldsymbol{Z}}_{\boldsymbol{6t}}, \hat{\boldsymbol{Z}}_{\boldsymbol{18t}}\right)$ for $t = [t_{ta}, t_{ta} + 24]$.	0.46, 0.74	0.77
3	Variance of disturbance change at dawn.	$MAD\left(\mathrm{d}\hat{\boldsymbol{Z}}_{\boldsymbol{6t}}/\mathrm{d}\boldsymbol{t}\right)$ for $t = [t_{ta}, t_{ta} + 24]$ (nT/hr).	0.14, 0.22	0.63
4	Solar cycle phase.	$\boldsymbol{\phi}_{\boldsymbol{sc}} = \mathrm{d}\boldsymbol{H}_{\boldsymbol{11y},\boldsymbol{t_{ta}}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{ta}}$ (nT/hr).	0.45, 0.29	−0.42
5	Recurrence.	$\boldsymbol{R}_{\boldsymbol{t_{ta}}} \in [0, 1, 2]$. The event 0: does not recur, 1: recurs once with period 27.28 days ± 26 hr, or 2: recurs twice.	0.5, 0.5	−0.32
6	Robustness of disturbance peak location.	Number of instances $\underset{j}{argmax} \quad \lvert Z_{jt}\rvert \leq 17$ for $t = [t_{ta} - 48, t_{ta} - 24]$.	0.58, 0.67	0.28
7	Largest spatial disturbance asymmetry.	$\max\left(\boldsymbol{ASY}_{\boldsymbol{max},\boldsymbol{t}}\right)$ for $\boldsymbol{t} = [\boldsymbol{t}_{\boldsymbol{t_a}} - 24, \boldsymbol{t}_{\boldsymbol{t_a}}]$ (nT).	0.13, 0.22	0.31
8	Speed of dawn-dusk asymmetry decrease.	FDHM $ASY_{\boldsymbol{dd},\boldsymbol{t}}$ with respect to t_{ta} (hr).	0.05, 0.05	0.16
9	Mean intensity change for recovery phase.	$\overline{\mathrm{d}\boldsymbol{H}_{\boldsymbol{t_{rec}}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{rec}}}$ for $\boldsymbol{t}_{\boldsymbol{rec}} = \left[\underset{t}{argmax}\left\lvert \mathrm{d}\boldsymbol{H}_{\boldsymbol{t_p}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{p}}\right\rvert, \boldsymbol{t}_{\boldsymbol{ta}} + 12\right]$, $t_p = [t_{ta}, t_{ta} + 12]$ (nT/hr).	0.07, 0.12	
10	Range in $k = 2$ for time-shifted disturbances.	Range in $r\left(\hat{\boldsymbol{Z}}_{\boldsymbol{6t_l}}, \hat{\boldsymbol{Z}}_{\boldsymbol{18t}}\right)$ for $t = [t_{ta}, t_{ta} + 24]$, $t_l = t + [-12, \ldots, 12]$.	0.51, 0.65	
11	Difference in intensity changes for main and recovery phases.	$\overline{\mathrm{d}\boldsymbol{H}_{\boldsymbol{t_{rec}}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{rec}}} - \overline{\mathrm{d}\boldsymbol{H}_{\boldsymbol{t_{main}}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{main}}}$ for $\boldsymbol{t}_{\boldsymbol{main}} = [\boldsymbol{t}_{\boldsymbol{ta}} - 12, \underset{}{argmax} \quad \lvert \mathrm{d}\boldsymbol{H}_{\boldsymbol{t_p}}/\mathrm{d}\boldsymbol{t}_{\boldsymbol{p}}\rvert]$, $t_p = [t_{ta} - 12, t_{ta}]$ (nT/hr).	0.64, 0.61	

Note. Features $k = 1, \ldots, 11$ sorted according to their importance from top to bottom with medians for corotating or stream interaction regions- and interplanetary coronal mass ejections-driven training events ($\tilde{\boldsymbol{X}}_{\boldsymbol{tr}}[\boldsymbol{i}_{\boldsymbol{tr0}}, k]$, $\tilde{\boldsymbol{X}}_{\boldsymbol{tr}}[\boldsymbol{i}_{\boldsymbol{tr1}}, k]$; Figure 3) and corresponding model coefficients (α_k; section 3.2). $\boldsymbol{t}_{\boldsymbol{ta}}$: storm peak times of target events; $\boldsymbol{H}_{\boldsymbol{ny}}$: low-pass-filtered HMC with cutoff period of n years; r: Pearson's correlation coefficient; *MAD*: Median Absolute Deviation; *FDHM*: Full Duration at Half Maximum.

capture PRC-related characteristics of the disturbances at dawn and dusk. Apparently, the disturbances in these MLT-sectors are better correlated ($k = 2$) and the temporal change at dawn is more variable ($k = 3$) for ICMEs than for C/SIRs over a period of 24 hr after the HMC peak. The former seems to be compatible with the finding of Yakovchouk et al. (2012) that strong storms (mostly CME driven) are relatively more symmetric than weak ones (mostly CIR related). The related feature $k = 10$ describes to what extent feature $k = 2$ changes if the disturbances are time lagged and is found to be of limited use. Feature $k = 6$ quantifies whether the disturbance peak is located "abnormally," that is, not in the dusk sector, which is more prevalent for ICMEs. Features $k = 7, 8$ are related to the asymmetry measures (equation (1)), while features $k = 9, 11$ describe the rate of change in HMC for the storm main and recovery phases.

Features $k = 1, \ldots, N_f$ form the $N_{\mathrm{tr}\vee ta} \times N_f$ matrices $\boldsymbol{X}_{\boldsymbol{tr}\vee\boldsymbol{ta}}$ for the training and target events. Because they require data prior to the storm peaks (e.g., $k = 6$), they are defined for just $N_{\mathrm{ta}} = 7{,}546$ out of the 7,547 selected target events.

