Automated High-frequency Geomagnetic Disturbance Classifier: A Machine Learning Approach to Identifying Noise while Retaining High-Frequency Components of the Geomagnetic Field

Brett A. McCuen[1], Mark B. Moldwin[1], Erik S. Steinmetz[2], Mark J. Engebretson[2]

[1]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA
[2]Department of Physics, Augsburg University, Minneapolis, MN, USA

Key Points:

- A new technique is presented to identify high-frequency perturbations of the surface geomagnetic field
- A support vector machine predicts with high accuracy the source of high-frequency events as geophysical or caused by noise interference
- High-frequency magnetic signals with geomagnetic sources provide detail on small-scale features of magnetosphere-ionosphere current systems

Corresponding author: Brett A. McCuen, bmccuen@umich.edu

Abstract

We present an automated method to identify high-frequency geomagnetic disturbances in ground magnetometer data and classify the events by the source of the perturbations. We developed an algorithm to search for and identify changes in the surface magnetic field, dB/dt, with user-specified amplitude and timescale. We used this algorithm to identify transient-large-amplitude dB/dt events that have timescale less than 60 seconds and amplitude > 6 nT/s. Because these magnetic variations have similar amplitude and time characteristics to instrumental or man-made noise, the algorithm identified a large number of noise-type signatures as well as geophysical signatures. We manually classified these events by their sources (noise-type or geophysical) and statistically characterized each type of event; the insights gained were used to more specifically define a transient-large-amplitude geophysical event and greatly reduce the number of noise-type dB/dt identified. Next, we implemented a support vector machine classification algorithm to classify the remaining events in order to further reduce the number of noise-type dB/dt in the final data set. We examine the performance of our complete dB/dt search algorithm in widely-used magnetometer databases and the effect of a common data processing technique on the results. The automated algorithm is a new technique to identify geomagnetic disturbances and instrumental or man-made noise, enabling systematic identification and analysis of space weather related dB/dt events and automated detection of magnetometer noise intervals in magnetic field databases.

1 Plain Language Summary

High-frequency (second-timescale) components of the surface geomagnetic field are not often included in studies on geomagnetically induced currents (GIC) because they do not pose a direct threat to technological infrastructure. However, high-frequency intervals occur prior to and within some larger space weather events that can lead to GICs. Because these perturbations are very similar to signals that arise due to noise-interference, we have developed an automated procedure to identify such high-frequency intervals and predict the source of the signal as geophysical or noise-type. It was found that common data processing techniques can reduce or remove high-frequency geophysical disturbances, but do not remove all noise-type intervals. Thus, the automated process provides an event list of one-hour event windows that contain high-frequency disturbances and the classification of the signals within. This list can be used to identify hour windows of data that are undesirable for space weather research as well as events that contain high-frequency geophysical disturbances that may provide insight to the small-scale features of space weather events.

2 Introduction

Space weather occurs due to solar disturbances such as solar flares and coronal mass ejections that activates magnetohydrodynamic and electromagnetic disturbances that propagate throughout the magnetosphere-ionosphere (M-I) system down to the surface of Earth. One ground manifestation of severe space weather events is geomagnetically induced currents (GIC), perhaps the most critical space weather concern. Flowing through man-made conductors on Earth like railways, pipelines and power grids, GICs can be large enough to cause damage to transformers resulting in major power outages and costly equipment damage (Pulkkinen et al., 2017). GICs are the result of a horizontal surface electric field **E** induced in Earth's surface that is driven by large changes of the surface magnetic field, dB/dt, via Faraday's law of induction. Thus, the dB/dt is often used as a proxy to study GIC.

While large GICs often occur during global space weather events like sudden commencements (SC) and geomagnetic storms that cause major changes in the global large-scale M-I currents, it has been known for some time that smaller-scale phenomena are

capable of causing GICs as well. An example of such phenomena are nighttime magnetic perturbation events (MPE), also known as nighttime geomagnetic disturbances (GMD), that are often associated with substorm activity and may be a result of other magnetotail phenomena that commonly occurs during substorms (Engebretson et al., 2019a). MPEs have 5-10 minute timescales, relatively small spatial scales (∼275 km) compared to global events and are related to localized ionospheric instabilities.

It was shown by Viljanen (1997) that smaller-scale ionospheric currents play a key role in producing very large dB/dt at the surface. Several recent studies also suggest that beyond the largest space weather disturbances, there are more rapid, localized and small-scale processes involved in generating some extreme GICs (Engebretson et al., 2021; Ngwira et al., 2015; Pulkkinen et al., 2015). Dimmock et al. (2020) found that the localized horizontal magnetic field derivative can vary by a factor of three times the spatial average and thus these regional extremes are not accurately represented in global geomagnetic activity indices. Further, Dimmock et al. (2020) found that enhancements in regional dB/dt are linked to increased energy deposition in the magnetosphere mapping to local ionospheric structures and thus play a key role in modeling GIC during strong storms.

Less is understood about rapid and regional dB/dt enhancements because magnetic field data with 1-minute temporal resolution has long been the accepted standard in space weather research. This is because higher-frequency, second-timescale variations are effectively low-pass filtered when computing the geoelectric field (Pulkkinen et al., 2006, 2013). However, these second-scale magnetic field changes may be especially important in understanding small-scale dynamics of space weather events. While magnetic disturbances in this Pi 1-2 frequency range do not cause GICs directly, they have been found to occur prior to and/or during some GIC-capable space weather events, nighttime MPEs in particular (McCuen et al., 2021).

We refer to rapid dB/dt enhancements as transient-large-amplitude (TLA) events: instances of high-frequency, short-timescale magnetic field variations (< 60 s) that have large dB/dt values over 6 nT/s and occur within a 1-hour window. McCuen et al. (2021) found that TLA dB/dt intervals identified in 2015 often occurred in the pre-midnight sector (magnetic local time, MLT), 30 minutes after a substorm onset and in association to many of the most extreme nighttime MPEs. Of 175 MPEs at four MACCS stations in 2015 (IGL, RBY, PGG, CDR), nearly half of the 52 largest events (maximum dB/dt values greater than 10 nT/s) had associated TLA dB/dt intervals. Figure 1 shows an example of TLA events that was published as Figure 1b of McCuen et al. (2021). The TLA events in Figure 1 occurred at five stations of the Magnetometer Array for Cusp and Cleft Studies (MACCS) that occurred on 11 November 2015. These TLA signatures occurred within nighttime MPEs that were included in the study of Engebretson et al. (2019a) (all but the GJO stations were included in the study).

Engebretson et al.(2019a) used a superposed epoch spherical elementary current systems (SECS) analysis on 21 strong events at the CDR station to conclude that they were associated with westward overhead currents that coincided with a region of shear between upward and downward field-aligned currents (FAC). The TLA event in Figure 1 is one of these strongest MPEs identified at CDR in the study of Engebretson et al. (2019a); the example shows many TLA intervals within the MPEs and appears to exhibit a westward moving disturbance as the minimum of the negative bays in the Bx component appear successively in each station from east to west (see map of MACCS stations in Section 3). A westward current in the ionosphere can generate a magnetic field with field lines that point northward above the current region and southward below, resulting in large negative depressions in the Bx component of the ground magnetometers. While there is some evidence for the processes responsible for generating MPEs, their exact physical mechanisms and the current systems involved are still under investigation. Analyzing these higher-frequency perturbations within the MPEs and observing

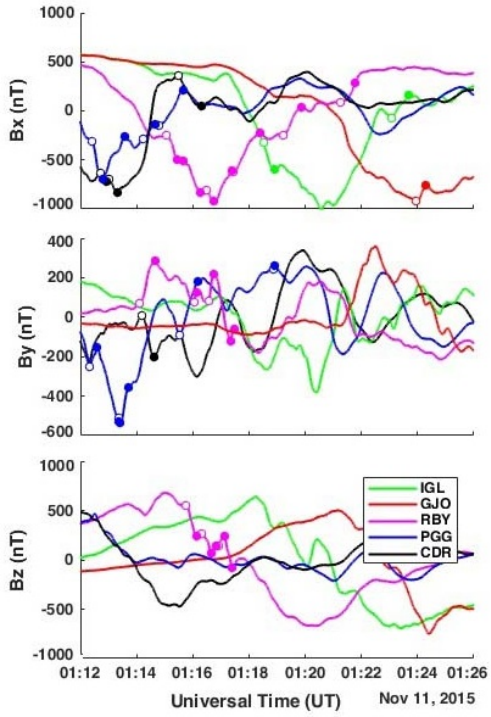

Figure 1. Station locations shown on a map of Nunavut, North-East Canada. Circles represent locations of MACCS stations, the square is the location of the CANMOS IQA station and the triangle signifies the AUTUMNX INUK station. Lines of latitude and longitude are in corrected geomagnetic coordinates for 2015.

the ionospheric behavior during such events allows for more detailed understanding of regional dB/dt enhancements, small-scale ionospheric currents, the dynamics of shear regions between upward and downward FAC and the potential connection to other magnetotail phenomena.

While substorms and MPEs have minutes to tens of minute timescales, there is clear evidence of higher-frequency (<60 s) behavior within many of these events. Because many of the MPEs that exhibited TLA signatures were amongst the most intense events but were not related to the most extreme space weather events (i.e. SCs and/or global geomagnetic storms), this suggests that more localized, small-scale ionospheric currents are involved in generating these large disturbances. Further, because the MPEs that exhibited TLA intervals prior to or within the overall disturbance were some of the largest MPEs of the dataset suggests that TLA signatures may be good indicators of the strongest small-scale events that have the capability to cause GICs.

Analysis of the second-timescale behavior of the surface magnetic field is a pathway to understanding the small-scale dynamics of M-I current systems that can give rise to GIC. Studying these high-frequency signatures will improve the understanding of rapid and localized magnetic field behavior and associated ionospheric currents. This more detailed knowledge of the fine-scale nature of the geomagnetic field can aid in improving modeling and forecasting of space weather events.

While it is necessary to analyze high-frequency TLA variations in ground magnetic field data in order to advance our understanding of small-scale M-I dynamics, the challenge in this task is retaining these high-frequency signatures in global magnetic field databases.

Advancements in technological capabilities (Love & Finn, 2017) and the need for improved accuracy in measuring dB/dt (Tõth et al., 2014) have motivated the shift toward using higher temporal resolution magnetic field measurements for space weather applications. However, common data processing methods often reduce or remove transient-large-amplitude signatures via their data cleaning or noise removal procedures because the signatures are similar in amplitude and timescale to that of magnetometer noise. The term magnetometer noise refers to two main sources of error in magnetometer readings: instrumental defect and/or magnetic deviation caused by interference of ferromagnetic materials in the vicinity of the magnetometer (Nguyen et al., 2020). Either of these sources can cause rapidly varying and irregular data measurements that have similar amplitude and timescale characteristics to TLA signatures.

Because of the similarity of noise-type data to TLA signatures, the geophysical TLA dB/dt are often reduced or removed with the noise signatures in common data processing procedures. Intermagnet, a worldwide magnetometer database commonly used for M-I and GIC research, uses a frequency band pass filter of 0.008-0.2 Hz (5-125 seconds) on 1-second data to remove error artifacts (St-louis et al., 2014). SuperMAG is a widely-used, global magnetic field data collaboration that provides uniformly processed data from over 300 ground based magnetometers (Gjerloev, 2012). SuperMAG offers 1-second (averaged if raw data has higher resolution) resolution magnetic field data that has undergone an automated data cleaning procedure. Both of these procedures can alter or remove higher-frequency variations of the field. Beyond data processing procedures by commonly used databases, many magnetic field data are averaged over 1-minute or more in practice for GIC and space weather studies. Even though many magnetic field arrays offer 1-second magnetic field data, the data averaging and processing techniques used often remove or modify TLA variations.

