MMS SITL Ground Loop: Automating the burst data selection process

Matthew R. Argall 1,*, Colin Small 2, Samantha Piatt 2, Liam Breen 2, Marek Petrik 2, Julie Barnum 3, Kim Kokkonen 3, Kristopher Larsen 3, Frederick D. Wilder 3, Mitsuo Oka 4, Roy B. Torbert 1,5, Robert E. Ergun 3, Tai Phan 4, Barbara L. Giles 6, and James L. Burch 7

1 Space Science Center, EOS, University of New Hampshire, 8 College Rd., Durham, USA
2 Department of Computer Science, University of New Hampshire, Durham, NH, USA
3 Laboratory for Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, CO, USA
4 University of California at Berkeley, Berkeley, CA, USA
5 EOS-SwRI, Southwest Research Institute, Durham, NH, USA
6 Goddard Space Flight Center, NASA, Greenbelt, MD, USA
7 Southwest Research Institute, San Antonio, TX, USA

Correspondence*: Matthew R. Argall, Institute for the Study of Earth, Oceans, and Space, Space Science Center, University of New Hampshire, 8 College Rd., Durham, NH, 03824, USA
matthew.argall@unh.edu

ABSTRACT

Global-scale energy flow throughout Earth’s magnetosphere is catalyzed by processes that occur at Earth’s magnetopause (MP) in the electron diffusion region (EDR) of magnetic reconnection. Until the launch of the Magnetospheric Multiscale (MMS) mission, only rare, fortuitous circumstances permitted a glimpse of the electron dynamics that break magnetic field lines and energize plasma. MMS employs automated burst triggers onboard the spacecraft and a Scientist-in-the-Loop (SITL) on the ground to select intervals likely to contain diffusion regions. Only low-resolution survey data is available to the SITL, which is insufficient to resolve electron dynamics. A strategy for the SITL, then, is to select all MP crossings. This has resulted in over 35 potential MP EDR encounters but is labor- and resource-intensive; after manual reclassification, just ~ 0.7% of MP crossings, or 0.0001% of the mission lifetime during MMS’s first two years contained an EDR. We introduce a Long-Short Term Memory (LSTM) Recurrent Neural Network (RNN) to detect MP crossings and automate the SITL classification process. The LSTM has been implemented in the MMS data stream to provide automated predictions to the SITL. This model facilitates EDR studies and helps free-up mission operation costs by consolidating manual classification processes into automated routines.

Keywords: Magnetospheric Multiscale (MMS), Scientist in the Loop (SITL), Burst Data Management, Magnetopause, Long Short-Term Memory (LSTM), Ground Loop
1 INTRODUCTION

Earth is a strongly magnetized planet whose internal dynamics are largely influenced by its interactions with the solar wind and the resulting cycle of magnetic reconnection [1]. Reconnection occurs initially at the magnetopause (MP), at the interface between the shocked solar wind and Earth’s magnetosphere (MSP), in what is known as the electron diffusion region (EDR). The EDR had been enigmatic, with few direct observations [2, 3, 4, 5, 6] because spacecraft lacked the spacial and temporal resolution to resolve electron-scale dynamics, a hindrance later overcome by the Magnetospheric Multiscale (MMS) mission [7]. Leading up to its launch, so little was known about the EDRs that it was unclear how or if EDRs could be identified in the data, which led to speculation into the best EDR indicator [8, 9, 10, 11, 12, 13, 14]. Since launch, however, MMS has identified > 50 EDRs (see Webster et al. [15] for a partial list) and greatly expanded our knowledge of what catalyzes the global reconnection cycle.

To do this, MMS made significant efforts to capture enough of the right data to achieve its mission goals. The amount of high time resolution burst data recorded onboard is such that only about 4% can be downlinked. We present the first machine learning (ML) model that has been fully implemented into the mission operations data flow in order to ensure the selected 4% of burst data is able to address MMS science objectives [16, 17]. Our efforts help to transfer mission operations resources to science and to work around external constraints to advance our understanding of energy flow within the MSP.

MMS is not the only mission to face data limitations. Missions such as WIND, THEMIS, Cluster, STEREO, and others, all have burst mode schemes that operate only when triggered. Some burst modes are triggered on a pre-determined duty cycle, others are triggered on an instrument-by-instrument basis when onboard measurements meet certain criteria, and still others are coordinated among multiple or all instruments. But while burst modes and their triggers are mentioned in the instrumentation literature, details about the algorithms and the criteria behind them are mostly omitted, and their efficacy is largely unknown. Some WIND and THEMIS triggers used to detect plasma boundaries are described by Phan et al. [18]. Triggers used on STEREO for shock detection, their evolution, and their efficacy are documented in [19]. The MMS mission, driven by the large data volumes required to study electron dynamics at Earth’s MP, is the first to fully document its burst system from the beginning of the mission. In this paper, we describe our efforts to build upon the early mission design work in order to automate the burst data selection process.

The MMS burst management system consists of the automated burst system (ABS) that selects burst intervals by passing 10 s averaged trigger data numbers (TDNs) to on-board tables that set the burst trigger criteria [20, 21], and a human Scientist-in-the-Loop (SITL) who examines all of the low-resolution survey data and manually selects and classifies burst intervals. When designing the ABS and SITL selection process, it was envisioned that at some point automated algorithms would supplement or replace them. Algorithms that use the survey data available to the SITL on the ground fit into the “ground loop”. Presented below is the design, implementation, and results of the first ground loop. It is an ML model and is currently operating in near-real time to provide input to the SITL.

Most applications of ML in magnetospheric physics problems to-date have been geared toward the prediction of catastrophic events, including geomagnetically induced currents [22] that threaten power grids, solar energetic particles that threaten space assets [23], and geomagnetic indices [24, 25, 26] that indicate when global geomagnetic activity could lead to such events. Most methods link upstream conditions at L1 to those at geosynchronous or on the ground because of the continuous data coverage linking upstream and downstream conditions. Unfortunately, such models tend to suffer precisely during the extreme events they are trying to predict because they lack knowledge of processes internal to the MSP.
Space weather prediction can be improved by creating several ML models with knowledge of various specific aspects of magnetospheric dynamics. We later describe how several MMS ground loops could be combined to identify complex geomagnetic processes and support special science campaigns.

The primary science goal of MMS is to study electron dynamics associated with magnetic reconnection. However, because electron dynamics are not resolved in the survey data, the strategy employed by the SITL during the dayside phase is to select all MP crossings, a prominent location for magnetic reconnection [18]. Past attempts to identify the MP in an automated fashion used gradients in plasma parameters such as density or flux [27, 18]. Density and temperature probability functions were used to classify data from the solar wind, magnetosheath (MSH), and MSP [28]. Other region classifiers use Support Vector Machines on density and temperature data [29], 3D Convolutional Neural Networks (3D-CNNs) on the full 3D ion distribution function [30], and Random Forests on magnetic field and plasma data [31]. These models were trained to identify the solar wind, MSH, and MSP. The MP is then inferred as the boundary between the MSH and MSP. We present the first model specifically trained to identify MP crossings, thereby automating the primary SITL task.

