Using multiple signatures to improve accuracy of substorm identification

John D. Haiducek[1], Daniel T. Welling[2,3], Steven K. Morley[4], Natalia Yu. Ganushkina[3,5], and Xiangning Chu[6]

[1]Postdoctoral fellow, U.S. Naval Research Laboratory, Washington, DC, USA
[2]University of Texas at Arlington, Arlington, TX, USA
[3]Climate and Space Sciences, University of Michigan, Ann Arbor, MI, USA.
[4]Space Science and Applications (ISR-1), Los Alamos National Laboratory, Los Alamos, NM, USA.
[5]Finnish Meteorological Institute, Helsinki, Finland
[6]Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO

Key Points:

- Combining substorm onsets from multiple types of observations can produce a more accurate list of onset times than any single list
- The resulting onset list exhibits expected behavior for substorms in terms of magnetospheric driving and response
- SWMF has a weak, but consistent and statistically significant skill in predicting substorms

Corresponding author: John Haiducek, `jhaiduce@umich.edu`

Abstract

We have developed a new procedure for combining lists of substorm onset times from multiple sources. We apply this procedure to observational data and to magnetohydrodynamic (MHD) model output from 1-31 January, 2005. We show that this procedure is capable of rejecting false positive identifications and filling data gaps that appear in individual lists. The resulting combined onset lists produce a waiting time distribution that is comparable to previously published results, and superposed epoch analyses of the solar wind driving conditions and magnetospheric response during the resulting onset times are also comparable to previous results. Comparison of the substorm onset list from the MHD model to that obtained from observational data reveals that the MHD model reproduces many of the characteristic features of the observed substorms, in terms of solar wind driving, magnetospheric response, and waiting time distribution. Heidke skill scores show that the MHD model has statistically significant skill in predicting substorm onset times.

Plain Language Summary

Magnetospheric substorms are a process of explosive energy release from the plasma environment on the night side of the Earth. We have developed a procedure to identify substorms that uses multiple forms of observational data in combination. Our procedure produces a list of onset times for substorms, where each onset time has been independently confirmed by two or more observational datasets. We also apply our procedure to output from a physical model of the plasma environment surrounding the Earth, and show that this model can predict a significant fraction of the substorm onset times.

1 Introduction

Geomagnetic substorms consist of an explosive release of stored solar wind energy from the magnetotail, much of which is deposited in the ionosphere. Originally they were observed as an auroral phenomenon (e.g. Akasofu, 1964), consisting of sudden brightening of auroral emissions accompanied by rapid changes in their spatial distribution. It is now recognized that a rapid reconfiguration of the night-side magnetic field, consisting of a plasmoid release and dipolarization, is a fundamental component of the substorm process. The plasmoid release coincides with the formation of field-aligned currents, termed the substorm current wedge, connecting the auroral zone to the magnetotail (e.g. Kepko et al., 2015). When the concept of the current wedge was first introduced, it was imagined as a pair of equal and opposite currents entering and exiting the ionosphere at the same latitude but different longitudes. More recent work has shown evidence that the upward and downward currents may overlap in longitude (Clauer & Kamide, 1985), and that the real structure may involve multiple filaments of upward and downward current (Forsyth et al., 2014), possibly organized into localized regions of flow-driven current termed "wedgelets" (Liu et al., 2013). However, some doubt has been cast on the wedgelet model (Forsyth et al., 2014), and the manner in which wedgelets might contribute to filamentation remains an open question (Kepko et al., 2015). Similarly, the behavior of the earthward flow upon arrival at the inner magnetosphere has not been clearly determined from observations (Sergeev et al., 2012).

Other open questions remain regarding the conditions that lead to substorm onset, and the timing of events leading to and following from substorm onset. For instance, the question of how substorm onset is influenced by solar wind conditions has not been fully resolved, with some holding that some or all substorms are "triggered" by changes in solar wind conditions (e.g. Caan et al., 1977; Lyons et al., 1997; Russell, 2000; Hsu & McPherron, 2003, 2004), and others claiming that the observed characteristics of substorms can be explained without invoking solar wind triggering (e.g. S. K. Morley & Freeman, 2007; Wild et al., 2009; Freeman & Morley, 2009; Newell & Liou, 2011; Johnson

& Wing, 2014). Similarly, the question of where a substorm originates in geospace (magnetotail, ionosphere, or somewhere else) has remained open for a number of years (e.g. Korth et al., 1991; Angelopoulos et al., 2008; Rae et al., 2009; Henderson, 2009).

A major factor limiting progress on these questions is a lack of sufficient observational data, due to the need for simultaneous observations in particular locations, or simply the need for more complete spatial coverage of the magnetosphere. However, addressing this problem directly requires launching additional satellites with the required instrumentation, and this is a long and costly process. Global magnetohydrodynamic (MHD) models have the potential to address the problem of limited observational coverage by providing predictions of currents, velocities, and magnetic fields throughout the magnetosphere. These predictions can provide insights into magnetospheric dynamics that would require an impractically large number of spacecraft to obtain using observations alone. The ability of MHD simulations to shed light on substorm dynamics has been demonstrated already by a number of studies (e.g. S.-i. Ohtani & Raeder, 2004; Birn & Hesse, 2013; El-Alaoui et al., 2009). The capability of MHD models to provide a global, spatially resolved picture of the magnetosphere has been used in previous studies to shed light on cause and effect relationships relating to the evolution of a substorm (e.g. Zhu et al., 2004; Raeder et al., 2010). However, such results have been limited to single event studies or idealized test cases, which leaves open questions about the degree to which MHD models can reproduce substorm dynamics consistently and reliably. Despite years of application of MHD models to substorms, no MHD model has been rigorously validated with regard to its ability to predict substorm onsets.

Validating any model (MHD or otherwise) for substorm prediction is complicated by the fact that substantial disagreement remains within the community about what constitutes a substorm. While a general consensus exists around several of the main features of substorms, the community has not developed a set of criteria for identifying substorm onsets that is unambiguous, comprehensive, and widely agreed upon. This remains the case despite decades of attempts to clarify the salient characteristics of substorms (e.g. Akasofu, 1964, 1968; Akasofu & Meng, 1969; R. L. McPherron, 1970; R. L. McPherron et al., 1973; Pytte, Mcpherron, & Kokubun, 1976; Pytte, McPherron, et al., 1976; Caan et al., 1978; Rostoker et al., 1980; Hones, 1984; Lui, 1991; Baker et al., 1996; Rostoker, 2002; Sergeev et al., 2012; Kepko et al., 2015). As a result, different researchers studying the same time period often come to substantially different conclusions about what events should be considered substorms.

A major factor contributing to the sometimes discordant results obtained is the fact that substorms produce numerous observational signatures, most of which have substantial limitations. Although a substorm is generally regarded as a global phenomenon, many of its effects are localized in a particular region. As a result, gaps in observational data can easily prevent detection of a substorm. For instance, the sparse distribution of ground-based magnetometers can result in negative bay onsets not being detected (Newell & Gjerloev, 2011a). In situ observations are subject to similar limitations: Dipolarizations and plasmoids can only be detected when a satellite is on the night side of the Earth and in the right range of distance, MLT sector, and latitude. Moreover, a plasmoid that propagates too slowly relative to the observing spacecraft might go unnoticed (Nishida et al., 1986). At the same time, many observational features used to identify substorms can be created by other processes, resulting in false positives. For instance, single-satellite observations may not be able to distinguish a plasmoid from other transient features in the current sheet (such as thickening, thinning, or bending) (Eastwood et al., 2005). A storm sudden commencement can result in a negative bay at auroral magnetometers (Heppner, 1955; Sugiura et al., 1968), as can a pseudobreakup (Koskinen et al., 1993; S. Ohtani et al., 1993; Aikio et al., 1999; Kullen et al., 2009). A discussion of the challenges faced by researchers in distinguishing different magnetospheric phenomena from each other can be found in R. L. McPherron (2015).

Differences in results obtained when different observational datasets are used can be substantial. An illustrative example is Boakes et al. (2009), which compared substorm onsets previously published by Frey et al. (2004) based on analysis of auroral images with energetic particle observations at geosynchronous orbit. Boakes et al. (2009) found that 26% of the auroral expansion onsets had no corresponding energetic particle injection even though a satellite was in position to detect such an injection, and suggested that such events might not be substorms.

The difficulty in positively identifying substorm onsets presents a problem for validation of substorm models. In the absence of a definitive substorm onset list against which to validate a model, those seeking to validate a substorm prediction model are left to choose among the published lists, or create a new one. Given the substantial differences between the existing onset lists, validation against any single onset list leaves open the question of whether the validation procedure is testing the model's ability to predict substorms, or merely the model's ability to reproduce a particular onset list, whose contents may or may not really be substorms.

One potential way to address the problems of onset list accuracy is to use multiple substorm signatures in combination, checking them against each other to remove false positives and avoid missed identifications. The resulting consensus list may prove more reliable than any of its constituent lists, providing a more comprehensive and trustworthy set of onsets. Comparing two or three substorm signatures by hand for individual events has been commonplace since the beginning of substorm research (e.g. Akasofu, 1960; Cummings & Coleman, 1968; Lezniak et al., 1968), and a number of researchers have produced statistics comparing onset lists for two or more substorm signatures (e.g. Moldwin & Hughes, 1993; Boakes et al., 2009; Liou, 2010; Chu et al., 2015; Forsyth et al., 2015; Kauristie et al., 2017). R. L. McPherron and Chu (2017) demonstrated that a better onset list could be obtained using the midlatitude positive bay (MPB) index and the SML index together than by using either dataset alone.

Despite an awareness within the community that multiple observational signatures are required to positively identify a substorm, R. L. McPherron and Chu (2017) has been the only work to date that uses multiple signatures to create a combined onset list, and no attempt to create an onset list using more than two different signatures has been published. This may in part be due to the complexities involved in doing so. As was discussed earlier, the absence of a particular signature does not always indicate the absence of a substorm, while at the same time some identified signatures may not in fact be substorms. Ideally a combined list should somehow allow for these possibilities and correct for them. Further complicating matters is the fact that different signatures may be identified at different times for the same substorm (e.g. Rae et al., 2009; Liou et al., 1999, 2000; Kepko, 2004).

In the present work we present a new procedure which uses multiple substorm signatures to identify substorm onsets. By using multiple datasets consisting of different classes of observations, we reduce the risk of missing substorms due to gaps in individual datasets. At the same time, the new procedure aims to reduce false identifications by only accepting substorm onsets that can be identified by multiple methods. Our procedure is generalizable to any combination of substorm onset signatures, and allows for the possibility that the signatures may not be precisely simultaneous. We demonstrate the technique on observational data from January, 2005. We present evidence that the procedure is successful at reducing false identifications while avoiding missed identifications due to observational data gaps, and that the resulting onset list is consistent with the known characteristics of substorms. Finally, we demonstrate the technique on output from an MHD simulation of the same January, 2005 time period, and show preliminary evidence of predictive skill on the part of the MHD model.

2 Methodology

2.1 Identification of substorm events from combined signatures

Our procedure for combining multiple substorm onset lists consists of first convolving each onset list with a Gaussian kernel. The result of this convolution is re-scaled using an error function (erf) in order to keep the values bounded by 1. The re-scaled convolutions of the onset lists are then summed together to produce a nominal "substorm score." For a series of onset times τ_{ij} from a set of onset lists i, this score is given by

$$f(t) = \sum_{i=1}^{n_{sigs}} \operatorname{erf}\left(\sum_{j=1}^{n_{onset}} \exp\left(-\frac{(t-\tau_{ij})^2}{2\sigma^2} \right) \right), \tag{1}$$

where σ is a tunable kernel width. The i's each represent a particular substorm onset list. The onset lists each represent a distinct substorm signature and are described in detail in Sections 2.4 and 2.5. The j's represent the onset times in each onset list. To obtain a list of onset times, we search for local maxima in the score $f(t)$, and keep any maxima that rise above a specified threshold T. We apply this procedure to the onset lists produced from the simulation, and separately apply the procedure to the observational data.

The process is illustrated in Figure 1 for the 24-hour time period of 31 January, 2005. Figure 1 was created using a kernel width $\sigma = 13.8$ minutes and a threshold $T = 1.6$. These values were selected using an optimization process that will be described later. The specifics of how the signatures were identified will be discussed in Section 2.4, but to illustrate the convolution procedures it suffices to say that a list of candidate onset times was identified separately for each signature. Figures 1a-1e show the scores obtained from the onset list obtained from each signature. Figure 1f shows the sum of the scores in Figures 1a-1e. The threshold value T is drawn in red, and vertical dashed lines mark the onset times identified from local maxima of the combined score that exceed the threshold. In order to exceed the threshold, signatures from two different lists must occur within a few minutes of each other, and this occurred seven times during the time period shown in Figure 1.

It is worth noting that the individual onset lists in Figure 1 are substantially different from each other, each identifying substorms at different times from the others, and two including candidate onset times that are not near those in any other list. As long as a value of $T > \operatorname{erf}(1)$ is used, our procedure rejects those onsets, such as the dipolarization around 1300 UT and the AL onset around 1400 UT, which appear only in one list. Onsets are then counted only if two or more occur close enough in time to each other that the score rises above the threshold T. For the value $T = 1.6$ used in this illustration, onsets from two different lists falling within approximately 0.89σ of each other will produce a peak that exceeds T. Reducing the threshold from $T = 1.6$ would tend to increase the total number of substorm identifications, while increasing it would tend to lower the number of substorm identifications. The implications of changing the threshold will be explored further in Section 3.2. Note also that if the score remains above the threshold for a period of time and multiple local maxima are found within that period, all of them are counted as substorm onsets. For example, the local maxima around 1130 UT and a second one just before 1200 UT are both counted as substorm onsets.

In general, increasing T will make the list more restrictive and shorter, while decreasing T will make the list less restrictive. However, any local maximum in $f(t)$ will have a value of at least $\operatorname{erf}(1) \approx 0.843$, so any threshold $T < \operatorname{erf}(1)$ will produce the least restrictive onset list possible for a given kernel width σ, and further reduction of T will have no affect on the resulting list. If we choose a threshold $T > \operatorname{erf}(1)$, we effectively require at least two signatures to identify a substorm onset. The temporal sep-

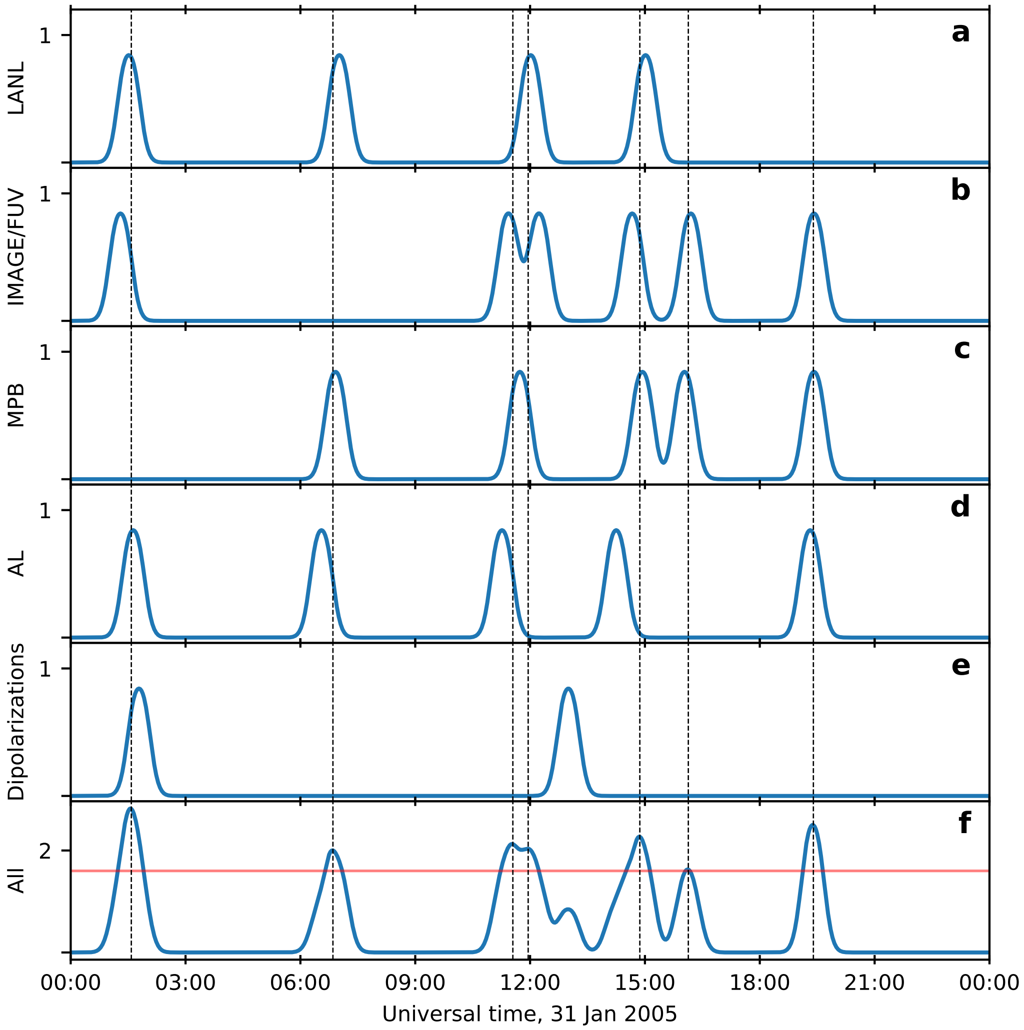

Figure 1. An illustration of the procedure used to combine multiple substorm onset lists into a single one. Panels (a-e) show scores obtained by convolving individual onset lists with a Gaussian kernel (using $\sigma = 13.8$ minutes), while (d) shows the combined score obtained by adding together the scores in panels (a-e). The threshold $T = 1.6$ is marked with a red horizontal line, and vertical dashed lines are drawn through local maxima of the combined score that exceed this threshold.

aration between these signatures must be small enough that their respective kernels overlap significantly. However, one cannot in general identify a specific maximum separation that determines this threshold. Rather, the threshold T determines the minimum height of the peak and therefore influences the maximum separation between signatures contributing to a single onset in the combined list.

Even if the threshold is set below $\mathrm{erf}(1)$, so that every local maximum in $f(t)$ is included in the combined list, the convolution process will result in combining some signatures that occur near each other. In order for two signatures to be counted as independent onsets (without any additional nearby signatures) they must be separated by more than approximately 2.55σ so that two local maxima can form in the resulting function $f(t)$. Smaller separations than this will result in a single local maximum that falls between the two signatures. If more than two signatures occur within the same vicinity, smaller separations can give two maxima in f. For instance, onsets at 0, 1.6σ, and 3.1σ from three separate lists will result in two local maxima in f. Thus, the number of subordinate onset lists, and the choice of T and σ interact with each other to influence the characteristics of the resulting onset list. The implications of the choice of threshold T and kernel width σ will be explored further later in the paper.

The convolution process effectively acts as a low-pass filter, with the choice of σ determining the minimum time between successive onsets. As discussed in the introduction, different substorm signatures may not be detected simultaneously even if they are related to the same substorm. For instance, Liou et al. (1999) and Liou et al. (2000) found geosynchronous energetic particle injections tended to lag the onset of auroral breakup by 1-3 minutes, while the high-latitude magnetic bay can be delayed up to tens of minutes relative to the onset of auroral breakup. Some of the findings of Liou et al. (2000) were challenged by Kepko and McPherron (2001) and Kepko (2004), but even Kepko (2004) found that Earthward plasma flows could precede auroral onset by 1-3 minutes. These results and others suggest that a kernel width of $\sigma \approx 3$ minutes represents a lower bound for appropriate values of σ, unless the analysis is restricted to a set of observational signatures that have been shown to occur nearly simultaneously. An upper end of the appropriate range for σ can be identified by noting that previous research has shown that successive substorms rarely occur within 30 minutes of each other (e.g. Borovsky et al., 1993; Frey, 2010). This suggests that σ should be chosen to be under 30 minutes, but leaves substantial room for tuning.

Some of the underlying onset lists could have onsets occurring close enough that their kernel functions overlap substantially. Scaling the convolved scores using the error function $\mathrm{erf}(x)$ helps prevent an onset list with closely spaced signatures from contributing too strongly to the combined list. If two signatures occur simultaneously in the same onset list, this could indicate a greater confidence in the signature, but this should arguably not be weighted as strongly as two independent signatures from separate datasets. The erf function is approximately linear for small values, so that the general shape of the Gaussian kernel is retained except for an approximately 15.7% reduction in the height of the peak. If two signatures occur at the same time in the same list, the resulting peak height is only 0.995, a 15.3% increase from the single-signature case. If three or more signatures occur simultaneously in the same underlying list, the result is an even smaller increase as the peak height asymptotically approaches 1. Thus an isolated signature in one of the underlying onset lists contributes significantly to the total score, but multiple closely-spaced detections of the same signature do not cause that signature to dominate the combined onset list.

2.2 Event description

To test our technique we selected the month of January, 2005. S. K. Morley (2007) and S. Morley et al. (2009) had previously identified substorms from this time period, and from the data analyzed in those papers this time period was determined to have a sufficient number of substorms to enable statistical analysis. The substorm database provided by the SuperMag collaboration (http://supermag.jhuapl.edu/substorms/) (Gjerloev, 2012), which contains onsets identified from the SML index (Newell & Gjerloev, 2011b) using the Newell and Gjerloev (2011a) algorithm, lists 322 substorms during this period, placing it in the top 3% of 31-day periods included in that dataset. The substorm onset lists from Borovsky and Yakymenko (2017) include 124 AL onsets and 109 energetic particle injections during January, 2005, placing that month in the top 3% in terms of AL onsets and in the top 7% in terms of energetic particle injections, compared with other 31-day periods from the same onset lists. Frey et al. (2004) (whose list has subsequently been updated to include 2003-2005 and published online at http://sprg.ssl.berkeley.edu/image/) lists 97 substorms in January 2005, placing the month in the top 13% of 31-day periods in that dataset. Chu et al. (2015) found 167 onsets during this month, placing it in the top 9% of 31-day intervals analyzed in that paper. Forsyth et al. (2015) found 356 onsets during this month, placing it in the top 6% of 31-day intervals in that dataset (here, we use the middle of three lists included in the supporting information of that paper, with an expansion threshold of 75%). In addition, two of the "supersubstorms" ($AL < -2500$ nT) identified by Hajra et al. (2016) occurred during this time period.

Three geomagnetic storms occurred during this month: One on January 7 with a minimum Sym-H of -112 nT, one on January 16 with a minimum Sym-H of -107 nT, and one on January 21 with a minimum Sym-H of -101 nT. A table of the minima, maxima, and quartiles of various observed quantities over the course of the month can be found in Haiducek et al. (2017). Of particular note is the consistently high solar wind speed (median solar wind speed was 570 km/s), which may have contributed to the relatively high frequency of substorms during this period.