The problem remains, TLA variations that are important to retain for space weather studies can be removed or reduced in common data cleaning and processing, but are difficult to distinguish from noise in raw data. Numerous methods have been used to characterize and statistically analyze noise in magnetometer data (Khomutov et al., 2017; Nguyen et al., 2020) but challenges in anomaly detection have motivated the use of more modern machine learning techniques to identify and remove outliers from magnetometer data (Mitra et al., 2020; Xu et al., 2020). The data cleaning process for large magnetic field databases usually requires an experienced magnetologist to determine whether some signals are natural or noisy. In the case of TLA signatures that are similar in frequency and amplitude to error artifacts, machine learning algorithms can be especially useful for making these types of determinations without the need for human supervision.

In this paper, we present the full methodology for a geomagnetic disturbance classifier that identifies occurrences of high-frequency (0.017-1 Hz) signals in magnetic field data and classifies whether they are a result of noise interference or geophysical sources. This process utilizes statistical characteristics of both noise-type and geophysical dB/dt signatures to define a high frequency geomagnetic disturbance event and implements a machine learning classification algorithm to classify the dB/dt signatures by their sources.

This paper is organized as follows. Section 3 describes the magnetometer data used in this study and section 4 outlines the dB/dt search algorithm. Section 5 discusses and illustrates the noise-shapes identified in data from MACCS, and section 6 describes the statistical characteristics of the noise-type and TLA dB/dt intervals and events. The filters implemented into the search algorithm based on the analysis of sections 5 and 6 are explained in section 7. In section 8, the machine learning approach used to fully automate the search algorithm is described and the results discussed (the cross-validation process is detailed in Supporting Information). Section 9 examines the effect of a common data processing procedure on the high frequency signatures being studied and discusses the data products provided by the procedure developed. Finally, section 10 discusses our results and the implications for space weather studies followed by our conclusions.

3 Data Sets

This study uses magnetic field data from three geomagnetic and space physics magnetometer databases, as well as data processed through the SuperMAG data service that includes all three databases. The MACCS data are used for the initial identification of TLA dB/dt signatures and the noise classification for algorithm improvement. Then, we use data from a magnetometer site within the Athabasca University THEMIS UCLA Magnetometer Network eXtension (AUTUMNX) (Connors et al., 2016) as well as data from the CANadian Magnetic Observatory System (CANMOS) (Nikitina et al., 2016) to compare how well the dB/dt search process performs on magnetic field data from different systems.

The geographic and geomagnetic coordinates of the magnetic observatories used in this study are listed in Table 1 and shown on the map in Figure 2 with lines of corrected geomagnetic (CGM) latitude and longitude for 2015. The CGM coordinates were calculated using the AACGM-v2 Calculator (available at http://sdnet.thayer.dartmouth.edu/aacgm/aacgm_calc.php#AACGM) for epoch 2015.

Station	Geographic Latitude	Geographic Longitude	Corrected Geomagnetic Latitude	Corrected Geomagnetic Longitude
IGL	69.3	278.2	77.6	355
GJO	68.6	264.2	76.8	329.8
RBY	66.5	273.8	75.2	347.2
PGG	66.1	294.2	73.2	19.9
CDR	64.2	283.4	72.6	3.0
IQA	63.8	291.5	71.4	15.2
INUK	58.8	281.9	67.6	0.02
NAN	56.4	298.3	63.1	22.5

Table 1. Location coordinates of stations used in this study.

The ground-based stations used in this study are all in the near vicinity of Inuit communities in arctic Nunavut, Canada. Many of the MACCS stations are located at the local airport, configured such that the computing instrumentation is kept inside the airport or nearby facility and the sensor is located away from the building inside a small, enclosed box on the ground. The IGL magnetometer sensor is located right within the local town of Igloolik near the Igloolik Research Centre where the rest of the station equipment is held. The PGG magnetometer is located ∼1 km outside of town near the Pangnirtung water reservoir. In all cases of the MACCS magnetometer stations, their locations make them susceptible to man-made noise interference from multiple sources (cars, snowmobiles, nearby facilities, etc.). The CANMOS station and AUTUMNX station are also susceptible to local interference from human activity, however these observatories are dedicated solely to magnetic field data acquisition and do not rely on local facilities like an airport to house instrumentation. This allows the CANMOS and AUTUMNX observatories to be located further from town centers and aids in prevention of noise contamination.

The magnetometers used in this study at the MACCS and the IQA station of CANMOS are Narod ringcore fluxgate magnetometers designed and supplied by Dr. Barry Narod of Narod Geophysics, Ltd., Vancouver, B.C., Canada (Hughes & Engebretson, 1997). The AUTUMNX instruments are THEMIS-class fluxgate magnetometers provided by UCLA (Russell et al., 2008) and based on the design for the earlier Sino Magnetic Ar-

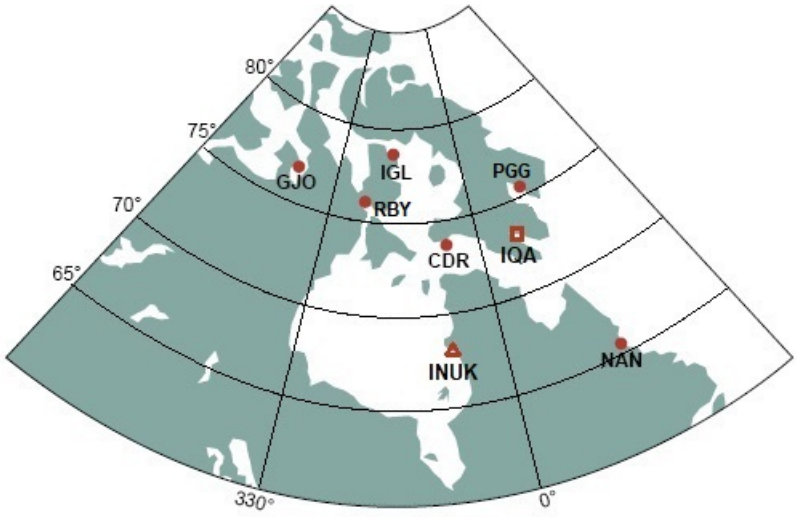

Figure 2. Station locations shown on a map of Nunavut, North-East Canada. Circles represent locations of MACCS stations, the square is the location of the CANMOS IQA station and the triangle signifies the AUTUMNX INUK station. Lines of latitude and longitude are in corrected geomagnetic coordinates.

ray at Low Latitudes (SMALL) terrestrial vector fluxgate magnetometers (Gao et al., 2000).

The Narod magnetometers collect 8 samples per second in three axes, then average and record the data at two samples per second for MACCS data and one sample per second for the CANMOS data. The AUTUMNX magnetometers record the magnetic field at 2 Hz. The data used from AUTUMNX and CANMOS observatories have resolution of 0.01 nT, the MACCS data have a 0.025 nT data resolution, and all three have timing accuracy of at least 1 ms. The high-resolution, sampling rate and timing accuracy are sufficient to detect short-timescale Pi 1-2 pulsations. The magnetometer data used from MACCS and AUTUMNX are in geomagnetic coordinates: H (geomagnetic north-south), D (geomagnetic east-west) and Z (vertical). The data from CANMOS is in geographic coordinates: X (geographic north-south), Y (geographic east-west) and Z (vertical).

4 dB/dt search algorithm

We developed an initial algorithm to identify changes of the magnetic field with user-specified magnitude and duration. The initial algorithm works in the following main steps: 1) calculate the change in magnetic field strength (ΔB) divided by the timestep (Δt): dB/dt (or slope) between each pair of successive data points and label the sign of the slope (labeled as a -1 for negative slope, +1 for positive slope and zero for zero slope), 2) mark the points when the sign of the slope changes for at least two measurement cycles (i.e. local minima and maxima) and 3) recalculate the new dB/dt between the local minima and maxima and return the information if the signature also meets the user-specified criteria for timescale, minimum and maximum ΔB and dB/dt.

Because the search criteria are such that the slope must have the same sign for two measurement cycles (step 2), the algorithm relies on the sampling frequency of the data and should be used for magnetic field data with 1-second or higher temporal resolution

for high-frequency studies. However, the same dB/dt search procedure can be performed on data averaged over a longer time period to identify dB/dt signatures with varying timescales (i.e. performing the dB/dt search algorithm on 1-minute averaged data will identify dB/dt signatures that last at least 2 minutes).

There is also an intermediate step after Step 1 that deals with the instances of zero slope that last only one measurement cycle: if a zero slope occurs only once in between two like-sign slope values, the sign of the slope is changed to match those slope values. This measure is taken so that a change in slope will only be marked in cases of zero slope if it persists for at least two measurement cycles and is consistent with the minimum dB/dt search to be intervals that last twice the sampling frequency. The final product returned from the algorithm is a nine column matrix; each row represents an individual dB/dt interval and provides the start and end time of the interval, start and end B value, the time elapsed: dt, the change in magnetic field amplitude: ΔB, and the total perturbation: dB/dt. The final two columns indicate the component that the interval was identified in and the station at which the interval took place.

The algorithm was developed to identify high-frequency (0.017-1 Hz) transient-large-amplitude (TLA) events in the magnetometer data. We define an event as any number of geomagnetic signatures with < 60 seconds timescale and dB/dt > 6 nT/s within a 1-hour event window. The hour windows are defined by the UT clock and determined by measurement frequency (i.e. number of data points in one hour) and are divided consecutively. For example, for a measurement frequency of 2 Hz, the first 7200 data points define the first hour window, and the next hour window is the following 7200 data points. The minimum dB/dt threshold was chosen as it is comparable to magnetic field measurements during the March 1989 geomagnetic storm that caused the HydroQuebec power grid failure (Kappenman, 2006). This was the most severe geomagnetic storm of the twentieth century and maximum magnetic field changes during that storm were on the order of 8 nT/s (but lasting much longer than just seconds), so dB/dt of this magnitude are considered large-amplitude.

Unfortunately, these specifications also describe the signals that can occur as a result of instrumentation error or interference by ferromagnetic material, i.e. "noise". The term noise is relative to the specific goal of the measurement or problem to be solved. For this study, we use the term noise to refer to two main sources of error in magnetometer readings: instrumental defect and/or magnetic deviation caused by interference of ferromagnetic materials in the vicinity of the magnetometer (Nguyen et al., 2020).

5 Noise Shapes Identified in MACCS Data

In order to capture all such magnetic signals of interest with this timescale, we set the initial criteria for the dB/dt search to signatures with 1-60 second timescale, ΔB 6-10,000 nT and dB/dt from 6-1000 nT/s. The similarity of TLA events to noisy signals resulted in the algorithm identifying a majority of signals that were due to noise rather than natural geophysical processes. Because the size and timescale of these signatures fall into the same ranges, and common data cleaning techniques can alter or remove TLA signatures, determining whether a given signal is of geophysical nature or a result of noise was done by examining the shape of the signal, the behavior of the magnetic field prior to and after the signal, and the amplitude characteristics of the interval. Thus, we manually separated the noise signals from the natural geophysical perturbations by comparing with the magnetic noise characterization of Khomutov et al. (2017).