This paper serves two purposes: 1) to document the burst management system and infrastructure and 2) demonstrate the performance of the first ground loop ML model. It is laid out as follows. First, the mechanisms that exist to make burst selections, including the SITL, ABS, and GLS, are described in §2. Next, an overview of the tools and processes developed to support the GLS infrastructure is provided in §3. Then, a description of the data and the GLS model are given in §4 and §5. In §6, we present the model results and performance. Section 7 is the Discussion, and §7.1 outlines the GLS Hierarchy, a framework needed to fully automate the SITL selection process. Finally, a summary is given in §8. Those interested in only the model and its results are referred to §3.3, §5, and §6.

2 BURST MANAGEMENT SYSTEMS

MMS has three systems for selecting intervals of burst data for downlink: the Scientist-in-the-Loop (SITL), the Automated Burst System (ABS), and the Ground Loop System (GSL).

2.1 Scientist-in-the-Loop

The SITL is a role that rotates among MMS team members. Currently, there are 73 participating SITL scientists that have made selections on over 1090 orbits of data. Each orbit contains Sub-Regions of Interest (SROIS) that encompass the most probable MP location, the bow shock, and other regions of scientific interest. SITLs make selections from the SROIS within a SITL window, a timeframe in which the MMS satellites make contact with the Deep Space Network (DSN) and incrementally downlink data. Data is passed through to the Science Data Center (SDC) (§3.1) where preliminary calibrations are applied and the data is made available to the SITL. The SITL then uses the EVA tool (§3.2) to interactively select data intervals for downlink.

SITLs follow guidelines set by mission PIs and Super SITLs (SITLs with super-user privileges) to help standardize the selection process. Those related to the MP are provided in Table 1. Each selection is given a Figure of Merit (FOM), a ranking between 0 and 255 split into five categories, to prioritize which selections are downlinked first. Priorities change based on the type of MP crossing. For example, complete, high-shear MP crossings receive a category 1 ranking (FOM 150-199), indirect reconnection signatures such as FTEs receive a category 2 ranking (FOM 100-149), and boundary layer encounters receive a category 3 ranking...
### Table 1. SITL guidelines during the Phase 5A dayside pass that are relevant to magnetopause encounters.

Categories 1-4 are given FOMs of 150-199, 100-149, 50-99, and 0-49, respectively. A “+” or “−” after a given category indicated a selection at the upper- or lower-range of the category. FOMs ≥ 200 are reserved for special events, such as calibration intervals or a definitive EDR encounter.

(FOM 50-99). Using these guidelines, the SITL makes a median of 30 selections per orbit, with a maximum to-date of 200 selections in a single orbit – a clear indicator of the time savings that can be achieved by automating the process.

### 2.2 Automated Burst System

The ABS applies configurable tables of weights and offsets to 10 s averaged burst quantities from each instrument, named Trigger Data Numbers (TDNs), to assign a Cycle Data Quality (CDQ) index to each 10 s buffer of burst data. The four CDQ values provided by the four spacecraft are downlinked, multiplied by another weighting factor, then summed to provide an overall Mission Data Quality (MDQ) index. The MDQ index is used to prioritize data for download [20].

There are 34 TDN terms available to the ABS. Early in the mission, the system was configured to look only for large changes in the magnetic field $B_z$ component. Reconnection events identified by scientists during the first two years of the mission have subsequently been used to determine which of the TDNs efficiently parameterize reconnection and to determine their relative importance.

For the dayside magnetopause, a set of 6 parameters is now employed in the search for reconnection. These parameters and their corresponding weights were selected based on their ability to efficiently identify the 32 potential dayside EDRs identified by Webster et al. [15]. The ABS as it is now configured would have selected 31 of the 32 Webster et al. [15] events for download with efficiency comparable to that of the SITL.
For the magnetotail, a different set of parameters is used. Six trigger terms that respond strongly to reversals of the magnetic field and bulk velocity were identified in data acquired during the 2017 magnetotail phase of the mission. Weights and gains were optimized through linear regression. The current configuration would have captured two well-substantiated magnetotail EDR encounters \cite{32, 33} with total download of burst intervals equivalent to the actual number of SITL selections.

### 2.3 Ground Loop System

The GLS is designed to be a system of ML or emperical models that automate the event classification process using all of the data available to the SITL. This last point is a key distinction as the SITL-level data is of lower quality than the Level-2 science-quality data freely available to the public – a ground loop trained with L2 data may not perform well within the GLS.

The first GLS model (§5) uses the text description given to each burst selection by the SITL as the ground-truth manual classification. SITL classifications significantly reduce the time required to train a supervised learning model, and the variety of selections made can facilitate a hierarchical ground loop infrastructure (§7.1), thereby eliminating the need for the SITL entirely.

### 3 Infrastructure and Tools

#### 3.1 Science Data Center

The MMS Science Data Center (SDC) is a collection of virtual machines and software applications that collectively support the science data processing and data access requirements for the MMS mission. It has been running since mission launch in March 2015 and currently manages a collection of over 11 million science data files accessible to MMS mission team members and 4 million files available to the public.

One of the key activities for the GLS is the ability to process the data used as input to the ground loop prediction models. This activity starts with a fixed time schedule or an external event set to trigger a science data processing job. The event sets relevant to the GLS are Deep Space Network (DSN) contacts that transfer spacecraft telemetry data to a ground station. The ground station transfers the raw data files to a NASA facility which then uploads them to the LASP Payload Operations Center (POC), where they are ingested to a raw telemetry database. A spacecraft-specific processing task is scheduled at the end of each DSN contact, delayed enough to allow the various data transfer and ingest tasks to complete. Each MMS instrument has a set of associated processing tasks for different data rates (survey vs. burst) and data levels. Processing tasks for the SITL ground loop are associated with survey data and the lowest data levels.

The GLS model-evaluation task is delayed an additional amount to allow completion of the various science data products that it needs to evaluate the model. Each GLS job produces a \texttt{csv} file containing the time range and FOM for each of the automated selections. A dropbox manager transfers the ground-loop selections into main SDC storage and indexes it for web-service access by the remote scientist’s EVA tool (§3.2), which then can plot the ground-loop selections alongside those of the ABS and the science data products, allowing the SITL to make informed selections.

The SITL must make selections within 12 hours of observation time to ensure that spacecraft commands used to “lock” the selected memory buffers are received before valuable observations are overwritten by newer ones. Although current MMS orbit periods are about 84 hours, spacecraft memory can hold only about 48 hours of burst data. The spacecraft contacts have variable schedules and cannot be optimized just
for MMS. The SDC completes the GLS processing within about two hours of the end of each DSN contact, well within required time limits.

The SDC mails reports to a broad team of experienced MMS SITL scientists, allowing review and comment on the latest selections. If it becomes practical to make fully automated selections based on algorithmic analysis, the SDC could short-circuit the human loop and transfer GLS results directly to the POC without human intervention.

