

# MACHINE LEARNING IN HELIOPHYSICS

EDITED BY: Thomas Berger, Enrico Camporeale, Bala Poduval,  
Veronique A. Delouille and Sophie A. Murray

PUBLISHED IN: Frontiers in Astronomy and Space Sciences and  
Frontiers in Physics



#### Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714  
ISBN 978-2-88971-671-5  
DOI 10.3389/978-2-88971-671-5

#### About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

#### Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

#### Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

#### What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](http://frontiersin.org/about/contact)

# MACHINE LEARNING IN HELIOPHYSICS

Topic Editors:

**Thomas Berger**, University of Colorado Boulder, United States

**Enrico Camporeale**, University of Colorado Boulder, United States

**Bala Poduval**, University of New Hampshire, United States

**Veronique A. Delouille**, Royal Observatory of Belgium, Belgium

**Sophie A. Murray**, Trinity College Dublin, Ireland

**Citation:** Berger, T., Camporeale, E., Poduval, B., Delouille, V. A., Murray, S. A., eds. (2021). Machine Learning in Heliophysics. Lausanne: Frontiers Media SA.  
doi: 10.3389/978-2-88971-671-5

# Table of Contents

- 05** *Inferring Plasma Flows at Granular and Supergranular Scales With a New Architecture for the DeepVel Neural Network*  
Benoit Tremblay and Raphaël Attie
- 15** *Classification of Magnetosheath Jets Using Neural Networks and High Resolution OMNI (HRO) Data*  
Savvas Raptis, Sigiava Aminalragia-Giamini, Tomas Karlsson and Martin Lindberg
- 30** *Spectral Deconvolution With Deep Learning: Removing the Effects of Spectral PSF Broadening*  
Momchil Molnar, Kevin P. Reardon, Christopher Osborne and Ivan Milić
- 37** *Analyzing AIA Flare Observations Using Convolutional Neural Networks*  
Teri Love, Thomas Neukirch and Clare E. Parnell
- 45** *Incorporating Physical Knowledge Into Machine Learning for Planetary Space Physics*  
Abigail R. Azari, Jeffrey W. Lockhart, Michael W. Liemohn and Xianzhe Jia
- 59** *Dynamic Time Warping as a New Evaluation for Dst Forecast With Machine Learning*  
Brecht Laperre, Jorge Amaya and Giovanni Lapenta
- 76** *Improvement of Plasma Sheet Neural Network Accuracy With Inclusion of Physical Information*  
Brian M. Swiger, Michael W. Liemohn and Natalia Y. Ganushkina
- 85** *Neural Network Based Identification of Energy Conversion Regions and Bursty Bulk Flows in Cluster Data*  
Vlad Constantinescu and Octav Marghitu
- 96** *MMS SITL Ground Loop: Automating the Burst Data Selection Process*  
Matthew R. Argall, Colin R. Small, Samantha Piatt, Liam Breen, Marek Petrik, Kim Kokkonen, Julie Barnum, Kristopher Larsen, Frederick D. Wilder, Mitsuo Oka, William R. Paterson, Roy B. Torbert, Robert E. Ergun, Tai Phan, Barbara L. Giles and James L. Burch
- 112** *Automatic Classification of Plasma Regions in Near-Earth Space With Supervised Machine Learning: Application to Magnetospheric Multi Scale 2016–2019 Observations*  
Hugo Breuillard, Romain Dupuis, Alessandro Retino, Olivier Le Contel, Jorge Amaya and Giovanni Lapenta
- 125** *Convolutional Neural Networks for Predicting the Strength of the Near-Earth Magnetic Field Caused by Interplanetary Coronal Mass Ejections*  
Anna Malanushenko, Natasha Flyer and Sarah Gibson
- 140** *Visualizing and Interpreting Unsupervised Solar Wind Classifications*  
Jorge Amaya, Romain Dupuis, Maria Elena Innocenti and Giovanni Lapenta