The noise shapes described in Khomutov et al. (2017) are compiled from Intermagnet data from observatories located in the mid-latitude, eastern hemisphere. The observatories have various types of fluxgate magnetometers with measurement frequency from 0.2-2 Hz (5-0.5 seconds). The main sources of noisy signals in magnetometer data are

both external and internal. Externally, there are large-scale noise sources like DC railways that can impact magnetic field data at large distances, and there are more local sources of ferromagnetic and/or conductive material within the nearby vicinity of the magnetometer sensor (e.g. cars, technological devices, other instrumentation). Internally, noisy signals can arise from instrumentation error. While the exact source of a specific noisy signal can vary, the main sources are consistent across observatories and databases. For these main sources of interference with fluxgate magnetometer systems, the characteristic shapes and sizes of the resulting noisy signals in the data are common (Neska et al., 2013; Santarelli et al., 2014; Khomutov et al., 2017). The four most common shapes of noise and their characteristics reported by Khomutov et al. (2017) are defined and illustrated as follows.

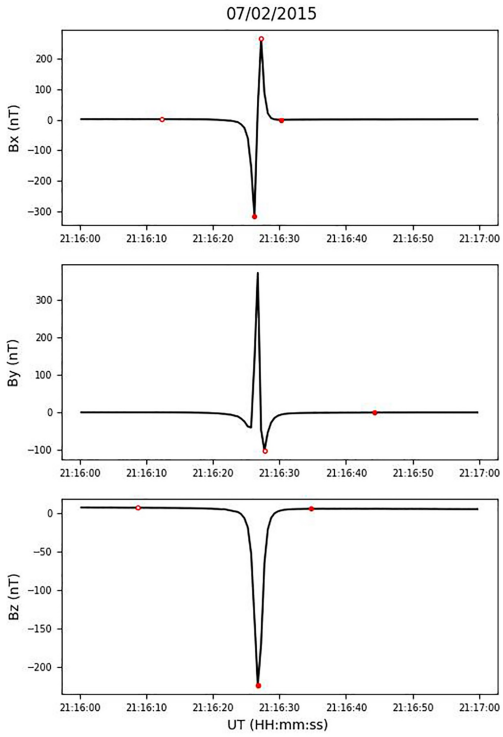

Figure 3. A spike in the magnetometer data that occurred on 2 July 2015 at the PGG station. The hollow circles mark the start of each dB/dt signature and the solid dots mark the end. Note that the consecutive solid red dots in the Bz plot (bottom) signify that the negative peak of this spike is both the end of the interval prior and the start of the interval following.

Spikes in magnetometer data are large-amplitude ($\sim$ tens of nT), relatively short signals (generally lasting less than a few seconds) with well-defined leading and back edges that have similar amplitudes. Isolated spikes, spikes with large amplitude (many tens of nT), and spikes that last only one measurement cycle have a low probability of being caused by geophysical sources. An example of a spike is shown in Figure 3: 1-minute of MACCS magnetometer data taken at the PGG station on 2 July, 2015. The hollow red circles represent the start of a dB/dt interval that is > 6 nT/s and the solid red dots represent the end of the signature. The mean B value of each component in the interval shown is subtracted from the data, but this does not change the ΔB and dB/dt amplitudes or the timescales of the intervals from the original data. The entire spike signature lasts about 20 seconds with each interval of large dB/dt lasting 3.5-10.5 seconds. The maximum amplitude of the spike is about 318 nT (dBx/dt beginning at 21:16:17.75

and decreasing for 8.5 seconds). We further define spikes in this noise classification to be instances of three or less large dB/dt signatures (with < 60 s timescale and magnitude > 6 nT/s) occurring within a 1-minute interval.

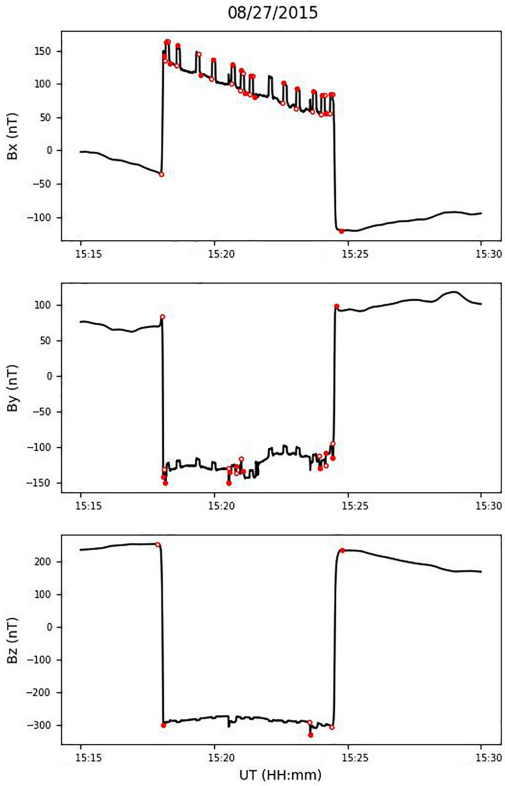

Figure 4. A noise jump that occurred at the CDR station on 27 August 2015. Hollow circles mark the start of a dB/dt signature and solid dots mark the end. The mean B value of each component in the interval shown is subtracted from the data.

Figure 4 shows an example of a jump shape in the MACCS data. Jumps are much like spikes but with a continuous interval between the leading and back edges. The timescales of these jumps vary; in this study, we specify jumps to have a minimum 1-minute interval sustained between the leading and trailing edges in which the magnetic field does not increase/decrease beyond the starting value of the leading edge (i.e. the very first hollow circle in all three panels of Figure 4 at approx. 15:18 UT). Jumps often occur due to changes of the magnetic field distribution via ferromagnetic material.

Random-like noise is usually caused by man-made disturbances which add randomized variations to the background magnetic field. These look like patches of highly frequent dB/dt intervals with randomized shape and amplitude. An example of random-like noise is shown in Figure 5. This patch of random noise lasted about 7-minutes; the algorithm identified 93 dB/dt signatures from the three components combined. Figure 4b is a zoomed view of a section of this event from 10:16:10-10:17:10 UT showing how some of these variations are presented on a 1-minute timescale. Figure 4b shows that, on a 1-minute timescale, these magnetic field variations have dissimilar shapes to classic spikes as defined above although they may appear to be a group of frequent spikes when observed on a slightly longer timescale. While the shape of these magnetic field changes cannot be defined as spikes or a jump, we determine that they are noise variations because of 1) the highly frequent nature and the randomized shapes of the dB/dt

intervals, 2) the jagged behavior of these variations on a second-scale (Figure 4b) and 3) the shape of the noise group on a minute-scale (Figure 5a) that appears to positively deviate from the background magnetic field in each axis.

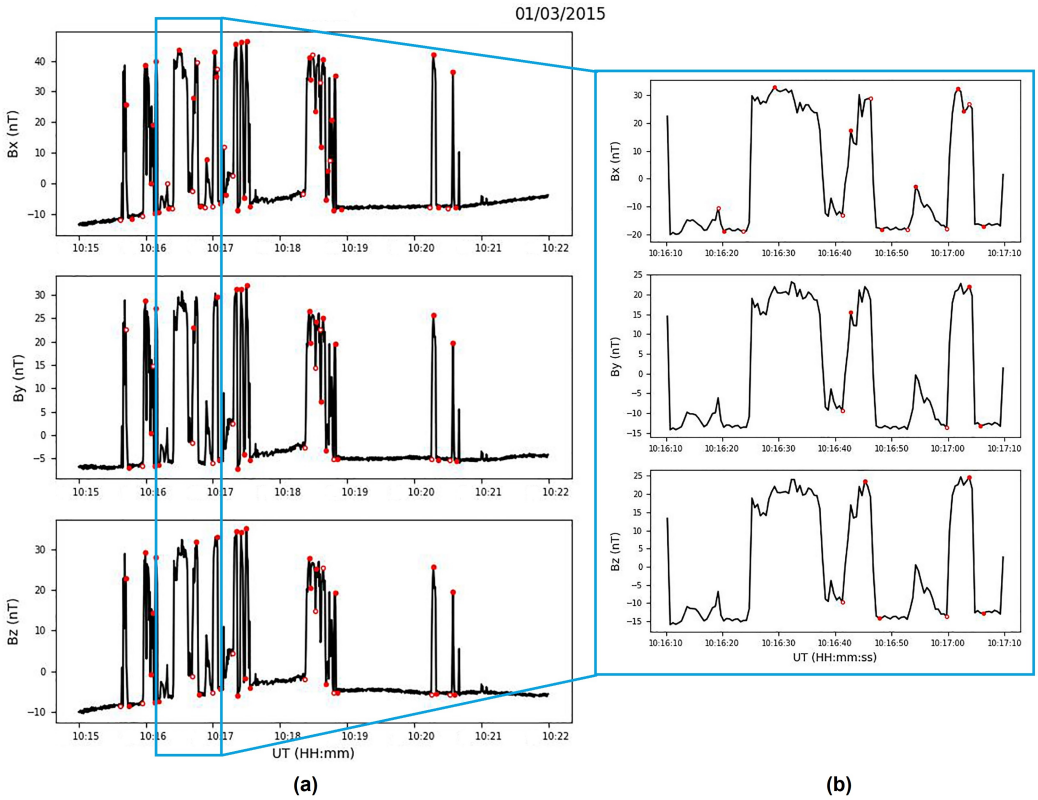

Figure 5. (a): Random-like noise that occurred at the IGL station on 3 January 2015. Hollow circles mark the start of a dB/dt signature and solid dots mark the end. (b): A zoomed-in view of 1-minute of the random-like noise-type event shown in (a). The mean B value of each component in the interval shown is subtracted from the data.

The last noise shape found in the MACCS magnetometers throughout 2015 is bay-like noise. An example of bay-like noise is shown in Figure 6b: a disturbance that occurred within 1-minute at the PGG station on 20 June 2015. The full high-frequency disturbance event (Figure 6a) consists of a bay-like disturbance as well as three separate spikes later in the hour window (note that just 25 minutes of this event are shown to emphasize the shape of the high-frequency intervals within this hour window). The bay-like disturbance is shown with a zoomed view in Figure 6b. The magnetic field changes in Figure 6b are near 50 nT in the x- and y- components but nearly 150 nT in the z-component. This is a common manifestation of noise in magnetometer data, usually caused by magnetic field changes near the instrument due to a moving ferromagnetic object (i.e. a vehicle or other instrumentation). It is shaped like a positive/negative magnetic bay that persists for the duration of the passing object (usually seconds). Bay-like noise often has sharp leading and trailing edges like spikes or jumps, but the behavior between these edges is more random and variable. These impulses can be difficult to distinguish from natural signals because negative and positive bays can also occur due to M-I sources. While bay-like noise events have similar shapes to TLA events, the distinction between them is that TLA events often occur within a bay that lasts 5-15 minutes (McCuen et al., 2021) while noise-type bays generally have a duration of just seconds. Further, this example

is decided to be noise because of the jagged magnetic field shapes on a second-timescale, as well as the very similar and smooth behavior of the magnetic field prior to and after the disturbance bay: a common characteristic of noise-type events in magnetic field data.