2.3 Model description

The simulations presented in this work were performed using the Block-Adaptive Tree Solar-Wind, Roe-Type Upwind Scheme (BATS-R-US) MHD solver (Powell et al., 1999; De Zeeuw et al., 2000). This was coupled to the Ridley Ionosphere Model (RIM, Ridley et al., 2003; Ridley et al., 2004) and the Rice Convection Model (RCM, Wolf et al., 1982; Sazykin, 2000; Toffoletto et al., 2003). The Space Weather Modeling Framework (SWMF, Tóth et al., 2005, 2012) provided the interface between the different models. The model settings and grid configuration for the simulation are described in detail in Haiducek et al. (2017), which includes results from the same simulation. (In Haiducek et al. (2017) the simulation was referred to as "Hi-res w/ RCM" to distinguish it from the other two simulations included in that paper.) The inputs to the model are solar wind parameters (velocity, magnetic field, temperature, and pressure) and F10.7 radio flux. Solar wind parameters were obtained from the OMNI dataset, supplemented with data from the ACE spacecraft as described in Haiducek et al. (2017). Data from the ACE SWEPAM instrument used in this process, as well as the solar wind input file used with SWMF, is provided in the supplemental data. The results of Haiducek et al. (2017) showed that the simulation produced good predictions of the Sym-H, AL, and Kp indices on average. On the other hand, the model was found to under-predict the frequency of occurrence for strongly negative AL values, suggesting a tendency to under-predict the strength or occurrence rate of substorms.

2.4 Identification of model signatures

The substorm process results in numerous observational signatures that can be leveraged for identification. These include plasmoid releases, magnetic perturbations observable in the auroral zone and at mid latitudes, dipolarization of night-side magnetic fields observable from geosynchronous orbit, Earthward injection of energetic particles, and auroral brightenings. Several of these can be synthesized using MHD as well. Unfortunately, as was discussed in the introduction, all of these signatures can be produced by other processes besides substorms, and this is true for both the observations and the model output. For instance, magnetospheric convection, pseudobreakups and poleward boundary intensifications can cause a negative bay response in the northward magnetic field component at auroral-zone magnetometers, which could be interpreted as substorm onsets (Pytte et al., 1978; Koskinen et al., 1993; S. Ohtani et al., 1993; Aikio et al., 1999; Kim et al., 2005). On the other hand, substorms could occur but not be identified because of the limited spatial coverage of observational data, as was shown by Newell and Gjerloev (2011a) for auroral-zone magnetic field. Substorms could also be missed simply because they produce a response below the threshold selected for analysis (e.g. Forsyth et al., 2015). Even for analysis of model output, many of these factors remain relevant, and we aim to mitigate this by using multiple signatures to identify our substorms. Specifically, we identify dipolarization signatures at 6-7 R_E distances (Nagai, 1987; Korth et al., 1991), negative bays in the AL index (Kamide et al., 1974; Newell & Gjerloev, 2011a; Borovsky & Yakymenko, 2017), positive bays in the midlatitude positive bay (MPB) index (Chu et al., 2015), and plasmoid releases (Hones et al., 1984; Ieda et al., 2001).

Figure 2 shows examples of substorm signatures from a substorm event on January 2, 2005. This substorm was selected for illustrative purposes because it can be identified by all four of the signatures used in the model output. A handful of previous researchers have identified substorm onsets during the time period shown in the plot (2000-2200 UT). Borovsky and Yakymenko (2017) found an AL onset at 2026 UT on this day, and a geosynchronous particle injection at 2130 UT. Chu et al. (2015) identified an MPB onset at 2112 UT. The SuperMag substorm database (populated using the Newell and Gjerloev (2011a) algorithm) contains onsets at 2016, 2038, and 2059 UT. Figures 2a-2c show time-series plots of B_z at $x = -7\ R_E$ (GSM), the AL index, and the MPB index. Apparent onset times identified from each curve are marked by triangles. Figures 2d-2f show the MHD solution within the x-z (GSM) plane at 5-minute intervals during a plasmoid release. The backgrounds of Figures 2d-2f are colored according to the plasma pressure. Closed magnetic field lines are plotted in white, and open field lines in black. The Earth is shown as a pair of black and white semicircles, and surrounded by a grey circle denoting the inner boundary of the MHD domain. The approximate location of the reconnection region is denoted by a red triangle, and a blue dot marks where x=-7 R_E along the noon-midnight line (this is the location from which the data in Figure 2a was obtained).

2.4.1 Plasmoid release

A fundamental characteristic of a substorm is the tailward release of a plasmoid(e.g. Hones et al., 1984), and this is the first substorm signature we will describe. In observations, plasmoids are identified by a bipolar variation of B_z as observed by a spacecraft near the central plasma sheet (e.g. Slavin et al., 1989, 1992; Ieda et al., 2001; Eastwood et al., 2005). MHD models provide data throughout the magnetosphere rather than being limited to a few point observations, and this enables several additional techniques for identifying plasmoids. One approach is to plot variables such as temperature, velocity, and magnetic field over time for different x coordinates along a line through the central plasma sheet at midnight. This produces a 2-D map showing the time evolution of the MHD solution in the plasma sheet, in much the same way that keograms are used to visualize the time evolution of auroral emissions (Raeder et al., 2010). Plasmoids appear in such maps as tailward propagating magnetic field perturbations, with corresponding tailward flow velocity. Another approach for identifying plasmoids was proposed by Honkonen et al. (2011), who used the magnetic field topology derived from an MHD simulation to identify a plasmoid, which they define as a set of closed field lines that enclose a region of reconnecting open field lines. Probably the most common method is to plot magnetic field lines in the x-z plane, looking for evidence of a flux rope in the form of wrapped up or self-closed field lines, as in e.g. Slinker et al. (1995).

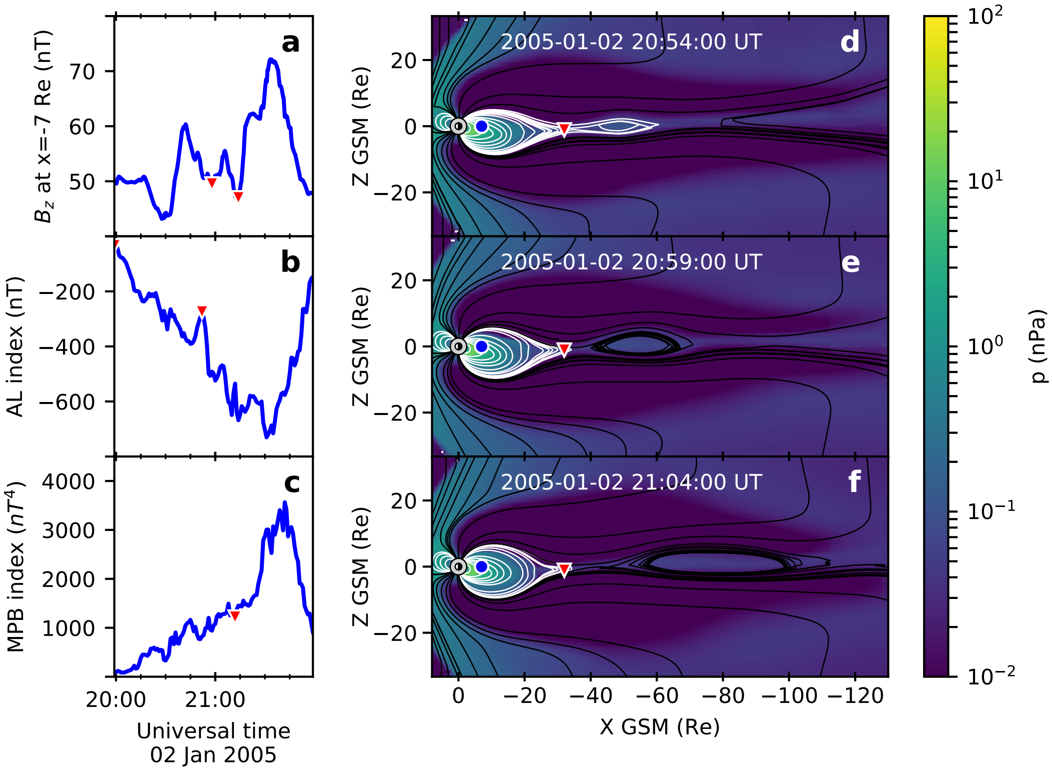

Figure 2. Model signatures for an example substorm. (a) B_z variations at $x = -7\ R_E$ along the GSM x axis. (b) AL index. (c) MPB index. Apparent substorm onset times are marked with triangles in (a-c). (d-f) $x-z$ (GSM) cut planes, at 5-minute intervals, colored by pressure. Closed magnetic field lines are drawn in white, and open field lines in black. Earth is drawn as a pair of black and white semicircles, surrounded by a grey circle denoting the inner boundary of the MHD domain. The location x = -7 R_E, from which the data in (a) was obtained, is marked a blue circle. The apparent X-line location is marked with a red triangle.

The method of visually identifying plasmoids by searching for regions of wrapped-up field lines is the one used in the present work. We require that such features be located in or near the central plasma sheet, and that they exhibit tailward motion. For each such plasmoid, we record the time of the first indication of tailward motion, and the x and z coordinates of the apparent X-line at that time. Plasmoids for which the X-line is beyond 35 R_E down-tail are ignored. Figures 2d-2f show examples of the images that are used for this analysis. For the event in Figure 2, the first apparent tailward motion occurred at 2054 UT, and this time is shown in Figure 2d. The X-line occurs at around x=-32 R_E, and the plasmoid extends from there to -60 R_E. Figures 2e and 2f show the same plasmoid 5 and 10 minutes after release. Tailward motion is clearly apparent, with the center of the plasmoid moving from $x \approx -55$ to $x \approx -80$ R_E in 10 minutes.

2.4.2 Dipolarization

While the plasmoid propagates tailward, the magnetic fields Earthward of the X-line undergo a dipolarization. Previous studies have identified dipolarizations by searching for sharp increases in B_z (e.g. Lee & Lyons, 2004; Runov et al., 2009; Birn et al., 2011; Runov et al., 2012; Liu et al., 2013; Frühauff & Glassmeier, 2017) or elevation angle

$$\theta = \tan^{-1}\left(\frac{B_z}{\sqrt{B_x^2 + B_y^2}}\right) \tag{2}$$

(e.g. R. L. McPherron, 1970; Coroniti & Kennel, 1972; Noah & Burke, 2013) within the night-side magnetotail. A number of studies have also used a decrease in

$$|B_r| = \left|\frac{xB_x + yB_y}{\sqrt{x^2 + y^2}}\right|, \tag{3}$$

coincident with the increase in B_z or θ, as criteria for identifying a dipolarization onset (e.g. Nagai, 1987; Korth et al., 1991; Schmid et al., 2011; Liou et al., 2002). Automated procedures for identifying dipolarizations have been developed by Fu et al. (2012) and Liu et al. (2013). We found the Fu et al. (2012) algorithm unsuitable for our purposes because it uses flow velocity as part of its criteria, for which we had no observational data from the GOES satellites used in the analysis. The Liu et al. (2013) algorithm was designed for THEMIS and uses B_z alone for event selection. Since our data was from 6-7 R_E from the Earth (where the fields differ substantially from those seen by THEMIS), we developed a new algorithm which uses variations in B_z, $|B_r|$, and θ to identify dipolarizations from the model output. The new procedure is described in detail in Appendix A. The algorithm was used to identify dipolarization signatures along the orbits of GOES 10 and 12, and at a fixed point located at $x = -7$ R_E in GSM coordinates on the sun-Earth line; this point is identified by a blue circle in Figures 2d-2f. A plot of B_z at $x = -7$ R_E is shown in Figure 2a, and two dipolarization onsets identified using our procedure are marked on the plot with triangles. The first of these is closely aligned with the plasmoid release time.

2.4.3 Auroral-zone negative bay

The dipolarization process can be interpreted as a partial redirection of cross-tail current into the ionosphere (e.g. Bonnevier et al., 1970; R. L. McPherron et al., 1973; Kamide et al., 1974; Lui, 1978; Kaufmann, 1987). The ionospheric closure of this current results in a negative bay in the northward component of the magnetic field on the ground in the auroral zone (Davis & Sugiura, 1966). As a result, substorm onsets can be identified by sharp negative diversions of the AL index. A number of algorithms have previously been developed for identifying substorm onsets from the AL index, including the Newell and Gjerloev (2011a) (SuperMag) algorithm and the Substorm Onsets and Phases from Indices of the Electrojet (SOPHIE) algorithm (Forsyth et al., 2015).

In the present paper we identify AL onsets using the algorithm presented in Borovsky and Yakymenko (2017). This algorithm was chosen for its simplicity and because it produces a distribution of inter-substorm timings that is consistent with that obtained from other signatures, as Borovsky and Yakymenko (2017) demonstrated through comparison with timings of energetic particle injections. We apply the Borovsky and Yakymenko (2017) algorithm to a synthetic AL index computed from the model output using virtual magnetometers as described in Haiducek et al. (2017). An example AL onset is shown in Figure 2b. A negative bay onset, marked by a triangle, occurs just before 2100 UT, just after the plasmoid release at 2054 UT.

2.4.4 Midlatitude positive bay

The integrated effect of the currents closing between the tail and auroral zone results in a northward diversion of the ground magnetic field in the mid latitudes, called a midlatitude positive bay (MPB, R. L. McPherron et al., 1973). Often MPB's are identified manually through examination of individual magnetometers (e.g. R. McPherron, 1972; R. L. McPherron et al., 1973; Caan et al., 1978; Nagai et al., 1998; Forsyth et al., 2015). However, the ASYM-H index may also be used (Iyemori & Rao, 1996; Nosé et al., 2009). More recently, Chu et al. (2015) and R. L. McPherron and Chu (2017) have developed procedures to compute what they call the MPB index, which is specifically designed to respond to a midlatitude positive bay, along with procedures for identifying substorm onsets using the MPB index. In the present paper we use the MPB index implementation described in Chu et al. (2015) and its accompanying onset identification procedure. To evaluate the MPB index from the model output, we use a ring of 72 virtual magnetometers placed at a constant latitude of 48.86° and evenly spaced in MLT. We compute estimated magnetic fields for the locations of these magnetometers by performing a Biot-Savart integral over the entire MHD domain, and to this add the contributions of the Hall and Pedersen currents computed using RIM; this procedure is described in Yu and Ridley (2008); Yu et al. (2010). Using the estimated magnetic fields at these virtual magnetometer locations, we compute the MPB index and associated substorm onsets using the procedures described in Chu et al. (2015). An example of the MPB response is shown in Figure 2c. The MPB onset time occurs roughly 10 minutes after the plasmoid release time, but is well aligned with the second of the two dipolarizations in Figure 2a.

2.5 Identification of substorm events from observational data

When possible, we use the same procedures to identify substorm signatures in the observational data as we do with the model output. This includes the dipolarizations, AL index, and MPB index. In some cases modifications are required due to limitations in the availability of observational data; for instance ground-based magnetometers are normally restricted to being placed on land with suitable terrain, and the locations of satellite observations are constrained by orbital mechanics. On the other hand, some observations rely on physical phenomena that cannot be modeled by the MHD code, such as energetic particle injections and auroral brightenings. In an effort to obtain the best possible identifications of observed substorms, we use as many observational datasets as possible, which for this time period included GOES magnetic field observations, the AL and MPB indices, energetic particle injections at geosynchronous orbit, and auroral brightenings.

We identify AL onsets by applying the procedure from Borovsky and Yakymenko (2017) to the SuperMag SML index (Newell & Gjerloev, 2011a). For simplicity, we will use the term AL throughout the paper to refer to both the observed SML index and the synthetic AL computed from the model output. For the observed MPB index and observed MPB onset times we use the values from the analysis previously published in Chu et al. (2015). We identify dipolarizations by applying the procedure described in Appendix A to measurements obtained with the magnetometers onboard GOES 10 and 12 (Singer et al., 1996).

In addition to the dipolarization, another substorm signature that can be observed at geosynchronous orbit is the Earthward injection of energetic electrons and protons (e.g. Lezniak et al., 1968; DeForest & McIlwain, 1971). Previous studies have identified a temporal association between such particle injections and auroral zone magnetic signatures (e.g. Lezniak et al., 1968; Kamide & McIlwain, 1974; Weygand et al., 2008), along with a connection between energetic particle injections and dipolarizations (e.g. Sauvaud & Winckler, 1980; Birn et al., 1998). In the present work we use energetic particle injections identified by Borovsky and Yakymenko (2017) using the Synchronous Orbit Particle Analyzer (SOPA) instrument (Cayton & Belian, 2007) on the LANL-1990-095, LANL-1994-085, and LANL-97A satellites. The list of particle injections found in the supplemental data of Borovsky and Yakymenko (2017) is used as-is.

Some of the energetic particles produced by the substorm enter the ionosphere and cause a brightening and reconfiguration of the aurora. These can be observed from the ground using all-sky imagers, or from cameras onboard spacecraft. For the month of January, 2005, observations from the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) spacecraft are available for this purpose. The IMAGE spacecraft was in a highly elliptical polar orbit with an apogee of 45,600 km and an orbital period of 14 hours, providing 8-10 hours per orbit of good conditions for imaging the northern auroral oval (Frey et al., 2004). Frey et al. (2004) examined images from the Far Ultraviolet Imager (FUV) instrument onboard IMAGE, and produced a list of northern hemisphere substorm onsets for the years 2000-2002, since updated to include 2003-2005 and available online at `http://sprg.ssl.berkeley.edu/sprite/ago96/image/wic_summary/substorms/`. We use the January, 2005 portion of this list as part of our substorm identification.

3 Results

3.1 Substorm waiting times

The distribution of substorm waiting times (the amount of time that passes between successive substorms) gives an indication of the occurrence frequency for substorms. A number of previous papers have examined waiting times, including Borovsky et al. (1993) which identified substorm onsets from energetic particle injections and found the modal waiting time to be around 2.75 hours. Chu et al. (2015) and R. L. McPherron and Chu (2017) analyzed MPB onsets and reported modal waiting times of 80 and 43 minutes, respectively. Kauristie et al. (2017) reported modal waiting times of 32 minutes for AL onsets identified by Juusola et al. (2011) and 23 minutes for SML onsets identified by the Newell and Gjerloev (2011a) procedure. Hsu and McPherron (2012) obtained a modal waiting time of about 1.5 hours for AL onsets, about 2 hours for onsets identified from tail lobe fields, and about 2.5 hours for Pi 2 onsets. Freeman and Morley (2004) reproduced the waiting time distribution from Borovsky et al. (1993) using a solar wind driven substorm model.

To visualize the distributions of waiting times, we use kernel density estimates (KDEs) (Parzen, 1962), which approximate the probability density function of a distribution by convolving samples from the distribution with a Gaussian kernel. The resulting curve can be interpreted in the same way as a normalized histogram. The width of the kernel is scaled using the standard deviation of the data multiplied by a scaling factor $b = 0.7$ (see Appendix D for details). Since the waiting times can take only positive values, while the Gaussian kernels used in the KDE give nonzero probabilities for negative values, we perform the KDE in logarithmic space and transform the result to linear space for plotting as described in Appendix C. For some of our KDE plots we have estimated confidence intervals using a bootstrapping procedure described in Appendix D. This provides a means to assess whether the waiting time distribution obtained from the model is significantly different from the observed distribution, in a statistical sense.

To test the sensitivity of the waiting time distributions to the choice of kernel width and threshold, we plotted waiting time distributions for a range of each parameter, as shown in Figure 3. Figure 3 shows the distribution of waiting times for the model and for the observations using three different choices of threshold and four different kernel widths, ranging from $\sigma = 5$ minutes to $\sigma = 30$ minutes. We found that values of $\sigma <$ 5 minutes resulted in a severe decrease in the number of substorms in the combined list, while $\sigma \gtrsim 30$ minutes risks merging unrelated substorm onsets together. The y-axis of each panel shows the probability densities of waiting time, and the x axis shows the waiting times. Figures 3a, 3b, and 3c show waiting time distributions from the observations, while Figures 3d, 3e, and 3f show waiting time distributions obtained from the MHD simulation. Figures 3a and 3d show thresholds of 1.0, Figures 3b and 3e show thresholds of 1.5, and Figures 3c and 3f show thresholds of 2.0. Within each plot, the kernel width

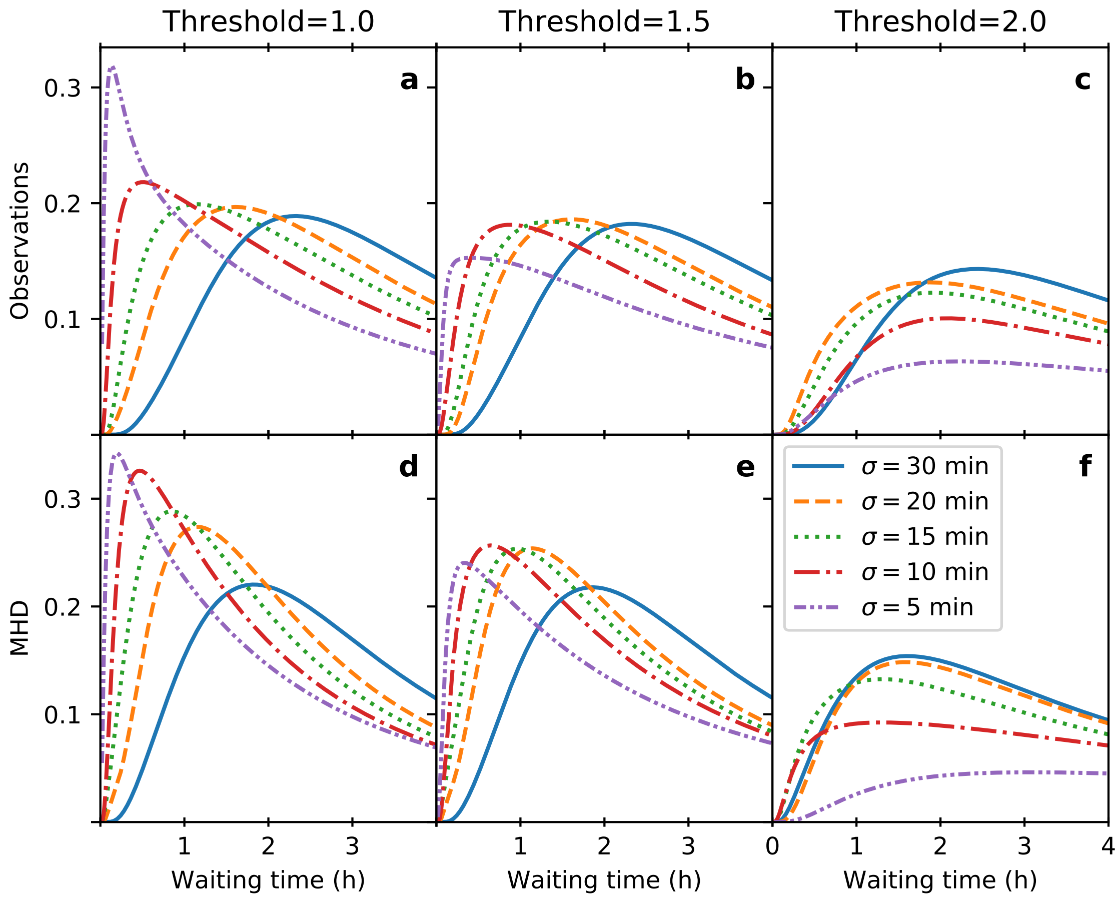

Figure 3. Distributions of substorm waiting times for a range of identification thresholds and kernel widths used in the identification procedure. a), b), and c): Observed waiting time distributions. d), e), and f): MHD waiting time distributions. a) and d): Threshold=1.0; b) and e): Threshold=1.5; c) and f): Threshold=2.0.