- 162** *Solar-Terrestrial Data Science: Prior Experience and Future Prospects*  
Daniel N. Baker
- 173** *Comparison of Deep Learning Techniques to Model Connections Between Solar Wind and Ground Magnetic Perturbations*  
Amy M. Keesee, Victor Pinto, Michael Coughlan, Connor Lennox, Md Shaad Mahmud and Hyunju K. Connor
- 181** *Data Mining Reconstruction of Magnetotail Reconnection and Implications for Its First-Principle Modeling*  
Mikhail Sitnov, Grant Stephens, Tetsuo Motoba and Marc Swisdak
- 206** *Exploring Three Recurrent Neural Network Architectures for Geomagnetic Predictions*  
Peter Wintoft and Magnus Wik
- 217** *Concurrent Empirical Magnetic Reconstruction of Storm and Substorm Spatial Scales Using Data Mining and Virtual Spacecraft*  
Grant K. Stephens and Mikhail I. Sitnov



# Inferring Plasma Flows at Granular and Supergranular Scales With a New Architecture for the DeepVel Neural Network

Benoit Tremblay<sup>1,2\*</sup> and Raphaël Attie<sup>3,4</sup>

<sup>1</sup> National Solar Observatory, Boulder, CO, United States, <sup>2</sup> Department of Physics, Université de Montréal, Montréal, QC, Canada, <sup>3</sup> NASA Goddard Space Flight Center, Greenbelt, MD, United States, <sup>4</sup> Department of Physics and Astronomy, George Mason University, Fairfax, VA, United States

## OPEN ACCESS

### Edited by:

Veronique A. Delouille,  
Royal Observatory of  
Belgium, Belgium

### Reviewed by:

Francesco Malara,  
University of Calabria, Italy  
Christian L. Vásconez,  
National Polytechnic School, Ecuador

### \*Correspondence:

Benoit Tremblay  
btremblay@nso.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 06 March 2020

**Accepted:** 06 May 2020

**Published:** 05 June 2020

### Citation:

Tremblay B and Attie R (2020)  
Inferring Plasma Flows at Granular  
and Supergranular Scales With a New  
Architecture for the DeepVel Neural  
Network.  
*Front. Astron. Space Sci.* 7:25.  
doi: 10.3389/fspas.2020.00025

The wealth of observational data available has been instrumental in investigating physical features relevant to solar granulation, supergranulation and Active Regions. Meanwhile, numerical models have attempted to bridge the gap between the physics of the solar interior and such observations. However, there are relevant physical quantities that can be modeled but that cannot be directly measured and must be inferred. For example, direct measurements of plasma motions at the photosphere are limited to the line-of-sight component. Methods have consequently been developed to infer the transverse plasma motions from continuum images in the case of the Quiet Sun and magnetograms in the case of Active Regions. Correlation-based tracking methods calculate the optical flows by correlating series of images locally while other methods like “Coherent Structure Tracking” or “Balltracking” exploit the coherency of photospheric granules to track them and use the group motions of the granules as a proxy of the average plasma flows advecting them. Recently, neural network computing has been used in conjunction with numerical models of the Sun to be able to recover the full velocity vector in photospheric plasma from continuum images. We experiment with a new architecture for the DeepVel neural network which takes inspiration from the U-Net architecture. Simulation data of the Quiet Sun and Active Regions are then used to evaluate the response at granular and supergranular scales of the aforementioned method.

**Keywords:** active region, granulation, photosphere, neural networks, simulations, sunspots, supergranulation, velocity fields

## 1. INTRODUCTION

The Quiet Sun (i.e., in the absence of significant magnetic activity), hereafter QS, is filled with patterns of flows at multiple spatial and temporal scales. Granules, which typically have a diameter of 1 Mm and a lifespan of the order of 6 min, are associated with hot plasma upwellings whereas the intergranular lanes surrounding them, smaller in comparison, are associated with cold plasma sinking back in the interior. Supergranular flows are found at greater scale, i.e., above 20 Mm in diameter and with lifespans that range from hours to nearly 2 full days. This supercell-like pattern in the Quiet Sun is only revealed indirectly, for example by analyzing Dopplergrams and magnetograms or by tracking the average motion of granulation. It is believed