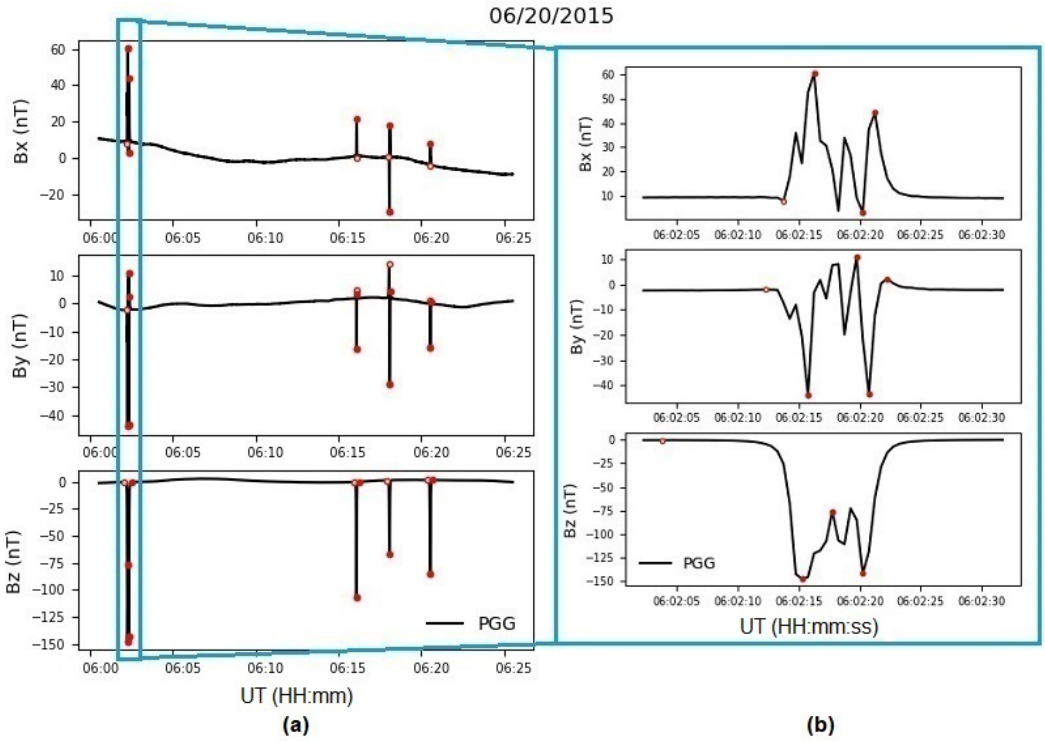

Figure 6. (a): A noise-type hour-event that occurred on 20, June 2015 at the PGG station consisting of a bay-like disturbance and three spikes. (b): Bay-like noise in MACCS magnetic field data. Hollow circles mark the start of a dB/dt signature and solid dots mark the end. The mean B value of each component in the interval shown is subtracted from the data.

Figure 7 shows an example of a TLA event that occurred on 10 November 2015. The figure shows dB/dt signatures that occurred at the PGG station but there were also TLA dB/dt observed at two other MACCS stations during this hour. These signatures at the PGG station occur prior to (in the Bx component) and within a large nighttime MPE that began at 00:36 UT. There are twelve total dB/dt signatures in the full event shown in Figure 7a with average ΔB of about 274 nT, mean Δt of 33.8 seconds and mean dB/dt of just under 8 nT/s. Figure 7b shows one-minute of zoomed-in data from this event from 00:41:30 to 00:42:30 with one TLA-type dB/dt signature in the x- and z- component each. The signature in the z-component of Figure 7b has the largest dB/dt amplitude of the event of 10.37 nT/s. Figure 7b shows that on a 1-minute timescale, these are smooth changes of the magnetic field rather than jagged edges of noisy data. This is a distinct characteristic of TLA events with geophysical sources: the magnetic field is smoothly varying on a second-timescale rather than rapidly changing with sharp edges as observed in noise-type events.

The common feature of these noise shapes in magnetometer data is that they are composed of some combination of second-timescale magnetic field changes with dB/dt $>$ 6 nT/s. These are characteristics equal to that of the geophysical TLA dB/dt that are meaningful in the context of small-scale M-I currents. However, the examples of TLA events in Figures 1 and 7 both show dB/dt intervals occurring prior to or within night-

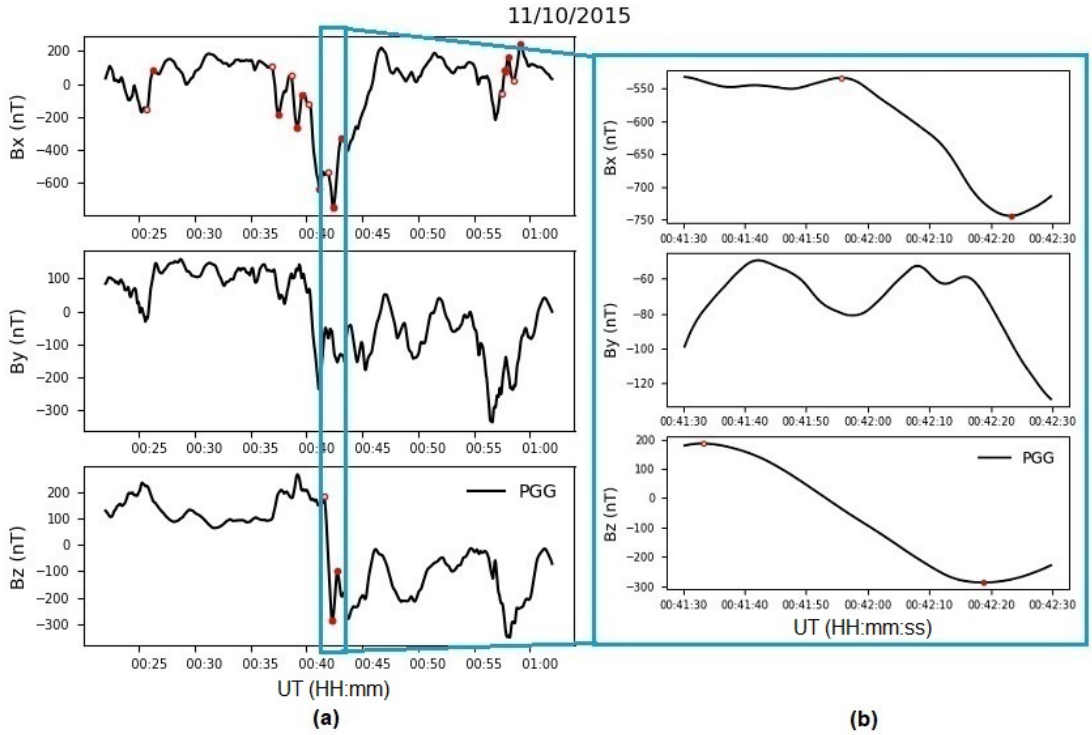

Figure 7. (a): A transient-large-amplitude (TLA) geomagnetic event that occurred on 10 November 2015 at the PGG station. Hollow circles mark the start of a dB/dt signature and solid dots mark the end. (b): A zoomed-in view of 1-minute of the TLA event shown in (a). The mean B value of each component in the interval shown is subtracted from the data.

time MPEs that are associated to small-scale ionospheric currents and these TLA dB/dt intervals show smooth variations on a 1-minute timescale.

There are some distinct differences between the dB/dt signatures that arise from noise sources and TLA dB/dt caused by M-I sources. Noise-type dB/dt events can often be identified by the shape of the event, the behavior of the magnetic field prior to and after the event (the highly similar and often steady nature of the magnetic field on either sides of the noise-type perturbation) and the smoothness of the magnetic field on a second-timescale (noise-type events often show sharp magnetic field changes on a second-timescale whereas TLA dB/dt events are always smoothly varying on such fine timescales). These criteria were used to manually separate noise-type and TLA events. This manual classification method based on the descriptions in the study by Khomutov et al. (2017) was expert-verified by one of the co-authors of this study. From the manual separation of events, the numerical characteristics of the dB/dt signatures of each event type were then used to create filters to automatically classify noise-type and TLA dB/dt signatures, discussed in greater detail in the following section.

6 Statistical characteristics of noise-type and TLA events

While some shapes of noise signals are more likely to result from either man-made sources or internal instrumentation issues, all four of the noise types described in section 5 can arise from both hardware and external sources. Determining the exact source of noise in magnetic field data can be a challenge, but separating geophysical magnetic

signatures from data contaminated with noise from outside interference is a more tangible task. After collecting all dB/dt signatures that satisfy the conditions for a high-frequency event (dB/dt > 6 nT/s, dt from 1-60 s), we manually classified the geophysical events, as well as each type of noise shape identified. Then we analyzed the statistical characteristics of these types of dB/dt events to improve the selection criteria for the search algorithm. The statistical characteristics that set geophysical TLA events apart from noise-type events are described and compared below.

Noise-type events, whether from instrumentation error or external interference, contribute significantly more dB/dt events than geophysical events. From the six MACCS stations throughout 2015, we identified 215 TLA dB/dt (making up 59 separate events) and 845,572 noise-type dB/dt signatures (making up nearly 5500 separate events). Figure 8 shows histograms of the number of noise-type dB/dt (orange) and the number of TLA dB/dt (blue) based on their timescale (dt), amplitude (ΔB) and magnitude (dB/dt). All three histograms show the number of events on a logarithmic scale. Figures 8b and 8c both include a zoomed-in view of the bottom left corner of the full distribution showing the portion containing the geophysical events. It can be seen from all three plots that the number of noise-type dB/dt identified is orders of magnitude larger than that of TLA dB/dt.

Figure 8a shows that noise-type dB/dt signatures were far more likely to last less than 10 seconds whereas TLA dB/dt had a relatively even spread of timescales from 3.5-60 seconds. Noise-type events at the six MACCS stations throughout 2015 had 99.8% of dB/dt intervals that lasted less than 10 seconds compared to just under 10% of the total TLA dB/dt. Further, all of the TLA dB/dt intervals that had dt < 10 seconds occurred within hour event windows that had longer dB/dt from 10-60 seconds, whereas most of the noise-type hour events consisted solely of dB/dt intervals lasting less than 10 seconds. The uniform distribution of Δt of the TLA intervals shows that there are a relatively consistent number of meaningful geophysical signatures over the second-timescale range.

Figure 8b shows that noise-type dB/dt signatures were far more likely to be less than 60 nT in amplitude (94.9% of noise-type dB/dt had $\Delta B < 60$ nT/s compared to just 5.5% of TLA dB/dt signatures), however the noise also contributed to outliers thousands of nT higher than any of the TLA dB/dt which had a maximum ΔB = 580.75 nT. A similar trend is seen in the histogram of dB/dt magnitudes (Figure 8c) where the TLA dB/dt occupy a small slice under the distribution of the noise-type dB/dt. The zoomed view of Figure 8c shows that the largest TLA dB/dt magnitude was ~33 nT/s compared to many noise-type dB/dt magnitudes exceeding 200 nT/s. Reasonable magnitudes for the most extreme second-timescale magnetic field changes are from 40-110 nT/s (Kataoka & Ngwira, 2016).

Noise-type dB/dt signatures occurred more often than TLA dB/dt overall and they also occurred in higher concentration per 1-hour event window. The random-type noise signature was the most frequently occurring. As is shown in Figure 5, random-noise events usually sustained longer intervals of highly variable magnetic field that contributed hundreds, sometimes thousands, of characteristic dB/dt signatures while geophysical TLA events often had just a few TLA dB/dt within a longer ~10-20 minute perturbation. We found that a noise event (within a 1-hour window) at an individual station had 154.6 dB/dt intervals on average while geophysical TLA events had an average of 3.2 dB/dt (maximums of 25 and 2370 dB/dt per 1-hour event window respectively). As previously mentioned, the hour windows are defined by the measurement frequency (i.e. number of data points in one hour) and are divided consecutively.