### 3.2 EVA

EVA is a graphical user interface (GUI) designed specifically for the MMS/SITL activity and provided as a part of the MMS plug-in for the Space Physics Environment Data Analysis Software (SPEDAS) package [34]. SPEDAS is a software package for the IDL language that provides scripts for convenient plotting of spacecraft time series data and particle distributions. The MMS plugin includes the EVA GUI software, as well as software routines to load and plot data from every instrument on board MMS. The main functions of EVA are:

1. Load and display reduced-resolution, survey data for the entire duration of Region-of-Interest (ROI)
2. Help the SITL to identify and prioritize scientifically-valuable time ranges for downlinking the full-resolution burst data, and
3. Send the list of selected time ranges and their FOM values back to SDC for commanding.

A key feature of EVA is that it provides some pre-defined parameter sets. Parameter sets consist of data products frequently used by the SITLs (many are listed in Table 2) in combinations tailored to individual instruments, specific investigations, or regions of space. For example, there are parameter sets for the magnetopause, magnetotail, bow shock, and solar wind. By selecting a specific parameter set, in addition to the desired spacecraft ID, date, and time period, a SITL scientist can go straight to the task of viewing data needed for the SITL activity.

Parameter sets are displayed as tiles of time-series data in an interactive window in which the SITL can add, edit, and delete burst selections. In the earlier phases of the mission, EVA would append an ABS selections panel to the bottom of the window and the SITL scientist would manually adjust them while inspecting the data. Today, ground loop selections are also available to better guide the SITL in their selection process. This helps reduce personal bias when selecting similar phenomena, such as MP crossings.

### 3.3 pymms

pymms is a software package written in Python and freely available on GitHub and PyPI that makes full use of the MMS SDC’s data API. It is able to download instrument data (including both SITL and L2 quality data), as well as the ABS, GLS, and SITL selections. The SITL provides an ASCII text description of each burst interval that they select, which can be easily downloaded and searched with pymms to train supervised learning models. The GLS model described in this paper (§5) was trained in this way.

Note that the SITL, GLS, or ABS selections can be submitted multiple times, as described in §3.1, often with changes, so the available selections files have duplicate and overlapping entries, and may not

---

1. [http://spedas.org/blog/](http://spedas.org/blog/)
2. [https://github.com/argallmr/pymms](https://github.com/argallmr/pymms)
necessarily be in chronological order. Also, because of downlink and storage limitations, selections of long duration are broken into smaller chunks. pymms has tools to deal with these issues.

4 DATA

The flow of data used to train the MP model is shown in Figure 1. The SDC provides data from the Analog Fluxgate (AFG) magnetometer [35], the Electric field Double Probes (EDP) [36, 37], and the Fast Plasma Investigation (FPI) Dual Ion Spectrometer (DIS) and Dual Electron Spectrometer (DES) [38]. Fast survey data from each instrument was subjected to a preliminary set of calibrations to produce SITL-quality data, which is suitable for making informed decisions about burst selections but not for deep scientific scrutiny. This is necessary because of the urgency with which selections need to be made.

From these products, the 121 features listed in Table 2 were chosen to be inputs into the ML model. Most are standard products for the instruments, such as the B and E fields and their magnitudes, and the plasma energy spectrograms and moments. Others, like the temperature anisotropy, the custom $\gamma$ value, and the $Q$ values are metafeatures, features computed from the standard features. The $Q_\Delta$ features are gradient-based trigger terms used on Wind and THEMIS, are described in Phan et al. [39], and are calculated as $Q_\Delta x = |x - \bar{x}|$, where $\bar{x}_{j+1} = [\bar{x}_j(2^M - 1) + x_j]/2^M$, and $M = 2$ sets the amount of smoothing. This set of features represents a large portion of the data available to the SITL.

Data from 2017-01-01 to 2017-01-30 was used to train the model. The amount of training data was limited by resources on the platform on which model training was performed. Once downloaded, all data outside of the intervals selected and labeled by the SITL were discarded to ensure the accuracy of the ground truth. The data was then interpolated onto the 4.5 s cadence of the FPI instruments, then scaled and regularized. Because of an imbalance between the number of measurements that were selected as MP...
Table 2. Features used by the GLS magnetopause model.

| Instr. | Feature | Description |
|--------|---------|-------------|
| AFG    | $(B_x, B_y, B_z)$ | X, Y, and Z-components of the magnetic field in DMPA coordinates |
|        | $|B| = \sqrt{B_x^2 + B_y^2 + B_z^2}$ | Magnitude of the magnetic field |
|        | $Q_{\Delta B_z}$ | Quality value for $B_z$ |
|        | $P_B = |B|^2/(2\mu_0)$ | Magnetic pressure |
|        | $\theta_C = \arctan(B_y/B_x)$ | Clock angle. |
| EDP    | $(E_x, E_y, E_z)$ | X-, Y-, and Z-component of the DC electric field in DSL coordinates |
|        | $|E| = \sqrt{E_x^2 + E_y^2 + E_z^2}$ | Magnitude of the electric field |
| DIS    | $\varepsilon_i$ | Omni-directional energy spectrogram |
|        | $N_i$ | Number density |
|        | $(V_{i,x}, V_{i,y}, V_{i,z})$ | Bulk velocity in DBCS coordinates |
|        | $(Q_{i,x}, Q_{i,y}, Q_{i,z})$ | Heat-flux vector in DBCS coordinates |
|        | $T_{i,\|}/T_{i,\perp}$ | Parallel and perpendicular temperatures |
|        | $p_i = \frac{1}{3} \sum_{j=1}^{3} P_{i,j,k}$ | Scalar pressure |
|        | $Q_{\Delta N_i}$ | Quality value for $N_i$ |
|        | $Q_{\Delta V_{i,z}}$ | Quality value for the $V_{i,z}$ |
| DES    | $\varepsilon_e$ | Omni-directional energy spectrogram |
|        | $N_e$ | Number density |
|        | $(V_{e,x}, V_{e,y}, V_{e,z})$ | Bulk velocity in DBCS coordinates |
|        | $(Q_{e,x}, Q_{e,y}, Q_{e,z})$ | Heat-flux vector in DBCS coordinates |
|        | $T_{e,\|}/T_{e,\perp}$ | Parallel and perpendicular temperatures |
|        | $P_{e,j,k}$ | Pressure tensor where $(j,k) = (x,y,z)$ components in DBCS coordinates |
|        | $A_e = T_{e,\|}/T_{e,\perp} - 1$ | Temperature anisotropy |
|        | $T_e = (T_{e,\|} + 2T_{e,\perp})/3$ | Scalar temperature |
|        | $T_{e,\|}/T_{e,\perp}$ | Temperature ratio |
|        | $Q_{\Delta N_e}$ | Quality value for $N_e$ |
|        | $Q_{\Delta V_e,z}$ | Quality value for $V_{e,z}$ velocity |
|        | $Q_{\Delta N_e|V_e}$ | Quality value for ram pressure |
| Multiple | $\gamma = (T_i + T_e)/|E|$ | Custom feature |

Crossings compared to those that were not, class weights were applied to the “MP” and “not MP” classified data. We normalized all features with standardization, calculated as $x' = \frac{x-\bar{x}}{\sigma_x}$, where $x$ is the original vector for a given feature, $\bar{x}$ is the average of that vector, and $\sigma_x$ is its standard deviation. This method of normalization is widely used in machine learning applications to boost the performance of the model’s gradient descent while learning. Finally, each orbit was broken down into consecutive sequences of 250 measurements to reduce computational complexity.
5 GLS MAGNETOPAUSE MODEL

We develop an ML model that aids and, potentially, automates a key task performed by the SITL. Using the same low-resolution data as the SITLs, we identify time intervals that are likely to contain MP crossings. The ML models are trained using historical data annotated by SITL selections. The input is a low-resolution \((4\ s)\) time sequence of the data quantities outlined in Table 2. The goal is to predict, for each data point at time \(t\), whether the particular \(4\ s\) interval would be selected by the SITL and an MP event.