σ used in the substorm identification procedure is varied from $\sigma = 5$ minutes to $\sigma = 30$ minutes. $\sigma = 5$ minutes in purple with a dash-dot-dot pattern, $\sigma = 10$ minutes is plotted in red with a dash-dot pattern, $\sigma = 15$ minutes in green with dots, $\sigma = 20$ minutes in orange with dashes, and $\sigma = 30$ minutes in blue with a solid line.

From Figure 3, it is apparent that both the threshold and the kernel width affect waiting time distributions substantially. The modal waiting time varies from approximately 0.25 to 2.5, while the height of the peak varies from greater than 0.3 to less than 0.1. Note that, as discussed in Section 2.1, any threshold $T \lesssim 0.843$ will produce an identical onset list for a given kernel width σ; because of this we chose thresholds $T > 0.843$ for all parts of Figure 3. As the threshold is increased, we expect the waiting times to increase as onset times are removed from the combined list. Figure 3 shows that this is the case. For a given choice of σ, the modal waiting time tends to increase as the threshold is increased from 1 to 2. This is particularly noticeable for the shortest kernel width $\sigma = 5$. For $\sigma = 5$ and $T = 1.0$, the modal waiting time begins at less than a half hour in both the model and the observations. When T is increased to 2.0, the modal waiting increases to approximately two hours for the observations and three hours for the model. At the same time, the height of the peak decreases as shorter waiting times at the left of the peak give way to longer waiting times in the tail of the waiting time distribution.

The influence of σ on the waiting time distribution is somewhat more complicated and depends on the value of T. For the lower threshold of $T = 1.0$, increasing σ results in an increase in the modal waiting time and a decrease in the peak height. This suggests that larger values of σ are causing nearby peaks to merge. As noted in Section 2.1, the practice of selecting by local maxima results in a merging of signatures whose separation is less than a certain multiple of σ (for two signatures, they will be merged if they fall within 2.55σ). Increasing σ may cause more signatures to be merged in this way, and this can result in a decrease in the number of substorms and an increase in the waiting times, as seen in Figures 3a and 3d.

For higher values of T, increasing σ can sometimes cause an increase in the number of substorms rather than a decrease, and can decrease the waiting times as well. This is because as σ is increased, the height of the peaks tend to increase as the sphere of influence for each signature increases with σ. The effect of increasing σ causing nearby signatures to merge into a single onset still applies at the higher thresholds, but σ and T seem to interact to influence the waiting time distribution in sometimes complicated ways. While for a threshold of 1.5 (Figures 3b and 3e) the modal waiting time increases monotonically with increasing σ, for a threshold of 2.0 (Figures 3c and 3f) it does not. (Note, however, that for the $T = 2.0$ cases the total number of substorms contributing to the waiting time distributions is fewer than 100, so the lack of a consistent relationship between σ and the modal waiting time for $T = 2.0$ may simply be due to the waiting time distribution being poorly sampled.) The influence of σ on the height of the waiting time distribution for these higher threshold values is similarly complicated. With increasing σ the peak of the waiting time distribution initially becomes higher and the tail shorter as seen in Figures 3b, 3c, 3e, and 3f. However, for $T = 1.5$ the peak height levels off and decreases for the largest values of σ.

The somewhat complicated influence that σ has on the waiting time distribution can be explained in part by the fact that σ can affect both ends of the waiting time distribution simultaneously. As σ increases, signatures can combine to produce higher peaks that exceed the threshold where they could not for lower values of σ. This adds additional onsets to the combined list. In general, one expects such additions to lower the number of long waiting times and increase the number of short waiting times, resulting in a reduction of the tail of the waiting time distribution, a growth of the peak of the distribution, and a decrease in the modal waiting time. However, at same time an increase in σ can cause separate onsets already included in the list at smaller values of σ to be merged together, causing an increase in the modal waiting time. The latter effect appears to be dominant for $T = 1.0$, while the former becomes more significant as T increases.

In order to choose appropriate values of σ and T for the remainder of the analysis, we aimed to reproduce the mean and mode waiting times from the AL onset list published by Borovsky and Yakymenko (2017). Only the waiting times during January, 2005 were used. The Borovsky and Yakymenko (2017) AL onset list was chosen because it was near the middle of the currently published lists in terms of the total number of substorms during January, 2005 (see the substorm counts in Section 2.2 for comparison). The Borovsky and Yakymenko (2017) AL onset list contained 124 substorm onsets during this time, corresponding to a mean waiting time of 6.0 hours. This led to the choice of T_{obs}=1.60, $\sigma_{obs} = 13.8$ min, $T_{model} = 1.72$, and $\sigma_{model} = 20$ min.

Figure 4 shows the waiting time distribution obtained from the observational data (thick blue line) and the model (orange line), along with waiting time distributions from six previously published substorm onset lists that cover January, 2005. The 95% confidence interval of the observed distribution is denoted with light blue shading. The to-

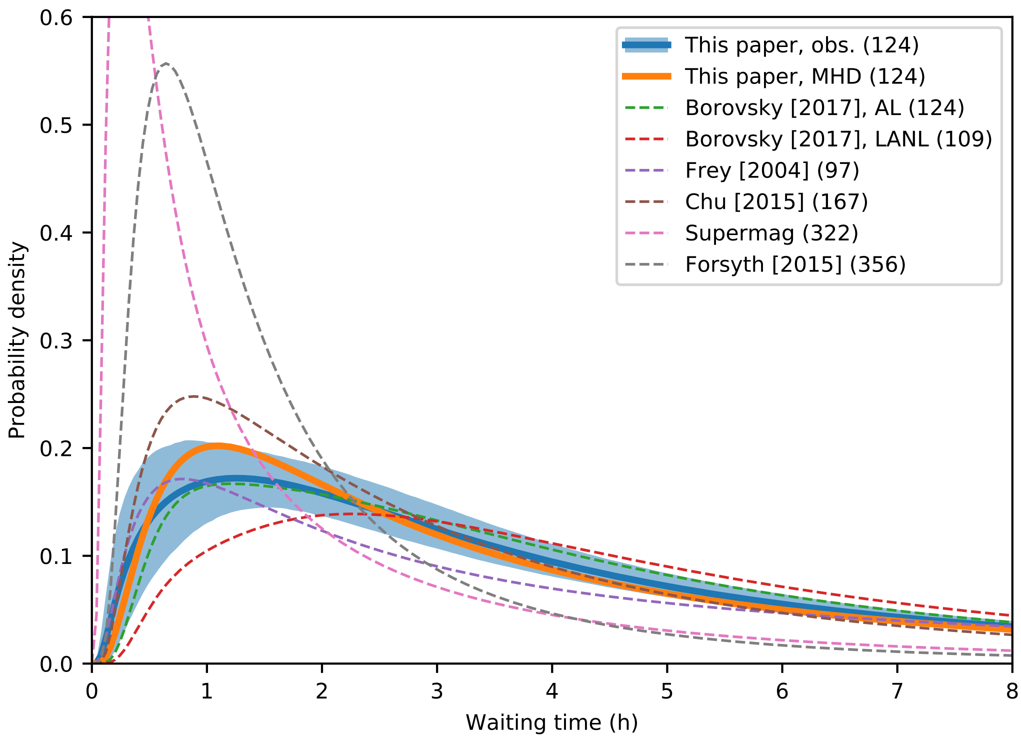

Figure 4. Distributions of substorm waiting times from the present paper (thick solid lines), compared with other published lists that cover the same time period (dashed lines). The shaded region denotes the 95% confidence interval for the observed waiting time distribution in the present work. The total number of substorms in each list (which corresponds to the mean waiting time) is given in parentheses in the legend.

tal number of substorms in each list, which corresponds to the mean waiting time, is listed in parentheses in the legend. The Supermag list was something of an outlier compared with the others, and its mode is not visible with the chosen axis limits. Figure B1 in the appendix shows the full Supermag waiting time distribution for January, 2005.

Figure 4 shows that the waiting time distribution of the Borovsky and Yakymenko (2017) AL list (the green dashed curve) falls near the middle of the published lists in terms of its waiting time distribution, not only in terms of the mean waiting time but also in terms of the mode and overall shape of the distribution. The observed onset list developed for the current paper (blue curve) produces a waiting time distribution that is very close to that of the Borovsky and Yakymenko (2017) AL list. The MHD model produces a waiting time distribution with a higher peak probability, but it falls entirely within the 95% confidence interval of the observed distribution.

Figure 5 compares the waiting time distributions of the combined lists with those of the individual onset lists used to create the combined lists. The observed onsets are shown in light blue, with the 95% confidence interval represented as a shaded region of lighter blue. The MHD results are shown in dark blue. Figure 5a shows the AL onsets, Figure 5b shows dipolarization onsets, Figure 5c shows MPB onsets, and Figure 5d shows all signatures in combination.

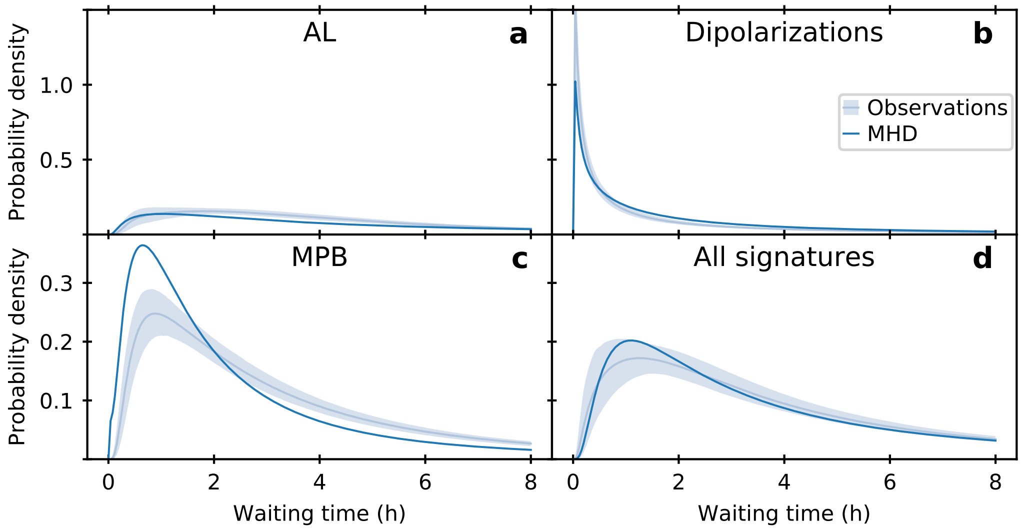

Figure 5. Substorm waiting times for MHD and observations. a) AL onsets only b) Dipolarizations only, and c) MPB onsets only d) All signatures combined.

The distributions of waiting time between AL onsets (Figure 5a) show a modal waiting time of around 1 hour for the simulation and 2 hours for the observations. This is shorter than the 2.75 hours reported by Borovsky et al. (1993), and longer than the results of Juusola et al. (2011) and Newell and Gjerloev (2011a), but it is comparable to the approximately 1 hour reported by Hsu and McPherron (2012). The model distribution for AL waiting time falls within the confidence intervals of the observed distribution for shorter (<1.5 hours) waiting times, though the model underestimates prevalence of 2-6 hour waiting times somewhat.

Dipolarizations produce a much narrower waiting time distribution (Figure 5b), with the modes of both the modeled and observed distributions occurring at less than one-half hour of waiting time. This suggests that the dipolarizations are substantially more frequent than AL onsets. Note that this modal waiting time is shorter than the modal waiting time from any of the previously published lists shown in Figure 4, which may indicate that many of the dipolarizations are not associated with substorms. The model reproduces the observed waiting time distribution reasonably well, straying only slightly outside the confidence bounds of the observed distribution.

The observed waiting time distribution for MPB onsets (Figure 5c) has a mode around 1 hour, in between those of the dipolarizations and AL onsets. The model waiting time distribution has its mode positioned fairly close to that of the observed distribution, but the height of the peak is noticeably higher, and well outside the confidence bounds of the observed distribution. This suggests that the model produces MPB onsets with similar dynamics to reality in terms of recovery time, but that the onsets occur more often. One possible reason for this is that the model MPB index was computed using virtual magnetometers distributed evenly across all longitudes, while the observed MPB index is necessarily computed using real magnetometers, for which substantial gaps in spatial coverage may have prevented some substorms from producing an MPB signature.

Figure 5d shows, for comparison, the same waiting time distributions already shown Figure 4 (they are shown as solid blue and orange curves in that figure). Note that the modal waiting times are close to those obtained from the AL and MPB onset lists (i.e., they are not reduced by the influence of the dipolarizations included in the analysis). As we noted earlier in the section, the model waiting time distribution for the combined onset list remains within the 95% confidence interval of the observed waiting time distribution, even though this was not the case for the individual signatures. This suggests a degree of consistency is achieved between the observations and model in the combined list, which is not the case for individual signatures.

3.2 Forecast metrics

In order to evaluate the predictive capabilities of the model, we first apply the procedure described in Section 2.1 to the onset lists from the model and separately to the observed onset lists, in order to produce a combined onset list for each. We next divide the month into 30-minute bins, and determine whether a substorm onset from each combined list was present in each bin. We then classify each bin according to whether a substorm was identified in the model, observations, neither, or both. The four categories are commonly displayed in a two-by-two table called a contingency table, as shown generically in Table 1: In the upper left corner (a) are true positives, the bins in which a substorm was found in both the model and the observations. Next are false positives (b), in which substorms were found in the model only. In the bottom row of the table are false negatives (c), in which substorms were found in the observations only, and true negatives (d), in which no substorm was found.

		Observations	
		Y	N
Predictions	Y	a	b
	N	c	d

Table 1. A generic contingency table.

To produce a contingency table using our data from January, 2005, we first produced lists of substorm onsets using the procedure described in Section 2.1, and the parameters T_{model}, T_{obs}, σ_{model}, and σ_{obs} set to the values given in Section 3.1.

Table 2 shows the contingency table produced from the onset lists obtained using our procedure. We obtained 124 positive bins from the model list, 25 of which were true positives. We obtained 122 positive bins from the observed list. Since the observed list contains 124 substorms, this indicates that two of the 30-minute bins contained two substorms from the observed list.

		Observations	
		Y	N
SWMF	Y	25	99
	N	97	1267

Table 2. Contingency table for SWMF vs. observations

From the values in the contingency table we compute several metrics summarizing the predictive abilities of the model. These include Probability of Detection (POD), Probability of False Detection (POFD), and the Heidke skill score (HSS), all of which are in common use in space weather applications (e.g. Lopez et al., 2007; Welling & Ridley, 2010; Pulkkinen et al., 2013; Ganushkina et al., 2015; Glocer et al., 2016; Jordanova et al., 2017; S. K. Morley et al., 2018). The POD, given by

$$\mathrm{POD} = \frac{a}{a+c}, \tag{4}$$

(Wilks, 2011) indicates the relative number of times a substorm was forecast when one occurred in observations. A model that predicts all the observed events will have a POD of 1. POFD, given by

$$\mathrm{POFD} = \frac{b}{b+d} \tag{5}$$

indicates the relative number of times that a substorm was forecast when none occurred. Smaller values of POFD indicate better performance, and a model with no false predictions will have a POFD of 0.

Skill scores are a measure of relative predictive accuracy (e.g. Wilks, 2011). The Heidke Skill Score (HSS) is based on the proportion correct (PC), defined as

$$\mathrm{PC} = \frac{a+d}{a+b+c+d}, \tag{6}$$

which measures the fraction of correct predictions relative to the total number of predictions. A perfect forecast would have a PC of 1. The HSS adjusts PC relative to a reference value, PC_{ref}, which is the value of PC that would be obtained by a random forecast that is statistically independent of the observations, and is given by

$$\mathrm{PC}_{ref} = \frac{(a+b)(a+c)+(b+d)(c+d)}{(a+b+c+d)^2}. \tag{7}$$

The HSS is obtained from PC_{ref} as

$$\mathrm{HSS} = \frac{\mathrm{PC}-\mathrm{PC}_{ref}}{1-\mathrm{PC}_{ref}} = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}. \tag{8}$$

The HSS ranges from -1 to 1, where 1 represents a perfect forecast, 0 is equivalent to a no-skill random forecast, and -1 represents the worst possible forecast.

All of the above metrics are subject to sampling uncertainties, meaning that any particular value could be obtained simply by chance, and might not be representative of the model's overall abilities. To address this, we estimate 95% confidence intervals for each metric. The 95% confidence interval is a range in which we estimate that each metric will fall for 95% of a given number of random samples of the dataset. Since no analytical formulas are known for computing confidence intervals for the HSS (Stephenson, 2000), we estimate the confidence interval using bootstrapping (e.g. Conover, 1999). This approach was used previously by S. K. Morley et al. (2018), and the procedure is described in detail in Appendix D.

We now apply the above forecast metrics to our substorm onset lists. Figure 6 shows receiver operating characteristic (ROC) curves for the MHD model. An ROC curve, by definition, shows the probability of detection (POD) of a predictive model as a function of the probability of false detection (POFD), as the threshold for event identification is varied (e.g. Ekelund, 2012; Carter et al., 2016). Such curves are commonly used in evaluating predictive models; a notable recent example from the space weather field is Liemohn

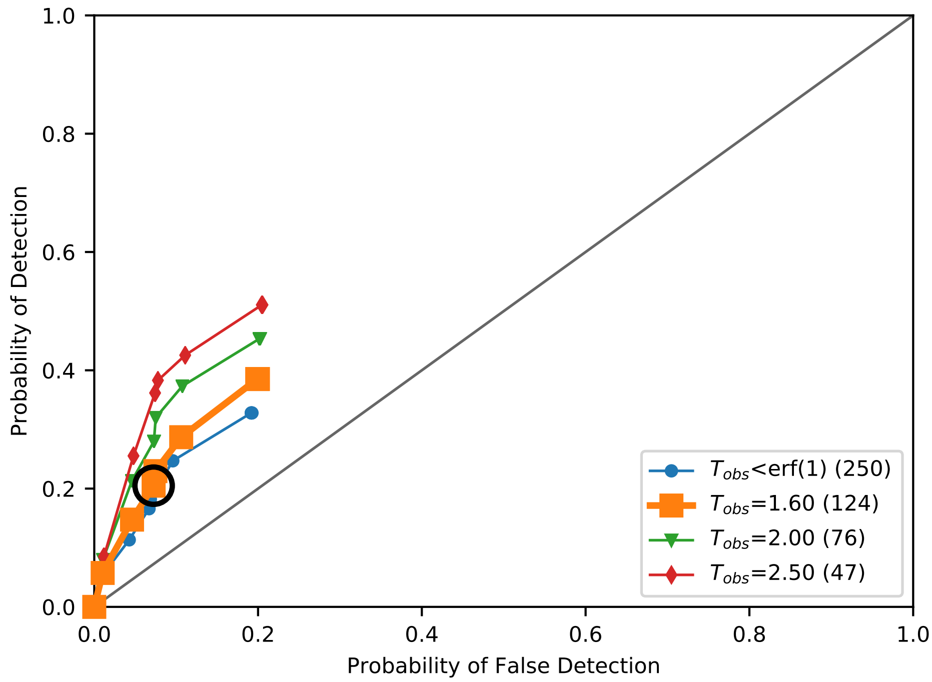

Figure 6. ROC curves for the MHD simulation. The threshold score for identifying substorms from the model output is varied to produce each curve, resulting in changes in the probability of detection (POD) and probability of false detection (POFD). Each curve is computed using a particular threshold score T_{obs} for identifying observed substorms; the thresholds and number of observed substorm identifications are listed in the legend. The case of the observed threshold equal to 1.6 is highlighted with a bold line, and the case of model threshold and the observed threshold equal to 1.72 along this line is highlighted with a black circle.

et al. (2018). For a perfect forecast, the ROC curve would pass through the upper left corner of the plot (POD=1 and POFD=0), so the closer the ROC curve comes to the upper left corner of the plot, the greater the overall accuracy of the forecast. To produce the curves in Figure 6, the threshold T_{model} used to identify a substorm in the model output is varied along the length of each curve, while the threshold T_{obs} for identifying an observed substorm is held fixed. Each curve is computed using a different threshold value T_{obs} for identifying an observed substorm. $T_{obs} = 0.5$ is shown in blue, $T_{obs} = 1.60$ is shown in orange, $T_{obs} = 2.0$ is shown in green, and $T_{obs} = 2.5$ is shown in red. The total number of observed substorms obtained with each threshold is shown in parentheses in the legend. The orange curve, corresponding to an observed threshold of 1.6, is drawn in bold since that is the threshold that was chosen for use throughout the paper, except for tests like this one in which the thresholds are varied. A black circle denotes the model threshold of 1.72 along this green curve. A diagonal grey line shows where POD equals POFD, indicating no skill. For a forecast, POD should exceed POFD, and this is the case along the entire length of each curve (except for the case POD = POFD = 0, where equality is expected).

Note that although a typical ROC curve continues to POD = POFD = 1, ours ends at POFD $\approx$ 0.2. The reason for this is that the practice of using local maxima in the substorm score places a ceiling on the POD and POFD based on the characteristics of the underlying substorm onset lists. If the substorm score has no local maxima within a given 30-minute window, no substorm will be identified regardless of what threshold is used. Also note that the curves corresponding to higher values of T_{obs} produce higher values of POD. While higher POD is desirable, in this case it comes at the cost of an unrealistically low total number of substorms in the observations (and correspondingly, an unrealistically high average waiting time). Rather than maximizing POD, we chose instead in the present work to choose thresholds T_{obs} and T_{model} that produce realistic statistics in terms of substorm waiting time.