The number of 1-hour windows containing TLA and/or noise-type dB/dt per station is shown in Table 2, as well as the number of individual dB/dt signatures identified at each station. In order to numerically describe the distinction between the con-

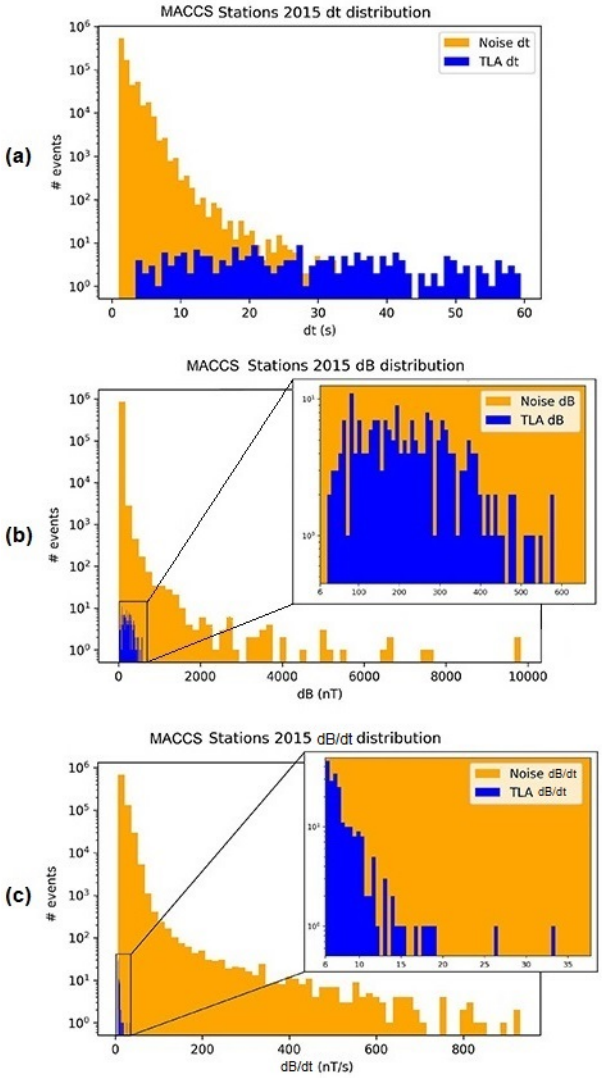

Figure 8. Histograms showing number of dB/dt signatures (separated by TLA and noise-type) from all six MACCS stations throughout 2015. (a): Distribution based on dt values, (b): distributions based on ΔB values and (c): distribution based on dB/dt values.

centration of dB/dt per hour window for noise-type and TLA events, we calculated the ratio of number of noise-type or TLA dB/dt per event to the total number of dB/dt (with any timescale and any amplitude) within the event hour. Table 2 contains the minimum and maximum of these ratios. While TLA and noise-type dB/dt events had similar minimum ratios (i.e. both event types exhibited events with very few or even singular high-frequency dB/dt intervals), the maximum ratios between TLA and noise-type dB/dt are very different from one another. TLA dB/dt never populated more than 4% of the total dB/dt within the respective hour event window, while noise-type events more often exhibited hour windows where the dB/dt composed of more than 10% and up to nearly 67% of all the magnetic field changes within the hour.

It is worth noting that the maximum ratio of noise-type or TLA type is not directly proportional to the total number of noise-type or TLA event windows. For instance, the GJO station had the least amount of noise-type event windows, but the highest max-

Station	Total dB/dt	# Noise windows	# TLA windows	Min ratio TLA:All	Max ratio TLA:All	Min ratio Noise:All	Max ratio Noise:All
IGL	33413	2159	4	0.0011	0.0069	0.0001	0.4377
GJO	1369	241	6	0.0019	0.0227	0.0004	0.6667
RBY	65800	991	7	0.0009	0.0085	0.0002	0.2117
PGG	1790	607	20	0.0006	0.0258	0.0002	0.0140
CDR	2353	695	15	0.0005	0.0355	0.0002	0.1998
NAN	741062	759	7	0.0008	0.0033	0.0002	0.5923

Table 2. Table showing number of 1-hour event windows that contain noise-type or TLA dB/dt, as well as the minimum and maximum ratios of TLA and noise-type to all dB/dt respectively.

imum ratio of noise-type to all dB/dt out of all the stations. This is to say that because the ratio is calculated based on the specific hour, it is dependent on the type of noise and how much there is and independent of the overall noise present in the station data.

To summarize the statistical characterization of geophysical TLA and noise-type events in this paper, there are three main distinctions between geophysical TLA dB/dt events and noise-type dB/dt events:

1. TLA events have at least one dB/dt signature > 6 nT/s that lasts 10 or more seconds within the 1-hour event window.
2. Large, second-timescale dB/dt are more likely to be of geophysical nature if they last from 10-60 seconds and have amplitude 60-1000 nT.
3. Large, second-timescale dB/dt are more likely to be noise if they occur in large concentration per 1-hour window (occupying more than 5% of the total magnetic field changes within the hour window). TLA-type events often have less than 20 dB/dt within an approximately 15-20 minute perturbation.

7 dB/dt Search Algorithm Filters

Following from the main characteristics described in the previous section, two main filters were applied to improve the dB/dt search algorithm and reduce the number of noise-type dB/dt identified by the routine. First: the dB/dt search is performed on consecutive 1-hour partitions of data and the requirements to determine a potential TLA event are specified as a 1-hour event window that contains at least one dB/dt that has magnitude 6-100 nT/s, ΔB from 60-1000 nT and timescale 10-60 seconds. Not only are the maximum values for ΔB and dB/dt decreased to the range observed for all TLA-type events, but the requirement that there be at least one signature lasting more than 10 seconds (and effectually having $\Delta B > 60$ nT) is implemented. If there are no signatures that meet this criteria in the hour window, the search procedure moves on to the next hour window. If there are any dB/dt that do fall within these values, the algorithm continues to the next stage.

In the second stage of the dB/dt search, dB/dt intervals with 6-100 nT/s, timescale 1-60 seconds and ΔB 6-1000 nT are identified (i.e. all of the high-frequency dB/dt signatures that could be TLA or noise-type), as well as the total number of dB/dt intervals with any amplitude and timescale within the hour. If the number of high-frequency signatures is more than 5% of the total number of the dB/dt within the hour, then the algorithm rejects all signatures identified. If this ratio is less than 5%, then the algorithm

removes any intervals that last less than 2-seconds (as the minimum dt for all TLA events identified from the MACCS stations in 2015 was 3.5 seconds) and returns the remaining dB/dt intervals as the final data product. In this case where all TLA criteria are satisfied, the dB/dt search is also performed for 1-minute prior to the start time of the hour and 1-minute after the start time of the hour (as well as for the two minutes framing the end time of the hour) so that no dB/dt intervals are lost by being split by the hour partition.

The ratio method allows for the 5% threshold to depend on the individual station data and 1-hour environment which can be highly variable across magnetometer arrays, dates and times. In other words, if a stations data are overall highly variable (higher number of total dB/dt on average per 1-hour) then the 5% threshold allows for a larger number of dB/dt comparative to the instrumentation and/or the surrounding magnetic environment to be identified before rejecting the hour-window as containing only noise-type dB/dt. This ratio method is a general metric to reduce noise in magnetometer data based on the concentration of short-lived (< 60 s) and large-amplitude (> 6 nT/s) dB/dt intervals per 1-hour event window at an individual station.

To summarize the algorithm filters, the filtered dB/dt search returns magnetic field intervals with dt from 2-60 seconds, ΔB from 6-1000 nT and dB/dt from 6-100 nT *only if:* at least one of these signatures within the 1-hour event window lasts 10 seconds or more, and if these high-frequency intervals (along with those that last less than 2 seconds) do not populate more than 5% of the total dB/dt within the hour window. Implementation of the above conditions into the dB/dt search process returned all of the same 215 TLA dB/dt and reduced the number of noise-type dB/dt returned by 99.6% (from 845680 to 2970 noise-type dB/dt). The numbers of both TLA and noise-type dB/dt prior to and after the filters are listed in Table 3. The filters removed all noise-type dB/dt from the RBY and NAN station, the latter of which had the most noise-type dB/dt in the unfiltered search. The IGL station had the most noise-type dB/dt remaining after the filtered search with 2,669 dB/dt.

Station	Total pre-filter	# TLA dB/dt	# Noise-type pre-filter	Total post-filter	# Noise-type post-filter
IGL	33413	20	33393	2689	2669
GJO	1369	14	1355	50	36
RBY	65800	32	65768	32	0
PGG	2353	61	1729	151	90
CDR	1790	69	2284	242	173
NAN	741062	19	741043	19	0
IQA	92	71	19	71	0
INUK	392	301	87	303	2

Table 3. Table with number of dB/dt intervals from 2015 of both TLA and noise-type, before and after the filters described in this section

In order to better evaluate the performance of the dB/dt search algorithm and the performance filters, the dB/dt search routine was tested with and without the filters on one year of data from both a CANMOS observatory and an AUTUMNX ground magnetometer station. The IQA (Iqaluit) station from CANMOS and the INUK (Inukjuak) station from AUTUMNX were used for comparison because they are both in the same region of NE Nunavut as the other stations used in the original dB/dt study. We used all available data from 2015 (note that AUTUMNX magnetometers (IQA) record mag-

netic field variation data with a 1-second rather than half-second cadence). The unfiltered dB/dt search results were manually classified as noise-type or TLA events via the criteria described in Sections 5 and 6 in order to test the accuracy of the filters.

The results of these search algorithms with and without the filters are presented in Table 3. In the MACCS stations, all TLA intervals were retained and a vast majority of noise-type signatures were successfully removed. The filters removed all of the noise-type dB/dt from the IQA station and all but 2 noise-type signatures from the INUK station. It is important to note that at the IQA and INUK stations, the filtered dB/dt search removed two events at each station that were classified as geophysical rather than noise-type, but did not meet the TLA selection criteria of having a dB/dt with timescale of 10-60 seconds and a ΔB of at least 60 nT. These were the only events that were removed via the filters that were not classified as noise-type events nor TLA events; these four events make up six signatures total comprising just 1% of the total geophysical signatures (i.e. total of the "# TLA dB/dt unfiltered filtered" in Table 3) from all eight stations in 2015.

The filtered dB/dt set contains all of the same TLA-type dB/dt signatures as prior to the filters, however there are significantly less noise-type intervals after being filtered. The ΔB, Δt and dB/dt values of the intervals in the filtered data set are much more similar between TLA and noise-type, however the noise-type events still exhibit many more signatures in general, and many more with the smallest Δt and ΔB values from 2-10 s and 6-100 nT (2134 noise-type intervals compared to 20 TLA intervals). The distribution of dB/dt values after the filters has many noise-type signatures with large dB/dt values that only few TLA signatures have (over 500 noise-type intervals have dB/dt value from 20-100 nT/s compared to 2 TLA intervals), although it is still very possible for TLA signatures to have dB/dt intervals in this range from 20-100 nT/s.

The filtered dB/dt signatures have greatly narrowed dB, dt and dB/dt characteristics. The number of dB/dt signatures per noise-type and TLA event is also much more similar in the post-filtered data set. Prior to the filters, the average number of dB/dt signatures per noise-type hour event window (for the six MACCS station used for the noise characterization in section 5) was over 150 dB/dt, and after the filters, this average for the same six stations is just over 10 dB/dt intervals. Overall, the filters greatly reduced the total number of noise-type dB/dt but also narrowed the noise-type dB/dt to just those that are most similar to TLA events. However, it can be seen from Table 3 that there is still a large number of noise-type dB/dt in the filtered dB/dt set.