3.2. Model Setup

The binary logistic regression model takes a linear combination of the features, $\boldsymbol{X}_{\boldsymbol{i}}$, and delivers the logarithm of the odds ("Logit") that the storm event i is driven by an ICME (equation (5), Default Class 1). The odds are defined as the probability $P\left(Y_i = 1|\boldsymbol{X}_{\boldsymbol{i}}\right) \in [0, 1]$, divided by the complementary probability $1 - P\left(Y_i = 1|\boldsymbol{X}_{\boldsymbol{i}}\right)$.

$$\mathrm{Logit}\left[P\left(Y_i = 1|\boldsymbol{X}_{\boldsymbol{i}}\right)\right] = \ln\left(\frac{P\left(Y_i = 1|\boldsymbol{X}_{\boldsymbol{i}}\right)}{1 - P\left(Y_i = 1\middle|\boldsymbol{X}_{\boldsymbol{i}}\right)}\right) = \alpha_0 + \boldsymbol{X}_{\boldsymbol{i}}^T\alpha \tag{5}$$

$$P(Y_i = 1|\boldsymbol{X}_{\boldsymbol{i}}) = \frac{1}{1 + \exp\left[-\left(\alpha_0 + \boldsymbol{X}_{\boldsymbol{i}}^T\alpha\right)\right]} \tag{6}$$

The model is trained by using $\boldsymbol{X}_{\boldsymbol{tr}}$ and known class labels $\boldsymbol{Y}_{\boldsymbol{tr}}$ to solve equation (5) for the intercept $\alpha_0 = -0.17$ and the coefficients $\alpha_k, k = 1, \ldots, N_f$ (Table 1). The class ratio γ_{tr} is balanced by applying weights inversely proportional to the class frequencies, that is, $N_{\mathrm{tr}}/\left(2 \cdot N_{\mathrm{tr0}}\right) = 0.79$ for C/SIRs and $N_{\mathrm{tr}}/\left(2 \cdot N_{\mathrm{tr1}}\right) = 1.37$

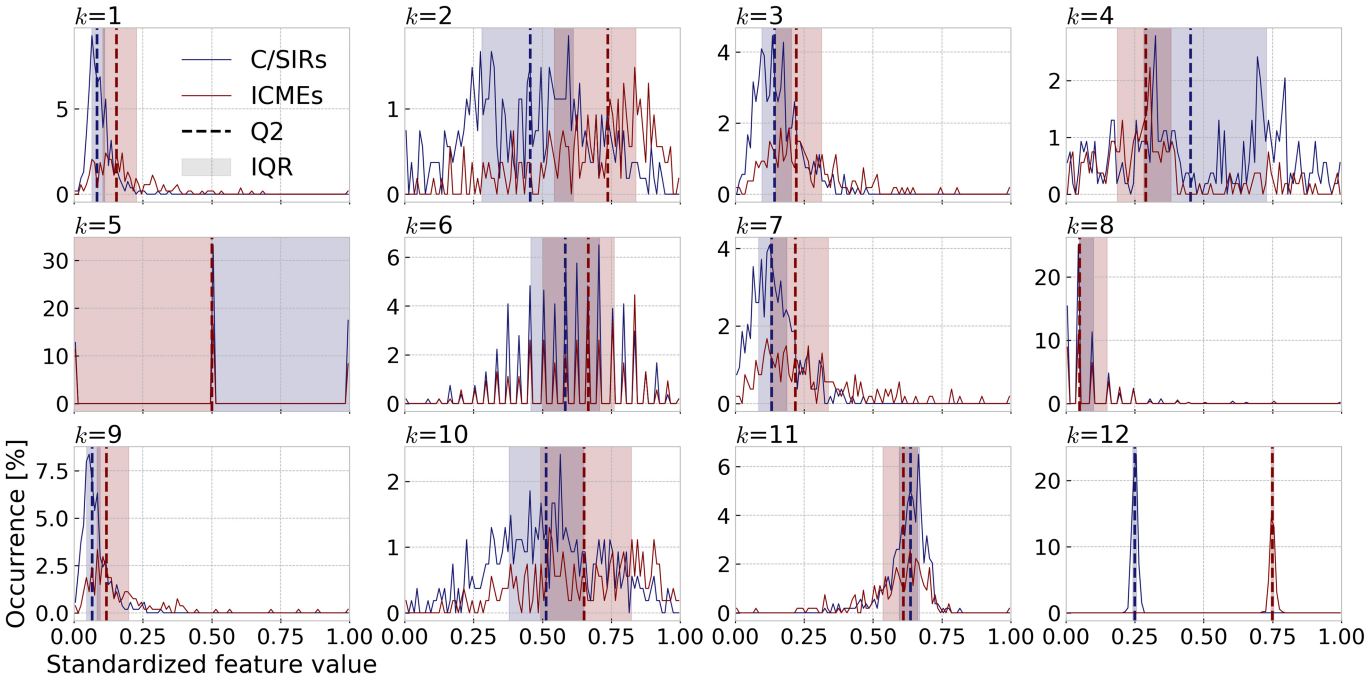

Figure 3. Feature distributions. Histograms of standardized features for C/SIR-driven (blue) and ICME-driven (red) training events with interquartile ranges (IQRs) and medians (Q2, Table 1). Feature k=12 is an ideal synthetic feature. Features are scaled to [0, 1] for visual comparability. C/SIRs = corotating or stream interaction regions; ICMEs = interplanetary coronal mass ejections.

for ICMEs. We choose the incremental gradient algorithm "SAGA" (Defazio et al., 2014), because it is the most versatile solver available in Python's scikit-learn library (https://scikit-learn.org/) that supports sparse regression and a potential future extension to multinomial cases.

3.3. Model Assessment

The most important features and optimum model hyperparameters are identified with respect to Matthew's correlation coefficient (MCC):

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad \in [-1, 1] \tag{7}$$

MCC, also known as the ϕ-coefficient (e.g., Parker, 2013), is particularly useful for data sets with imbalanced classes as it takes into account both True (T) and False (F) predictions of Negatives (N, $\boldsymbol{Y}_{\boldsymbol{tr}}[i_{\boldsymbol{tr0}}]$: C/SIRs) and Positives (P, $\boldsymbol{Y}_{\boldsymbol{tr}}[i_{\boldsymbol{tr1}}]$: ICMEs). A stratified K-fold cross-validation scheme is used (Figure 4a), in which the training set is split into K_o=4 folds, so that each consists of 134–135 data points with preserved driver class percentages (outer loop). The hyperparameters of the model are initialized with default values. The classifier is recursively trained on three of the folds, so that the feature with the lowest absolute α_k is pruned from $\boldsymbol{X}_{\boldsymbol{tr}}$ at each iteration, until $N_f = 1, \ldots, 11$ features are left. For each of these 11 runs, *MCC* is calculated four times on the different validation and training sets and then averaged over the sets ($MCC_{o,l}$) and the $l = 1, \ldots, N_o$=25 splits. Figure 4b shows that $N_f = 8$ features, namely, features $k = 1, \ldots, 8$ (Table 1), give the highest *MCC*, so that the feature matrices are fixed accordingly. We note that the training scores confirm this result, but that the absolute *MCC* values have no meaning yet, since they refer to an interim model.