Figure 7 shows the Heidke skill score (HSS) as a function of the frequency bias (the ratio of the total number of model substorm bins to the total number of observed substorm bins). Figure 7 was produced by varying the modeled and observed thresholds in the same manner as was done to produce Figure 6. This provides a means to test the sensitivity of HSS to changes in these thresholds. The x-axis value is obtained by dividing the total number of substorm bins obtained from model output by the total number of bins obtained from the observational data. Different observed thresholds are identified by color and shape in the same manner as Figure 6, with error bars denoting the 95% confidence interval for each skill score. Also like Figure 6, the case of the observed threshold equal to 1.6 is drawn with bold lines, and the case of the model threshold equal to 1.72 with the observed threshold equal to 1.6 is marked with a black circle.

For a perfect forecast, the model should produce the same number of substorms as occur in the observations, in which case the frequency bias on the x-axis of Figure 7 will equal one. Since we chose the thresholds T_{obs} and T_{model} so that they produce the same mean waiting time, the black circle corresponding to our chosen thresholds corresponds with a frequency bias very close to one.

For a skill score to represent a true predictive skill, it should be significantly greater than zero, in a statistical sense. This is indicated by the lower end of the 95% confidence interval being greater than zero. A forecast satisfying this criterion is estimated to produce an HSS greater than zero 95% of the time. Figure 7 shows that the skill scores obtained from the MHD model are significantly greater than zero in the majority of cases. The only exception is a single case where $T_{obs} = 2.5$, which as discussed earlier produced an unrealistically large mean waiting time in the observed onset list.

Figure 8 shows the same analysis as Figure 7, but with the kernel width σ_{model} decreased from 20 minutes to 10 minutes. This provides a means to test the sensitivity of HSS to the kernel width σ. The style and axes are the same as Figure 7, and the case of the modeled threshold set to 1.72 and observed threshold both set to 1.6 is again identified with a black circle. Figure 8 shows that the skill scores are sensitive to the choice of kernel width. Halving the kernel width reduces many of the skill scores by about half. However, a majority (all but five) remain significantly greater than zero as determined by their estimated 95% confidence intervals.

Table 3 shows the total number of events, POD, POFD, and HSS for each of the substorm onset lists obtained from the model output. The first row of the table, labeled "All," shows the metrics computed from all signatures, combined into a single onset list using the methodology in Section 2.1, while the remaining rows show results for individual signatures. With the exception of the last column of the table, all quantities are obtained by testing each signature in the model output with observed signatures of the same category (for instance, model AL is compared with observed AL). These numbers are absent for the plasmoids since there was no observational plasmoid data with which to compare. Two columns are shown for HSS. The first (labeled "HSS, same signature") is computed using model and observed substorm onset lists obtained using the signature identified at the beginning of that row (all signatures combined in the case of the first row). The second uses the same model onset list as the first, but the observed onset list

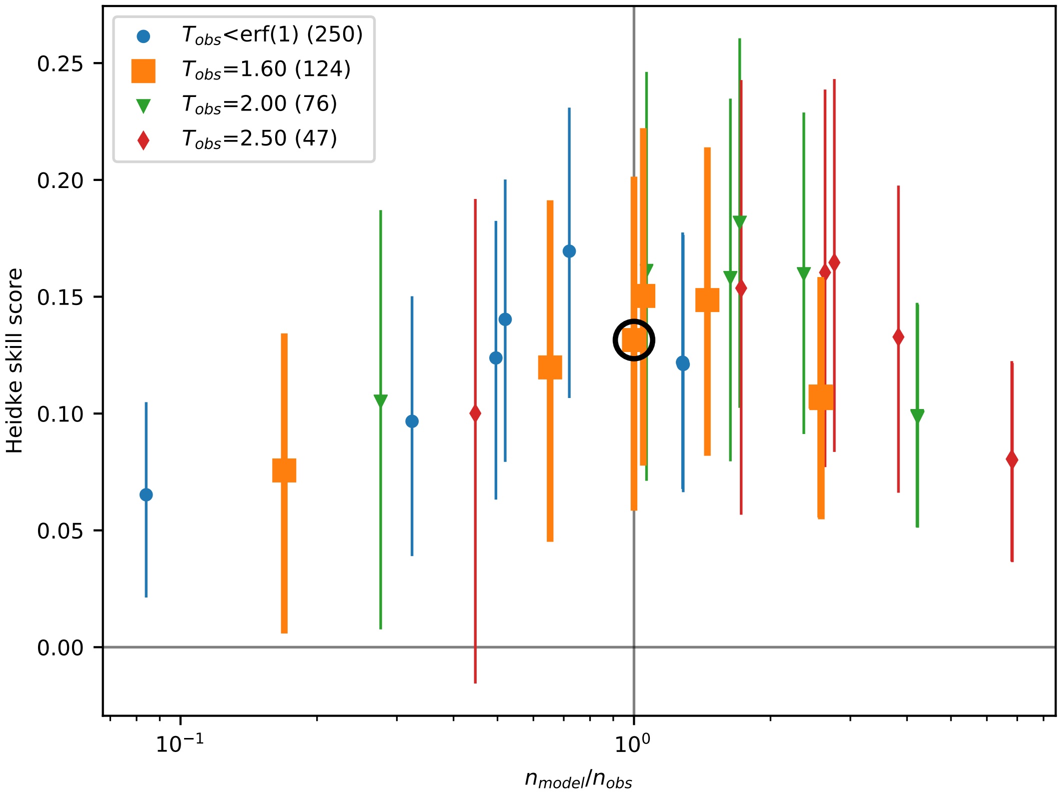

Figure 7. Heidke skill score as a function of the frequency bias (the ratio of the number of model substorm bins to the number of observed substorm bins). The threshold scores T_{obs} and T_{model} for identifying substorms have been varied to test the sensitivity of skill scores and frequency biases to these thresholds. Each color and shape corresponds to a particular threshold score T_{obs} for identifying observed substorms; the thresholds and number of observed substorm bins are listed in the legend. For a given observed threshold, different skill scores and frequency biases are obtained by varying the threshold for identifying a model substorm. Error bars represent the 95% confidence interval for each skill score. The case of observed threshold equal to 1.6 is drawn in bold, and the case of the model threshold equal to 1.72 with the observed threshold equal to 1.6 is marked with a black circle.

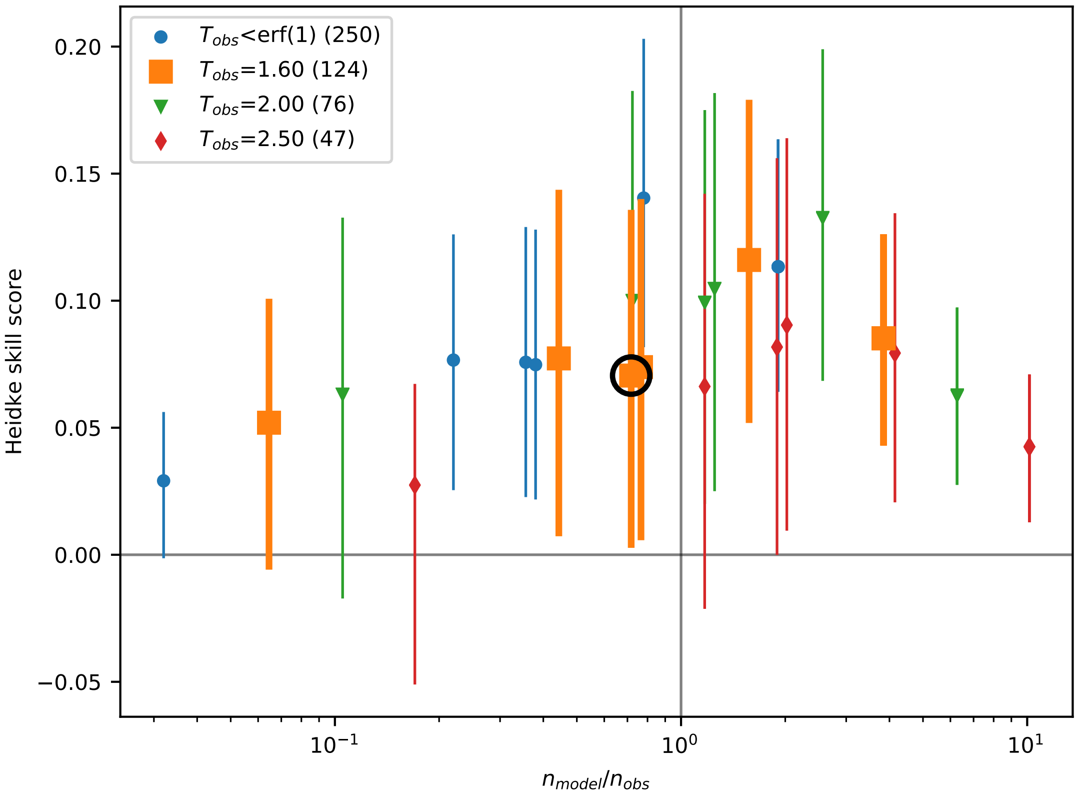

Figure 8. Heidke skill score as a function of frequency bias, using a kernel width $\sigma_{model} = 10$ minutes instead of the $\sigma_{model} = 20$ minutes width used elsewhere. The format is the same as Figure 7.

is the one obtained using all signatures combined together. This gives an indication of how well the individual model signature predicts the combined (all signatures) observed substorm onsets. For the POD, POFD, and HSS, a bar over the number identifies the last significant digit, as determined by the limits of the 95% confidence interval. For the skill scores, the limits of the confidence intervals are shown in brackets. The lower limits of the confidence intervals are positive for every case except the plasmoids, indicating that the skill scores are significantly greater than zero.

	SWMF events	Obs. events	POD	POFD	HSS, same signature	HSS, all signatures
All	124	124	$0.\bar{2}0$	$0.0\bar{7}2$	$0.1\bar{3}1$ $[0.061, 0.20]$	$0.1\bar{3}1$ $[0.062, 0.20]$
AL	85	130	$0.\bar{1}8$	$0.0\bar{4}5$	$0.1\bar{6}6$ $[0.089, 0.24]$	$0.1\bar{2}5$ $[0.052, 0.20]$
MPB	201	167	$0.\bar{2}7$	$0.1\bar{1}1$	$0.1\bar{4}8$ $[0.085, 0.21]$	$0.1\bar{2}9$ $[0.065, 0.19]$
dipolarizations	166	96	$0.\bar{2}6$	$0.0\bar{8}9$	$0.1\bar{2}1$ $[0.052, 0.19]$	$0.0\bar{8}3$ $[0.02, 0.1]$
plasmoids	447	–	–	–	–	$0.0\bar{4}2$ $[-9 \times 10^{-4}, 0.09]$

Table 3. Forecast metrics for each signature

Of all the signatures, the plasmoids releases do the least well at predicting the observed substorms. The AL and MPB signatures produce higher skill scores than the dipolarizations, but the confidence intervals for all three overlap so the differences between them may not be statistically significant.

Far more plasmoid releases (447 in total) were identified than any other substorm signature, with the next most common signature being MPB onsets with only 166 occurrences. This strongly implies that the plasmoid release list contained a large number of false positives. While we have confidence that all the plasmoids were real (in the sense that they occurred within the simulation), the much smaller number of AL and MPB onsets (85 and 201, respectively) suggests that only a few of them were substorm related. The total number of events in the combined substorm list obtained from the simulation is only 124. This means that more than two thirds of the plasmoid releases were rejected by our substorm identification procedure, and indicates that the procedure used to combine signatures is largely successful at eliminating false positive identifications.

3.3 Relative contribution of signatures

Although we included multiple substorm signatures in the analysis, not all contribute equally. To assess the relative contributions of different signatures to the combined list, we performed counts of the number of substorms in the combined list to which each signature contributed, and a count of the number of signatures that contributed to each onset. For the purpose of this analysis, we count a signature as contributing to an onset in the combined list if it accounts for more than 5% of the total value of $f(t)$ at the time of the onset. Table 4 breaks down the substorms by the number of observational signatures contributing toward the identification of each substorm in the combined list. The columns of the table are organized according to the signature count, or the number of signatures contributing more than 5% of $f(t)$ for each substorm. The signature counts are listed on the first row of Table 4, with a final column containing the total number of substorms independent of the signature count. The next five rows show the number of substorms for which each individual onset list contributed more than 5%, again broken down by the total number of contributing signatures for each substorm. The final row shows the total number of substorms having each signature count.

Signature count	2	3	4	5	Any
LANL	15	31	31	6	83
IMAGE/FUV	12	23	29	6	70
MPB	25	39	30	6	100
AL	16	32	34	6	88
Dipolarizations	6	13	16	6	41
Combined onsets	37	46	35	6	124

Table 4. Counts of substorms for which each signature contributed more than 5% of the total score $f(t)$, broken down by the total number of signatures exceeding 5% of the total score for each substorm in the combined onset list. The last column is a sum of the preceding columns. The last row contains the total number of substorms in the combined onset list having the number of contributing signatures corresponding to that column.

As an example, the first row of Table 4 shows that the LANL energetic particle data contributed at least 5% to 83 substorms in the combined list. Of these, 15 had two signatures (including LANL) contributing to the total $f(t)$, 31 had three signatures, and so on. 37 of the substorms in the combined list had two signatures contributing, 46 had three contributing, and so on.

From Table 4 it is apparent that the dipolarizations contributed appreciably less to the combined list than did the other signatures. In total, only 41 (33%) of the substorms in the combined list had corresponding dipolarization signatures. The MPB list contributed to the greatest number of substorms at 100 (80.6%) of the 124 substorms in the combined list. The number of signatures contributing to each substorm was quite variable. A plurality (46) of the substorms had three contributing signatures, but a substantial number had two or four as well.

Table 5 shows the number of substorms for which each signature from the model output contributed more than 5% of the total substorm score $f(t)$. The counts are presented in the same format as Table 4, with the information again separated columnwise according to the number of signatures exceeding the 5% level for each substorm in the combined list. Table 5 shows that the plasmoids contributed to largest fraction (112 or 90.3%) of substorms in the combined list, while the AL onsets contributed to the smallest portion (59 or 48%) of the combined list.

Signature count	2	3	4	Any
Plasmoids	31	54	27	112
MPB	25	46	27	98
AL	5	27	27	59
Dipolarizations	17	47	27	91
Combined onsets	39	58	27	124

Table 5. Contingency table for SWMF vs. observations

Interpreting Tables 4 and 5 is complicated by the interaction between different lists as part of the selection process. Although the plasmoids contribute to a majority of onsets in the combined list obtained from model output, it does not necessarily follow that the plasmoids were the most influential in determining what events are included in the model-derived onset list, because the plasmoids were also the most numerous of all the signatures obtained from the model. The high fraction of substorms for which the plasmoids contributed to the total score may therefore simply reflect a high frequency of occurrence for plasmoids, rather than a high correlation with actual substorm onsets. This can be illustrated more clearly by considering hypothetically the addition of a randomly distributed list containing a very large number of onsets into the analysis. Such a random onset list would serve to increase $f(t)$ approximately uniformly, and would therefore have the same effect as reducing the threshold T. The randomly distributed signature would contribute significantly to the total score for every onset, but the contents of the list would be determined primarily by the other signatures and not the randomly distributed one. In much the same way, the plasmoids, whose number exceeded the number of onsets in the combined list by a factor of 4, were likely not the most important factor determining what onsets were included in the combined list. Instead, the other signatures were likely be more influential in determining the contents of the combined list because of their role in restricting which onsets are included. Similarly, the fact that MPB contributed to 80.6% of the observed onsets does not necessarily indicate that the MPB index was most influential in determining the contents of the observed onset list.

What does seem to follow from Tables 4 and 5 is that no single signature dominates the combined lists on its own, judging from the fact that a majority of onsets had three or more contributing signatures. To further test whether any signatures were dominating the list, we computed the relative contributions of individual signature scores to the total score $f(t)$. We identified the relative contribution of the largest contributing signature for each onset in the combined list, and took the median of this value for all substorms in the list. This median was found to be 36.6% for the observational list and 37.3% for the model. This indicates that the largest contribution of any single signature to $f(t)$ was equal to or less than this median value for a majority of substorms. Since the median value is well below 50%, this provides additional confirmation that the method is successful in finding substorm onset times that can be identified by multiple signatures. We also computed the maximum relative contribution to the total score $f(t)$ of any single signature was 54.2% for the observational list and 54.3% for the model onset list. This means that even in the few cases where one signature contributed a majority of the score, other signatures were essential to producing the total score that was obtained.

3.4 Superposed epoch analysis

We now present superposed epoch analyses (SEAs) of parameters related to the solar wind driving during substorms and to the geomagnetic signatures of the substorms. SEA consists of shifting a set of time-series data $y(t)$ to a set of epoch times t_k, producing a group of time-series $y_k = y(t - t_k)$ from which properties common to the epoch times can be estimated (e.g. Samson & Yeung, 1986). Common properties of the SEA may be estimated and visualized in a variety of ways. For instance, S. K. Morley et al. (2010) plotted shaded regions representing the 95% confidence interval for the median and interquartile range, and Katus and Liemohn (2013) plotted 2-D histograms colored according to the number of SEA members passing through each cell of the histogram, while Hendry et al. (2013) created images colored according to the total electron flux observed by the Medium Energy Proton and Electron Detector among all SEA members, binned by epoch time and L-shell. Probably the most common approach to visualizing an SEA is to use a measure of central tendency such as the mean or median to obtain a new time-series $\hat{s}(t)$ that estimates the typical behavior of $y(t)$ in the vicinity of the epoch times t_k. In the present work we will use the median of y_k to accomplish this. The epoch times t_k will come from one of two lists of substorm onset times (one derived from the MHD simulation and the other from the observations).

Computing an SEA using our substorm onset times serves as a diagnostic to determine whether the onset times identified by our selection procedure are consistent with previously reported behavior for substorms, in terms of both the solar wind driving and the geomagnetic response. With the model substorm onsets, the SEAs also provide a means to test how closely the model's behavior during substorms follows the observed behavior of the magnetosphere.

Figure 9 shows SEAs of the observational data and the model output, with the epoch times corresponding to substorm onset times obtained using each of the methods described in Section 2.5. SEAs obtained using the combined onset list (produced as described in Section 2.1 with the parameters given in Section 3.1) are shown as a thick blue curve, along with all the individual signatures: MPB onsets (orange), IMAGE/FUV (green), plasmoids (red), AL (purple), LANL (brown), and dipolarizations (pink). The left column (Figures 9a-9d) shows observed results, while the right column (Figures 9e-9h) shows the MHD results. The variables plotted on the y axes are IMF B_z (Figures 9a and 9e), solar wind ϵ (Figures 9b and 9f), the AL index (Figures 9c and 9g), and the MPB index (Figures 9d and 9h). IMF B_z is in GSM coordinates. ϵ provides an estimation of the rate at which solar wind energy is entering the magnetosphere (Perreault & Akasofu, 1978), and is given by

$$\epsilon = |u_x| \frac{|\mathbf{B}|^2}{\mu_0} \sin\left(\frac{\theta_{clock}}{2}\right)^4, \tag{9}$$

where u_x is the sunward component of solar wind velocity, $\mathbf{B}$ is the IMF, and θ_{clock} is the IMF clock angle.

From the SEA of IMF B_z (Figures 9a and 9e), it is apparent that the observed substorms are typically preceded by a decrease in IMF B_z, with the minimum B_z occurring just before the onset time and a recovery back to near-zero B_z following the onset. Similar behavior is present in both the model and the observations, but the decrease in B_z is somewhat sharper for the model onsets (with the exception of the plasmoids, which have a particularly weak decrease in B_z). The decrease is evident for all of the onset lists. In addition to the plasmoids, the AL onsets stand out significantly. When using AL onsets for the epoch times (both for observations and model) the minimum B_z occurs slightly later, which may be an indication that the AL onsets precede the other signatures on average. The model AL onsets are preceded by a 1-2 nT increase 1-2 hours prior to onset, and a particularly sharp decrease just prior to onset. The tendency of substorms to occur near a local minimum in IMF B_z has been previously reported, and our results for both observations and MHD are qualitatively similar to those obtained by SEA in previous studies (e.g. Caan et al., 1975, 1978; Newell et al., 2001; Freeman & Morley, 2009; Newell & Liou, 2011; Walach & Milan, 2015).

Figures 9b and 9f show that all onset lists correspond with an increase in ϵ prior to onset, with a maximum occurring prior to onset, or in the case of AL, just after onset. A separate SEA of the solar wind velocity component u_x (not shown) showed no appreciable trend, which indicates that the trend in ϵ is driven almost entirely by variation in IMF B_z. However, despite a lack of change in u_x before and after onset, we found that some classes of onsets seem to be associated with higher or lower u_x; most notably dipolarizations were associated with higher u_x than any other signature type, and this is responsible for the higher ϵ values associated with dipolarizations. As with B_z, ϵ undergoes a sharp transition prior to the model AL onsets, and the plasmoid release times are associated with only a very weak increase and decrease in ϵ.