Our machine learning model is based on neural networks (NNs) [40]. A conceptual understanding of the MP model is built up from the NNs shown in Figure 2. The most simple of NNs is a perceptron, which takes a linear combination of the input features, \(x_i\), for a given time sample and passes the results through a sigmoid (“S”-shaped) activation function that maps the result to the interval \([0,1]\) to predict the output. Training occurs via backpropagation, a process by which the error in the prediction is used to adjust the weights applied to the inputs, usually via some gradient of the sigmoid function. After iterating, the NN output, \(y\), converges to a “yes” or “no” prediction. Perceptrons identify linear relationships between inputs and outputs.

Feed Forward NNs (FFNNs), Recurrent Neural Networks (RNNs), and Long-Short Term Memory (LSTM) Neural Networks evolve the perceptron to learn more complex, non-linear concepts. FFNNs do so by adding hidden layers whose weights determine the relationship between the input features themselves. They are trained via backpropagation in the same way as perceptrons. Like perceptrons, FFNNs make predictions using data from a single time sample. RNNs take the output of a FFNN at time \(t-1\) and combine it with the input of the FFNN at time \(t_0\), thereby incorporating the context inherent to time series data. Training backpropagates errors not only through the hidden layers, but also through time. Vanishing gradients in long prediction chains cause RNNs to have short-term memory. LSTMs create
long-term memory by applying gates to information carried forward from past predictions. Our model uses an adaptation of the LSTM to identify MP crossings.

The GLS MP model is composed of two bidirectional Long-Short Term Memory (LSTM) layers [40]. The output activation functions are $\tanh$ and the recurrent activation functions passed to units in $t_{+1}$ and $t_{-1}$ time steps are logistic. Each LSTM layer is followed by a drop-out layer with a drop probability of 0.4 as a means of forgetting information. Dropout layers help to reduce over-fitting when training the network. The output layer is a single unit with a logistic activation function. Additional details about the LSTM model can be found in the supplemental information ??.

Contiguous data points with positive predictions are combined to determine the time interval and duration of a suspected MP crossing. These suspected MP crossings are then presented to SITLs during their selection process to quicken, improve, and, ultimately, replace the manual selection process. Figure 1 shows a graphical representation of the data flow in our proposed automated SITL model. Data is downloaded from the SDC using pymms (§3.3) and pre-processed (§4) before being fed to our model to identify predicted MP crossings. These predictions are saved to csv files and stored on the SDC’s servers until finally transferred to the EVA team for a SITL to view when making selections.

6 MODEL PERFORMANCE

6.1 Case Studies

On 19 Oct. 2019, the model was installed and executed at the SDC for the first time and has been providing guidance to the SITL ever since. Figure 3 shows MP crossings from SROI1 on orbits 1061 (left)
and 1062 (center), and SROI3 on orbit 1075 (right), along with the selections made by the ABS, GLS, and SITL. The MP is identified in the ion and electron energy spectrograms (rows 1 and 2) as the transition between the hot, tenuous plasma of the magnetosphere and the colder, denser plasma of the magnetosheath. MMS transitions outward from the MSP to the MSH during the intervals shown for orbits 1061 and 1062, and inward from the MSH to the MSP for orbit 1075. During the transition, the MSH and MSP plasmas are observed simultaneously, $B_z$ (row 3) often rotates from negative to positive, and the density (row 4) transitions from $\sim 1\text{ cm}^{-3}$ in the MSP to $> 10\text{ cm}^{-3}$ in the MSH. SITL selections for these intervals are displayed in the bottom row while the SITL-provided description of each selection is given in Table 3. Rows 5 and 6 show similarities and differences between the selections made by the SITL and those made by the ABS and GLS.

Orbits 1061 and 1062 SROI1, and orbit 1075 SROI3 were chosen because they are cases when the ABS and GLS both select the SITL-identified MP crossings, only the GLS selects the MP crossings, and only the ABS selects the MP crossings, respectively. Orbit 1061 SROI1 was a low-shear, extended MP crossing lasting roughly 2.5 hours. MSH and MSP plasma were visible simultaneously, $B_z > 0$, and density took on intermediate values between the MSP at 1900 and the MSH at 2300 UT. Because the SITL was aware of telemetry and memory restrictions, they selected the final transition into the MSH and only the most promising-looking transitions within the long interval. The ABS made a similar set of selections to the SITL but at a significantly lower FOM. Meanwhile, the GLS, having been trained to identify the MP,

| Orbit | Date       | Start     | End       | FOM | Description                                      |
|-------|------------|-----------|-----------|-----|-------------------------------------------------|
| 1061  | 2019-10-22 | 19:09:53  | 19:11:03  | 100 | partial MP crs with deep B min                  |
| 1061  | 2019-10-22 | 21:14:33  | 21:15:53  | 100 | partial MP CRs min B 5 nT                       |
| 1061  | 2019-10-22 | 21:27:23  | 21:29:53  | 100 | partial MP CR                                   |
| 1061  | 2019-10-22 | 21:40:43  | 21:46:33  | 150 | partial & full MP Crs                           |
| 1061  | 2019-10-22 | 22:03:03  | 22:08:23  | 135 | Partial MP crs with K-H?                         |
| 1061  | 2019-10-22 | 22:50:23  | 22:51:33  | 125 | partial Mp                                      |
| 1061  | 2019-10-22 | 22:53:53  | 22:57:03  | 150 | mutliple MP CRs                                 |
| 1061  | 2019-10-22 | 22:57:43  | 23:00:03  | 150 | MP cr with mirror mode but no whistlers         |
| 1062  | 2019-12-17 | 18:16:03  | 18:25:43  | 120 | partial MP                                      |
| 1062  | 2019-12-17 | 18:25:53  | 18:27:33  | 40  | fill                                            |
| 1062  | 2019-12-17 | 18:27:43  | 18:33:13  | 120 | partial MP                                      |
| 1062  | 2019-12-17 | 18:35:53  | 18:43:33  | 120 | low shear MP                                    |
| 1075  | 2020-01-17 | 20:00:13  | 20:01:43  | 105 | Magnetosheath IMF rotation with bifurcated      |
| |            |           |           |       | signature - unresolved exhaust                  |
| 1075  | 2020-01-17 | 20:02:23  | 20:02:53  | 125 | Potential msheath flux rope                     |
| 1075  | 2020-01-17 | 20:08:43  | 20:13:03  | 175 | High-shear complete MP                          |
| 1075  | 2020-01-17 | 20:13:13  | 20:15:23  | 125 | Partial MPs with $V_z < 0$ jetting               |
| 1075  | 2020-01-17 | 20:15:33  | 20:15:53  | 45  | Fill                                            |
| 1075  | 2020-01-17 | 20:16:03  | 20:17:23  | 125 | Partial MPs                                     |
| 1075  | 2020-01-17 | 20:17:33  | 20:18:33  | 45  | Fill                                            |
| 1075  | 2020-01-17 | 20:18:43  | 20:20:23  | 125 | Partial MPs                                     |
| 1075  | 2020-01-17 | 20:20:33  | 20:21:33  | 45  | Fill                                            |
| 1075  | 2020-01-17 | 20:21:43  | 20:24:23  | 125 | Partial MPs                                     |
| 1075  | 2020-01-17 | 20:24:33  | 20:25:23  | 45  | Fill                                            |
| 1075  | 2020-01-17 | 20:25:33  | 20:26:33  | 125 | Partial MPs                                     |