We optimize two of the hyperparameters, which are the inverse regularization strength C (default: $C = 1$) and the norm used in the penalization L (default: L = L2 norm). An inner loop is introduced to the cross-validation scheme with K_i=3 and $N_i = 25$ (Figure 4a) in order to prevent an artificial overestimation of *MCC* (e.g., Krstajic et al., 2014). *MCC* is calculated for ten combinations of hyperparameters

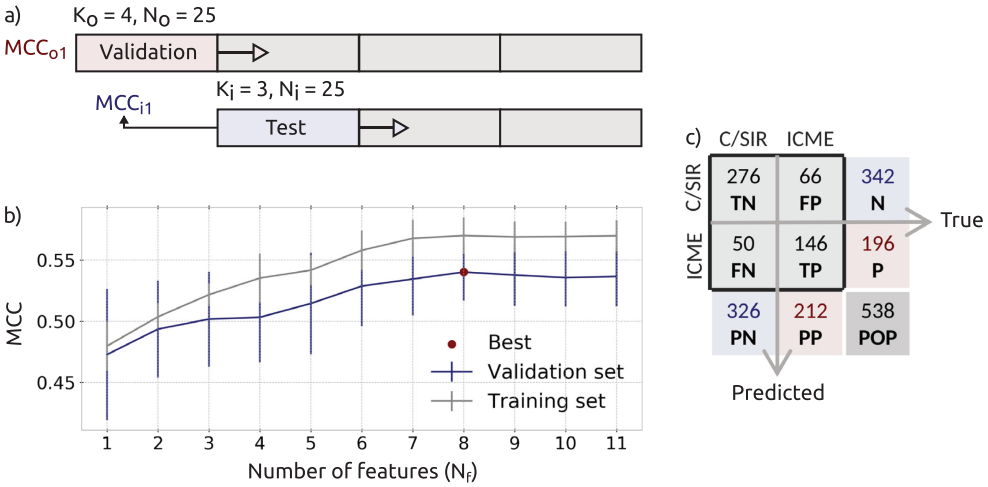

Figure 4. Model assessment. (a) *K*-fold cross-validation scheme on the training set (gray) with four folds in the outer and three folds in the inner loop, each repeated 25 times. (b) Cross-validated *MCC* with uncertainty (two standard deviations) from models using the N_f best features. (c) Confusion matrix from final model. *MCC*: Matthew's correlation coefficient; *P*: Positives; *N*: Negatives; *POP* = *N*+*P*: Population; *T*: True; *F*: False; *P*[*P*/*N*] = Predicted *P*/*N*; C/SIRs = corotating or stream interaction regions; ICMEs = interplanetary coronal mass ejections.

$C \in \{1, 0.1, 0.01, 1e-3, 1e-4\}$ and $L \in \{\text{L1-norm}, \text{L2-norm}\}$ on the test sets of the inner loop (7500 scores each) and averaged over the inner sets/splits (MCC_i) and the outer sets/splits. The combination of $C = 0.1$ and $L =$ L2-norm gives the highest averaged *MCC* (supporting information Figure S2) and is thus fixed in the final model, which is assessed on the validation sets of the outer loop.

From the confusion matrix of the final model (Figure 4c) a cross-validated *MCC* score of 0.54 ± 0.013 is calculated, which means that the classifier is skillful and performs significantly better than chance ($MCC = 0$). For completeness and in order to facilitate comparisons, other frequently cited scores are listed in Appendix B.

Apart from the classifier's skill, we can also assess its probabilistic predictions (equation (6)), that is, its reliability. By calculating that fraction of events with a certain predicted ICME-driver probability that are truly ICME-driven, we find that the classifier's probabilistic predictions are slightly overestimated, but reliable within a tolerance of 15% (supporting information Figures S3b and S3c).

4. Classification Result

In the prediction step, the model coefficients α_0, $\boldsymbol{\alpha}$ (section 3.2 and Table 1), and the feature matrix $\boldsymbol{X_{ta}}$ are inserted into equation (6) to give the ICME-driver probabilities $P(Y_{ta,i} = 1)$ for the target events. The class labels, $\boldsymbol{Y_{ta}}$, are set according to the decision rule

$$Y_{ta,i} = \begin{cases} 0 & \text{if} \quad P(Y_{ta,i} = 1) \leq 0.5, \\ 1 & \text{otherwise.} \end{cases} \tag{8}$$

Applying the default decision boundary $P(Y_{ta,i} = 1) = 0.5$ is appropriate, because it delivers the best classification result as judged by the mean of several performance scores (supporting information Figure S3a).

The ICME driver probabilities are color coded in Figure 5a (left), which groups the target events in a "time plane" spanned by the decimal year and the solar rotation phase. In this view, CIR-driven events are vertically aligned. Of all target events, ≈63% are driven by C/SIRs, giving a class ratio $\gamma_{\text{ta}} \approx 1.7$. On average, the uncertainties of the predictions are similar for both drivers with median class probabilities $\tilde{P}_0 = 1 - \tilde{P}(\boldsymbol{Y_{ta}}[\boldsymbol{i_{ta0}}] = 1) = 0.74$ for C/SIRS and $\tilde{P}_1 = \tilde{P}(\boldsymbol{Y_{ta}}[\boldsymbol{i_{ta1}}] = 1) = 0.73$ for ICMEs.

Throughout time, ICME-dominated periods (red) alternate with C/SIRs-dominated ones (blue), while γ_{ta} seems to increase from the past to the present (cf. supporting information Figure S4). The stacked histograms