In the SEA of observed AL (Figure 9c), a sharp decrease occurs at onset. This occurs for the combined onset list and for all of the individual signatures except for the dipolarizations. Dipolarizations are associated with a downward trend in AL but the decrease begins earlier and is more gradual. The behavior of the observed AL index is qualitatively similar to what was obtained by previous authors. The approximately 2 hour re-

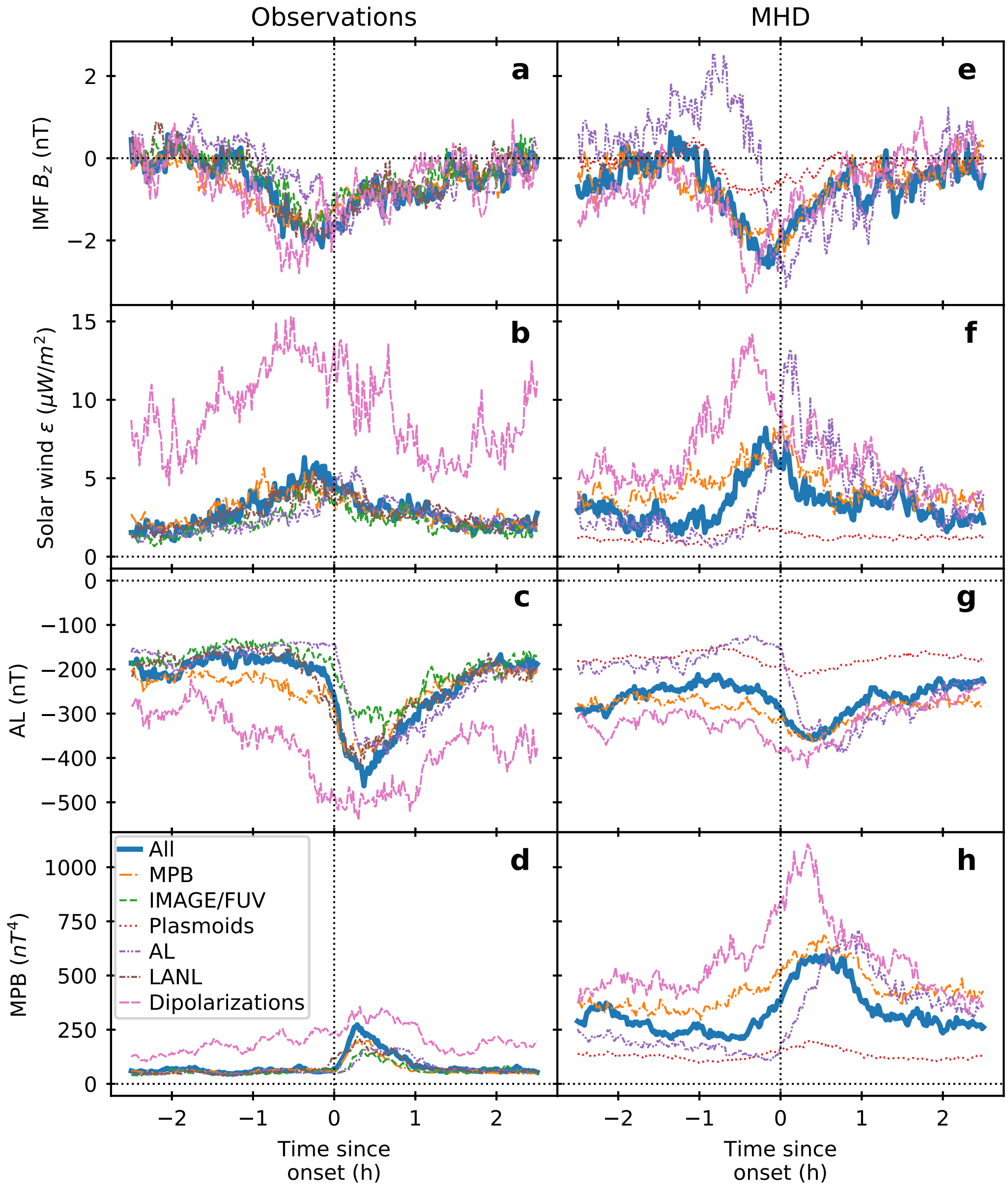

Figure 9. Superposed epoch analyses of IMF B_z, ϵ, AL, and MPB, comparing onsets identified from the model and from the observations. The left column shows SEAs computed using epoch times from the observations, while the right column shows SEAs computed using epoch times from the simulation. The AL and MPB data come from the respective datasets used to create the onsets (observations or model run), and the other values come from the solar wind data input to the model. The lines show the median value for all epoch times as a function of the time offset. The thick blue line (labeled "All" in the legend) shows the SEA computed with epoch times from the combined onset list using all signatures, while thinner colored lines show SEAs obtained using epoch times from the individual signatures.

covery time is similar to the results of e.g. Caan et al. (1978); Forsyth et al. (2015), but the -500 nT minimum is lower than their results. Both Caan et al. (1978) and Forsyth et al. (2015) analyzed multi-year time periods, and the lower minimum AL obtained here may simply be due to the fact that the analysis covers a much shorter time period which was chosen for its relatively large amount of substorm activity. In the model output (Figure 9g), AL onsets are also associated with a sharp decrease at onset, but the MPB onsets, dipolarizations, and plasmoids are associated with gradual decreases in AL. When AL onsets alone are used for the onset list, an increase occurs in the hour prior to onset, followed by a decrease similar to that obtained from the SEA of observed AL onsets. When all the model signatures are combined, the increase 1 hour prior to onset is absent (although a more gradual, possibly unrelated increase occurrs 1-3 hours prior to onset), and the associated decrease in AL is weaker than occurs in observations.

It is notable that while the combined signature list from the observations produces a robust decrease at onset in the SEA of AL, the same cannot be said of the combined onset list obtained from the model. A possible explanation is that combining signatures does not preferentially eliminate weak substorms, but rather tends to eliminate those that are too far from the average for a given input dataset. The fact that the average in the model involves a weaker onset reflects the fact that the model produces weaker variations in AL in general, as was noted for the same simulation in Haiducek et al. (2017). The weak association between dipolarizations and AL onsets in the observations may be due in part to the fact that only two satellites are used to identify dipolarizations (versus three for the LANL energetic particle injections). The model output uses dipolarizations identified from a third location (which is ideally positioned on the sun-Earth line), and in the model output the dipolarizations do not contrast as strongly from the other datasets in terms of their associated AL response.

From Figure 9d, it can be seen that all of the observed signatures are associated with an increase in MPB beginning at onset. Dipolarizations are associated with an additional gradual increase prior to onset, with the rate of increase becoming greater at the onset time. When all signatures are combined, the associated increase in MPB is noticeably stronger than for any single signature alone. For all curves except the one produced using dipolarizations as the signature, the shape is qualitatively similar to the superposed epoch analysis shown in Chu et al. (2015) for MPB onsets, which similar to our results showed peaks between 50 and 250 nT and recovery times on the order of 1 hour. With the model output (Figure 9h), all of the signatures are also associated with an increase in MPB. However, the magnitude of this increase varies substantially from one signature to another. Plasmoid releases are associated with the weakest increase in MPB, while AL onsets are associated with the strongest increase. Combining all signatures together does not intensify the associated MPB response as it does for the observations: The combined MPB curve falls in between those of the AL, dipolarization, and MPB onsets.

It is worth noting that plasmoid releases are only very weakly associated with changes in driving conditions (IMF and ϵ) or in response indicators (AL and MPB). This is related to the fact that many more plasmoid releases were identified than any other signature (see Table 3), which means that many plasmoid releases may have no associated auroral or geosynchronous response, or the response might be below the threshold for selection. Such plasmoids may be too weak or too far down-tail to have a substantial effect close to the Earth. The state of the fields and plasmas in the inner magnetosphere may also influence how much energy from the plasmoid release is transported Earthward. Similarly, dipolarizations are also only weakly associated with changes in driving conditions and magnetospheric response, though they are more strongly associated than plasmoids are. Like the plasmoids, dipolarizations are observed in the magnetosphere and most likely some of them occur without a strong coupling to the ionosphere that would produce a typical substorm response.

4 Discussion

In the present paper we have demonstrated a procedure to combine multiple substorm onset lists into a single list. We applied this procedure to observational data and to MHD output from the same one-month period. By performing superposed epoch analysis we demonstrated that the resulting onset list is consistent with previous results in terms of the solar wind driving and the geomagnetic response as measured by ground-based magnetometers. We showed that the total number of substorms and the waiting time distributions are also consistent with previous results. Finally, we showed preliminary evidence that our MHD model has statistically significant predictive skill and is able to reproduce the observed waiting time distribution, as well as some of the observed features in terms of driving and response.

4.1 Effectiveness of combining signatures

The method appears to be effective in identifying substorm onsets that are identifiable by multiple methods. The thresholds used were high enough to ensure each substorm could be identified by at least two signatures, and a majority of onsets in both of the combined lists were identifiable by three or more signatures. For a majority of observed substorms the largest contributing score of any single signature was less than 36.6% of the total score for the onset (37.3% for the model substorms), with no signature contributing more than 54.2% of the total score (54.3% for the model substorms). We found no indication that any one signature plays a dominant role in determining the contents of the combined onset list.

The approach of combining onset lists obtained using different techniques into a single combined list appears to at least partially address the problems of false identifications and data gaps. More than twice as many plasmoid releases were identified from the model output than were obtained by analyzing any single observational signature, yet the total number of substorms identified in the model output is far smaller than the number of plasmoid releases, indicating that the vast majority of plasmoid releases were rejected for lack of an associated AL, MPB, or dipolarization signature. At the same time, data gaps in the observations account for significant under-counting of dipolarization signatures, but the total number of observed substorms in the combined list is significantly higher than the total number of dipolarizations. This suggests that the combined inputs from other observed signatures were able to compensate for the lack of continuous nightside magnetic field observations in geosynchronous orbit.

In addition to differing in terms of their total numbers, both dipolarizations and plasmoids exhibited noticeably different statistics compared with other signatures in terms of waiting time distributions and in terms of SEA behavior when both were used as epoch times. In both the model and the observations, the waiting time distribution for the dipolarizations is noticeably different from MPB, AL, or combined onset lists. Similarly, SEAs using dipolarizations and plasmoid releases to determine epoch times produced results that differed substantially both from epoch times obtained using other signatures, and from behavior expected based on previous research. This suggests that dipolarizations and plasmoid releases may be relatively poor indicators of substorm onset, perhaps because both regularly occur independently of substorms. Nonetheless, the waiting time distributions and SEAs obtained from the combined onset appear not to be overly influenced by the statistics of the dipolarization and plasmoid timings.

We chose tuning parameters so that the resulting onset list has a mean and mode waiting time that is on par with previously published results for the same time period. The resulting waiting time distribution is qualitatively similar to previously published results (by e.g. Borovsky et al., 1993; Chu et al., 2015; Kauristie et al., 2017; Borovsky & Yakymenko, 2017). The modal waiting time of around 1-1.5 hours is consistent with previously published results covering January, 2005, and the distribution shape is very close to that of the Borovsky and Yakymenko (2017) results for that time period, reproducing not only the mean and mode for which we optimized, but also the shape of the distribution. We also find that SEAs of our combined onset lists reproduce many of the expected behaviors for substorms, such as a local maximum in IMF B_z (e.g. Caan et al., 1975, 1978; Newell et al., 2001; Freeman & Morley, 2009; Newell & Liou, 2011; Walach & Milan, 2015) and a negative bay in AL (e.g. Kamide et al., 1974; Caan et al., 1978; Forsyth et al., 2015) that occur around the substorm onset time. This indicates that, on average, the magnetosphere exhibited dynamics previously reported for substorms around the times included in the combined onset lists.

4.2 Paths for improving the substorm identifications

We have demonstrated that the mean and mode waiting time of substorms identified by our method can be controlled by adjusting its tuning parameters: The detection threshold T and the kernel width σ. While we chose to optimize these parameters to reproduce the waiting time distribution of a previously published substorm onset list, this may not be the best approach in all situations. In general it is possible to determine a range of values for each parameter beyond which reasonable results are no longer expected. For instance, we showed in Section 2.1 that values of $T < \mathrm{erf}(1)$ will all produce identical results, while values of T exceeding the number of underlying onset lists will produce an empty onset list. Similarly, setting the kernel width too low can greatly reduce the number of events selected by reducing the kernel overlap for nearby signatures, and in extreme cases can result in no events being selected at all. An overly large kernel width could cause unrelated signatures to be merged together, potentially causing spurious onsets to appear in the combined list between the contributing signatures while removing correct onset times. We selected kernel widths σ of 13.8 and 20 minutes, respectively, for the observational and model datasets, but kernel widths as small as 5 minutes and as large as 25 minutes might be considered reasonable. Similarly, the threshold T can have a substantial effect on the total number of events selected, as was illustrated in Figures 6 and 7 in which the total number of observed events varies from 47 to 250 as the detection threshold is varied.

The relationship between the threshold T, kernel width σ, and what events are selected depends on the number of signatures used as well as the statistical characteristics of each signature, such as their waiting time distributions. As a result, the threshold needs to be adjusted whenever signatures are added or removed. In the present work we optimized T and σ to produce a waiting time distribution that is comparable with previously published results. However, this approach is only possible for time periods that have existing published lists to which to compare. An alternative approach might be to construct a heuristic based on the number of onset lists that are combined. A simple way to do this would be to scale the threshold according to the number of onset lists used. The threshold might be adjusted down for time periods in which one or more signatures is known to contain a data gap.

While we used all available signatures, there might be merit in excluding one or more signatures from consideration in future efforts. We found indications that dipolarizations and plasmoids exhibited substantially different statistics compared to other substorm signatures, possibly indicating that many of these signatures are not substorm associated. The relative importance of a signature might be tested by selectively removing signatures from the list to determine its relative importance to the combined onset list. Or, as an alternative to removing a signature entirely from the list, we could instead apply weight factors to the signatures prior to adding them together. Lacking an objective means to determine appropriate weight factors, we have decided not to apply weights to the individual signatures in the present work, and instead all signatures were weighted equally. However, in the future it might be appropriate to introduce such weight factors. One way to do this is to compute weighting factors based on the average waiting time in each onset list. This would weight signatures such as plasmoids that occur very frequently (and probably are not always associated with substorms) less heavily than those that occur infrequently. Another approach might be to develop a reliability measure of some sort, which could be applied to each signature and used to compute its weight factor. For some signatures, it might be appropriate to weight individual onsets according to a measure of event strength associated with that signature. For instance, the amount of change in AL within a specified time after onset could be used as a measure of AL onset strength, and AL onsets with large changes could be weighted more strongly than those with small changes.

In Section 3.2 we noted that some of the data in Figure 9 suggests a tendency for the AL onsets to precede the other signatures by a few minutes. Such a tendency could result in onset times that are slightly too early in the combined list, and could also result in some onsets not being counted (due to falling below threshold with signatures being poorly aligned in time). A severe temporal bias could result in some substorm events being double counted. The temporal bias we noted in Figure 9 appears to be smaller than σ so the effects resulting from it are likely to have a fairly small affect on the results. However, in the future it might be possible to adapt the method to remove or reduce such effects. This could be done by replacing the Gaussian kernel function with a non-Gaussian shape. This would remove the temporal symmetry imposed by the Gaussian kernel. A non-Gaussian kernel shape could be developed individually for each signature based on its tendency to lead or follow other signatures.

The tunability of our procedure, along with the possible modifications described in this section, give it a significant amount of flexibility. This enables it to be optimized to produce desired characteristics in terms of what events are identified. An obvious approach to optimization is to adjust the tuning parameters to best fit established criteria for identifying substorms. However, the lack of a community consensus on precise procedures, benchmarks, or tests for correct substorm identification precludes this approach. This lack of such a consensus has been an issue in the community for a while, and has been noted by a number of authors (e.g. Rostoker et al., 1980; R. L. McPherron & Chu, 2017, 2018). While we can readily compare our list against existing ones, as has been done by a number of researchers (e.g. Moldwin & Hughes, 1993; Boakes et al., 2009; Liou, 2010; Chu et al., 2015; Forsyth et al., 2015; Kauristie et al., 2017), fundamentally such comparisons tell us about the similarities and differences between the lists and not which list is most correct. In the meantime, optimizing for known characteristics of substorms, rather than a specific list, is probably the best approach.

If our identification procedure is used applied for operational purposes, another important consideration in choosing detection thresholds is the needs of forecast customers. In this case, factors such as the costs and risks associated with false positive and false negative detections should be considered. Is the cost of responding to a false positive prediction greater or less than the cost incurred when a substorm arrives unannounced? Of course, this probably depends on the strength of an event, and ideally the procedure should be tuned in a manner that makes stronger events more likely to be identified.

4.3 Substorm prediction with MHD

One of the possible operational applications for our identification procedure is the development of a substorm forecast product. This could be done using an MHD model as we demonstrated in the present work, although the technique of combining multiple types of signatures can certainly be applied to other types of models. The ability to simulate a substorm with an MHD model has been demonstrated previously (e.g. Lyon et al., 1981; Slinker et al., 1995; Raeder et al., 2001; Wang et al., 2010). However, previous efforts simulating substorms with MHD have covered time periods lasting no more than a few days and at most several substorms, preventing a rigorous analysis of the model's predictive skill. In the present paper we used a one-month simulation including over 100 substorms, which is sufficient to enable computation of forecast accuracy metrics such as POD, POFD, and HSS. To our knowledge, this is the first attempt to rigorously evaluate an MHD model for its ability to predict substorms.

In our test, the MHD model demonstrated consistently positive predictive skill, with zero or negative skill scores occurring only in extreme cases of high or low detection thresholds. The skill scores achieved are significantly greater than zero, but they are closer to zero (no skill) than they are to one (perfect skill). This certainly leaves room for improvement, and also begs the question of whether scores on this level are sufficiently high to be of practical use. Looking to evaluations of existing operational models, one can find some examples of tropospheric models that deliver performance on this level, particularly for long lead time forecasts of difficult to predict phenomena such as precipitation (e.g. Barnston et al., 1999). However, such comparisons are of limited utility not only because of the differences in the system being modeled, but also difference in the lead time and the temporal and spatial granularity of the forecast. Ultimately, an assessment of operational usefulness depends on the manner in which the forecast is used by customers, including the operational impact and mitigation strategies available.

4.4 Paths for improved MHD modeling of substorms

An obvious path forward with the MHD model is to explore whether this initial demonstration of predictive skill can be improved upon. The first step would be to conduct tests of different configurations of the model to determine the sensitivity of results to parameters such as grid resolution and boundary conditions. Another possible path for improvement is the incorporation of non-ideal MHD and other physical processes that were not incorporated in the simulation shown here. A likely candidate for this is the inclusion of additional resistive terms. It has long been recognized that resistivity plays an important role in controlling magnetotail dynamics such those associated with substorms. Birn and Hones Jr. (1981), for instance, demonstrated that an X-line formation and plasmoid release could be induced in an MHD simulation by abruptly increasing the amount of resistivity. In the present work, as with many efforts involving MHD simulation, we rely entirely on numerical resistivity to enable reconnection to occur. Our results show that numerical resistivity can produce substorms at a realistic rate, as evidenced by the fact that the total number of substorms is in line with other lists from the same time period, and the waiting time distribution produced by the model is close to that produced by the observations. This means that our numerical resistivity is realistic enough that the model can capture important aspects of the system dynamics. However, improved prediction of substorms may require a more realistic resistivity model. One approach is to introduce Hall resistivity, which has been shown by observations to play a role in magnetotail reconnection (Øieroset et al., 2001). Hall MHD has been implemented in SWMF (Tóth et al., 2008), but has not been tested in the context of substorm prediction. Another approach that may improve substorm-related reconnection physics is the use of a particle-in-cell (PIC) model in place of MHD in and near the reconnection region. This has been demonstrated by Tóth et al. (2016) and Chen et al. (2017) for magnetospheric simulations, but again has not been tested for substorm prediction. On the other hand, the PIC approach, while promising for its ability to capture aspects of reconnection physics that are not incorporated in ideal MHD, is likely too computationally expensive for operational use in the near term.

Besides night-side reconnection, coupling between the magnetosphere and ionosphere plays an important role in the substorm process. For instance, ionospheric conductivity influences the strength and spatial distribution of field-aligned currents within the magnetosphere (e.g. Ridley et al., 2004). However, there is considerable room for improvement in the models of this conductance, particularly in the auroral zone. SWMF currently estimates auroral-zone conductance using an empirical relationship based on the strength of field-aligned currents, since MHD does not directly estimate the precipitating fluxes that determine the conductivity in reality (Ridley et al., 2004). Welling et al. (2017) showed that SWMF is frequently used to simulate conditions that fall outside the range of validity for the existing conductance model. Efforts are currently ongoing to develop an improved empirical model for this purpose (Mukhopadhyay et al., 2018). However, this approach has limitations because the conductance depends on other factors besides the field-aligned current, including particle precipitation, that are not modeled by MHD. An alternative might be to estimate the conductivity using the particle distributions in an inner magnetosphere model such as RCM, but this would likely require the development of new empirical relationships between precipitating fluxes and conductivity. Other improvements to the MHD model that could influence magnetosphere-ionosphere coupling include the use of anisotropic pressure (Meng et al., 2012, 2013), polar outflow (Glocer, Tóth, Gombosi, & Welling, 2009), and multi-fluid MHD (Glocer, Tóth, Ma, et al., 2009), all of which have been implemented in BATS-R-US and demonstrated in magnetospheric simulations, but none of which have been tested for their effect on substorm prediction. The initial tests of anisotropic pressure and polar outflow in SWMF (Meng et al. (2012) and Glocer, Tóth, Gombosi, and Welling (2009), respectively) both showed that simulations using those models have increased tail stretching compared with BATS-R-US simulations that do not use them, and increased tail stretching could have a significant influence on substorm dynamics since the substorm growth stage is associated with magnetotail stretching (e.g. Kaufmann, 1987; Sergeev et al., 1990).

Of the enhancements mentioned above, ionospheric outflow may be particularly important because it has been shown to be associated with substorms. For instance Øieroset et al. (1999) and Wilson et al. (2004) both found that ionospheric outflow increases by a factor of two on average from quiet time to substorm onset, and that stronger substorms are associated with higher rates of ionospheric outflow. Modeling results have shown that ionospheric outflow can influence magnetospheric dynamics in general (e.g. Winglee et al., 2002; Wiltberger et al., 2010) and substorm strength and onset times in particular (e.g. Welling et al., 2016). Such results suggest that exploration of ionospheric outflow may be a fruitful path toward improved substorm prediction.

5 Conclusions

The conclusions of the paper can be summarized as follows:

1. We have demonstrated a new technique for substorm identification that combines multiple substorm signatures to reduce false positive identifications as well as reduce missed identifications.
2. The technique can be tuned to produce a mean and mode waiting time that are comparable to previously published results.
3. The magnetospheric driving and response at the substorm onset times identified using our technique is consistent with expected behavior during substorms.
4. When our substorm identification technique is applied to output from an MHD simulation, we obtain a distribution of waiting times that is comparable to the observational data, driving conditions that are similar to those at the observed epoch times, and a magnetospheric response that is qualitatively similar to (though quantitatively different from) the observed response.
5. The MHD simulation has weak, but statistically significant, skill in predicting substorms.

Appendix A Procedure for identifying dipolarizations

Our procedure aims to find points that satisfy the following criteria:

- Local minimum of θ
- Onset of a rapid increase in B_z and θ
- Near a local maximum of $|B_r|$

The procedure consists of first finding local minima in θ by searching for points that are less than both of their immediate neighbors (endpoints in the data are not considered). Neighboring points around each of these local minima are checked against a set of thresholds to determine whether they satisfy the criteria given above. Given a minimum in θ, denoted by the subscript i, we specify a set of ranges $m : n$ relative to i, and a threshold B_z or $|B_r|$ must satisfy within that range in order for i to be considered a dipolarization candidate. The thresholds are defined as follows:

$$\begin{array}{rcl} max(B_{z_{i:i+10}}) & > & B_{z_i} + 2 \\ max(B_{z_{i:i+30}}) & > & B_{z_i} + 10 \\ max(B_{z_{i:i+60}}) & > & B_{z_i} + 16 \\ min(|B_r|_{i-10:i-2}) & < & |B_r|_i - 0.25 \\ min(|B_r|_{i+2:i+20}) & < & |B_r|_i - 0.5 \\ min(|B_r|_{i+10:i+40}) & < & |B_r|_i - 2 \end{array} \tag{A1}$$

The thresholds for B_z require an immediate increase in B_z (2 nT in 10 minutes), which proceeds to at least 10 nT within 30 minutes and 16 nT within 60 minutes. This is not a particularly fast increase; the thresholds are designed to identify all dipolarizations and not only the strong ones.