that supergranulation originates from the deeper layers in the convection zone (see Rieutord and Rincon, 2010, and references therein). More advanced techniques can provide actual images of supergranular cells (Potts and Diver, 2008; Attie et al., 2009, 2016). Sunspots, which appear as dark spots in intensitygrams with diameters ranging from 1 to 50 Mm and lifespans of the order of days, are signatures of the magnetic activity of the Sun (i.e., Active Regions, hereafter ARs), as revealed by magnetograms. The strong magnetic field in sunspots inhibits the convective motions of the plasma. Centered around sunspots are moat flows which start from the penumbra and expand horizontally and radially outward (i.e., away from the sunspot). This flow was originally revealed through the analysis of spectroheliograms and moving magnetic features in magnetograms (Sheeley, 1969; Hagenaar, 2005). When the moat flow encounters the neighboring supergranular cells of the quiet sun, it forms a ring-shaped boundary around the sunspot that looks like a rather dynamic, yet coherent supergranule-like cell structure (Attie et al., 2018).

Dopplergrams measure only the line-of-sight component of the aforementioned flows (Welsch et al., 2013). Methods were thus developed to recover the missing transverse component of the velocity vector by either inversion of the velocity vector as is done here or by finding the electric fields instead and then applying the ideal MHD Ohm's law to find velocity (Kazachenko et al., 2014). Subsequent applications include the estimation of energy fluxes (Kazachenko et al., 2015), the inference of boundary conditions from which to drive simulations of the Sun (Fisher et al., 2015) or the inference of synthetic observations such as velocity field reconstructions for data assimilation in a model of the solar photosphere, as suggested in Abbott and Fisher (2010). Reconstruction algorithms for ARs solve the magnetic induction equation with the vertical velocity and magnetic field vector being assigned Dopplergrams and vector magnetograms (e.g., Longcope, 2004; Schuck, 2005, 2006, 2008). Intensitygrams are used to track optical flows in the Quiet Sun, i.e., the displacement that needs to be applied to one image subfield to recover the image subfield at the following timestep. Optical flows were shown to be highly correlated with actual plasma motions averaged over spatial scales of the order of a megameter or larger and timescales of the order of 30 min (Rieutord et al., 2001). Correlation-based tracking methods such as "Local Correlation Tracking" (LCT: November and Simon, 1988) and "Fourier-based Local Correlation Tracking" (FLCT: Fisher and Welsch, 2008) calculate optical flows by correlating series of images locally while other methods like "Coherent Structure Tracking" (CST: Rieutord et al., 2007) or "Balltracking" (Potts et al., 2004) exploit the coherency of photospheric granules to track them and use the group motions of the granules as a proxy of the average plasma flows advecting them. Recently, a deep-learning algorithm was trained with computations performed by a radiative magnetohydrodynamics (MHD) simulation of the solar photosphere to emulate the physics that relate the continuum intensity to the velocity vector that appears in the model equations (i.e., not an optical flow). DeepVel<sup>1</sup> (Asensio

Ramos et al., 2017) is a fully-convolutional neural network that infers instantaneous depth-dependent transverse plasma motions from pairs of intensitygrams (i.e., the same inputs as tracking methods). Outputs are dependent on the spatial resolution, cadence and physics of the model presented during training. Versions of the neural network have been trained for the QS (e.g., Asensio Ramos et al., 2017; Tremblay et al., 2018) and ARs (e.g., Tremblay et al., submitted) and have been used to generate synthetic SDO/HMI observations, i.e., estimates of the plasma motions that reflect a numerical simulation but appear as though they were derived from the *Helioseismic Magnetic Imager* (HMI: Schou et al., 2012) onboard the *Solar Dynamics Observatory* (SDO: Hoeksema et al., 2014).

A comparison between a sample of intensity-based methods as a function of spatial scales identified DeepVel as best capturing the physics of the Quiet Sun at granular scales whereas it appeared less effective at supergranular scales (Tremblay et al., 2018). In this paper, we use simulation data at the HMI instrument spatial resolution to test whether adapting the architecture of the DeepVel neural network to that of a U-Net (Ronneberger et al., 2015) could further improve reconstructions at supergranular scales in the QS and ARs. Additionally, the architecture is modified to accept a combination of intensitygrams, magnetograms, and Dopplergrams as input to account for the physics and spatial features they encompass and their impact on the inferred flows.