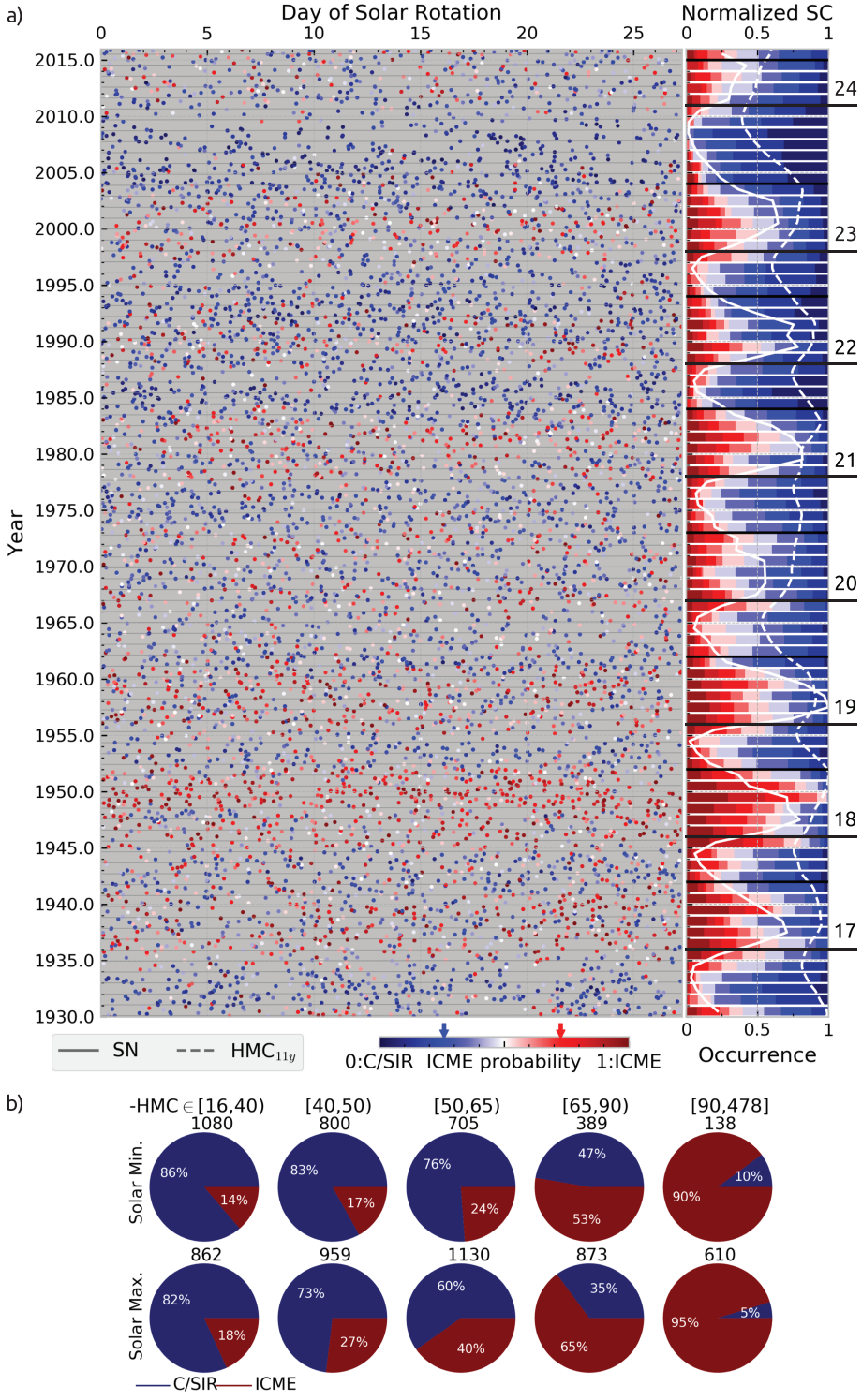

Figure 5. Classification result. (a) Left: Probabilities $P(Y_{ta,i}=1)$ for target events in a year versus day of solar rotation plot. Median class probabilities $\tilde{P}_0 = 0.74$, $\tilde{P}_1 = 0.73$ are marked on the color bar. Right: Stacked histograms for 0.1 probability intervals per year with normalized sunspot number (SN, white solid) and 11-year low-pass-filtered HMC (dashed) representing the solar cycle (SC; numbers in black). (b) Driver fractions during solar min. (top) and max. years (bottom), as separated by horizontal lines in the histograms, grouped into five intensity levels defined by HMC (nT). Group definitions and event numbers are given above the pie plots. C/SIRs = corotating or stream interaction regions; ICMEs = interplanetary coronal mass ejections; HMC = Hourly Magnetospheric Currents index.

of 10 probability intervals $P(Y_{ta,i} = 1) \in [0.0, 0.1), \ldots, [0.9, 1.0]$ per year (Figure 5a, right) accentuate that the occurrence pattern is tightly linked to the SC. The latter is manifest in either the 11-year low-pass-filtered HMC (dashed line) or in the independent yearly mean total sunspot number SN (solid line). Both these time series are normalized, that is, divided by their largest absolute value. On average, ICME drivers are most frequent at SC maximum years, while C/SIRs prevail during the declining phase and SC minimum years. The dependency of the driver occurrence statistics on the intensity of the storm events and the SC phase is shown in Figure 5b. Five intervals of HMC values between −16 nT (weak) and −478 nT (strong) are defined, such that they include a roughly comparable number of events. Since the HMC distribution of target events is strongly skewed toward smaller absolute values (Figure 2a), the intervals get wider as the event intensity increases. Nevertheless, the number of events per class still decreases as intensity increases, especially for solar minimum years (upper row). We define "solar minimum" years as years for which SN≤70 and all other years as "solar maximum years." This threshold was chosen, so that solar minima (40 years) and maxima (46 years) are roughly balanced and the transitions visually coincide with ascending or decreasing flanks of SN (Figure 5a, black horizontal lines). As can be seen in Figure 5b, ICMEs predominantly drive intense geomagnetic storms and their fraction is generally larger at solar maximum years.

5. Discussion

The binary classifier we developed can be seen as a proof of concept, demonstrating that it is possible to identify storm drivers from ground-level data without the need for satellite observations. However, the performance of the classifier could probably be improved technically, that is, without changing the fundamental type of input information. First, there is much more geomagnetic data available from mid latitudes than we use, especially for recent years, which can improve the coverage of MLTs (blue dots in Figure 1d). Our choice is motivated by the fact that it is fairly stable throughout time, we have assured a decent quality of the observatory time series during early years and it is consistent with the input data used for HMC. Second, the definition of features could be avoided altogether by using data-interpolating empirical orthogonal functions to decompose the data set into spatiotemporal patterns, which replace the features (e.g., Holappa et al., 2014, 2015; Shore et al., 2018). Lastly, more sophisticated machine learning techniques could be applied, for example, Support Vector Machines, Random Forest Classifiers, or Neural Networks, which would however preclude a simple ranking of features. As opposed to the aforementioned points, the following issues are relevant to the scientific interpretability of our results.