Table 3. Selections made by the SITL during the intervals shown in Figure 3. "Partial", "high-shear", "full", and "complete" refer to classes of MP crossings ("crs") that receive different FOM values. By selecting a "fill" interval at low-FOM, the SITL provides context to adjacent events that is saved on board and can be increased to higher FOM later by a Super-SITL. "FPI Burst Cal" = calibration, "K-H" = Kelvin-Helmholtz.
selects all of the partial and full crossings throughout the interval. Even though the GLS is significantly over-selecting compared to the SITL, it is correctly identifying the MP.

During orbit 1062 SROI1, the SITL selected two partial MP crossings (\(\sim 1820\) and \(\sim 1830\) UT) and a full, low-shear MP crossing (\(\sim 1840\) UT) – see Table 3. None of the MP intervals were selected by the ABS while the GLS captures all three. One notable difference between the GLS and the SITL is apparent for the full crossing at 1840 UT. Whereas the SITL selected a large portion of the MSP and MSH on either side of the MP crossing to provide future scientists analyzing the data with relevant contextual information, the GLS selected only the MP transition. In this case, the GLS is under-selecting when compared to the SITL, but again is correctly identifying the MP (i.e. it is not classifying the surrounding MSP and MSH as the MP like the SITL did).

Orbit 1075 SROI3 is one in which the SITL and ABS select the MP but the GLS does not. The SITL selects a complete, high-shear MP crossing at 2010 UT followed by several partial crossings (Table 3), as well as reconnection-like signatures in the MSH near 2000 UT. The ABS also selects the high-shear crossing but only some of the partial crossings. It also captures the reconnection signatures in the MSH. For this time interval, the ABS over-selects in the MSH and under-selects at the MP. As for the GLS, further testing on this interval reveals that the GLS does select the MP if the LSTM model is run on a limited interval surrounding the MP, as opposed to the entire SROI. More generally, the GLS selects the majority of MP crossings during all SROI1 intervals but very rarely selects MP crossings during SROI3. These two facts could indicate that more care is needed when training the LSTM.

6.2 Statistical Study

To more broadly assess model performance, we make comparisons between the GLS and SITL, ABS and SITL, and GLS and ABS for all selections made in SROI1 between 19 Oct. 2019 and 25 March 2020. Figure 4 histograms the percent overlap between individual GLS (left) or ABS (center) burst selections with selections made by the SITL, while overlap between the GLS and ABS is on the right. Panels (a-f) depict the overlap using all selections. Panels (h-m) filter the SITL segments to contain only those that were classified as MP encounters. ABS segments were filtered so that only those that overlapped with SITL MP crossings were kept. Similar plots for SROI3 and both SROI1 and SROI3 together are included in the supplemental material (Figs. ?? and ??).

6.2.1 GLS-SITL Comparison

Focusing on the GLS and SITL first, we see that if the GLS and SITL make the same selection, the SITL generally selects 100% of the GLS segment (Fig. 4a). Overall, the SITL selects 69% of all 362 GLS selections. When we take the inverse case, however, the GLS selects only 28% of SITL selections (Fig. 4b), a clear indication that the SITL is selecting a lot more than just MP crossings.

Since the GLS is trained to select only MP crossings, we next examine how many of the GLS selections overlap with intervals identified by the SITL as MP crossings. Figure 4h shows that 37% of the GLS segments were classified as MP crossings by the SITL. This can be explained partly because the SITL is aware external control factors such as telemetry restrictions, as was the cause during Orbit 1061 (Fig. 3), and partly because some GLS segments are MP-like but are not classified as MP crossings by the SITL. Such selections include flux transfer events, boundary layers, etc. in which the plasma environment is a mixture of MSP and MSH plasmas. They also include bow shock crossings, which have field and plasma gradients similar to those present at MP crossings. While not MP crossings, these extra selections made by
Figure 4. Comparison between of the GLS, SITL, and ABS selections showing that the GLS outperforms the ABS at selecting SITL-classified MP crossings. Columns 1-3 show overlap between the GLS and SITL, ABS and SITL, and GLS and ABS, respectively, for all selections (a-f) and only those classified as MP crossings by the SITL (g-l).

Another indicator of good model performance is that the GLS selected 75% of the 172 segments classified as MP crossings by the SITL (Fig 4k). It serves as an indicator that individual SITL classification tasks (e.g. Table 1) can be automated by ML models and that a combination of models (§7.1) could reduce or eliminate operations costs associated with the SITL.
6.2.2 ABS-SITL Comparison

Next, we compare the ABS to all SITL selections (Fig. 4b,e), and to only those SITL selections that were classified as MP crossings ((Fig. 4i,l). The SITL selects a larger percentage (80% vs. 70%) of ABS segments than GLS segments (Fig. 4b), but the ABS selects as few SITL segments as the GLS (Fig. 4e), again indicating that the SITL is making more selections than the ABS. Examining the MP crossings, only 16% of ABS segments were classified as MP by the SITL (Fig 4h). Conversely, only 35% of SITL-classified MP crossings were selected by the ABS (Fig 4k). So, while a majority of ABS selections are of interest to the SITL, the ABS is significantly under-selecting both in a general sense and with respect to MP crossings.

The differences between the ABS and GLS are most likely due to how they were trained. The ABS was tuned to select EDR candidate events, which generally exhibit larger activity levels than MP crossings that do not occur within the EDR. Such high-activity levels in a false-positive EDR selection still occur in events that are of interest to the SITL (Fig 4b), but most MP crossings do not exhibit such activity (e.g. the MP crossings in Orbit 1062 shown in Fig. 3), leading to many false-negative cases ((Fig 4k).

6.2.3 GLS-ABS Comparison

Lastly, we compare GLS selections with those of the ABS, and with the subset of ABS selections that overlapped with SITL MP selections. Figures 4c,f show that there is little overlap between the GLS and ABS. This is partly because of the 282 burst segments selected by the ABS, only 46 overlapped with SITL-classified MP crossings (Fig. 4l). When both the SITL and ABS select a MP crossing, the GLS does so also 78% of the time. These results indicate that, although they under-select compared to all SITL selections, the GLS and ABS are not redundant; they each make useful, complementary selections.