The thresholds for $|B_r|$ require an increase of at least 0.25 nT within the 10 minutes preceding the candidate onset, a decrease of 0.5 nT within the following 20 minutes, and a decrease of 2 nT within the following 40 minutes. These are fairly weak criteria, and are designed to select candidate onsets occurring near a local maximum, without requiring the maximum be particularly strong nor that the onset candidate occur exactly at the local maximum in $|B_r|$.

An additional procedure aims to prevent counting multiple onset times for a single dipolarization event. If an onset j is followed by an onset k within the preceding 60 minutes, then we require

$$max(B_{z_{j:k}}) > 0.25max(B_{z_{k:k+60}}); \tag{A2}$$

that is, the maximum B_z between j and k must exceed 25% of the maximum B_z reached following onset k. If this threshold is not satisfied, the onset having the lowest value of θ is kept and the other is discarded. Finally, for a candidate dipolarization to be included in the final list, the satellite providing the observations must be located on the night side; that is, MLT¡6 or MLT¿18.

The chosen thresholds are not particularly stringent individually, but in combination produce a set of dipolarizations that resembles what has been previously reported for ensembles of dipolarizations. To demonstrate this, we performed a superposed epoch analysis (SEA) of the magnetic fields for the two GOES satellites in the observations. This is shown in Figure A1, which shows superposed epoch analyses of $|B_r|$, B_z, and θ for dipolarization onsets identified from the observational data and each of the three model runs. In this figure, and throughout the paper, plots comparing the model runs to each other and to observations use a common color scheme: Observations are shown in light blue, the Hi-res w/ RCM simulation in medium blue, the Hi-res w/o RCM simulation in orange, and the SWPC simulation in green. The lines in Figure A1 represent the median of the SEA. The number of dipolarizations identified for each dataset is shown in parentheses in the legend. Although the thresholds specified allow for as little as a 16 nT increase in 60 minutes, the median increase is much faster, closer to 20 nT in 20 minutes. This is similar to what has been reported in previous studies such as Liou et al. (2002). The peaks in $|B_r|$ are less pronounced than what occurs in Liou et al. (2002). This could probably be addressed with more stringent criteria for $|B_r|$, at the cost of possibly missing some dipolarizations.

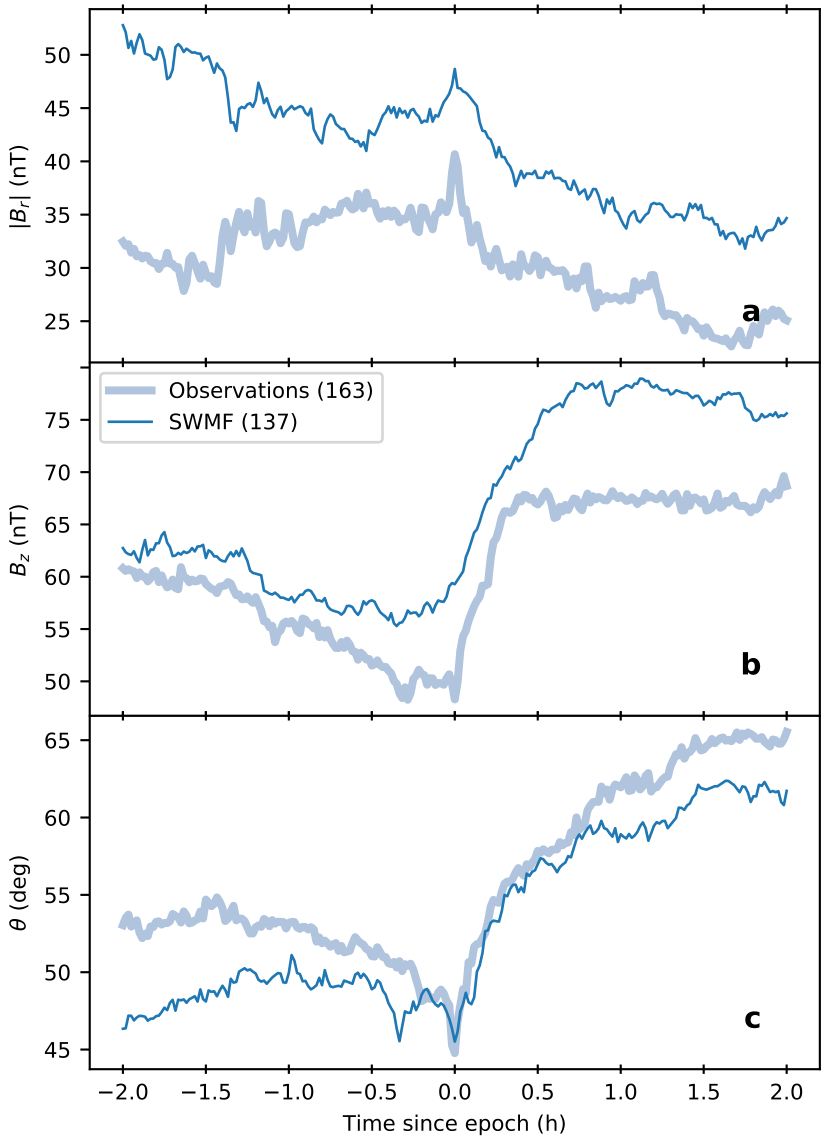

Figure A1. Superposed epoch analysis of B_r, B_z, and inclination angle θ for all dipolarization onset times.

Appendix B Comparison of inter-substorm intervals obtained using the Borovsky and Newell algorithms

Figure B1 shows distributions of waiting times for AL onsets identified using the Borovsky and Yakymenko (2017) algorithm (blue curve), for AL onsets identified using the Supermag algorithm (Newell & Gjerloev, 2011a) (orange curve) and for energetic particle injections identified from LANL satellite data by Borovsky and Yakymenko (2017) (green curve). The Supermag algorithm stands out with a modal 1-hour waiting time, while both the AL onsets and the LANL particle injections from Borovsky and Yakymenko (2017) produce a modal 3-hour waiting time. The fact that the Borovsky and Yakymenko (2017) algorithm produces a waiting time distribution that resembles that obtained using particle injections contributed to the decision to use the Borovsky and Yakymenko (2017) algorithm for substorm identification in the present work.

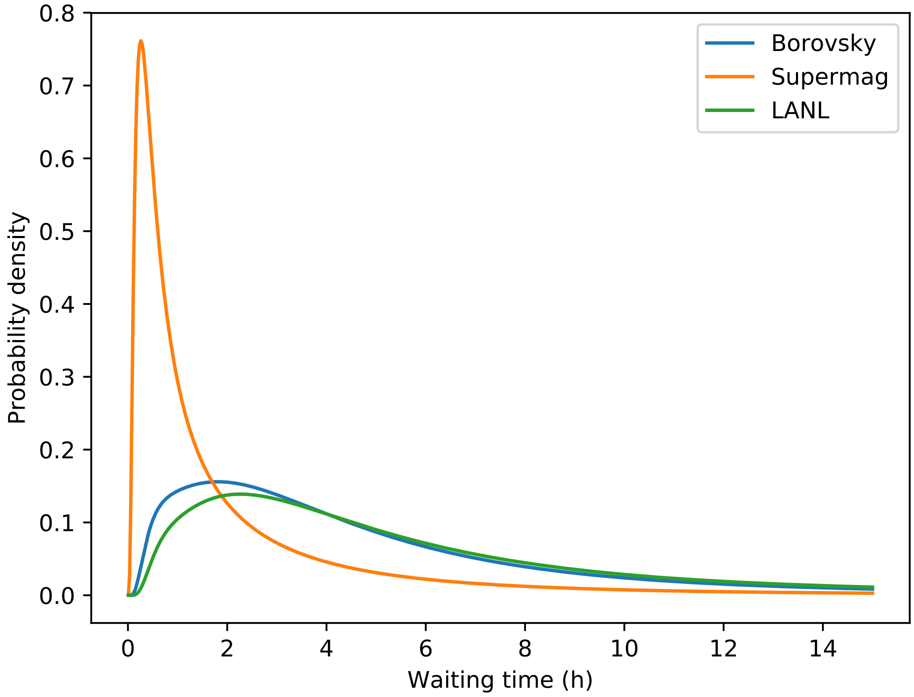

Figure B1. Substorm waiting times for onsets obtained using the Borovsky (blue curve) and Supermag (orange curve).

Appendix C Log-space computation of KDE

In Section 3.1 we visualize distributions of substorm waiting times using kernel density estimation (KDE). A KDE estimates a probability density function (PDF) by convolving samples of the PDF with a kernel function. For a set of n samples X_i and a kernel function $K(x)$, the KDE is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{C1}$$

We evaluate this using the Scipy python library, which computes h as

$$h = \frac{1}{b^2 \Sigma}, \tag{C2}$$

where Σ is the covariance of X_i and b is a scaling factor.

In this paper we take $K(x)$ to be a Gaussian. However, this introduces a difficulty because the waiting times can take only positive values (meaning that the underlying PDF is nonzero only for positive x), while $K(x)$ takes nonzero values everywhere (including negative x). To correct for this, we compute the KDE of $\log X_i$, and evaluate this KDE for $\log x$. Since this log-space transform alters the spacing (and in turn the estimated densities), we must correct this by multiplying the resulting KDE by $\frac{1}{x}$ (the derivative of $\log x$):

$$\hat{f}'(x) = \frac{1}{x}\hat{f}(\log x). \tag{C3}$$

Appendix D Bootstrapping procedure to estimate confidence intervals for forecast metrics and probability densities

The sampling distribution for the HSS is not known (Stephenson, 2000), and this means that no analytical formula is available to estimate the confidence interval. We instead employ a bootstrapping procedure (e.g. Conover, 1999), which involves randomly sampling the binary event sequence in order to obtain an estimated distribution for the skill score. This is done as follows: Given a sequence of n observed bins o_i and n predicted bins p_i, we take a sequence of n random samples, with the same indices taken from both sequences. For instance, if $n = 9$, we might have

$$o = [0, 0, 1, 1, 0, 0, 1, 0, 1] \tag{D1}$$

and

$$p = [0, 1, 0, 1, 0, 0, 0, 1, 1]. \tag{D2}$$

We then generate a sequence of n random integers representing indices to be sampled from o and p, for instance we might randomly obtain the indices [8,1,4,4,2,6,5,0,3], which would result in

$$o' = [1, 1, 1, 1, 1, 0, 0, 1, 0] \tag{D3}$$

and

$$p' = [1, 0, 0, 1, 0, 1, 0, 1, 1], \tag{D4}$$

from which we can compute a new HSS. We repeat this process N times (typically we use $N = 4000$). The 95% confidence interval for HSS is the 2.5th and 97.5th percentiles of the N skill scores obtained from the N sampled distributions. The same procedure is applied to estimate confidence intervals for POD and POFD.

To obtain a confidence interval for a kernel density estimate, a similar procedure is applied: Given a sequence of n values x_i for which a KDE is to be computed, n we generate a sequence of n random integers to be used as indices for x_i to produce a new sequence x'_j. A KDE $f_j(y)$ is computed from each sequence x'_j, and these points are evaluated at a series of points y_k. This process is repeated $N = 2000$ times, producing $n \times N$ probability density estimates $p_{jk} = f_j(y_k)$. For each y_k, the 95% confidence interval of the KDE is estimated as the 2.5th and 97.5th percentile of the p_j values obtained for that evaluation point y_k.

Acknowledgments

Thanks to our reviewers who provided detailed feedback leading to a much improved manuscript.

Thanks to Ruth Skoug of Los Alamos National Laboratory for finding solar wind data from the Advanced Composition Explorer (ACE) satellite to cover gaps in the publicly available Level 2 datasets. This data is included in the supporting information.

Supporting information for this work can be accessed at DOI 10.5281/zenodo.3625071.

Thanks to Sarah Cannon for suggesting the use of kernel density estimation for visualizing the probability density functions, which served as the inspiration for the use of a convolved Gaussian kernel function as a tool for multi-signature substorm identification.

We gratefully acknowledge the SuperMAG collaborators (`http://supermag.jhuapl.edu/info/?page=acknowledgement`) for providing the SML index data.

GOES magnetometer data and OMNI solar wind data were obtained from CDAWeb (https://cdaweb.sci.gsfc.nasa.gov/).

The Spacepy python library (S. K. Morley et al., 2011; S. Morley et al., 2014; Burrell et al., 2018) was used to read and write data in various formats (including HDF5 and the various formats used by SWMF), to interpolate the SWMF pressures and trace the magnetic field lines shown in Figure 2, and to compute superposed epoch analyses. Spacepy is available at `https://github.com/spacepy/spacepy` or DOI 10.5281/zenodo.3470304.

The Scipy python library (`https://scipy.org/`, DOI 10.5281/zenodo.3240707) was used to compute the kernel density estimations, to perform linear interpolation, to find local maxima in $f(t)$, and to evaluate the erf function.

Portions of the work of J. Haiducek leading to these results were performed while he held a National Research Council (NRC) Research Associateship award at the U.S. Naval Research Laboratory.

Portions of the work of J. Haiducek leading to these results were financially supported by the U.S. Veterans Administration under the Post/911 GI Bill.

Contributions by S. K. Morley were performed under the auspices of the US Department of Energy and were partially supported by the Laboratory Directed Research and Development program (award 20170047DR).

References

Aikio, A. T., Sergeev, V. A., Shukhtina, M. A., Vagina, L. I., Angelopoulos, V., & Reeves, G. D. (1999, jun). Characteristics of pseudobreakups and substorms observed in the ionosphere, at the geosynchronous orbit, and in the midtail. *Journal of Geophysical Research: Space Physics*, *104*(A6), 12263–12287. Retrieved from `http://doi.wiley.com/10.1029/1999JA900118` doi: 10.1029/1999JA900118

Akasofu, S.-I. (1960, sep). Large-scale auroral motions and polar magnetic disturbancesI A polar disturbance at about 1100 hours on 23 september 1957. *Journal of Atmospheric and Terrestrial Physics*, *19*(1), 10–25. Retrieved from `http://www.sciencedirect.com/science/article/pii/0021916960901033` doi: 10.1016/0021-9169(60)90103-3

Akasofu, S.-I. (1964, apr). The development of the auroral substorm. *Planetary and Space Science*, *101*(4), 273. Retrieved from `https://www.sciencedirect.com/science/article/pii/0032063364901515` doi: 10.1016/0032-0633(64)

90151-5

Akasofu, S.-I. (1968). *Polar and Magnetospheric Substorms.* Dordrecht, The Netherlands: Springer Netherlands.

Akasofu, S. I., & Meng, C. I. (1969, jan). A study of polar magnetic substorms. *Journal of Geophysical Research*, *74*(1), 293–313. Retrieved from http://doi.wiley.com/10.1029/JA074i001p00293 doi: 10.1029/JA074i001p00293

Angelopoulos, V., McFadden, J. P., Larson, D., Carlson, C. W., Mende, S. B., Frey, H., ... Kepko, L. (2008, aug). Tail reconnection triggering substorm onset. *Science (New York, N.Y.)*, *321*(5891), 931–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18653845 doi: 10.1126/science.1160495

Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W., & McPherron, R. L. (1996, June). Neutral line model of substorms: Past results and present view. *Journal of Geophysical Research: Space Physics*, *101*(A6), 12975-13010. doi: 10.1029/95JA03753

Barnston, A. G., Leetmaa, A., Kousky, V. E., Livezey, R. E., O'Lenic, E., Van den Dool, H., ... Unger, D. A. (1999, sep). NCEP Forecasts of the El Niño of 1997-98 and Its U.S. Impacts. *Bulletin of the American Meteorological Society*, *80*(9), 1829–1852. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1520-0477%281999%29080%3C1829%3ANFOTEN%3E2.0.CO%3B2 doi: 10.1175/1520-0477(1999)080⟨1829:NFOTEN⟩2.0.CO;2

Birn, J., & Hesse, M. (2013, jun). The substorm current wedge in MHD simulations. *Journal of Geophysical Research: Space Physics*, *118*(6), 3364–3376. Retrieved from http://doi.wiley.com/10.1002/jgra.50187 doi: 10.1002/jgra.50187

Birn, J., & Hones Jr., E. W. (1981). Three-Dimensional Computer Modeling of Dynamic Reconnection in the Geomagnetic Tail. *J. Geophys. Res.*, *86*(A8), 6802–6808. Retrieved from http://dx.doi.org/10.1029/JA086iA08p06802 doi: 10.1029/JA086iA08p06802

Birn, J., Nakamura, R., Panov, E. V., & Hesse, M. (2011, jan). Bursty bulk flows and dipolarization in MHD simulations of magnetotail reconnection. *Journal of Geophysical Research: Space Physics*, *116*(A1), A01210. Retrieved from http://doi.wiley.com/10.1029/2010JA016083 doi: 10.1029/2010JA016083

Birn, J., Thomsen, M. F., Borovsky, J. E., Reeves, G. D., McComas, D. J., Belian, R. D., & Hesse, M. (1998, may). Substorm electron injections: Geosynchronous observations and test particle simulations. *Journal of Geophysical Research: Space Physics*, *103*(A5), 9235–9248. Retrieved from http://doi.wiley.com/10.1029/97JA02635 doi: 10.1029/97JA02635

Boakes, P. D., Milan, S. E., Abel, G. A., Freeman, M. P., Chisham, G., & Hubert, B. (2009). A statistical study of the open magnetic flux content of the magnetosphere at the time of substorm onset. *Geophysical Research Letters*, *36*(4), L04105. Retrieved from http://dx.doi.org/10.1029/2008GL037059 (L04105) doi: 10.1029/2008GL037059

Bonnevier, B., Boström, R., & Rostoker, G. (1970, jan). A three-dimensional model current system for polar magnetic substorms. *Journal of Geophysical Research*, *75*(1), 107–122. Retrieved from http://doi.wiley.com/10.1029/JA075i001p00107 doi: 10.1029/JA075i001p00107

Borovsky, J. E., Nemzek, R. J., & Belian, R. D. (1993, mar). The occurrence rate of magnetospheric-substorm onsets: Random and periodic substorms. *Journal of Geophysical Research: Space Physics*, *98*(A3), 3807–3813. Retrieved from http://doi.wiley.com/10.1029/92JA02556 doi: 10.1029/92JA02556

Borovsky, J. E., & Yakymenko, K. (2017, mar). Substorm occurrence rates, substorm recurrence times, and solar wind structure. *Journal of Geophysical Research: Space Physics*, *122*(3), 2973–2998. Retrieved from http://doi.wiley.com/10.1002/2016JA023625 doi: 10.1002/2016JA023625

Burrell, A. G., Halford, A. J., Klenzing, J., Stoneback, R. A., Morley, S. K., Annex,

A. M., ... Ma, J. (2018, November). Snakes on a Spaceship - An Overview of Python in Heliophysics. *Journal of Geophysical Research: Space Physics*, *123*(12), 10384-10402. doi: 10.1029/2018JA025877

Caan, M. N., McPherron, R. L., & Russell, C. T. (1975, jan). Substorm and interplanetary magnetic field effects on the geomagnetic tail lobes. *Journal of Geophysical Research*, *80*(1), 191–194. Retrieved from http://doi.wiley.com/10.1029/JA080i001p00191 doi: 10.1029/JA080i001p00191

Caan, M. N., McPherron, R. L., & Russell, C. T. (1977, oct). Characteristics of the association between the interplanetary magnetic field and substorms. *Journal of Geophysical Research*, *82*(29), 4837–4842. Retrieved from http://doi.wiley.com/10.1029/JA082i029p04837 doi: 10.1029/JA082i029p04837

Caan, M. N., McPherron, R. L., & Russell, C. T. (1978, mar). The statistical magnetic signature of magnetospheric substorms. *Planetary and Space Science*, *26*(3), 269–279. Retrieved from https://www.sciencedirect.com/science/article/pii/0032063378900922 doi: 10.1016/0032-0633(78)90092-2

Carter, J. V., Pan, J., Rai, S. N., & Galandiuk, S. (2016, June). ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*, *159*(6), 1638-1645. doi: 10.1016/j.surg.2015.12.029

Cayton, T., & Belian, R. D. (2007). *Numerical modeling of the Synchronous Orbit Particle Analyzer (SOPA, Version 2) that flew on S/C 1990-095* (Tech. Rep.). Los Alamos, NM: Los Alamos National Laboratory. Retrieved from http://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-14335

Chen, Y., Tóth, G., Cassak, P., Jia, X., Gombosi, T. I., Slavin, J. A., ... Henderson, M. G. (2017, oct). Global Three-Dimensional Simulation of Earth's Dayside Reconnection Using a Two-Way Coupled Magnetohydrodynamics With Embedded Particle-in-Cell Model: Initial Results. *Journal of Geophysical Research: Space Physics*, *122*(10), 10,318–10,335. Retrieved from http://doi.wiley.com/10.1002/2017JA024186 doi: 10.1002/2017JA024186

Chu, X., McPherron, R. L., Hsu, T. S., & Angelopoulos, V. (2015, apr). Solar cycle dependence of substorm occurrence and duration: Implications for onset. *Journal of Geophysical Research A: Space Physics*, *120*(4), 2808–2818. Retrieved from http://doi.wiley.com/10.1002/2015JA021104 doi: 10.1002/2015JA021104

Clauer, C. R., & Kamide, Y. (1985). DP 1 and DP 2 current systems for the March 22, 1979 substorms. *Journal of Geophysical Research*, *90*(A2), 1343. Retrieved from http://doi.wiley.com/10.1029/JA090iA02p01343 doi: 10.1029/JA090iA02p01343

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). Hoboken, NJ: Wiley.