## 2. METHODOLOGY

U-nets are widely used for image segmentation, e.g., the segmentation of coronal holes in solar data (Illarionov and Tlatov, 2018). This neural network inherits its name from the shape of its architecture which features a contracting branch, a bottleneck and an expansive branch (Ronneberger et al., 2015). The contracting branch identifies the dominant features at a given spatial scale through convolutional layers and downsampling operations to halve the resolution. The number of channels is doubled at each level. The expansive branch mirrors the contracting branch, upsampling from the low spatial resolution output of the bottleneck to higher resolution to provide context. Skip-connections concatenate the outputs of the contracting branch with inputs of each level of the expansive branch to localize features (Ronneberger et al., 2015). Moreover, U-nets have the ability to train for specific spatial scales by freezing (i.e., stop training) the weights and biases of other layers.

Photospheric flows range from subgranular scales (< 1 Mm) to supergranular scales (> 10 Mm). We adapt the architecture of DeepVel to that of a U-net to probe spatial scales that range from the pixel-size to the size of the sub-images presented to the neural network during training. We refer to Asensio Ramos et al. (2017) for a detailed description of the original DeepVel architecture. Each level of the contracting branch features in succession a 2D convolutional layer with a kernel of 3 by 3 pixels<sup>2</sup>, batch normalization (Ioffe and Szegedy, 2015), a ReLU activation function and a Dropout layer of 50 % to avoid overtraining. Downsampling is then performed using 2D convolutional layers

<sup>1</sup>DeepVel is an open-source neural network: <https://github.com/aasensio/deepvel>.

with strides of two pixels in all directions. The expanding branch mirrors the contracting branch, with the output of each level being concatenated with the output from the contracting branch at the same spatial scale. Upsampling in the expansion branch is performed by repeating each row and each column twice. Unlike the classic U-net architecture, the number of channels is maintained throughout all layers (i.e., 128 channels), with the exception of the bottleneck (i.e., 256 channels), to diminish the number of free parameters to adjust.

Versions of the U-net architecture (hereafter DeepVelU) are trained, validated and tested using the Tremblay et al. (2018) dataset for the Quiet Sun (hereafter the QS dataset) and the Tremblay et al. (submitted) dataset which features a mixture of ARs and QS (hereafter the ARs dataset). Training of the DeepVel and DeepVelU neural networks relies on a similar approach to that of measuring optical flows: two consecutive images (i.e., intensitygrams, Dopplergrams, and/or magnetograms) are presented to the network and their differences and dominant features are used to reconstruct a velocity vector. The ARs dataset is required in addition to the QS dataset as a neural network trained exclusively with examples of the QS will fail to extrapolate the flows in Sunspots whose physical mechanisms governing plasma motions in relation to the observed light intensity is very different from those in the QS Tremblay et al. (submitted). On the other hand, although it is presented a dataset that differs from the one used for training, the neural network is capable of generalizing the behavior for granulation (Tremblay et al., 2018; Tremblay et al., submitted).

The QS dataset was derived from the STAGGER magnetoconvection simulation<sup>2</sup> of solar granulation (Stein, 2012; Stein and Nordlund, 2012). It features maps of the continuum intensity  $I_c$  at 500 nm and the velocity vector  $\vec{v}$  at optical depths  $\tau = \{1, 0.1, 0.01\}$  with a field of view of dimensions 96.768 by 96.768 Mm<sup>2</sup>, spatial resolution  $\Delta x = \Delta y = 96$  km pixel<sup>-1</sup> and time step  $\Delta t = 60$  s for a total duration of 6 h (i.e., many turnover times). The ARs dataset for  $I_c$ ,  $\vec{v}$  and  $\vec{B}$  was generated by the MURaM simulation of a Sunspot (Rempel and Cheung, 2014). The Sunspot has a diameter of  $\approx 25$  Mm and is featured at the center of a field of view of 98.304 by 98.304 Mm<sup>2</sup>. The spatial resolution of the dataset is the same as the QS dataset (i.e.,  $\Delta x = \Delta y = 96$  km pixel<sup>-1</sup>), but the cadence  $\Delta t$  is higher at 45 s which coincides with the cadence of SDO/HMI level-2 products (Hoeksema et al., 2014).