7 DISCUSSION

The GLS MP model presented above is the first ML model implemented into the MMS burst management system to automate critical mission operation tasks. To fully automate the burst selection process, the GLS and ABS systems need to be expanded to:

1. Identify the variety of phenomena listed in the Seasonal SITL Guidelines
2. Assign appropriate FOM values to each phenomena
3. Include an appropriate amount of context around each selection
4. Respond to external control factors

To classify all phenomena within the SITL Guidelines (Item 1), models could be developed that train on all SITL selections. However, as the SITL rotates and mission priorities change, however, model performance would suffer. A better approach would be to create a hierarchy of classification models, as described in §7.1.

To assign appropriate FOM values (Item 2), two basic approaches could be considered: use a regression instead of a classification model, or create another model that classifies only on sub-types of MP crossings. In terms of the LSTM MP model, radial basis functions could be used instead of sigmoid functions for activation. Unlike sigmoid functions, radial basis functions map inputs to a continuous output variable so could be trained to predict FOM values. These models, however, would have to be retrained whenever mission priorities changed. As an example from MMS, during Phase 3B, low-shear MP crossings were classified as Category 3 events, whereas in all other phases they have been Category 1. To be more adaptable to changing mission priorities, events classified as MP crossings by the MP model could go through another
stage of classification that identifies their sub-type (complete/partial, high-/low-shear, etc). The sub-type, then, passes through a look-up table to assign a more appropriate FOM value.

Adding contextual information (Item 3) is relevant to the selections made on Orbit 1062 ROI1 in Figure 3. Model predictions could go through some post-processing to simply expand the selection forward and backward in time by a fixed amount or by some percentage of the selection duration.

Outside of the science considerations are operational control factors (Item 4), such as the amount of on-board memory available to store selections. Such considerations limited the number of selections that the SITL made during orbit 1061 SROI1. Once GLS selections are made (Items 1 and 3), they can be passed through the GLS Guideline look-up table and assigned an appropriate FOM (Item 2), then filtered through a system monitor that is aware of the state of on-board memory and can make decisions regarding the current set of selections. In some sense, the FOM prioritization does this intrinsically; however, selecting 2.5 hrs of a low-shear, slow MP crossing could potentially overwrite many other selections.

The GLS is an example of progressive autonomy [41]. It follows similar efforts undertaken by NASA to reduce mission costs through greater autonomy in ground control and spacecraft operations [42]. Autonomy can alleviate mission complexity and provide real-time decision making when communications latency exists [43]. The GLS MP model represents a key advancement toward reducing mission complexity by 1) facilitating larger data rates and more spacecraft through consolidation of event selection processes into a near real-time expandable and adaptable machine learning framework, and 2) accurately identifying and classifying events associated with prime science objectives.

7.1 Ground Loop Hierarchy

In the design phase, it was always envisioned that the ABS and GLS would eventually replace the SITL. So now that the first ground loop is in place, what can be done to expand the ground-loop infrastructure for that purpose? The answer is the to create a Ground Loop Hierarchy. The Hierarchy follows leaders in industry that found that combining many specialized models often out-performs one comprehensive model (e.g. Rascoff and Humphries [44]) For the GLS, this means training a hierarchy of models, as shown in Figure 5. At its lowest level, The Hierarchy consists of region classifiers that segregates data from topologically distinct regions of space. Tier 2 of the hierarchy consists of event classifiers that identify phenomena that are peculiar to a specific region. A third tier could distinguish between similar events to assign more appropriate FOM values, as suggested by the SITL Guidelines for MP crossings (Table 1). The final tier then activates sets of event classifiers from Tier 2 to answer science questions.

Applying the Ground Loop Hierarchy to the dayside region of Earth, all of the SITL data would be passed through the region classifiers to identify which data was recorded in the solar wind, magnetosheath, and magnetosphere, with the bow shock and magnetopause data being classified as the transitions between regions. Next, solar wind data would be passed to the Hot Flow Anomaly (HFAs) event classifier to identify HFAs. A similar process would be applied to all event classifiers, such as to identify mirror mode structures in the magnetosheath, plasmaspheric plumes in the magnetosphere, etc. In this way, the model that classifies HFAs does not have to know anything about the magnetosheath or magnetosphere. Finally, as mission objectives evolve from the primary mission through extended missions, different combinations of event classifiers can be activated to adapt to changing science goals or to strategic science campaigns.

As an example of a science campaign, we build upon a recent MMS discovery of micro-injections at the dusk flank MP in conjunction with ULF wave activity [45]. Simulations proposed that Kelvin-Helmholtz waves (KHWs) on the MP surface were the cause [46]. To gain more insight into this multi-scale process,
one could create event classifiers for micro-injections [47], field line resonances, and KHWs. Formulating a science campaign around micro-injections would entail activating each model. The models not only allow the mission to detect a complex series of events, they can also provide additional insights into the nature of the phenomena. Results from such automated science campaigns can be distributed to the wider scientific community in near real time, increasing the potential scientific impact and return of the data.

Work on the Ground Loop Hierarchy is already underway. Several models that could serve as region classifiers have already been developed [48, 30, 31, 29], and one is being adapted for that purpose [48]. The LSTM RNN model described above could serve as either a region or an event classifier. Other event classifiers have been developed (e.g., Claudepierre et al. [47]), but more are needed. Fortunately, the SITL has done the work to manually classify many events in many years of MMS data, and the tools provided as a product of this endeavor further reduce the effort required to make additions to the GLS. Soon there should be enough event classifiers to create the first automated science campaigns, thereby establishing the Ground Loop Hierarchy.

8 SUMMARY

MMS is providing key insights into the electron dynamics that catalyze the global flow of energy throughout the magnetosphere. Mission-critical science objectives depend on selecting a subset (~4%) of the high time resolution data that fit into its telemetry budget. A burst management system consisting of the Scientist-in-the-Loop (SITL), Automated Burst System, and Ground Loop System (GLS) ensure that the right ~4% of data makes it to the ground. This paper documents the tools and infrastructure of the
burst management system and demonstrates the performance of the first machine learning (ML) model implemented into the GLS to automate the SITL selection tasks. The GLS model is a Long Short-Term Memory Recurrent Neural Network trained on historical SITL selections to classify the magnetopause (MP), a primary task for the SITL as the MP is a key location for studying electron dynamics associated with magnetic reconnection. Since being implemented into the near real-time data stream, the GLS MP model has selected 75% of SITL-identified MP crossings in the outbound leg of its orbit, 40% better than the ABS. This represents the first attempt to introduce ML into critical mission operation tasks. By expanding the GLS into a hierarchy of ML models, MMS progresses toward full autonomy in its burst management system, thereby reducing operations costs and transferring information and resources back to answering fundamental science questions.