Coroniti, F. V., & Kennel, C. F. (1972, jul). Changes in magnetospheric configuration during the substorm growth phase. *Journal of Geophysical Research*, *77*(19), 3361–3370. Retrieved from http://doi.wiley.com/10.1029/JA077i019p03361 doi: 10.1029/JA077i019p03361

Cummings, W. D., & Coleman, P. J. (1968, jul). Simultaneous Magnetic Field Variations at the Earth's Surface and at Synchronous, Equatorial Distance. Part I. Bay-Associated Events. *Radio Science*, *3*(7), 758–761. Retrieved from http://doi.wiley.com/10.1002/rds196837758 doi: 10.1002/rds196837758

Davis, T. N., & Sugiura, M. (1966, feb). Auroral electrojet activity index AE and its universal time variations. *Journal of Geophysical Research*, *71*(3), 785–801. Retrieved from http://dx.doi.org/10.1029/JZ071i003p00785http://doi.wiley.com/10.1029/JZ071i003p00785 doi: 10.1029/JZ071i003p00785

De Zeeuw, D., Gombosi, T., Groth, C., Powell, K., & Stout, Q. (2000). An adaptive MHD method for global space weather simulations. *IEEE Transactions on Plasma Science*, *28*(6), 1956–1965. Retrieved from http://ieeexplore.ieee

.org/document/902224/ doi: 10.1109/27.902224

DeForest, S. E., & McIlwain, C. E. (1971, jun). Plasma clouds in the magnetosphere. *Journal of Geophysical Research*, *76*(16), 3587–3611. Retrieved from http://doi.wiley.com/10.1029/JA076i016p03587 doi: 10.1029/JA076i016p03587

Eastwood, J. P., Sibeck, D. G., Slavin, J. A., Goldstein, M. L., Lavraud, B., Sitnov, M., … Dandouras, I. (2005, jun). Observations of multiple X-line structure in the Earth's magnetotail current sheet: A Cluster case study. *Geophysical Research Letters*, *32*(11), L11105. Retrieved from http://doi.wiley.com/10.1029/2005GL022509 doi: 10.1029/2005GL022509

Ekelund, S. (2012, March). ROC Curves—What are They and How are They Used? *Point of Care*, *11*(1), 16-21. doi: 10.1097/POC.0b013e318246a642

El-Alaoui, M., Ashour-Abdalla, M., Walker, R. J., Peroomian, V., Richard, R. L., Angelopoulos, V., & Runov, A. (2009, aug). Substorm evolution as revealed by THEMIS satellites and a global MHD simulation. *Journal of Geophysical Research*, *114*(A8), A08221. Retrieved from http://doi.wiley.com/10.1029/2009JA014133 doi: 10.1029/2009JA014133

Forsyth, C., Fazakerley, A. N., Rae, I. J., J. Watt, C. E., Murphy, K., Wild, J. A., … Zhang, Y. (2014, feb). In situ spatiotemporal measurements of the detailed azimuthal substructure of the substorm current wedge. *Journal of Geophysical Research: Space Physics*, *119*(2), 927–946. Retrieved from http://doi.wiley.com/10.1002/2013JA019302 doi: 10.1002/2013JA019302

Forsyth, C., Rae, I. J., Coxon, J. C., Freeman, M. P., Jackman, C. M., Gjerloev, J., & Fazakerley, A. N. (2015, dec). A new technique for determining Substorm Onsets and Phases from Indices of the Electrojet (SOPHIE). *Journal of Geophysical Research: Space Physics*, *120*(12), 10,592–10,606. Retrieved from http://doi.wiley.com/10.1002/2015JA021343 doi: 10.1002/2015JA021343

Freeman, M. P., & Morley, S. K. (2004, jun). A minimal substorm model that explains the observed statistical distribution of times between substorms. *Geophysical Research Letters*, *31*(12), L12807. Retrieved from http://doi.wiley.com/10.1029/2004GL019989http://dx.doi.org/10.1029/2004GL019989 doi: 10.1029/2004GL019989

Freeman, M. P., & Morley, S. K. (2009, nov). No evidence for externally triggered substorms based on superposed epoch analysis of IMF B z. *Geophysical Research Letters*, *36*(21), L21101. Retrieved from http://doi.wiley.com/10.1029/2009GL040621 doi: 10.1029/2009GL040621

Frey, H. U. (2010). Comment on "Substorm triggering by new plasma intrusion: THEMIS all-sky imager observations" by Y. Nishimura et al. *Journal of Geophysical Research: Space Physics*, *115*(A12), A12232. doi: 10.1029/2010JA016113

Frey, H. U., Mende, S. B., Angelopoulos, V., & Donovan, E. F. (2004, oct). Substorm onset observations by IMAGE-FUV. *Journal of Geophysical Research*, *109*(A10), A10304. Retrieved from http://doi.wiley.com/10.1029/2004JA010607 doi: 10.1029/2004JA010607

Frühauff, D., & Glassmeier, K.-H. (2017, nov). The Plasma Sheet as Natural Symmetry Plane for Dipolarization Fronts in the Earth's Magnetotail. *Journal of Geophysical Research: Space Physics*, *122*(11), 11,373–11,388. Retrieved from http://doi.wiley.com/10.1002/2017JA024682 doi: 10.1002/2017JA024682

Fu, H. S., Khotyaintsev, Y. V., Vaivads, A., André, M., & Huang, S. Y. (2012, may). Occurrence rate of earthward-propagating dipolarization fronts. *Geophysical Research Letters*, *39*(10), L10101. Retrieved from http://doi.wiley.com/10.1029/2012GL051784 doi: 10.1029/2012GL051784

Ganushkina, N. Y., Amariutei, O. A., Welling, D., & Heynderickx, D. (2015, jan).

Nowcast model for low-energy electrons in the inner magnetosphere. *Space Weather*, *13*(1), 16–34. Retrieved from http://doi.wiley.com/10.1002/2014SW001098 doi: 10.1002/2014SW001098

Gjerloev, J. W. (2012, September). The SuperMAG data processing technique. *Journal of Geophysical Research: Space Physics*, *117*(A9). doi: 10.1029/2012JA017683

Glocer, A., Rastätter, L., Kuznetsova, M., Pulkkinen, A., Singer, H. J., Balch, C., ... Wing, S. (2016, jul). Community-wide validation of geospace model local K-index predictions to support model transition to operations. *Space Weather*, *14*(7), 469–480. Retrieved from http://doi.wiley.com/10.1002/2016SW001387 doi: 10.1002/2016SW001387

Glocer, A., Tóth, G., Gombosi, T., & Welling, D. (2009, may). Modeling ionospheric outflows and their impact on the magnetosphere, initial results. *Journal of Geophysical Research: Space Physics*, *114*(A5), A05216. Retrieved from http://doi.wiley.com/10.1029/2009JA014053 doi: 10.1029/2009JA014053

Glocer, A., Tóth, G., Ma, Y., Gombosi, T., Zhang, J.-C., & Kistler, L. M. (2009, dec). Multifluid Block-Adaptive-Tree Solar wind Roe-type Upwind Scheme: Magnetospheric composition and dynamics during geomagnetic storms-Initial results. *Journal of Geophysical Research: Space Physics*, *114*(A12), A12203. Retrieved from http://doi.wiley.com/10.1029/2009JA014418 doi: 10.1029/2009JA014418

Haiducek, J. D., Welling, D. T., Ganushkina, N. Y., Morley, S. K., & Ozturk, D. S. (2017, dec). SWMF Global Magnetosphere Simulations of January 2005: Geomagnetic Indices and Cross-Polar Cap Potential. *Space Weather*, *15*(12), 1567–1587. Retrieved from http://doi.wiley.com/10.1002/2017SW001695 doi: 10.1002/2017SW001695

Hajra, R., Tsurutani, B. T., Echer, E., Gonzalez, W. D., & Gjerloev, J. W. (2016, August). Supersubstorms (SML $<$ -2500 nT): Magnetic storm and solar cycle dependences. *Journal of Geophysical Research: Space Physics*, *121*(8), 7805-7816. doi: 10.1002/2015JA021835

Henderson, M. G. (2009). Observational evidence for an inside-out substorm onset scenario. *Annales Geophysicae*, *27*(5), 2129–2140. Retrieved from http://www.ann-geophys.net/27/2129/2009/angeo-27-2129-2009.pdf doi: 10.5194/angeo-27-2129-2009

Hendry, A. T., Rodger, C. J., Clilverd, M. A., Thomson, N. R., Morley, S. K., & Raita, T. (2013, March). Rapid Radiation Belt Losses Occurring During High-Speed Solar Wind Stream–Driven Storms: Importance of Energetic Electron Precipitation. In *Dynamics of the Earth's Radiation Belts and Inner Magnetosphere* (Vol. 199, p. 213-224). Washington, DC: American Geophysical Union (AGU). doi: 10.1029/2012GM001299

Heppner, J. P. (1955, mar). Note on the occurrence of world-wide S.S.C.'s during the onset of negative bays at College, Alaska. *Journal of Geophysical Research*, *60*(1), 29–32. Retrieved from http://doi.wiley.com/10.1029/JZ060i001p00029 doi: 10.1029/JZ060i001p00029

Hones, E. W. (1984). Plasma sheet behavior during substorms. In *Magnetic Reconnection in Space and Laboratory Plasmas* (Vol. 30, p. 178-184). Washington, DC: American Geophysical Union (AGU). doi: 10.1029/GM030p0178

Hones, E. W., Birn, J., Baker, D. N., Bame, S. J., Feldman, W. C., McComas, D. J., ... Tsurutani, B. T. (1984, oct). Detailed examination of a plasmoid in the distant magnetotail with ISEE 3. *Geophysical Research Letters*, *11*(10), 1046–1049. Retrieved from http://dx.doi.org/10.1029/GL011i010p01046http://doi.wiley.com/10.1029/GL011i010p01046 doi: 10.1029/GL011i010p01046

Honkonen, I., Palmroth, M., Pulkkinen, T. I., Janhunen, P., & Aikio, A. (2011, jan). On large plasmoid formation in a global magnetohydrodynamic simulation. *Annales Geophysicae*, *29*(1), 167–179. Retrieved from http://www.ann-geophys.net/29/167/2011/ doi: 10.5194/angeo-29-167-2011

Hsu, T.-S., & McPherron, R. L. (2003, jul). Occurrence frequencies of IMF triggered and nontriggered substorms. *Journal of Geophysical Research*, *108*(A7), 1307. Retrieved from http://doi.wiley.com/10.1029/2002JA009442 doi: 10.1029/2002JA009442

Hsu, T.-S., & McPherron, R. L. (2004, jul). Average characteristics of triggered and nontriggered substorms. *Journal of Geophysical Research*, *109*(A7), A07208. Retrieved from http://doi.wiley.com/10.1029/2003JA009933 doi: 10.1029/2003JA009933

Hsu, T.-S., & McPherron, R. L. (2012, nov). A statistical analysis of substorm associated tail activity. *Advances in Space Research*, *50*(10), 1317–1343. Retrieved from https://www.sciencedirect.com/science/article/pii/S0273117712004371 doi: 10.1016/J.ASR.2012.06.034

Ieda, A., Fairfield, D. H., Mukai, T., Saito, Y., Kokubun, S., Liou, K., … Brittnacher, M. J. (2001, mar). Plasmoid ejection and auroral brightenings. *Journal of Geophysical Research: Space Physics*, *106*(A3), 3845–3857. Retrieved from http://doi.wiley.com/10.1029/1999JA000451http://dx.doi.org/10.1029/1999JA000451 doi: 10.1029/1999JA000451

Iyemori, T., & Rao, D. R. K. (1996). Decay of the Dst field of geomagnetic disturbance after substorm onset and its implication to storm-substorm relation. *Annales Geophysicae*, *14*(6), 608–618. Retrieved from http://www.ann-geophys.net/14/608/1996/ doi: 10.1007/s00585-996-0608-3

Johnson, J. R., & Wing, S. (2014). External versus internal triggering of substorms: An information-theoretical approach. *Geophysical Research Letters*, *41*(16), 5748-5754. doi: 10.1002/2014GL060928

Jordanova, V., Delzanno, G., Henderson, M., Godinez, H., Jeffery, C., Lawrence, E., … Horne, R. (2017, nov). Specification of the near-Earth space environment with SHIELDS. *Journal of Atmospheric and Solar-Terrestrial Physics*, *177*, 148-159. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364682617302729 doi: 10.1016/j.jastp.2017.11.006

Juusola, L., Østgaard, N., Tanskanen, E., Partamies, N., & Snekvik, K. (2011, oct). Earthward plasma sheet flows during substorm phases. *Journal of Geophysical Research: Space Physics*, *116*(A10), A10228. Retrieved from http://doi.wiley.com/10.1029/2011JA016852 doi: 10.1029/2011JA016852

Kamide, Y., & McIlwain, C. E. (1974, nov). The onset time of magnetospheric substorms determined from ground and synchronous satellite records. *Journal of Geophysical Research*, *79*(31), 4787–4790. Retrieved from http://doi.wiley.com/10.1029/JA079i031p04787 doi: 10.1029/JA079i031p04787

Kamide, Y., Yasuhara, F., & Akasofu, S.-I. (1974, aug). On the cause of northward magnetic field along the negative X axis during magnetospheric substorms. *Planetary and Space Science*, *22*(8), 1219–1229. Retrieved from https://www.sciencedirect.com/science/article/pii/0032063374900063 doi: 10.1016/0032-0633(74)90006-3

Katus, R. M., & Liemohn, M. W. (2013, August). Similarities and differences in low- to middle-latitude geomagnetic indices. *Journal of Geophysical Research: Space Physics*, *118*(8), 5149-5156. doi: 10.1002/jgra.50501

Kaufmann, R. L. (1987). Substorm currents: Growth phase and onset. *Journal of Geophysical Research*, *92*(A7), 7471. Retrieved from http://doi.wiley.com/10.1029/JA092iA07p07471 doi: 10.1029/JA092iA07p07471

Kauristie, K., Morschhauser, A., Olsen, N., Finlay, C. C., McPherron, R. L., Gjerloev, J. W., & Opgenoorth, H. J. (2017, mar). On the Usage of Geomagnetic Indices for Data Selection in Internal Field Modelling. *Space Science Reviews*, *206*(1-4), 61–90. Retrieved from http://link.springer.com/10.1007/s11214-016-0301-0 doi: 10.1007/s11214-016-0301-0

Kepko, L. (2004). Relative timing of substorm onset phenomena. *Journal of Geophysical Research*, *109*(A4). doi: 10.1029/2003JA010285

Kepko, L., & McPherron, R. L. (2001). Comment on "Evaluation of low-latitude Pi2 pulsations as indicators of substorm onset using Polar ultraviolet imagery" by K. Liou, et al. *Journal of Geophysical Research: Space Physics*, *106*(A9), 18919-18922. doi: 10.1029/2000JA000189

Kepko, L., McPherron, R. L., Amm, O., Apatenkov, S., Baumjohann, W., Birn, J., ... Sergeev, V. (2015, jul). Substorm Current Wedge Revisited. *Space Science Reviews*, *190*(1-4), 1–46. Retrieved from http://link.springer.com/10.1007/s11214-014-0124-9 doi: 10.1007/s11214-014-0124-9

Kim, K.-H., Takahashi, K., Lee, D., Sutcliffe, P. R., & Yumoto, K. (2005). Pi2 pulsations associated with poleward boundary intensifications during the absence of substorms. *Journal of Geophysical Research*, *110*(A1), A01217. Retrieved from http://doi.wiley.com/10.1029/2004JA010780 doi: 10.1029/2004JA010780

Korth, A., Pu, Z. Y., Kremser, G., & Roux, A. (1991, mar). A Statistical Study of Substorm Onset Conditions at Geostationary Orbit. In J. R. Kan, T. A. Potemra, S. Kokubun, & T. Iijima (Eds.), *Magnetospheric substorms* (Vol. 64, pp. 343–351). Washington, D. C.: American Geophysical Union (AGU). Retrieved from http://doi.wiley.com/10.1029/GM064p0343 doi: 10.1029/GM064p0343

Koskinen, H. E. J., Lopez, R. E., Pellinen, R. J., Pulkkinen, T. I., Baker, D. N., & Bösinger, T. (1993, apr). Pseudobreakup and substorm growth phase in the ionosphere and magnetosphere. *Journal of Geophysical Research: Space Physics*, *98*(A4), 5801–5813. Retrieved from http://doi.wiley.com/10.1029/92JA02482 doi: 10.1029/92JA02482

Kullen, A., Ohtani, S., & Karlsson, T. (2009, April). Geomagnetic signatures of auroral substorms preceded by pseudobreakups. *Journal of Geophysical Research: Space Physics*, *114*(A4), A04201. doi: 10.1029/2008JA013712

Lee, D.-Y., & Lyons, L. R. (2004, apr). Geosynchronous magnetic field response to solar wind dynamic pressure pulse. *Journal of Geophysical Research*, *109*(A4), A04201. Retrieved from http://doi.wiley.com/10.1029/2003JA010076 doi: 10.1029/2003JA010076

Lezniak, T. W., Arnoldy, R. L., Parks, G. K., & Winckler, J. R. (1968, jul). Measurement and Intensity of Energetic Electrons at the Equator at 6.6 Re. *Radio Science*, *3*(7), 710–714. Retrieved from http://doi.wiley.com/10.1002/rds196837710 doi: 10.1002/rds196837710

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., ... Vasile, R. (2018, November). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, *16*(12), 2079-2102. doi: 10.1029/2018SW002067

Liou, K. (2010, dec). Polar Ultraviolet Imager observation of auroral breakup. *Journal of Geophysical Research: Space Physics*, *115*(A12), A12219. Retrieved from http://doi.wiley.com/10.1029/2010JA015578 doi: 10.1029/2010JA015578

Liou, K., Meng, C.-I., Lui, A. T. Y., Newell, P. T., & Wing, S. (2002). Magnetic dipolarization with substorm expansion onset. *Magnetospheric Physics: Magnetosphereinner*, *107*(A12), 1131. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001JA000179 doi: 10.1029/2001JA000179

Liou, K., Meng, C.-I., Lui, T. Y., Newell, P. T., Brittnacher, M., Parks, G., ... Yumoto, K. (1999). On relative timing in substorm onset signatures. *Journal of Geophysical Research: Space Physics*, *104*(A10), 22807-22817. doi: 10.1029/1999JA900206

Liou, K., Meng, C.-I., Newell, P. T., Takahashi, K., Ohtani, S.-I., Lui, A. T. Y.,

... Parks, G. (2000). Evaluation of low-latitude Pi2 pulsations as indicators of substorm onset using Polar ultraviolet imagery. *Journal of Geophysical Research: Space Physics*, *105*(A2), 2495-2505. doi: 10.1029/1999JA900416

Liu, J., Angelopoulos, V., Runov, A., & Zhou, X.-Z. (2013, may). On the current sheets surrounding dipolarizing flux bundles in the magnetotail: The case for wedgelets. *Journal of Geophysical Research: Space Physics*, *118*(5), 2000–2020. Retrieved from http://doi.wiley.com/10.1002/jgra.50092 doi: 10.1002/jgra.50092

Lopez, R. E., Hernandez, S., Wiltberger, M., Huang, C. L., Kepko, E. L., Spence, H., ... Lyon, J. G. (2007, jan). Predicting magnetopause crossings at geosynchronous orbit during the Halloween storms. *Space Weather*, *5*(1), S01005. Retrieved from http://doi.wiley.com/10.1029/2006SW000222 doi: 10.1029/2006SW000222

Lui, A. T. Y. (1978, oct). Estimates of current changes in the geomagnetotail associated with a substorm. *Geophysical Research Letters*, *5*(10), 853–856. Retrieved from http://doi.wiley.com/10.1029/GL005i010p00853 doi: 10.1029/GL005i010p00853

Lui, A. T. Y. (1991). Extended Consideration of a Synthesis Model for Magnetospheric Substorms. In *Magnetospheric Substorms* (Vol. 64, p. 43-60). Washington, DC: American Geophysical Union (AGU). doi: 10.1029/GM064p0043

Lyon, J. G., Brecht, S. H., Huba, J. D., Fedder, J. A., & Palmadesso, P. J. (1981, apr). Computer Simulation of a Geomagnetic Substorm. *Physical Review Letters*, *46*(15), 1038–1041. Retrieved from https://link.aps.org/doi/10.1103/PhysRevLett.46.1038 doi: 10.1103/PhysRevLett.46.1038

Lyons, L. R., Blanchard, G. T., Samson, J. C., Lepping, R. P., Yamamoto, T., & Moretto, T. (1997, dec). Coordinated observations demonstrating external substorm triggering. *Journal of Geophysical Research: Space Physics*, *102*(A12), 27039–27051. Retrieved from http://doi.wiley.com/10.1029/97JA02639 doi: 10.1029/97JA02639

McPherron, R. (1972, sep). Substorm related changes in the geomagnetic tail: the growth phase. *Planetary and Space Science*, *20*(9), 1521–1539. Retrieved from https://www.sciencedirect.com/science/article/pii/0032063372900542 doi: 10.1016/0032-0633(72)90054-2

McPherron, R. L. (1970, oct). Growth phase of magnetospheric substorms. *Journal of Geophysical Research*, *75*(28), 5592–5599. Retrieved from http://doi.wiley.com/10.1029/JA075i028p05592 doi: 10.1029/JA075i028p05592

McPherron, R. L. (2015, jan). Earth's Magnetotail. In A. Keiling, C. M. Jackman, & P. A. Delamere (Eds.), *Magnetotails in the solar system* (pp. 61–84). Hoboken, NJ: American Geophysical Union (AGU). Retrieved from http://doi.wiley.com/10.1002/9781118842324.ch4 doi: 10.1002/9781118842324.ch4

McPherron, R. L., & Chu, X. (2017, mar). The Mid-Latitude Positive Bay and the MPB Index of Substorm Activity. *Space Science Reviews*, *206*(1-4), 91–122. Retrieved from http://link.springer.com/10.1007/s11214-016-0316-6 doi: 10.1007/s11214-016-0316-6

McPherron, R. L., & Chu, X. (2018, apr). The Midlatitude Positive Bay Index and the Statistics of Substorm Occurrence. *Journal of Geophysical Research: Space Physics*, *123*(4), 2831-2850. Retrieved from http://doi.wiley.com/10.1002/2017JA024766 doi: 10.1002/2017JA024766

McPherron, R. L., Russell, C. T., & Aubry, M. P. (1973, jun). Satellite studies of magnetospheric substorms on August 15, 1968: 9. Phenomenological model for substorms. *Journal of Geophysical Research*, *78*(16), 3131–3149. Retrieved from http://doi.wiley.com/10.1029/JA078i016p03131 doi: 10.1029/JA078i016p03131

Meng, X., Tóth, G., Glocer, A., Fok, M.-C., & Gombosi, T. I. (2013, sep). Pressure anisotropy in global magnetospheric simulations: Coupling with ring current models. *Journal of Geophysical Research: Space Physics*, *118*(9), 5639–5658. Retrieved from http://doi.wiley.com/10.1002/jgra.50539 doi: 10.1002/jgra.50539

Meng, X., Tóth, G., Liemohn, M. W., Gombosi, T. I., & Runov, A. (2012, aug). Pressure anisotropy in global magnetospheric simulations: A magnetohydrodynamics model. *Journal of Geophysical Research: Space Physics*, *117*(A8). Retrieved from http://doi.wiley.com/10.1029/2012JA017791 doi: 10.1029/2012JA017791

Moldwin, M. B., & Hughes, W. J. (1993, jan). Geomagnetic substorm association of plasmoids. *Journal of Geophysical Research: Space Physics*, *98*(A1), 81–88. Retrieved from http://doi.wiley.com/10.1029/92JA02153 doi: 10.1029/92JA02153

Morley, S., Koller, J., Welling, D., Larsen, B., & Niehof, J. (2014, January). SpacePy: Python-Based Tools for the Space Science Community. *Astrophysics Source Code Library*, ascl:1401.002.