The simulation data was convolved with the SDO/HMI point spread function (PSF; Wachter et al., 2012) before being resampled to the SDO/HMI spatial resolution near disk center ( $\Delta x = \Delta y \approx 368$  km pixel<sup>-1</sup>) using nearest-neighbor sampling. Patches of 48 by 48 pixels<sup>2</sup> were then extracted at random positions in planes of constant  $\tau$  and at consecutive times  $t = \{t_i, t_i + \Delta t\}$  where  $i$  is a randomly-selected timestep. The QS dataset includes 2,000 training examples and 200 validation examples for  $I_c(x, y, \tau = 1, t = \{t_i, t_i + \Delta t\})$  and  $\vec{v}_t(x, y, \tau = \{1, 0.1, 0.01\}, t = t_i)$ . The ARs dataset is comprised of 3200 training examples and 900 validation examples for

$I_c(x, y, \tau = 1, t = \{t_i, t_i + \Delta t\})$ ,  $\vec{v}_t(x, y, \tau = 1, t = t_i)$ , the vertical velocity  $v_z(x, y, \tau = 1, t = \{t_i, t_i + \Delta t\})$  and the vertical magnetic field  $B_z(x, y, \tau = 1, t = \{t_i, t_i + \Delta t\})$ . Computations were performed on a NVIDIA-GTX-960 GPU using the Keras library with the *Tensorflow* backend. Weights and biases were only updated when the mean square error for the validation set improved.

## 3. RESULTS

Full field-of-view maps in sequences of 30 to 80 time steps from the QS and ARs datasets are used as test sets to evaluate the performance of DeepVelU at scales equal to and greater than those analyzed in Tremblay et al. (2018).

### 3.1. Metrics

Agreement between the simulation vector field  $[\vec{v}_{\text{ref}}]$  (i.e., the reference case) and the reconstructed vector field  $[\vec{v}_D]$  is quantified by the mean absolute errors

$$E_{\text{abs}} = \langle (v_{\text{ref},x} - v_{D,x})^2 + (v_{\text{ref},y} - v_{D,y})^2 \rangle, \quad (1)$$

and mean relative errors

$$E_{\text{rel}} = \left\langle \frac{(v_{\text{ref},x} - v_{D,x})^2 + (v_{\text{ref},y} - v_{D,y})^2}{(v_{\text{ref},x}^2 + v_{\text{ref},y}^2)} \right\rangle, \quad (2)$$

where  $\langle \cdot \rangle$  is the spatial average operator. Additionally, the Pearson linear correlation coefficient between  $\vec{v}_{\text{ref}}$  and  $\vec{v}_D$ , denoted  $[C]$ , is introduced as a measure of similarity (or discrepancy) that takes into account different spatial scales of the vector amplitude. The averaged normalized dot product

$$A \equiv \left\langle \frac{\vec{v}_{\text{ref}} \cdot \vec{v}_D}{||\vec{v}_{\text{ref}}|| ||\vec{v}_D||} \right\rangle, \quad (3)$$

is used to assess the global orientation of the inferred velocity vectors with respect to the simulation, with values  $A = \pm 1$  for parallel/anti-parallel vectors and  $A = 0$  for perpendicular vectors.

Response as function of the spatial scales is evaluated through the power spectrum of the total kinetic energy of the transverse plasma motions. Energy densities  $[E(k)]$  are computed following the definition of Rieutord et al. (2010) for a square dataset (i.e.,  $n_x = n_y$  and  $\Delta x = \Delta y$ ):

$$E(k) = \frac{n_x \Delta x}{2\pi} \sum_{\forall k' \in [k, k+dk]/2\pi} \frac{\bar{v}_x^2(k') + \bar{v}_y^2(k')}{2}, \quad (4)$$

where  $\bar{v}_x$  and  $\bar{v}_y$  are the discrete Fourier transform of  $v_x$  and  $v_y$ , and the wavenumber  $[k]$  is an inverse measure of the spatial scale.