CONTRIBUTION TO THE FIELD

This article describes the first application of machine learning into the near real-time data flow to select mission-critical burst data. One major limitation that all space missions face is data telemetry restrictions. To combat these, past missions used burst triggers to activate faster data sampling. Such burst triggers and their efficacy are not well documented in the science or instrumentation literature. The Magnetospheric Multiscale (MMS) mission captures burst data continuously whenever it is in its region of interest and has three burst management systems to select segments for downlink: a traditional on-board Automated Burst System (ABS), a human Scientist-in-the-Loop (SITL), and a Ground Loop System (GLS). We describe the GLS and describe how a Long Short-Term Memory (LSTM) neural network was trained using historical SITL selections to detect magnetopause crossings, a primary task of the SITL. We assess the efficacy of the GLS model by comparing its near real-time selections to those made by the SITL and ABS. Our work represents a step toward automating the burst memory management system to reduce mission operations costs.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

M.R.A, C.S., S.P, L.B., and M.P. developed machine learning models and analyzed data. J.B., K.K., and K.L. developed the processes at the MMS SDC to generate, store, and distribute MMS data and burst selections. R.E.E., F.D.W. and M.O. wrote the SPEDAS and EVA software, with contributions from many of the other authors. R.B.T., R.E.E., T.P., B.L.G., J.L.B., F.D.W., and M.O. are super-SITLs responsible for the management of the MMS SITL infrastructure. M.R.A, C.S., M.P., M.O., F.D.W., and K.K. contributed to the writing of the paper. All authors help build the ground loop infrastructure.

FUNDING

This work was supported by NASA grant 80NSSC19K1203 and contract NNG04EB99C.
ACKNOWLEDGMENTS

M.R.A would like to thank M. G. Bobra for inspiration and fruitful discussions, S. C. Zaffke for helpful commentary, and all of the MMS SILTers for their collective work in classifying magnetopause crossings.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study, including the Level 2 data products plotted in Figure 3 and the ABS, GLS, and SITL selections are publicly available through the MMS SDC (https://lasp.colorado.edu/mms/sdc/public/) and are accessible via the pymms Python package (https://github.com/argallmr/pymms) or the MMS-Plugin for SPEDAS (http://spedas.org/blog/). SITL-level data used to train the GLS MP model is not suitable for science and so is not publicly available but can be made available by request.

REFERENCES

[1] Dungey J. Interplanetary Magnetic Field and the Auroral Zones. Physical Review Letters 6 (1961) 47–48. doi:10.1103/PhysRevLett.6.47.

[2] Nagai T, Shinohara I, Fujimoto M, Matsuoka A, Saito Y, Mukai T. Construction of magnetic reconnection in the near-Earth magnetotail with Geotail. Journal of Geophysical Research: Space Physics 116 (2011). doi:10.1029/2010JA016283.

[3] Nagai T, Zenitani S, Shinohara I, Nakamura R, Fujimoto M, Saito Y, et al. Ion and electron dynamics in the ion-electron decoupling region of magnetic reconnection with Geotail observations. Journal of Geophysical Research: Space Physics 118 (2013) 7703–7713. doi:10.1002/2013JA019135.

[4] Oka M, Phan TD, Øieroset M, Angelopoulos V. In situ evidence of electron energization in the electron diffusion region of magnetotail reconnection. Journal of Geophysical Research: Space Physics 121 (2016) 1955–1968. doi:10.1002/2015JA022040.

[5] Scudder JD, Holdaway RD, Daughton WS, Karimabadi H, Roytershteyn V, Russell CT, et al. First Resolved Observations of the Demagnetized Electron-Diffusion Region of an Astrophysical Magnetic-Reconnection Site. Physical Review Letters 108 (2012) 225005. doi:10.1103/PhysRevLett.108.225005.

[6] Tang X, Cattell C, Dombeck J, Dai L, Wilson LB, Breneman A, et al. THEMIS observations of the magnetopause electron diffusion region: Large amplitude waves and heated electrons. Geophysical Research Letters 40 (2013) 2884–2890. doi:10.1002/grl.50565.

[7] Burch JL, Moore TE, Torbert RB, Giles BL. Magnetospheric Multiscale Overview and Science Objectives. Space Science Reviews 199 (2015) 5–21. doi:10.1007/s11214-015-0164-9.

[8] Mozer FS. Criteria for and statistics of electron diffusion regions associated with subsolar magnetic field reconnection. Journal of Geophysical Research: Space Physics 110 (2005). doi:10.1029/2005JA011258.

[9] Scudder JD, Holdaway RD, Glassberg R, Rodriguez SL. Electron diffusion region and thermal demagnetization. Journal of Geophysical Research 113 (2008) 1–14. doi:10.1029/2008JA013361.

[10] Scudder J, Daughton W. "Illuminating" electron diffusion regions of collisionless magnetic reconnection using electron agyrotropy. Journal of Geophysical Research: Space Physics 113 (2008). doi:10.1029/2008JA013035.
[11] Zenitani S, Hesse M, Klimas A, Black C, Kuznetsova M. The inner structure of collisionless magnetic reconnection: The electron-frame dissipation measure and Hall fields. *Physics of Plasmas* **18** (2011) 122108. doi:10.1063/1.3662430.

[12] Swisdak M. Quantifying gyrotropy in magnetic reconnection. *Geophysical Research Letters* **43** (2016) 43–49. doi:10.1002/2015GL06980.

[13] Aunai N, Hesse M, Kuznetsova M. Electron nongyrotropy in the context of collisionless magnetic reconnection. *Physics of Plasmas (1994-present)* **20** (2013). doi:http://dx.doi.org/10.1063/1.4820953.

[14] Hesse M, Aunai N, Sibeck D, Birn J. On the electron diffusion region in planar, asymmetric, systems. *Geophysical Research Letters* **41** (2014) 8673–8680. doi:10.1002/2014GL061586.

[15] Webster JM, Burch JL, Reiff PH, Daou AG, Genestreti KJ, Graham DB, et al. Magnetospheric Multiscale Dayside Reconnection Electron Diffusion Region Events. *Journal of Geophysical Research: Space Physics* **123** (2018) 4858–4878. doi:10.1029/2018JA025245.

[16] Burch JL, Torbert RB, Phan TD, Chen LJ, Moore TE, Ergun RE, et al. Electron-scale measurements of magnetic reconnection in space. *Science* **352** (2016). doi:10.1126/science.aaf2939.

[17] Torbert RB, Russell CT, Magnes W, Ergun RE, Lindqvist PA, LeContel O, et al. The FIELDS Instrument Suite on MMS: Scientific Objectives, Measurements, and Data Products. *Space Science Reviews* **199** (2016) 105–135. doi:10.1007/s11214-014-0109-8.

[18] Phan TD, Shay MA, Eastwood JP, Angelopoulos V, Oieroset M, Oka M, et al. Establishing the Context for Reconnection Diffusion Region Encounters and Strategies for the Capture and Transmission of Diffusion Region Burst Data by MMS. *Space Science Reviews* **199** (2015) 631–650. doi:10.1007/s11214-015-0150-2.

[19] Jian LK, Russell CT, Luhmann JG, Curtis D, Schroeder P. Burst mode trigger of STEREO in situ measurements. *AIP Conference Proceedings* **1539** (2013) 195–198. doi:10.1063/1.4811021.