What remains after the filters are noise-type and TLA signatures that are most similar in their amplitude and timescale characteristics, as well as the total number of dB/dt intervals within an hour event window. The data in Table 3 shows that the specific selection criteria imposed on the TLA dB/dt search algorithm greatly improved the efficiency of the results, removing over 99% of the noise-type dB/dt while retaining all TLA intervals that meet the formal definition of TLA events described in this section and excluding only four geophysical events that did not meet the criteria for a TLA event.

8 Support vector machine classification of noise-type and TLA dB/dt

While the filters described in section 7 improved the accuracy of the dB/dt search algorithm, there were still thousands of noise-type dB/dt (mostly found in the more commonly noisy stations IGL and CDR) which required further separation from the TLA dB/dt. Because the noise-type and TLA dB/dt intervals have very similar statistical characteristics after being filtered, they cannot be further separated with a linear approach and a more complex method of distinguishing the intervals is needed. As a final measure of separation, we implemented a machine learning classification technique to classify the dB/dt intervals returned from the filtered algorithm as TLA or noise-type. The

primary goal with a machine learning classifier was to identify and remove as many noise-type dB/dt as possible while retaining as many TLA-type dB/dt as possible.

The classifier used to identify TLA and noise-type dB/dt from the data set is called a support vector machine (SVM). In recent works, the SVM has been utilized for various space weather applications (e.g., prediction of solar flares using magnetic field data (Bobra & Couvidat, 2015) and prediction of high-latitude ionospheric scintillation with multiple types of solar wind and geomagnetic field data (McGranaghan et al., 2018)). This classifier was tuned and trained using all of the post-filter dB/dt signatures from 2015 and all eight stations (i.e. all of the dB/dt in the post-filter column of Table 3). The features used to tune and train the model are the dB, dt and dB/dt (values scaled to between 0-1), the geomagnetic latitude of the station represented as a fraction of 90 degrees, the time represented as a day fraction, and the day of year represented as a year fraction of 365 days (while also accounting for leap years). Thus, all of these features are scaled so that all values are between zero and one.

An SVM is a supervised machine learning technique often used for binary classification (Cortes & Vapnik, 1995). The objective of an SVM is to classify samples by determining the optimal hyperplane- or decision boundary- to separate the samples within the feature space (Suthaharan, 2016). The feature space for a training data set is the N-dimensional vector space that contains all of the feature values of the training set. The optimal hyperplane is determined by maximizing the space from the decision boundary to the nearest data points- or support vectors- in the feature space. If a data set is not linearly separable within the feature space (as in the case of the 2015 dB/dt set), the features are transformed into a higher-dimensional feature space where a linear hyperplane can be derived as decision boundary between classes. This transformation of the features to a higher dimensional space is performed using a kernel function.

The SVM used to classify dB/dt intervals in this study is from the scikit-learn library and uses the radial basis function (RBF) kernel (Pedregosa et al., 2011). The hyper-parameter C is used in the SVM model that introduces a penalty for incorrectly classified samples, the severity of the penalty determined by how large the scalar C is. A large value for C means a higher consequence for misclassified samples, this results in a decision boundary with smaller margins and can lead to overfitting of the training data. A C value that is too low results in very large margins and, in turn, more misclassified samples. The RBF kernel function also uses the hyper-parameter gamma, γ, that defines how much influence a single training example has. A large value of γ means that the similarity radius of each training point is larger and thus more points can be grouped together in the feature space, whereas a small value of γ means that the data points have to be much closer to one another in the feature space in order to be grouped together in the classification.

In the tuning and testing process, we used three main metrics to evaluate the performance of the SVM model: accuracy score, Probability Of Detection (POD) score, and Heidke skill score (HSS). The accuracy score represents the number of correct classifications (both TLA and noise-type) divided by the total number of predictions. Often the accuracy score does not best represent the performance of the model, so more complex metrics are utilized.

The latter two metrics are based on the model evalution guidelines of Liemohn et al. (2018) and they use the outcomes of the predictions made by the model in the tuning and testing process: H (hits i.e. correct classifications of TLA events), M (misses i.e. TLA events incorrectly classified as noise-type), F (false alarms i.e. noise-type events incorrectly classified as TLA events) and N (correct negatives i.e. noise-type events correctly classified as noise-type events). These metrics make up the contingency table for the model and are also commonly referred to as true positives, false negatives, false positives, and true negatives, respectively. The POD score gives a more specific evaluation

of how well the model performs at classifying TLA events, it is given by Eq. (9) of Liemohn et al. (2018):

$$POD = \frac{H}{H+M} \tag{1}$$

The POD score is a useful metric here because our purpose is to retain as many TLA events as possible. It ranges between 0 and 1 with higher values being better scores. The Heidke skill score (Heidke, 1926) represents all of the values in the contingency table and gives an evaluation of how well the model performs while excluding the classifications made by random chance (Eq (8) of Liemohn et al. (2018)):

$$HSS = \frac{2[(H \cdot N) - (M \cdot F)]}{[(H+M)(M+N) + (H+F)(F+N)]} \tag{2}$$

The HSS is highest at a value of 1 if the model perfectly classifies all of the hits and correct negatives and can result in a negative value if the model has no ability to classify TLA events.

In order to determine the optimal values for γ and C, the SVM model was cross-validated by first splitting the 2015 dB/dt data set into ten separate sets- or "stratified cross-folds"- with equal proportion of each type of sample (TLA and noise-type, of these sets may contain overlapping samples). Then each of these ten folds is split into training (80%) and testing (20%) sets and 49 SVMs are trained and tested for each of these ten data folds. Each of the 49 SVMs have a different combination of seven γ values (from 0.0001-100 in multiples of ten) and seven C values (from 0.001-1000 in multiples of ten). Thus, 49 combinations of γ and C were used to train SVMs on each of 10 separate folds of data for a total of 490 fits to the model. The SVM that has the highest accuracy and POD score averaged across all ten test folds is chosen to have the optimal hyper-parameters.

The results of the cross-validation process are shown in Figure 9: two grids showing the average accuracy and average POD of the ten folds for each C and γ value. In both cases, the hyper-parameters in the SVM that scored the highest average accuracy (0.989) and POD (0.944) scores across the ten folds of training data are $\gamma = 1$ and C = 10.

After the optimal values for γ and C were determined, these hyper-parameters were used to train the final SVM using all of the 2015 dB/dt data. In order to test the model performance, the initial, unfiltered dB/dt search as well as the filtered dB/dt search were performed on all of the same eight stations but for the year of 2016. All dB/dt identified from 2016 were manually classified as noise-type or TLA based on the criteria described in Sections 5 and 6 (i.e. comparison of shapes and amplitudes of the perturbations with those described in Khomutov et al. (2017) and statistical characteristics of events at MACCS stations in 2015) in order to assess the accuracy of the model predictions. The filtered dB/dt search was successful in removing a majority of noise-type dB/dt intervals while retaining all of the TLA signatures. Then the SVM classification was performed on the filtered dB/dt intervals.

The SVM model was chosen because it exhibited the best classification accuracy and POD scores out of four supervised machine learning classification algorithms. The details of the other three algorithms and their scores are provided in Supporting Information Text S1 and Table S1. Table 4 lists the number of TLA and noise-type dB/dt returned from the unfiltered and filtered dB/dt search for the year of 2016 as well as the results from the SVM classification. Because the classification is performed on individual dB/dt intervals and many events consist of multiple dB/dt grouped within a 1-hour window, the dB/dt label predictions are grouped if they occur within a 1-hour event window of one another and the final SVM classification of all the dB/dt intervals in the event

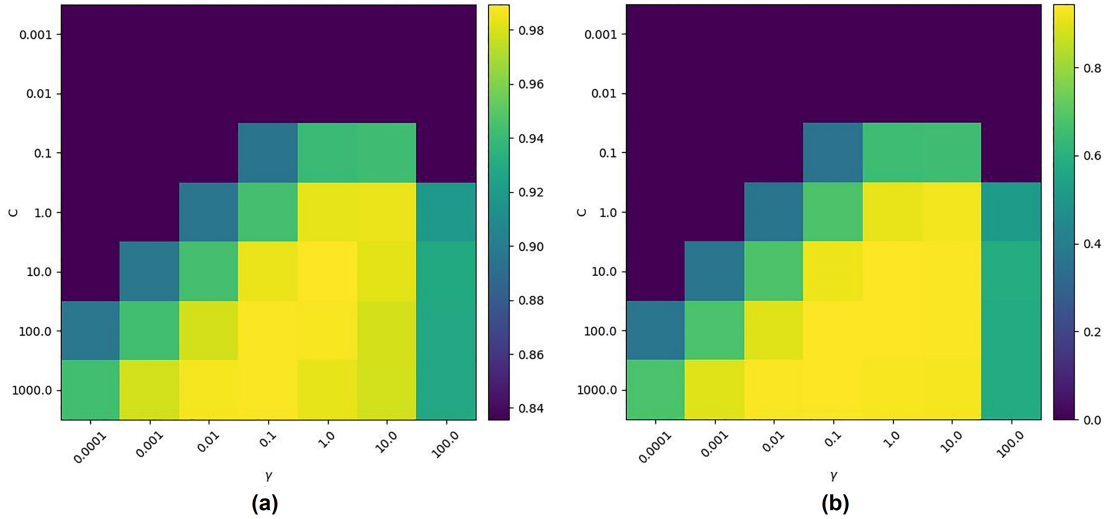

Figure 9. (a): Cross-validation grid showing the average accuracy score as the color of each square for each C and γ value for all 49 folds in the tuning process. (b): Same cross-validation grid as in (a) but for the average POD score for all 49 folds. Note that color bars are different for (a) and (b).

window is the majority vote of the predictions. If there are an equal number of dB/dt classified as noise-type and TLA within an event window, all dB/dt are labeled as geophysical TLA in order to reduce the number of TLA events removed by the SVM classification.

Table 4 shows that there were a total of 543,597 high-frequency dB/dt intervals identified in the 2016 data. These events were manually separated via the criteria described in section 5 to obtain a total of 543,159 noise-type dB/dt and 438 TLA dB/dt. After imposing the filters described in section 7, just 6349 intervals remain including 5911 noise-type and the same 438 TLA type (the manual classification found no geophysical events that did not meet the criteria for a TLA event). The filtered dB/dt intervals are those that go on to be classified with the SVM.

From the filtered dB/dt search, there are 5911 noise-type dB/dt signatures making up 327 event hours and 438 TLA dB/dt signatures making up 137 event hours. At the bottom of Table 4 are the number of dB/dt for each prediction type of the SVM classification. Out of 6349 total dB/dt signatures from the filtered dB/dt search for these eight stations throughout 2016, there are a total of 6299 correct predictions (i.e. H, "hits") resulting in an accuracy score for individual dB/dt signatures of 0.9923. Further, for the individual dB/dt interval SVM classifications, the POD score is 0.9361 and the HSS is 0.9383.

The dB/dt set returned from the full automated process of filtered dB/dt search with SVM majority-vote classification consists of 410 TLA dB/dt signatures making up 130 TLA hour-events and 22 noise-type dB/dt signatures making up 8 hour-events. In addition to the individual dB/dt predictions, Table 4 also includes the SVM prediction results of the hour-event windows. Out of the initial 464 event hours, 448 were classified correctly as having either TLA or noise-type dB/dt within, for an SVM classification accuracy score of full-hour event windows of 0.9655, POD score of 0.9416 and HSS of 0.9171.