Finally, for each velocity field, we compute the unsigned “shear” component of the vertical Poynting flux  $[S_z]$  and then integrate it over the field-of-view (Liu and Schuck, 2012; Welsch, 2015):

$$S_z = -\frac{1}{4\pi} (\vec{v}_D \cdot \vec{B}_{\text{ref},t}) B_{\text{ref},z}, \quad (5)$$

<sup>2</sup>Computations performed by the STAGGER code are available for download: <http://steinr.pa.msu.edu/~bob/96averages/>.

where  $[\vec{B}_{\text{ref},t}]$  and  $[B_{\text{ref},z}]$  are the transverse component and the vertical component of the magnetic field, respectively, computed from the reference simulation. Note that the total Poynting flux also includes the “emergence” term, which contains Doppler velocity. However the division into “shearing” and “emergence” terms is done conceptually, since both terms involve emergence of magnetized plasma across the photosphere (Welsch, 2015).

### 3.2. Quiet Sun Test Set

Plasma motions  $\vec{v}_{D,t}(\tau \approx 1)$  inferred by DeepVelU are consistent with  $\vec{v}_{\text{ref},t}$ , producing divergent velocity vectors at the center of granules and converging vectors in the intergranular lanes (**Figure 1A**). Furthermore, the new architecture results in reduced errors for  $\vec{v}_{D,t}(\tau \approx 1)$ ,  $(\vec{\nabla} \times \vec{v}_{D,t})_z$  and  $\vec{\nabla}_t \cdot \vec{v}_{D,t}$  in comparison to DeepVel (**Table 1**). Both methods underestimate flow amplitudes, as suggested by the scatterplot in **Figure 1B**, but the similarity with the reference velocity amplitudes increases from  $C = 0.841$  to  $0.947$  when transitioning from DeepVel to the DeepVelU architecture (**Table 1**). Furthermore, the RMSE,  $E_{\text{abs}}$  and  $E_{\text{rel}}$  at  $\tau \approx 1$  decrease from  $0.777 \text{ km s}^{-1}$ ,  $0.694 \text{ km s}^{-1}$  and  $70.6\%$  to  $0.501 \text{ km s}^{-1}$ ,  $0.442 \text{ km s}^{-1}$  and  $43.7\%$ , respectively. Similar improvements are noted for the divergence and the curl of the flow fields (**Table 1**). The global orientation of the velocity vectors is also improved, with  $A$  increasing from  $0.786$  when using DeepVel to  $0.914$  when using DeepVelU. Further analysis identified regions of downflows as the largest sources of errors (not shown). These are typically associated with intergranular lanes which feature more complex flow structures confined in small areas.

Similar conclusions are drawn for the inversion of flows at optical depths  $\tau \approx \{0.1, 0.01\}$  (i.e., at higher geometrical heights above the surface) from  $I_c(\tau \approx 1)$  (**Figure 1** and **Table 1**). More specifically,  $\tau \approx 0.1$  is just beyond the height where the reversal of the granulation pattern occurs (i.e., a few hundred kilometers above the surface: Cheung et al., 2007), with the center of granules becoming colder than the intergranular network. At  $\tau \approx 0.01$ , this pattern is more diffused but remains well-correlated with the surface granulation pattern (not shown). By extension, the structures in  $\vec{v}_t(\tau \approx \{0.1, 0.01\})$  correlate with  $I_c(\tau \approx 1)$ .

The power spectrum of the kinetic energy (Equation 4) as a function of the spatial scales shows an improved response for DeepVelU at supergranular scales ( $k^{-1} \approx 100 \text{ Mm}$ ) with respect to the test set (green and blue curves in **Figure 2A**). The transition from a loss to a gain in signal with respect to DeepVel (orange curve) occurs close to the spatial scale that is achieved through downsampling at the bottleneck of the U-net architecture ( $k^{-1} \approx 2.944 \text{ Mm}$ ). This change is interpreted as the new architecture favoring signal at supergranular scales over the pixel-size and granular scales to further optimize the cost function during the training process.