5.1. Driver Class Statistics

Although the time period from which we derive the geomagnetic driver statistics exceeds that of previous studies by roughly three SCs (e.g., Echer et al., 2011; Richardson & Cane, 2012), the C/SIR versus ICME occurrence statistics are in general agreement (see section 4). This supports the validity of our model beyond the mathematical perspective taken in section 3.3. However, an analysis of the confusion matrix (Figure 4c) reveals that the model suffers from False Negatives (*FNs*), that is, the misidentification of ICMEs as C/SIRs, more than from False Positives (*FPs*). A decrease of *FNs* would lower the rate of misses (∼26%) and false discoveries (∼31%), and thereby improve *MCC* ≈1.3 times more than the same decrease of *FPs*.

The validity of the classifier is probably challenged most by the weakest events of the target distribution. Owing to feature $k = 1$ (Table 1), it is likely that for these events (1) slow solar wind or unclear drivers (as considered in Richardson & Cane, 2012) are erroneously forced into the C/SIR class and (2) ICME drivers are misidentified (*FNs*). There are several instances of decreases in ICME-driven storm activity near SC maximum (e.g., cycle 20), which could either be due to such *FNs*, or to an actual phenomenon related to the "Gnevyshev gap" (Feminella & Storini, 1997), or a mixture of both.

Additionally, weaknesses in the features' definitions mitigate the classifier's skill. Feature $k = 4$ ("SC phase") is defined as the derivative of the low-pass-filtered HMC at t_{ta}, which lags behind SN (HMC_{11y}, Figure 5a). Therefore, driver preferences for SC phases are "blurred," such that negative values represent both SN maxima, favorable for ICMEs, and part of the declining phase, favorable for C/SIRs. Feature $k = 5$ ("Recurrence") gives information on whether or not an event recurs in the target set. There is no mechanism to prevent a positive response for ICME-driven events, in which case they are "unphysically" labeled as recurring. On

the other hand, SIRs are likely to be confused as ICMEs since both are nonrecurring. Although overall less effective, splitting the C/SIR class into two would likely reduce *FP*s.

Finally, the SC dependence of the drivers is reproduced even if the classification is done without features $k = 4, 5$ (supporting information Figure S5). However, these results show significantly reduced median class probabilities (≈ 0.6), underlining that, despite their weaknesses, both features contribute to the classifier's skill.

5.2. Geoeffectiveness of Training Events

One aspect not explicitly taken into account is the mechanism and efficiency with which the interplanetary drivers of the training set couple to the magnetosphere. Thus, there is no guarantee that each solar wind structure identified at L1 is geoeffective at midlatitudes, that is, the region to which our analysis is sensitive. This explains, why ≈38% of the reference events could not be included in the training set (see section 2.2). Among other criteria involving, for example, v_{sw}, a southward directed $B_{\mathrm{sw},z}$ is decisive for the storm generation (see section 1). ICMEs with clear signatures of MCs often comply with this favorable condition, making them the most geoeffective subset of ICMEs (e.g., Nikolaeva et al., 2011). Other cases include instances where the spacecraft misses the centre of the flux rope or there are successive ICMEs interacting with each other (complex ejecta). Similarly, C/SIRs driving shock waves at their leading edges and followed by Alfvénic high speed streams are more geoeffective than others (Snekvik et al., 2013; Zhang et al., 2008).

Since HMC is particularly sensitive to the RC, the degree of HMC disturbance depends on that of RC buildup, resulting from the above mentioned geoeffective drivers. The latter is controlled by the plasma sheet density (e.g., Jordanova et al., 2009) and RC composition, particularly the concentration of oxygen ions (e.g., Denton et al., 2006).

Consequently, one could enhance the physical expressiveness of the study by selecting a subset of training events showing unambiguous C/SIR and ICME signatures, which promote RC buildup. However, the price of such action is a reduced training set, less appropriate for machine learning methods. On the other hand, one could abandon the attempt to link ground-level disturbances to specific magnetospheric current systems and open the study to input data from high-latitude magnetometers sensitive to different kinds of current systems and disturbances (e.g., substorms). Here, we have steered a middle course on the basis of which both approaches can be further developed.

6. Conclusions

Our binary classifier reliably determines how likely a given geomagnetic storm is driven by an ICME as opposed to a C/SIR. The inputs (features) for the classifier are derived solely from the MLT-resolved geomagnetic disturbance field at ground level. Among them are two high-priority features, which have not been formulated previously in this context. We provide the driver probabilities for 7,546 storms between 1930 and 2015, whose statistics in dependence of event intensity and solar activity are in general agreement with previous satellite-based studies covering about four SCs. Our results can be taken as a proof of concept for a further exploitation of geomagnetic observatory data with regard to an interplanetary driver identification. To this end, several ideas for future refinement or modification of the presented method are discussed.

Appendix A: Illustration of Event Selection Scheme

Figure A1 illustrates the event selection scheme by the example of September 1981, when Dst, SYM-H (Iyemori, 1990), and Kp indices are available for comparison. The offset between HMC and Dst results from Dst's erroneous baseline (see section 1).

Appendix B: Performance Scores

Figure B1 shows the row- and column-normalized confusion matrices (Figure 4c), whose entries are used in Table B1 to calculate several performance scores (see also supporting information Figure S6). Note that the terms used here have long-standing equivalents in the field of weather forecasting (e.g., Woodcock, 1976), which have been adopted for space weather research (e.g., Liemohn et al., 2018).