Station	# Noise-type unfiltered dB/dt	# Noise-type filtered dB/dt	# TLA dB/dt unfiltered & filtered	# Noise-type post-SVM dB/dt	# TLA post-SVM dB/dt
IGL	131526	5126	13	7	12
GJO	3078	1	10	0	5
RBY	192525	249	37	0	32
PGG	3695	351	23	5	23
CDR	410	61	53	8	53
NAN	211736	0	2	0	2
INUK	7	2	194	2	194
IQA	182	121	106	0	89

	Total unfiltered	Total SVM-classified	# Correct noise-type	# Incorrect noise-type	# Incorrect TLA	# Correct TLA
dB/dt:	543597	6349	5889	22	28	410
Hour events:	3010	464	319	8	8	129

Table 4. Table with number of dB/dt signatures of both TLA and noise-type returned from the unfiltered and filtered dB/dt search algorithm and after the SVM classification

There are 22 incorrectly classified noise-type dB/dt signatures (making up 8 separate hour-event windows) that remain in the final data set and 28 incorrectly classified TLA-type dB/dt signatures (making up 8 event-windows) that are removed from the final data set after the SVM classification. All of the noise-type events mislabeled as TLA events consist of 1, 2 or 3 dB/dt in each component of the field that are part of a spike lasting less than 5 minutes; the average Δt and ΔB of the incorrectly classified noise-type intervals is longer and larger than that of the correctly classified noise-type intervals. The TLA events mislabeled as noise also have few dB/dt signatures (6 of 8 have less than 5 dB/dt signatures total) and all occur within a negative bay that lasts 20 minutes or more. The average Δt and dB/dt of the missed TLA events are similar to that of the correctly classified, however the average ΔB for the missed TLA intervals is about 40 nT smaller than that of the correctly classified TLA events. These details suggest that the most difficult events to distinguish are those with very few dB/dt intervals within the hour window: often spikes with longer than average timescale and amplitude, or TLA events with smaller than average amplitude. Because there are still eight hour events with noise-type dB/dt in the final dB/dt set, the final step of this complete dB/dt search process requires that the signatures are still plotted and the TLA-type events manually confirmed. However, the results of the full process in Table 4 show that the final dB/dt set is significantly narrowed to a majority of TLA-type events and only a few noise-type events.

The test scores of the SVM classifier on the filtered dB/dt intervals have all been presented above and show that the majority-vote SVM classification performs very well at identifying high-frequency disturbance events and classifying them as noise-type or geophysical.

In addition to providing the characteristics of the individual dB/dt signatures that meet the TLA event filter criteria and the SVM classification, the complete automated process provides a complete high-frequency disturbance event list for a magnetic field data set. The high-frequency event flagging process identifies all hour event windows that have any high-frequency dB/dt (defined as a dB/dt interval with 1-60 second timescale, dB/dt $>$ 6 nT/s and subsequent minimum ΔB of 6 nT) and initially classifies the hour

as a noise-type event. Then, if the requirements are met for these dB/dt to be a potential geophysical TLA event (i.e. the filter criteria: at least one dB/dt interval lasting more than 10 seconds and ratio of high-frequency dB/dt to all dB/dt within the hour being less than 0.05), the SVM majority-vote classification is performed. If the SVM classifies a majority of the high-frequency dB/dt as geophysical, then the classification of the hour window is changed to geophysical event rather than noise-type event. The resulting list is compiled of all of the hour event windows within a data set that contain high-frequency perturbations and includes the SVM majority-vote classification of the hour event as a zero if the dB/dt signatures are determined to be noise-type and a one if they are determined to be of geophysical nature. Thus, the complete high-frequency geomagnetic disturbance classifier can be used to retrieve information on the individual TLA dB/dt signatures as well as to identify hour event windows in the data that contain high-frequency signals and determine the geophysical or noise-type nature of those signals with high accuracy.

To concisely illustrate the performance of the fully automated geomagnetic disturbance classifier (initial dB/dt search, filters, and SVM classification), the contingency matrix for the 2016 test data is shown in Figure 10. This contingency matrix shows the four types of classification (H, F, M, N) for the entire set of high-frequency dB/dt intervals identified in the 2016 test data. The statistics in this Figure 10 are compiled from the Table 4 and show more clearly how well the complete process performs at identifying all second-timescale, high-frequency dB/dt intervals and classifying them as noise-type or geophysical TLA events. The test results for the full data set are listed below the contingency matrix. The accuracy score is quite high, but represents some possibility of correct classifications by random chance because there is such a larger proportion of noise-type dB/dt and event hours compared to TLA. The POD and HSS scores are more indicative of the actual performance of the automated process. The POD and HSS scores are all near 0.94 an show that the fully automated geomagnetic disturbance classifier performs quite well.

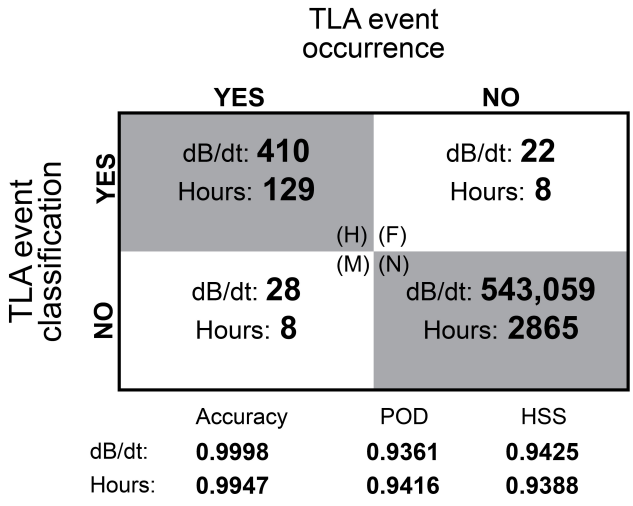

Figure 10. Contingency matrix and test scores for fully automated geomagnetic disturbance classifier performing on the 2016 test data.

This automated high-frequency geomagnetic disturbance classifier can be implemented on large-scale magnetic field databases. As a usable research artifact, we have provided the high-frequency event lists for the six MACCS stations used in this study for the year of 2017 to our data repository (doi.org/10.7302/78zf-yw59). From these lists,

we can identify that at the CDR station, 30 of the 104 MPEs that occurred during 2017 (of Engebretson et al. (2019a)) had TLA high-frequency variations associated to them and these are among the largest MPEs that occurred that year (> 10 nT/s). With these event lists, we can cross-reference these events with those from the other stations to identify what other stations to compare the spatial scales and relative strengths of these perturbations in this region which can help identify the M-I phenomena involved. Further, these event lists enable to us to avoid the hours of data that are highly likely to be contaminated with noise-type dB/dt events.

9 Effect of data processing on high-frequency geomagnetic signatures

We have identified both noise-type and geophysical TLA signals in raw data from MACCS, AUTUMNX and CANMOS magnetic field data as well as processed data from SuperMAG. While further data processing measures like averaging the data over 1-minute- or even 1-second- or using a band-pass filter may remove these signatures altogether, these techniques could also remove TLA signatures that are necessary for the study of small-scale M-I currents.

To briefly examine the effect of a common data processing and resampling procedure on high-frequency signals, we compared dB/dt signatures identified from raw, unprocessed MACCS data with those identified from processed data from the SuperMAG data service for two separate events that occurred at the PGG station in 2015. SuperMAG collects data from contributors (MACCS, AUTUMNX and CANMOS included) and processes it uniformly with the procedure described in Gjerloev (2012). SuperMAG offers 1-second averaged magnetic field data that has undergone the data cleaning (automated and manual) and baseline removal process: separation of the background magnetic field from sources in the M-I system by determining both the yearly trend and diurnal variations of the magnetic field (Gjerloev, 2012), as well as resampling the 2 Hz data to 1 Hz.

The MACCS, AUTUMNX and CANMOS magnetometer stations are all part of the SuperMAG network, so it is convenient to compare raw data from MACCS with processed data from SuperMAG for the same events. The filtered dB/dt search was conducted on both the raw MACCS data and the processed data from SuperMAG for two events at PGG during 2015. One of these events is the bay-like noise-type event that occurred on 20 June 2015, this event is shown in the unprocessed MACCS data in Figure 6 and in the processed SuperMAG data in Figure 11. The other event is a TLA event on 10 November 2015, shown in Figure 7.

With the unprocessed MACCS data, the noise-type event on 20 June exhibited 17 high-frequency dB/dt signatures among the four disturbances within the hour. These dB/dt signatures have an average ΔB of 69.8 nT, average Δt of 6.2 seconds, and average dB/dt of 13.1 nT/s. With the processed SuperMAG data (1-second averaged, cleaned and baseline removed) there are just 10 dB/dt signatures that have average ΔB, Δt and dB/dt of 68.7 nT, 6.9 seconds and 11.1 nT/s, respectively. Figure 11a shows that all four of the noise shapes are still present in the processed data, however there are less dB/dt signatures that meet the criteria for a high-frequency disturbance (second-timescale, dB/dt >6 nT/s and $\Delta B > 60$ nT). Further, the zoomed view of the bay-like disturbance in Figure 11 shows that the processed data removes some of the high-frequency behavior between the leading and trailing edges of the bay in all three components, but some of the high-frequency dB/dt signatures are still present.

The TLA event on 10 November 2015 at the PGG station exhibited 12 dB/dt signatures in the unprocessed MACCS data (shown in Figure 7) and 9 dB/dt signatures in the processed SuperMAG data. This event, like the noise-type event on 20 June 2015, had slightly lower average ΔB (273 nT) and dB/dt (7.6 nT/s) but slightly longer aver-

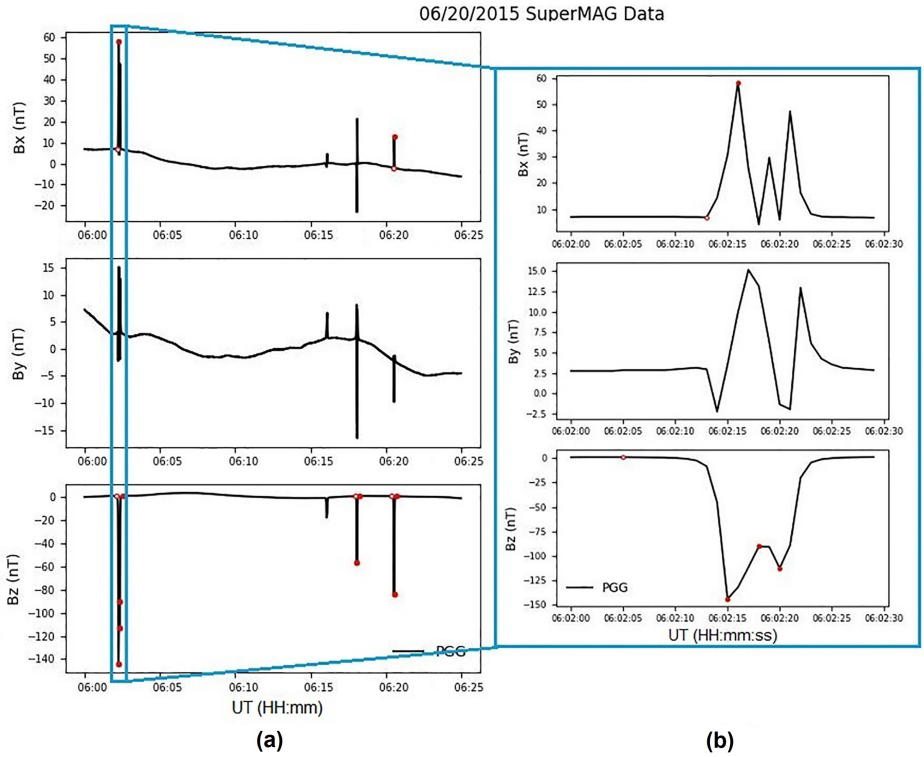

Figure 11. Bay-like noise in MACCS magnetic field data that has been processed with the SuperMAG data processing technique. The event occurred on 20, June 2015 at the PGG station. Hollow circles mark the start of a dB/dt signature and solid dots mark the end. The mean B value of each component in the interval shown is subtracted from the data (Note that this mean B value is different than that subtracted from the raw data in Figure 6 because all of the values are altered in the SuperMAG data processing.

age dt (34.7 s) in the cleaned and processed SuperMAG data. In both noise-type and TLA events, the processed data from SuperMAG exhibits fewer high-frequency dB/dt signatures overall, however in both cases some of these intervals are still present.