In addition, **Figure 2C** shows a significant improvement for DeepVelU in the correlation between the inferred velocities and the simulation at all time and spatial scales for the QS. DeepVel’s flows have the surprising disadvantage of losing correlation at increasing time averages by a few percents, whereas the correlation increases for DeepVelU. Similarly, DeepVel’s correlations significantly decreases at greater spatial averages,

e.g., from (resp.)  $C \approx 0.85$  to  $C \approx 0.62$  between  $368 \text{ km}$  and  $5 \text{ Mm}$  (resp.), which is not the case for DeepVelU which consistently correlates very well with the simulation at  $C > 0.9$ . The latter, however, plateaus near  $3 \text{ Mm}$  which coincides with the pixel-size in the bottleneck of the U-net architecture. Future work will test if increasing the field-of-view of the training images affects favorably the ability to improve the correlation further at greater spatial scales.

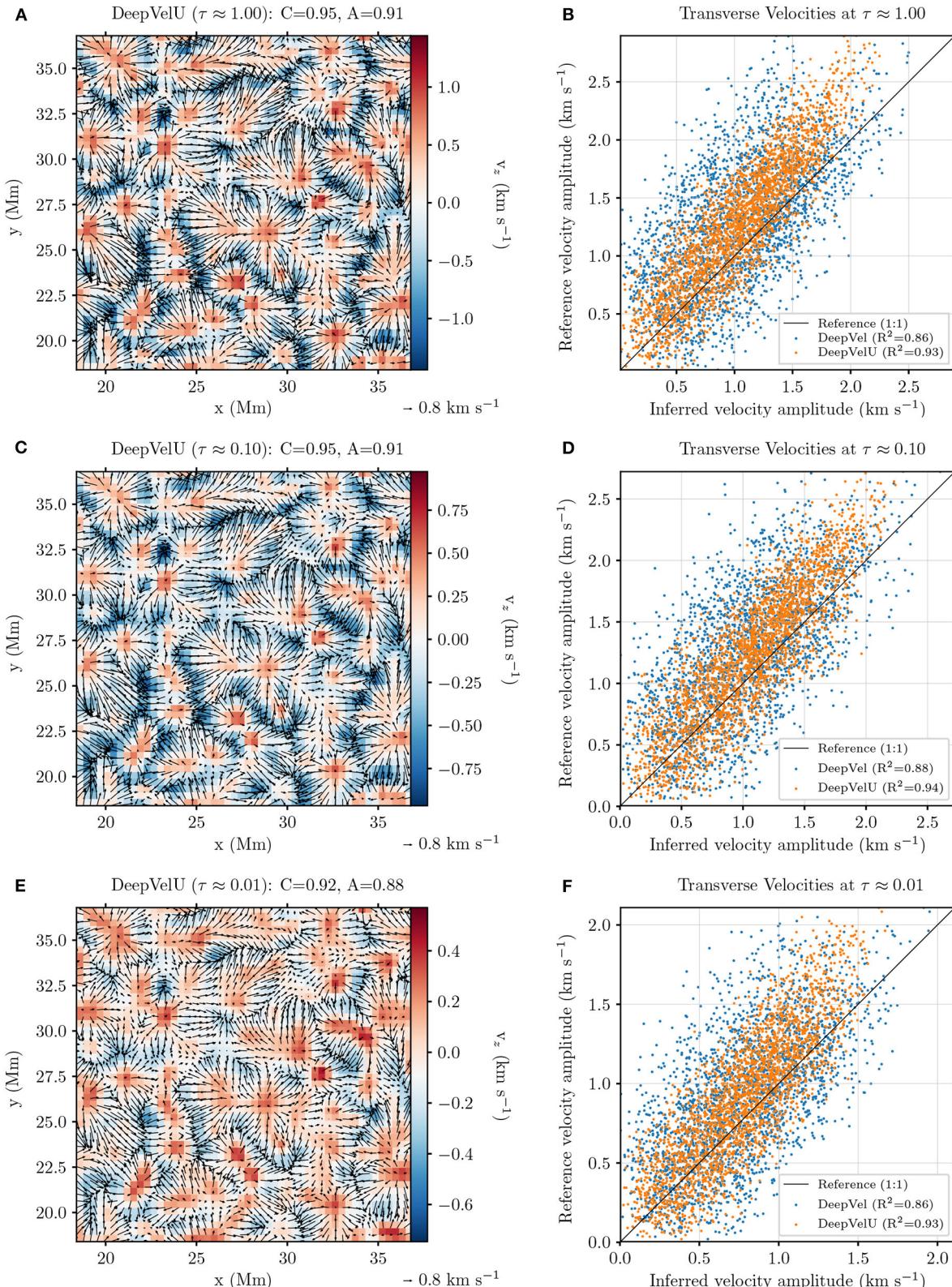
### 3.3. Active Region Test Set

**Figures 3A,C,E** show a subset of velocity field inversions generated by DeepVelU from single-quantity inputs that relate to SDO/HMI level-2 data products i.e., intensitygrams, line-of-sight magnetograms, and Dopplergrams, respectively. Only a sub-field of  $50 \times 50$  pixels<sup>2</sup> is shown for clarity. The position of this patch in the field-of-view was selected to highlight distinct flow structures in the presence of ARs: the center of the sunspot (upper right corner), the flows in the penumbra, the moat-like flows around the latter (close to the diagonal connecting the upper left corner to the lower right corner), and finally, the granulation like in the QS dataset (lower left corner). Scatterplots in **Figure 3** were computed over the sub-field whereas the metrics compiled in **Table 1** were computed over the entire field-of-view and include all tested combinations of inputs. An arbitrary threshold of  $|B_{\text{ref},z}| \geq 100 \text{ G}$  is set to compute metrics that are specific to the AR and the magnetic field network. Regions where  $|B_{\text{ref},z}| < 100 \text{ G}$  are interpreted as QS.