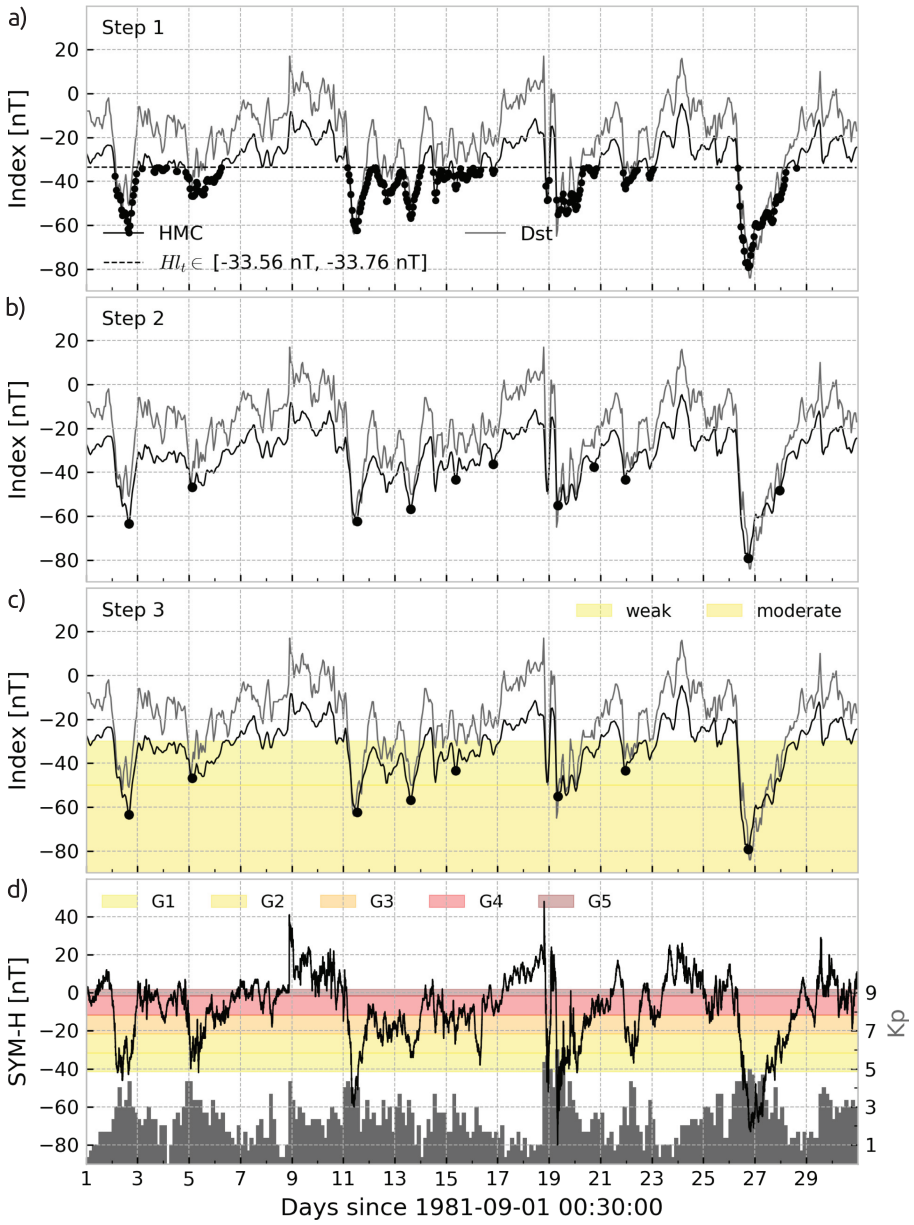

Figure A1. Event selection. Black dots mark (a) hours at which HMC falls below the detection threshold (dashed), which varies by 0.2 nT during this interval; (b) local HMC minima, separated by at least 27 hr; (c) final event peaks, at which HMC lies at least 7 nT below the background. (d) SYM-H (left) and Kp (right) indices. Colored bands in panels (c) and (d) indicate commonly used Dst- and Kp-based geomagnetic storm intensity scales (see section 2.2). HMC = Hourly Magnetospheric Currents index.

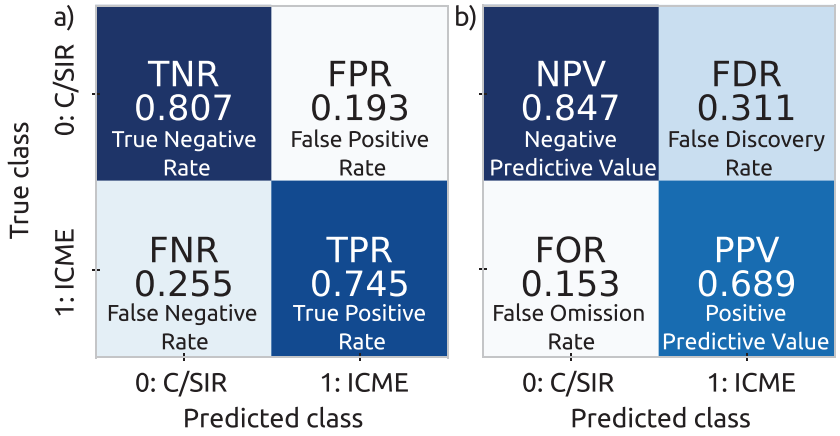

Figure B1. Normalized confusion matrices. Confusion matrices of (a) row-normalized type using $N = 342$ Negatives and $P = 196$ Positives and (b) column-normalized type using $PN = 326$ Predicted Negatives and $PP = 212$ Predicted Positives. C/SIRs = corotating or stream interaction regions; ICMEs = interplanetary coronal mass ejections.

Table B1
Performance Scores

Metric	Definition	Range	Scores validation set training set	Std validation set training set
Matthew's	$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}}$	[−1,1]	0.54	0.013
Correlation			0.57	0.005
Coefficient				
Heidke Skill	$HSS = \frac{2(TP \cdot TN - FN \cdot FP)}{(FN+TN)\cdot(TP+FN)+(TP+FP)\cdot(FP+TN)}$	[−1,1]	0.54	0.012
Score			0.57	0.005
Accuracy	$ACC = \frac{TP+TN}{POP}$	[0,1]	0.78	0.005
			0.80	0.003
Diagnostic	$DOR = \frac{TPR \cdot TNR}{FPR \cdot FNR}$	$[0, \infty]$	13.29	1.296
odds ratio			14.57	0.503
F1 score	$F1 = 2\frac{PPV \cdot TPR}{PPV+TPR}$	[0,1]	0.71	0.009
			0.73	0.003
Frequency	$FB = \frac{TP+FP}{TP+FN}$	$[0, \infty]$	1.08	0.018
bias			1.09	0.007
Youden's J	$J = TPR + TNR - 1$	[−1,1]	0.55	0.014
statistic			0.58	0.005
(Informedness)				
Deltap	$Deltap = PPV + NPV - 1$	[−1,1]	0.53	0.011
(Markedness)			0.56	0.006

Note. Cross-validated performance scores for the validation (Columns 4 and 5, top row) and training sets (bottom row). P: Positives; N: Negatives; $POP{=}N{+}P$: Population; T: True; F: False; R: Rate; $[P/N]PV$: $[P/N]$ Predictive Value; see Figures 4c and B1.

Acknowledgments
The authors thank Yuri Shprits, Joachim Vogt, and Robert Shore for useful discussions and two anonymous reviewers for helping to improve the manuscript. L. P. and M. K. acknowledge funding by the Deutsche Forschungsgemeinschaft under Project SPP1788 "Dynamic Earth" (DFG KO 2870/6-1). F. E. acknowledges support from a group linkage program with South Africa, funded by the Alexander von Humboldt Foundation. I. Z. was supported by Geo.X, the Research Network for Geosciences in Berlin and Potsdam, under Grant SO_087_GeoX. The sunspot number (http://www.sidc.be/silso/datafiles) and Dst/Kp/SYM-H indices (http://wdc.kugi.kyoto-u.ac.jp/) are publicly available. All input data, the results, and the corresponding software can be accessed via GFZ Data Services (Pick, 2019; http://doi.org/10.5880/GFZ.2.3.2019.003).