This comparative analysis shows that the SuperMAG data processing technique can reduce the amplitude of and even remove some high-frequency dB/dt signatures, but it does not remove the high-frequency noise-type events altogether. The same effect is observed for TLA events. Therefore, it is necessary to implement the automated high-frequency geomagnetic disturbance classifier on unprocessed data to identify intervals where high-frequency disturbances are present and classify them as noise-type or geophysical.

10 Conclusions

In this paper, we have outlined a basic dB/dt search algorithm and detailed the characteristics of the TLA and noise-type dB/dt identified by performing the search algorithm on data from six stations of the MACCS array during 2015. Then, we discussed the filters that were implemented to improve the dB/dt search process based on the characterization of the manually identified noise-type and TLA events and the SVM majority-vote classification of noise-type and TLA dB/dt signatures. Finally, we present an automated high-frequency geomagnetic disturbance classifier for magnetic field data.

The high-frequency geomagnetic disturbance classifier is a new technique that identifies intervals of unprocessed magnetic field data with 1-second or higher temporal resolution that contain high-frequency signals and determines if they are a result of noise or geophysical sources. The full dB/dt search process can identify these event windows and determine the correct source (noise-type or geophysical) with over 96% accuracy.

Because we found that both noise-type and geophysical high-frequency events are present in processed 1-second SuperMAG data, it is recommended that the SuperMAG data processing method incorporate this automated high-frequency event classifier on the raw, unprocessed magnetic field data and include this list of hour events containing high-frequency intervals and their classifications in the database. This list indicates windows of data that are likely contaminated with noise and undesirable for use in official space weather research, and identifies windows of data that contain high-frequency signals that are likely due to geophysical sources. The detailed information on these dB/dt intervals allows for analysis on the high-frequency behavior of space weather events and small-scale M-I currents.

11 Open Research

The data used for this analysis as well as the fully automated geomagnetic disturbance classifier are available on the University of Michigan's Deep Blue data repository (doi.org/10.7302/78zf-yw59).

Acknowledgments

Work supported by NSF grants (2013433 and 1848724) and NASA grant 80NSSC20K1779 to the University of Michigan and NSF grant (2013648) to Augsburg University.

The authors thank the MACCS team for data, available at http://space.augsburg.edu/maccs/.

We thank the Geological Survey of Canada (GSC). CANMOS data can be found at https://geomag.nrcan.gc.ca/obs/canmos-en.php.

We thank INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org).

We acknowledge NASA contract NAS5-02099 and V. Angelopoulos for use of data from the THEMIS Mission. Specifically: Martin Connors and C.T. Russell and the rest of the AUTUMN/AUTUMNX team for use of the GMAG data. AUTUMNX data are available at https://autumn.athabascau.ca/.

We gratefully acknowledge the SuperMAG collaborators (https://supermag.jhuapl.edu/info/?page=acknowledgement).

References

Bobra, M. G., & Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophysical Journal*, *798*(2). doi: 10.1088/0004-637X/798/2/135

Breiman, L. (1996). Bagging predictors. *Machine Learning*(24), 123–140. doi: 10.3390/risks8030083

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. doi: 10.1007/978-3-030-62008-0_35

Connors, M., Schofield, I., Reiter, K., Chi, P. J., Rowe, K. M., & Russell, C. T. (2016). The AUTUMNX magnetometer meridian chain in Québec, Canada. *Earth, Planets and Space*, *68*(1). doi: 10.1186/s40623-015-0354-4

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*(20), 273–297. doi: 10.1109/64.163674

Engebretson, M. J., Pilipenko, V. A., Ahmed, L. Y., Posch, J. L., Steinmetz, E. S., Moldwin, M. B., ... Vorobev, A. V. (2019). Nighttime Magnetic Perturbation Events Observed in Arctic Canada: 1. Survey and Statistical Analysis. *Journal of Geophysical Research: Space Physics*, *124*(9), 7442–7458. doi: 10.1029/2019JA026794

Engebretson, M. J., Pilipenko, V. A., Steinmetz, E. S., & Moldwin, M. B. (2021). Nighttime magnetic perturbation events observed in Arctic Canada : 3 . Occurrence and amplitude as functions of magnetic latitude , local time , and magnetic disturbance indices.

Gao, Y. F., Chi, P. J., Le, G., Russell, C. T., Yang, D. M., Zhou, X., ... Chun, F. K. (2000). Sino-Magnetic Array at Low Latitudes (SMALL) including initial results from the sister sites in the United States. *Advances in Space Research*, *25*(7-8), 1343–1351. doi: 10.1016/S0273-1177(99)00643-2

Gjerloev, J. W. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research: Space Physics*, *117*(9), 1–19. doi: 10.1029/2012JA017683

Heidke, P. (1926). Berechnung des erfolges und der gte der windstrkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, *8*(4), 301-349. doi: 10.1080/20014422.1926.11881138

Hughes, W., & Engebretson, M. (1997). MACCS: Magnetometer array for cusp and cleft studies. *Satellite - Ground Based Coordination Sourcebook. ESA SP-1198*(1), 119.

Kappenman, J. G. (2006). Great geomagnetic storms and extreme impulsive geomagnetic field disturbance events - An analysis of observational evidence including the great storm of May 1921. *Advances in Space Research*, *38*(2), 188–199. doi: 10.1016/j.asr.2005.08.055

Kataoka, R., & Ngwira, C. (2016). Extreme geomagnetically induced currents. *Progress in Earth and Planetary Science*, *3*(1). Retrieved from `http://dx.doi.org/10.1186/s40645-016-0101-x` doi: 10.1186/s40645-016-0101-x

Khomutov, S. Y., Mandrikova, O. V., Budilova, E. A., Arora, K., & Manjula, L. (2017). Noise in raw data from magnetic observatories. *Geoscientific Instrumentation, Methods and Data Systems*, *6*(2), 329–343. doi: 10.5194/gi-6-329-2017

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., ... Vasile, R. (2018). Model Evaluation Guidelines for Geomagnetic Index Predictions. *Space Weather*, *16*(12), 2079–2102. doi: 10.1029/2018SW002067

Love, J. J., & Finn, C. A. (2017). Real-time geomagnetic monitoring for space weather-related applications: Opportunities and challenges. *Space Weather*, *15*(7), 820–827. doi: 10.1002/2017SW001665

McCuen, B. A., Moldwin, M. B., & Engebretson, M. (2021). Characterization of Transient-Large-Amplitude Geomagnetic Perturbation Events. *Geophysical Research Letters*, *48*(15). doi: 10.1029/2021GL094076

McGranaghan, R. M., Mannucci, A. J., Wilson, B., Mattmann, C. A., & Chadwick, R. (2018). New Capabilities for Prediction of High-Latitude Ionospheric Scintillation: A Novel Approach With Machine Learning. *Space Weather*, *16*(11), 1817–1846. doi: 10.1029/2018SW002018

Mitra, P. P., Araya-polo, M., Byrd, D., & Akhiyarov, D. (2020). Machine Learning-based Anomaly Detection with Magnetic Data. doi: 10.20944/preprints202012.0092.v1

Neska, A., Reda, J., Neska, M., & Sumaruk, Y. (2013). On the influence of DC railway noise on variation data from Belsk and Lviv geomagnetic observatories. *Acta Geophysica*, *61*(2), 385–403. doi: 10.2478/s11600-012-0058-0

Nguyen, N., Muller, P., & Collin, J. (2020). The Statistical Analysis of Noise in Triaxial Magnetometers and Calibration Procedure. *2019 16th Workshop on Positioning, Navigation and Communications (WPNC)*, 1–6. doi: 10.1109/wpnc47567.2019.8970255

Ngwira, C. M., Pulkkinen, A. A., Bernabeu, E., Eichner, J., Viljanen, A., & Crowley, G. (2015). Characteristics of extreme geoelectric fields and their possible causes: Localized peak enhancements. *Geophysical Research Letters*, *42*(17), 6916–6921. doi: 10.1002/2015GL065061

Nikitina, L., Trichtchenko, L., & Boteler, D. H. (2016). Assessment of extreme values in geomagnetic and geoelectric field variations for Canada. *Space Weather*, *14*(7), 481–494. doi: 10.1002/2016SW001386

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pulkkinen, A., Bernabeu, E., Eichner, J., Viljanen, A., & Ngwira, C. (2015). Regional-scale high-latitude extreme geoelectric fields pertaining to geomagnetically induced currents. *Earth, Planets and Space*, *67*(1). Retrieved from `http://dx.doi.org/10.1186/s40623-015-0255-6` doi: 10.1186/s40623-015-0255-6

Pulkkinen, A., Bernabeu, E., Thomson, A., Viljanen, A., Pirjola, R., Boteler, D., ... MacAlester, M. (2017). Geomagnetically induced currents: Science, engineering, and applications readiness. *Space Weather*, *15*(7), 828–856. doi: 10.1002/2016SW001501

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

Pulkkinen, A., Viljanen, A., & Pirjola, R. (2006). Estimation of geomagnetically induced current levels from different input data. *Space Weather*, *4*(8). doi: 10.1029/2006SW000229

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Russell, C. T., Chi, P. J., Dearborn, D. J., Ge, Y. S., Kuo-Tiong, B., Means, J. D., ... Snare, R. C. (2008). THEMIS ground-based magnetometers. *Space Science Reviews*, *141*(1-4), 389–412. doi: 10.1007/s11214-008-9337-0

Santarelli, L., Palangio, P., & De Lauretis, M. (2014). Electromagnetic background noise at L'Aquila Geomagnetic Observatory. *Annals of Geophysics*, *57*(2). doi: 10.4401/ag-6299

Song, Y.-y., & Lu, Y. (2015). *Decision tree methods: applications for classification and prediction.* Shanghai Arch Psychiatry. doi: 10.11919/j.issn.1002-0829.215044

St-louis, C. B., Rasson, J., Schwarz, G., & Shanahan, T. (2014). INTERMAGNET Technical Note. , 1–7.

Suthaharan, S. (2016). *Machine Learning Models and Algorithms for Big Data Classification* (R. Sharda & S. Voß, Eds.). Springer. Retrieved from `http://www.springer.com/series/6157`

Tõth, G., Meng, X., Gombosi, T. I., & Rastätter, L. (2014). Predicting the time derivative of local magnetic perturbations. *Journal of Geophysical Research: Space Physics*, *119*(1), 310–321. doi: 10.1002/2013JA019456

Viljanen, A. (1997). The relation between geomagnetic variations and their time derivatives and implications for estimation of induction risks. *Geophysical Research Letters*, *24*(6), 631–634. doi: 10.1029/97GL00538

Xu, X., Huang, L., Liu, X., & Fang, G. (2020). DeepMAD: Deep Learning for Magnetic Anomaly Detection and Denoising. *IEEE Access*, *8*, 121257–121266. doi: 10.1109/ACCESS.2020.3006795