Morley, S., Rouillard, A., & Freeman, M. (2009). Recurrent substorm activity during the passage of a corotating interaction region. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(10), 1073–1081. doi: 10.1016/j.jastp.2008.11.009

Morley, S. K. (2007). Observations of magnetospheric substorms during the passage of a corotating interaction region. In W. Short & I. Cairns (Eds.), *Proceedings from 7th Australian Space Science Conference, 2007* (p. 118-129). Sydney: National Space Society of Australia Ltd.

Morley, S. K., & Freeman, M. P. (2007, April). On the association between northward turnings of the interplanetary magnetic field and substorm onsets. *Geophysical Research Letters*, *34*(8), L08104. doi: 10.1029/2006GL028891

Morley, S. K., Friedel, R. H. W., Spanswick, E. L., Reeves, G., Steinberg, J. T., Koller, J., ... Noveroske, E. (2010, November). Dropouts of the outer electron radiation belt in response to solar wind stream interfaces: Global positioning system observations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *466*(2123), 3329-3350. doi: 10.1098/rspa.2010.0078

Morley, S. K., Welling, D. T., Koller, J., Larsen, B. A., Henderson, M. G., & Niehof, J. (2011). SpacePy - A Python-based Library of Tools for the Space Sciences. *Proceedings of the 9th Python in Science Conference*, 39-45.

Morley, S. K., Welling, D. T., & Woodroffe, J. R. (2018, August). Perturbed Input Ensemble Modeling with the Space Weather Modeling Framework. *Space Weather*, *16*(9), 1330-1347. doi: 10.1029/2018SW002000

Mukhopadhyay, A., Welling, D. T., Liemohn, M. W., Zou, S., & Ridley, A. J. (2018, December). *Challenges in Space Weather Prediction: Estimation of Auroral Conductance.* Washington, DC.

Nagai, T. (1987, mar). Field-aligned currents associated with substorms in the vicinity of synchronous orbit: 2. GOES 2 and GOES 3 observations. *Journal of Geophysical Research*, *92*(A3), 2432. Retrieved from http://doi.wiley.com/10.1029/JA092iA03p02432 doi: 10.1029/JA092iA03p02432

Nagai, T., Fujimoto, M., Saito, Y., Machida, S., Terasawa, T., Nakamura, R., ... Kokubun, S. (1998, mar). Structure and dynamics of magnetic reconnection for substorm onsets with Geotail observations. *Journal of Geophysical Research: Space Physics*, *103*(A3), 4419–4440. Retrieved from http://doi.wiley.com/10.1029/97JA02190 doi: 10.1029/97JA02190

Newell, P. T., & Gjerloev, J. W. (2011a, dec). Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. *Journal of Geophysical Research: Space Physics*, *116*(A12), A12211. Retrieved from http://doi.wiley.com/10.1029/2011JA016779 doi: 10.1029/2011JA016779

Newell, P. T., & Gjerloev, J. W. (2011b, dec). Substorm and magnetosphere characteristic scales inferred from the SuperMAG auroral electrojet indices. *Journal of Geophysical Research: Space Physics*, *116*(A12), A12232. Retrieved from http://doi.wiley.com/10.1029/2011JA016936 doi: 10.1029/2011JA016936

Newell, P. T., & Liou, K. (2011, mar). Solar wind driving and substorm triggering. *Journal of Geophysical Research: Space Physics*, *116*(A3), A03229. Retrieved from http://doi.wiley.com/10.1029/2010JA016139 doi: 10.1029/2010JA016139

Newell, P. T., Liou, K., Sotirelis, T., & Meng, C.-I. (2001, dec). Auroral precipitation power during substorms: A Polar UV Imager-based superposed epoch analysis. *Journal of Geophysical Research: Space Physics*, *106*(A12), 28885–28896. Retrieved from http://doi.wiley.com/10.1029/2000JA000428 doi: 10.1029/2000JA000428

Nishida, A., Scholer, M., Terasawa, T., Bame, S. J., Gloeckler, G., Smith, E. J., & Zwickl, R. D. (1986, apr). Quasi-stagnant plasmoid in the middle tail: A new preexpansion phase phenomenon. *Journal of Geophysical Research*, *91*(A4), 4245. Retrieved from http://doi.wiley.com/10.1029/JA091iA04p04245 doi: 10.1029/JA091iA04p04245

Noah, M. A., & Burke, W. J. (2013, aug). Sawtooth-substorm connections: A closer look. *Journal of Geophysical Research: Space Physics*, *118*(8), 5136–5148. Retrieved from http://doi.wiley.com/10.1002/jgra.50440 doi: 10.1002/jgra.50440

Nosé, M., Iyemori, T., Takeda, M., Toh, H., Ookawa, T., Cifuentes-Nava, G., ... Curto, J. J. (2009). New substorm index derived from high-resolution geomagnetic field data at low latitude and its comparison with AE and ASY indices. In J. Love (Ed.), *Proc. xiii iaga workshop* (pp. 202–207). Golden, CO. Retrieved from https://geomag.usgs.gov/downloads/publications/Proceedings_202-207.pdf

Ohtani, S., Anderson, B. J., Sibeck, D. G., Newell, P. T., Zanetti, L. J., Potemra, T. A., ... Russell, C. T. (1993, nov). A multisatellite study of a pseudo-substorm onset in the near-Earth magnetotail. *Journal of Geophysical Research: Space Physics*, *98*(A11), 19355–19367. Retrieved from http://doi.wiley.com/10.1029/93JA01421 doi: 10.1029/93JA01421

Ohtani, S.-i., & Raeder, J. (2004, jan). Tail current surge: New insights from a global MHD simulation and comparison with satellite observations. *Journal of Geophysical Research*, *109*(A1), A01207. Retrieved from http://doi.wiley.com/10.1029/2002JA009750 doi: 10.1029/2002JA009750

Øieroset, M., Phan, T. D., Fujimoto, M., Lin, R. P., & Lepping, R. P. (2001). In situ detection of reconnection in the Earth's magnetotail. *Letters to Nature*, *412*(July), 414–416. doi: 10.1038/35086520

Øieroset, M., Yamauchi, M., Liszka, L., Christon, S. P., & Hultqvist, B. (1999, apr). A statistical study of ion beams and conics from the dayside ionosphere during different phases of a substorm. *Journal of Geophysical Research: Space Physics*, *104*(A4), 6987–6998. Retrieved from http://doi.wiley.com/10.1029/1998JA900177 doi: 10.1029/1998JA900177

Parzen, E. (1962, sep). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, *33*(3), 1065–1076. Retrieved from http://projecteuclid.org/euclid.aoms/1177704472 doi: 10.1214/aoms/1177704472

Perreault, P., & Akasofu, S.-I. (1978, sep). A study of geomagnetic storms. *Geophysical Journal International*, *54*(3), 547–573. Retrieved from https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.1978.tb05494.x doi: 10.1111/j.1365-246X.1978.tb05494.x

Powell, K. G., Roe, P. L., Linde, T. J., Gombosi, T. I., & De Zeeuw, D. L. (1999, sep). A Solution-Adaptive Upwind Scheme for Ideal Magnetohydrodynamics. *Journal of Computational Physics*, *154*(2), 284–309. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S002199919996299X doi: 10.1006/jcph.1999.6299

Pulkkinen, A., Rastätter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., ... Weigel, R. (2013, jun). Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space Weather*, *11*(6), 369–385. Retrieved from http://doi.wiley.com/10.1002/swe.20056 doi: 10.1002/swe.20056

Pytte, T., Mcpherron, R., & Kokubun, S. (1976, December). The ground signatures of the expansion phase during multiple onset substorms. *Planetary and Space Science*, *24*(12), 1115-IN4. doi: 10.1016/0032-0633(76)90149-5

Pytte, T., McPherron, R. L., Hones, E. W., & West, H. I. (1978). Multiple-satellite studies of magnetospheric substorms: Distinction between polar magnetic substorms and convection-driven negative bays. *Journal of Geophysical Research*, *83*(A2), 663. Retrieved from http://doi.wiley.com/10.1029/JA083iA02p00663 doi: 10.1029/JA083iA02p00663

Pytte, T., McPherron, R. L., Kivelson, M. G., West, H. I., & Hones, E. W. (1976, December). Multiple-satellite studies of magnetospheric substorms: Radial dynamics of the plasma sheet. *Journal of Geophysical Research*, *81*(34), 5921-5933. doi: 10.1029/JA081i034p05921

Rae, I. J., Mann, I. R., Murphy, K. R., Milling, D. K., Parent, A., Angelopoulos, V., ... Russell, C. T. (2009, jan). Timing and localization of ionospheric signatures associated with substorm expansion phase onset. *Journal of Geophysical Research: Space Physics*, *114*(A1), A00C09. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2008JA013559 doi: 10.1029/2008JA013559

Raeder, J., McPherron, R. L., Frank, L. A., Kokubun, S., Lu, G., Mukai, T., ... Slavin, J. A. (2001, jan). Global simulation of the Geospace Environment Modeling substorm challenge event. *Journal of Geophysical Research*, *106*(A1), 381–395. Retrieved from http://doi.wiley.com/10.1029/2000JA000605 doi: 10.1029/2000JA000605

Raeder, J., Zhu, P., Ge, Y., & Siscoe, G. (2010, dec). Open Geospace General Circulation Model simulation of a substorm: Axial tail instability and ballooning mode preceding substorm onset. *Journal of Geophysical Research*, *115*(A5), A00I16. Retrieved from http://doi.wiley.com/10.1029/2010JA015876 doi: 10.1029/2010JA015876

Ridley, A., Gombosi, T., & Dezeeuw, D. (2004, February). Ionospheric control of the magnetosphere: conductance. *Annales Geophysicae*, *22*, 567-584. doi: 10.5194/angeo-22-567-2004

Ridley, A. J., Gombosi, T. I., De~Zeeuw, D. L., Clauer, C. R., & Richmond, A. D. (2003). Ionospheric control of the magnetosphere: Thermospheric neutral winds. *J. Geophys. Res.*, *108*(A8), 1328. doi: 10.1029/2002JA009464

Rostoker, G. (2002, jul). Identification of substorm expansive phase onsets. *Journal of Geophysical Research*, *107*(A7), 1137. Retrieved from http://doi.wiley.com/10.1029/2001JA003504 doi: 10.1029/2001JA003504

Rostoker, G., Akasofu, S.-I., Foster, J., Greenwald, R., Kamide, Y., Kawasaki, K., ... Russell, C. (1980, apr). Magnetospheric substormsdefinition and signatures. *Journal of Geophysical Research*, *85*(A4), 1663. Retrieved from http://doi.wiley.com/10.1029/JA085iA04p01663 doi: 10.1029/JA085iA04p01663

Runov, A., Angelopoulos, V., Sitnov, M. I., Sergeev, V. A., Bonnell, J., McFadden, J. P., ... Auster, U. (2009, jul). THEMIS observations of an earthward-propagating dipolarization front. *Geophysical Research Letters*, *36*(14), L14106. Retrieved from http://doi.wiley.com/10.1029/2009GL038980 doi: 10.1029/2009GL038980

Runov, A., Angelopoulos, V., & Zhou, X.-Z. (2012, may). Multipoint observations of dipolarization front formation by magnetotail reconnection. *Journal of Geophysical Research: Space Physics*, *117*(A5), A05230. Retrieved from http://doi.wiley.com/10.1029/2011JA017361 doi: 10.1029/2011JA017361

Russell, C. T. (2000, oct). How northward turnings of the IMF can lead to substorm expansion onsets. *Geophysical Research Letters*, *27*(20), 3257–3259. Retrieved from http://doi.wiley.com/10.1029/2000GL011910 doi: 10.1029/2000GL011910

Samson, J., & Yeung, K. (1986, nov). Some generalizations on the method of superposed epoch analysis. *Planetary and Space Science*, *34*(11), 1133–1142. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0032063386900255 doi: 10.1016/0032-0633(86)90025-5

Sauvaud, J.-A., & Winckler, J. (1980). Dynamics of plasma, energetic particles, and fields near synchronous orbit in the nighttime sector during magnetospheric substorms. *Journal of Geophysical Research*, *85*(A5), 2043. Retrieved from http://doi.wiley.com/10.1029/JA085iA05p02043 doi: 10.1029/JA085iA05p02043

Sazykin, S. Y. (2000). *Theoretical Studies of Penetration of Magnetospheric Electric Fields to the Ionosphere* (Unpublished doctoral dissertation). Utah State University, Logan, Utah.

Schmid, D., Volwerk, M., Nakamura, R., Baumjohann, W., & Heyn, M. (2011, sep). A statistical and event study of magnetotail dipolarization fronts. *Annales Geophysicae*, *29*(9), 1537–1547. Retrieved from http://www.ann-geophys.net/29/1537/2011/ doi: 10.5194/angeo-29-1537-2011

Sergeev, V. A., Angelopoulos, V., & Nakamura, R. (2012, mar). Recent advances in understanding substorm dynamics. *Geophysical Research Letters*, *39*(5), L05101. Retrieved from http://doi.wiley.com/10.1029/2012GL050859 doi: 10.1029/2012GL050859

Sergeev, V. A., Tanskanen, P., Mursula, K., Korth, A., & Elphic, R. C. (1990, apr). Current sheet thickness in the near-Earth plasma sheet during substorm growth phase. *Journal of Geophysical Research*, *95*(A4), 3819. Retrieved from http://doi.wiley.com/10.1029/JA095iA04p03819 doi: 10.1029/JA095iA04p03819

Singer, H., Matheson, L., Grubb, R., Newman, A., & Bouwer, D. (1996, oct). Monitoring space weather with the GOES magnetometers. In E. R. Washwell (Ed.), *Proceedings of spie* (Vol. 2812, pp. 299–308). International Society for Optics and Photonics. Retrieved from http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1021197 doi: 10.1117/12.254077

Slavin, J. A., Baker, D. N., Craven, J. D., Elphic, R. C., Fairfield, D. H., Frank, L. A., … Zwickl, R. D. (1989, nov). CDAW 8 observations of plasmoid signatures in the geomagnetic tail: An assessment. *Journal of Geophysical Research*, *94*(A11), 15153. Retrieved from http://doi.wiley.com/10.1029/JA094iA11p15153 doi: 10.1029/JA094iA11p15153

Slavin, J. A., Smith, M. F., Mazur, E. L., Baker, D. N., Iyemori, T., Singer, H. J., & Greenstadt, E. W. (1992, apr). ISEE 3 plasmoid and TCR observations during an extended interval of substorm activity. *Geophysical Research Letters*, *19*(8), 825–828. Retrieved from http://doi.wiley.com/10.1029/92GL00394 doi: 10.1029/92GL00394

Slinker, S. P., Fedder, J. A., & Lyon, J. G. (1995, apr). Plasmoid formation and evolution in a numerical simulation of a substorm. *Geophysical Research Letters*, *22*(7), 859–862. Retrieved from http://doi.wiley.com/10.1029/95GL00300 doi: 10.1029/95GL00300

Stephenson, D. B. (2000, apr). Use of the Odds Ratio for Diagnosing Forecast Skill. *Weather and Forecasting*, *15*(2), 221–232. Retrieved from http://

journals.ametsoc.org/doi/abs/10.1175/1520-0434%282000%29015%3C0221%3AUOTORF%3E2.0.CO%3B2 doi: 10.1175/1520-0434(2000)015⟨0221:UOTORF⟩2.0.CO;2

Sugiura, M., Skillman, T. L., Ledley, B. G., & Heppner, J. P. (1968, nov). Propagation of the sudden commencement of July 8, 1966, to the magnetotail. *Journal of Geophysical Research*, *73*(21), 6699–6709. Retrieved from http://doi.wiley.com/10.1029/JA073i021p06699 doi: 10.1029/JA073i021p06699

Toffoletto, F., Sazykin, S., Spiro, R., & Wolf, R. (2003). Inner magnetospheric modeling with the Rice Convection Model. *Space Science Reviews*, *107*(1-2), 175–196. Retrieved from http://link.springer.com/10.1023/A:1025532008047 doi: 10.1023/A:1025532008047

Tóth, G., Jia, X., Markidis, S., Peng, I. B., Chen, Y., Daldorff, L. K. S., … Dorelli, J. C. (2016, feb). Extended magnetohydrodynamics with embedded particle-in-cell simulation of Ganymede's magnetosphere. *Journal of Geophysical Research: Space Physics*, *121*(2), 1273–1293. Retrieved from http://doi.wiley.com/10.1002/2015JA021997 doi: 10.1002/2015JA021997

Tóth, G., Ma, Y., & Gombosi, T. I. (2008, jul). Hall magnetohydrodynamics on block-adaptive grids. *Journal of Computational Physics*, *227*(14), 6967–6984. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0021999108002076 doi: 10.1016/j.jcp.2008.04.010

Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., … Kóta, J. (2005). Space Weather Modeling Framework: A new tool for the space science community. *Journal of Geophysical Research: Space Physics*, *110*(A12), A12226. Retrieved from http://dx.doi.org/10.1029/2005JA011126 doi: 10.1029/2005JA011126

Tóth, G., van der Holst, B., Sokolov, I. V., Zeeuw, D. L. D., Gombosi, T. I., Fang, F., … Opher, M. (2012). Adaptive Numerical Algorithms in Space Weather Modeling. *J. Comput. Phys.*, *231*(3), 870–903. doi: 10.1016/j.jcp.2011.02.006

Walach, M.-T., & Milan, S. E. (2015, mar). Are steady magnetospheric convection events prolonged substorms? *Journal of Geophysical Research: Space Physics*, *120*(3), 1751–1758. Retrieved from http://doi.wiley.com/10.1002/2014JA020631 doi: 10.1002/2014JA020631

Wang, H., Ma, S., & Ridley, A. J. (2010, mar). Comparative study of a substorm event by satellite observation and model simulation. *Chinese Science Bulletin*, *55*(9), 857–864. Retrieved from http://link.springer.com/10.1007/s11434-009-0282-4 doi: 10.1007/s11434-009-0282-4

Welling, D. T., Anderson, B. J., Crowley, G., Pulkkinen, A. A., & Rastätter, L. (2017, January). Exploring predictive performance: A reanalysis of the geospace model transition challenge. *Space Weather*, *15*(1), 192-203. doi: 10.1002/2016SW001505

Welling, D. T., Barakat, A. R., Eccles, J. V., Schunk, R. W., & Chappell, C. R. (2016). Coupling the Generalized Polar Wind Model to Global Magnetohydrodynamics: Initial Results. In *Magnetosphereionosphere coupling in the solar system* (Vol. 222, pp. 179–194). Washington, DC: American Geophysical Union (AGU). doi: 10.15142/T3C88J

Welling, D. T., & Ridley, A. J. (2010, mar). Validation of SWMF magnetic field and plasma. *Space Weather*, *8*(3), S03002. Retrieved from http://doi.wiley.com/10.1029/2009SW000494 doi: 10.1029/2009SW000494

Weygand, J. M., McPherron, R., Kauristie, K., Frey, H., & Hsu, T.-S. (2008, dec). Relation of auroral substorm onset to local AL index and dispersionless particle injections. *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*(18), 2336–2345. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364682608002563 doi: 10.1016/J.JASTP.2008.09.030

Wild, J. A., Woodfield, E. E., & Morley, S. K. (2009, sep). On the triggering of auroral substorms by northward turnings of the interplanetary magnetic field. *Annales Geophysicae*, *27*(9), 3559–3570. Retrieved from http://www.ann-geophys.net/27/3559/2009/ doi: 10.5194/angeo-27-3559-2009

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences.* Cambridge, MA: Elsevier/Academic Press.

Wilson, G. R., Ober, D. M., Germany, G. A., & Lund, E. J. (2004, feb). Nightside auroral zone and polar cap ion outflow as a function of substorm size and phase. *Journal of Geophysical Research: Space Physics*, *109*(A2), A02206. Retrieved from http://doi.wiley.com/10.1029/2003JA009835 doi: 10.1029/2003JA009835

Wiltberger, M., Lotko, W., Lyon, J. G., Damiano, P., & Merkin, V. (2010, oct). Influence of cusp O+ outflow on magnetotail dynamics in a multifluid MHD model of the magnetosphere. *Journal of Geophysical Research: Space Physics*, *115*(A10), A00J05. Retrieved from http://doi.wiley.com/10.1029/2010JA015579 doi: 10.1029/2010JA015579

Winglee, R. M., Chua, D., Brittnacher, M., Parks, G. K., & Lu, G. (2002, sep). Global impact of ionospheric outflows on the dynamics of the magnetosphere and cross-polar cap potential. *Journal of Geophysical Research*, *107*(A9), 1237. Retrieved from http://doi.wiley.com/10.1029/2001JA000214 doi: 10.1029/2001JA000214

Wolf, R. A., Harel, M., Spiro, R. W., Voigt, G.-H., Reiff, P. H., & Chen, C.-K. K. (1982). Computer simulation of inner magnetospheric dynamics for the magnetic storm of July 29, 1977. *Journal of Geophysical Research*, *87*(A8), 5949–5962. Retrieved from http://doi.wiley.com/10.1029/JA087iA08p05949 doi: 10.1029/JA087iA08p05949

Yu, Y., & Ridley, A. J. (2008, may). Validation of the space weather modeling framework using ground-based magnetometers. *Space Weather*, *6*(5), S05002. Retrieved from http://www.agu.org/pubs/crossref/2008/2007SW000345.shtmlhttp://doi.wiley.com/10.1029/2007SW000345 doi: 10.1029/2007SW000345

Yu, Y., Ridley, A. J., Welling, D. T., & Tóth, G. (2010, August). Including gap region field-aligned currents and magnetospheric currents in the MHD calculation of ground-based magnetic field perturbations. *Journal of Geophysical Research*, *115*(A8), A08207. doi: 10.1029/2009JA014869

Zhu, P., Bhattacharjee, A., & Ma, Z. W. (2004, nov). Finite k y ballooning instability in the near-Earth magnetotail. *Journal of Geophysical Research*, *109*(A11), A11211. Retrieved from http://doi.wiley.com/10.1029/2004JA010505 doi: 10.1029/2004JA010505