References

Akasofu, S.-I., & Chapman, S. (1964). On the asymmetric development of magnetic storm fields in low and middle latitudes. *Planetary and Space Science*, *12*(6), 607–626. https://doi.org/10.1016/0032-0633(64)90008-X

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

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

Defazio, A., Bach, F., & Lacoste-Julien, S. (2014). SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27, pp. 1646–1654). Montreal, Quebec: Curran Asociates, Inc.

Denton, M. H., Borovsky, J. E., Skoug, R. M., Thomsen, M. F., Lavraud, B., Henderson, M. G., et al. (2006). Geomagnetic storms driven by ICME- and CIR-dominated solar wind. *Journal of Geophysical Research*, *111*, A07S07. https://doi.org/10.1029/2005JA011436

Dubyagin, S., Ganushkina, N., Kubyshkina, M., & Liemohn, M. (2014). Contribution from different current systems to SYM and ASY midlatitude indices. *Journal of Geophysical Research: Space Physics*, *119*, 7243–7263. https://doi.org/10.1002/2014JA020122

Dumbović, M., Vršnak, B., Čalogović, J., & Župan, R. (2012). Cosmic ray modulation by different types of solar wind disturbances. *Astronomy & Astrophysics*, *538*, A28. https://doi.org/10.1051/0004-6361/201117710

Echer, E., Gonzalez, W., & Tsurutani, B. (2011). Statistical studies of geomagnetic storms with peak Dst ≤ -50 nT from 1957 to 2008. *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(11), 1454–1459. https://doi.org/10.1016/j.jastp.2011.04.021

Feldman, W. C., Asbridge, J. R., Bame, S. J., Fenimore, E. E., & Gosling, J. T. (1981). The solar origins of solar wind interstream flows: Near-equatorial coronal streamers. *Journal of Geophysical Research*, *86*(A7), 5408–5416. https://doi.org/10.1029/JA086iA07p05408

Feminella, F., & Storini, M. (1997). Large-scale dynamical phenomena during solar activity cycles. *Astronomy & Astrophysics*, *322*, 311–319.

Gosling, J. T. (1993). The solar flare myth. *Journal of Geophysical Research*, *98*(A11), 18,937–18,949. https://doi.org/10.1029/93JA01896

Gosling, J. T., Hildner, E., Macqueen, R. M., Munro, R. H., Poland, A. I., & Ross, C. L. (1975). Direct observations of a flare related coronal and solar wind disturbance. *Solar Physics*, *40*(2), 439–448. https://doi.org/10.1007/BF00162390

Greaves, W. M. H., & Newton, H. W. (1929). On the recurrence of magnetic storms. (Plate 9.) *Monthly Notices of the Royal Astronomical Society*, *89*(7), 641–646. https://doi.org/10.1093/mnras/89.7.641

Holappa, L., Mursula, K., & Asikainen, T. (2015). A new method to estimate annual solar wind parameters and contributions of different solar wind structures to geomagnetic activity. *Journal of Geophysical Research: Space Physics*, *119*, 9407–9418. https://doi.org/10.1002/2014JA020599

Holappa, L., Mursula, K., Asikainen, T., & Richardson, I. G. (2014). Annual fractions of high-speed streams from principal component analysis of local geomagnetic activity. *Journal of Geophysical Research: Space Physics*, *119*, 4544–4555. https://doi.org/10.1002/2014JA019958

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *Journal of Geomagnetism and Geoelectricity*, *42*(11), 1249–1265. https://doi.org/10.5636/jgg.42.1249

Jian, L. K., Russell, C. T., & Luhmann, J. G. (2011). Comparing solar minimum 23/24 with historical solar wind records at 1 AU. *Solar Physics*, *274*(1–2), 321–344. https://doi.org/10.1007/s11207-011-9737-2

Jian, L. K., Russell, C. T., Luhmann, J. G., & Skoug, R. M. (2006a). Properties of interplanetary coronal mass ejections at one AU during 1995–2004. *Solar Physics*, *239*(1–2), 393–436. https://doi.org/10.1007/s11207-006-0133-2

Jian, L. K., Russell, C. T., Luhmann, J. G., & Skoug, R. M. (2006b). Properties of stream interactions at one AU during 1995–2004. *Solar Physics*, *239*(1), 337–392. https://doi.org/10.1007/s11207-006-0132-3

Jordanova, V., Matsui, H., Puhl-Quinn, P., Thomsen, M., Mursula, K., & Holappa, L. (2009). Ring current development during high speed streams. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(10), 1093–1102. https://doi.org/10.1016/j.jastp.2008.09.043

Kamide, Y., & Fukushima, N. (1971). Analysis of magnetic storms with DR indices for equatorial ring-current field. *Radio Science*, *6*(2), 277–278. https://doi.org/10.1029/RS006i002p00277

Kilpua, E. K. J., Balogh, A., von Steiger, R., & Liu, Y. D. (2017). Geoeffective properties of solar transients and stream interaction regions. *Space Science Reviews*, *212*(3–4), 1271–1314. https://doi.org/10.1007/s11214-017-0411-3

Krieger, A. S., Timothy, A. F., & Roelof, E. C. (1973). A coronal hole and its identification as the source of a high velocity solar wind stream. *Solar Physics*, *29*(2), 505–525. https://doi.org/10.1007/BF00150828

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, *6*(10). https://doi.org/10.1186/1758-2946-6-10

Laundal, K. M., & Richmond, A. D. (2017). Magnetic Coordinate Systems. *Space Science Reviews*, *206*(1-4), 27–59. https://doi.org/10.1007/s11214-016-0275-y

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, *16*, 2079–2102. https://doi.org/10.1029/2018SW002067

Loewe, C. A., & Prölss, G. W. (1997). Classification and mean behavior of magnetic storms. *Journal of Geophysical Research*, *102*(A7), 14,209–14,213. https://doi.org/10.1029/96JA04020

Love, J. J., & Gannon, J. L. (2009). Revised Dst and the epicycles of magnetic disturbance: 1958-2007. *Annales Geophysicae*, *27*(8), 3101–3131. https://doi.org/10.5194/angeo-27-3101-2009