[20] Fuselier SA, Lewis WS, Schiff C, Ergun R, Burch JL, Petrinec SM, et al. Magnetospheric Multiscale Science Mission Profile and Operations. *Space Science Reviews* **199** (2014) 77–103. doi:10.1007/s11214-014-0087-x.

[21] Baker DN, Riesberg L, Pankratz CK, Panneton RS, Giles BL, Wilder FD, et al. Magnetospheric Multiscale Instrument Suite Operations and Data System. *Space Science Reviews* **199** (2016) 545–575. doi:10.1007/s11214-014-0128-5.

[22] Wintoft P, Wik M, Viljanen A. Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *J. Space Weather Space Clim.* **5** (2015) A7. doi:10.1051/swsc/2015008.

[23] Bouabram HI, Aydin B, Martens P, Angrýkov R. On the prediction of >100 MeV solar energetic particle events using GOES satellite data. *2017 IEEE International Conference on Big Data (Big Data)* (2017), 2533–2542. doi:10.1109/BigData.2017.8258212.

[24] Lundstedt H. Solar wind magnetosphere coupling: Predicted and modeled with intelligent hybrid systems. Tech. Rep. 7 (1997). doi:https://doi.org/10.1016/S0079-1946(97)00187-0.

[25] Bhaskar A, Vichare G. Forecasting of SYMH and ASYH indices for geomagnetic storms of solar cycle 24 including St. Patrick’s day, 2015 storm using NARX neural network. *J. Space Weather Space Clim.* **9** (2019) A12. doi:10.1051/swsc/2019007.

[26] Borovjke JE. Canonical correlation analysis of the combined solar wind and geomagnetic index data sets. *Journal of Geophysical Research: Space Physics* **119** (2014) 5364–5381. doi:10.1002/2013JA019607.

[27] Boardsen SA, Eastman TE, Sotirelis T, Green JL. An empirical model of the high-latitude magnetopause. *Journal of Geophysical Research: Space Physics* **105** (2000) 23193–23219. doi:10.1029/1998JA000143.
[28] Jelínek K, Němeček Z, Šafránková J. A new approach to magnetopause and bow shock modeling based on automated region identification. *Journal of Geophysical Research: Space Physics* **117** (2012). doi:10.1029/2011JA017252.

[29] da Silva D, Barrie A, Shuster J, Schiff C, Attie R, Gershman DJ, et al. Automatic Region Identification over the MMS Orbit by Partitioning n-T space (2020).

[30] [Dataset] Olshevsky V, Khotyaintsev YV, Divin A, Delzanno GL, Anderzen S, Herman P, et al. Automated classification of plasma regions using 3D particle energy distribution (2019).

[31] Nguyen G, Aunai N, de Welle B, Jeandet A, Fontaine D. Automatic detection of the Earth Bow Shock and Magnetopause from in-situ data with machine learning. *Annales Geophysicae Discussions* **2019** (2019) 1–22. doi:10.5194/angeo-2019-149.

[32] Torbert RB, Burch JL, Phan TD, Hesse M, Argall MR, Shuster J, et al. Electron-scale dynamics of the diffusion region during symmetric magnetic reconnection in space. *Science* (2018). doi:10.1126/science.aat2998.

[33] Zhou M, Deng XH, Zhong ZH, Pang Y, Tang RX, El-Alaoui M, et al. Observations of an Electron Diffusion Region in Symmetric Reconnection with Weak Guide Field. *The Astrophysical Journal* **870** (2019) 34. doi:10.3847/1538-4357/aaf16f.

[34] Angelopoulos V, Cruce P, Drozdov A, Grimes EW, Hatzigeorgiu N, King DA, et al. The Space Physics Environment Data Analysis System (SPEDAS). *Space Science Reviews* **215** (2019) 9. doi:10.1007/s11214-014-0057-3.

[35] Russell CT, Anderson BJ, Baumjohann W, Bromund KR, Dearborn D, Fischer D, et al. The Magnetospheric Multiscale Magnetometers. *Space Science Reviews* **199** (2014) 189–256. doi:10.1007/s11214-014-0057-3.

[36] Ergun RE, Tucker S, Westfall J, Goodrich KA, Malaspina DM, Summers D, et al. The Axial Double Probe and Fields Signal Processing for the MMS Mission. *Space Science Reviews* **199** (2014) 167–188. doi:10.1007/s11214-014-0115-x.

[37] Lindqvist PA, Olsson G, Torbert RB, King B, Granoff M, Rau D, et al. The Spin-Plane Double Probe Electric Field Instrument for MMS. *Space Science Reviews* **199** (2014) 137–165. doi:10.1007/s11214-014-0116-9.

[38] Pollock C, Moore T, Jacques A, Burch J, Gliese U, Saito Y, et al. Fast Plasma Investigation for Magnetospheric Multiscale. *Space Science Reviews* **199** (2016) 331–406. doi:10.1007/s11214-016-0245-4.

[39] Phan TD, Eastwood JP, Cassak PA, Øieroset M, Gosling JT, Gershman DJ, et al. MMS observations of electron-scale filamentary currents in the reconnection exhaust and near the X line. *Geophysical Research Letters* **43** (2016) 6060–6069. doi:10.1002/2016GL069212.

[40] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Adaptive Computation and Machine Learning series (MIT Press) (2016).

[41] Truszkowski W, Rouff C, Bailin S, Riley M. Progressive autonomy: a method for gradually introducing autonomy into space missions. *Innovations in Systems and Software Engineering* **1** (2005) 89–99. doi:10.1007/s11334-005-0018-9.

[42] Truszkowski WF, Hinchey MG, Rash JL, Rouff CA. Autonomous and autonomic systems: a paradigm for future space exploration missions. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **36** (2006) 279–291. doi:10.1109/TSMCC.2006.871600.

[43] Truszkowski W, Hinchey M, Rash J, Rouff C. NASA’s swarm missions: the challenge of building autonomous software. *IT Professional* **6** (2004) 47–52. doi:10.1109/MITP.2004.66.
[44] Rascoff S, Humphries S. *Zillow Talk: The New Rules of Real Estate* (Grand Central Publishing) (2015).

[45] Fennell JF, Turner DL, Lemon CL, Blake JB, Clemmons JH, Mauk BH, et al. Microinjections observed by MMS FIEPS in the dusk to midnight region. *Geophysical Research Letters* **43** (2016) 6078–6086. doi:10.1002/2016GL069207.

[46] Kavosi S, Spence HE, Fennell JF, Turner DL, Connor HK, Raeder J. MMS/FEEPS Observations of Electron Microinjections Due to Kelvin-Helmholtz Waves and Flux Transfer Events: A Case Study. *Journal of Geophysical Research: Space Physics* **123** (2018) 5364–5378. doi:10.1029/2018JA025244.

[47] Claudepierre S, Fennell J, Sengupta A, Kletzing C. Automated Identification of Electron Microinjections in MMS/FEEPS Measurements: Initial Results. *Earth and Space Science Open Archive* (2020) 10501706. doi:10.1002/essoar.10501706.1.

[48] Piatt S. Large-Scale Statistical Survey of Magnetopause Reconnection. *arXiv e-prints* (2019) arXiv:1905.11359.