Lühr, H., & Maus, S. (2010). Solar cycle dependence of quiet-time magnetospheric currents and a model of their near-Earth magnetic fields. *Earth, Planets and Space*, *62*(14), 843–848. https://doi.org/10.5047/eps.2010.07.012

Martini, D., Mursula, K., Orispaää, M., & Linthe, H.-J. (2015). Long-term decrease in the response of midlatitude stations to high-speed solar wind streams in 1914-2000. *Journal of Geophysical Research: Space Physics*, *120*, 2662–2674. https://doi.org/10.1002/2014JA020813

Maunder, E. W. (1904). Magnetic disturbances, 1882 to 1903, as recorded at the Royal Observatory, Greenwich, and their association with sun-spots. *Monthly Notices of the Royal Astronomical Society*, *65*(1), 2–18. https://doi.org/10.1093/mnras/65.1.2

Mursula, K., Martini, D., & Karinen, A. (2004). Did open solar magnetic field increase during the last 100 years? A reanalysis of geomagnetic activity. *Solar Physics*, *224*, 85–94. https://doi.org/10.1007/s11207-005-4981-y

Newell, P. T., & Gjerloev, J. W. (2012). SuperMAG-based partial ring current indices. *Journal of Geophysical Research*, *117*, A05215. https://doi.org/10.1029/2012JA017586

Nikolaeva, N. S., Yermolaev, Y. I., & Lodkina, I. G. (2011). Dependence of geomagnetic activity during magnetic storms on the solar wind parameters for different types of streams. *Geomagnetism and Aeronomy*, *51*(1), 49–65. https://doi.org/10.1134/S0016793211010099

Oughton, E. J., Hapgood, M., Richardson, G. S., Beggan, C. D., Thomson, A. W. P., Gibbs, M., et al. (2018). A risk assessment framework for the socioeconomic impacts of electricity transmission infrastructure failure due to space weather: An application to the United Kingdom. *Risk Analysis*, *39*(5), 1022–1043. https://doi.org/10.1111/risa.13229

Parker, C. (2013). On measuring the performance of binary classifiers. *Knowledge and Information Systems*, *35*(1), 131–152. https://doi.org/10.1007/s10115-012-0558-x

Pick, L. (2019). ClassifyStorms—An automated classifier for geomagnetic storm drivers based on machine learning techniques. V. 1.0.1. GFZ Data Services. https://doi.org/10.5880/GFZ.2.3.2019.003

Pick, L., & Korte, M. (2018). HMC Index. V. 1.0. GFZ Data Services. https://doi.org/10.5880/GFZ.2.3.2018.006

Pick, L., Korte, M., Thomas, Y., Krivova, N., & Wu, C.-J. (2019). Evolution of large-scale magnetic fields from near-Earth space during the last 11 solar cycles. *Journal of Geophysical Research: Space Physics*, *124*, 2527–2540. https://doi.org/10.1029/2018JA026185

Richardson, I. G., & Cane, H. V. (2012). Solar wind drivers of geomagnetic storms during more than four solar cycles. *Journal of Space Weather and Space Climate*, *2*, A01. https://doi.org/10.1051/swsc/2012001

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), 18,203–18,213. https://doi.org/10.1029/1999JA000400

Rostoker, G., & Fälthammar, C.-G. (1967). Relationship between changes in the interplanetary magnetic field and variations in the magnetic field at the Earth's surface. *Journal of Geophysical Research*, *72*(23), 5853–5863. https://doi.org/10.1029/JZ072i023p05853

Shen, X.-C., Hudson, M. K., Jaynes, A. N., Shi, Q., Tian, A., Claudepierre, S. G., et al. (2017). Statistical study of the storm time radiation belt evolution during Van Allen Probes era: CME- versus CIR-driven storms. *Journal of Geophysical Research: Space Physics*, *122*, 8327–8339. https://doi.org/10.1002/2017JA024100

Shore, R. M., Freeman, M. P., & Gjerloev, J. W. (2018). An empirical orthogonal function reanalysis of the northern polar external and induced magnetic field during solar cycle 23. *Journal of Geophysical Research: Space Physics*, *123*, 781–795. https://doi.org/10.1002/2017JA024420

Siscoe, G. L., Love, J. J., & Gannon, J. L. (2012). Problem of the Love-Gannon relation between the asymmetric disturbance field and Dst. *Journal of Geophysical Research*, *117*, A09216. https://doi.org/10.1029/2012JA017879

Snekvik, K., Tanskanen, E. I., & Kilpua, E. K. J. (2013). An automated identification method for Alfvénic streams and their geoeffectiveness. *Journal of Geophysical Research: Space Physics*, *118*, 5986–5998. https://doi.org/10.1002/jgra.50588

Snyder, C. W., Neugebauer, M., & Rao, U. R. (1963). The solar wind velocity and its correlation with cosmic-ray variations and with solar and geomagnetic activity. *Journal of Geophysical Research*, *68*(24), 6361–6370. https://doi.org/10.1029/JZ068i024p06361

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. *Annals of the International Geophysical Year*, *35*, 4–45.

Temerin, M., & Li, X. (2015). The Dst index underestimates the solar cycle variation of geomagnetic activity. *Journal of Geophysical Research: Space Physics*, *120*, 5603–5607. https://doi.org/10.1002/2015JA021467

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., et al. (2015). International Geomagnetic Reference Field: the 12th generation. *Earth, Planets and Space*, *67* (79). https://doi.org/10.1186/s40623-015-0228-9

Turner, N. E., Cramer, W. D., Earles, S. K., & Emery, B. A. (2009). Geoefficiency and energy partitioning in CIR-driven and CME-driven storms. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(10-11), 1023–1031. https://doi.org/10.1016/j.jastp.2009.02.005

Wilcox, J. M., & Ness, N. F. (1965). Quasi-stationary corotating structure in the interplanetary medium. *Journal of Geophysical Research*, *70*(23), 5793–5805. https://doi.org/10.1029/JZ070i023p05793

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

Yakovchouk, O. S., Mursula, K., Holappa, L., Veselovsky, I. S., & Karinen, A. (2012). Average properties of geomagnetic storms in 1932-2009. *Journal of Geophysical Research*, *117*, A03201. https://doi.org/10.1029/2011JA017093

Zhang, Y., Sun, W., Feng, X. S., Deehr, C. S., Fry, C. D., & Dryer, M. (2008). Statistical analysis of corotating interaction regions and their geoeffectiveness during solar cycle 23. *Journal of Geophysical Research*, *113*, A08106. https://doi.org/10.1029/2008JA013095