DeepVel and DeepVelU generate very similar velocity fields from consecutive intensitygrams, with DeepVel performing slightly better where  $|B_{\text{ref},z}| < 100 \text{ G}$  and DeepVelU (**Figure 3A**) improving absolute and relative errors slightly where  $|B_{\text{ref},z}| \geq 100 \text{ G}$  (**Table 1**) but underestimating low amplitude velocities (**Figure 3B**). In fact, the performances of the two neural networks are comparable to that of DeepVel with the QS dataset (**Table 1**). For both architectures, the signal for the continuum intensity inside the Sunspot results in less effective flow inversions where  $|B_{\text{ref},z}| \geq 100 \text{ G}$ . More specifically, the mean relative error almost doubles when transitioning from the QS to strong field regions and the metric  $A$  describing the global orientation of the vectors decreases to a value of about  $0.55$  (**Table 1**). For this reason, different combinations of physical inputs were tested to measure their impact on photospheric flows. Although these tests were performed using the DeepVelU architecture, we expect that similar conclusions can be drawn for the DeepVel neural network.

The use of line-of-sight magnetograms as inputs instead of intensitygrams slightly improves the errors where  $|B_{\text{ref},z}| \geq 100 \text{ G}$  (**Table 1**), i.e., where there is the most signal in the input data. Despite weak magnetic fields in the QS, the neural network is capable of distinguishing individual granules and their flow patterns (**Figure 3C**), but the errors are significantly larger with greatly underestimated flow amplitudes (**Figure 3D**) and misaligned vectors (**Table 1**). The overall performance is thus worse due to a larger fraction of the field-of-view being covered by QS (**Table 1**).

Dopplergrams, such as the ones depicted in the background of **Figures 3A,C,E**, are the inputs that best capture the behavior



**FIGURE 1 |** (Left) Patches of 50 by 50 pixels<sup>2</sup> extracted from the 30-timestep-averaged transverse velocity fields  $\vec{v}_{D,t}$  inferred by DeepVelU at optical depths **(A)**  $\tau \approx 1$ , **(C)**  $\tau \approx 0.1$ , and **(E)**  $\tau \approx 0.01$ . The 30-timestep-averaged vertical velocity  $v_{z,ref}(\tau \approx \{1, 0.1, 0.01\})$  computed by the STAGGER simulation and resampled to the SDO/HMI resolution is displayed as background (colorscale). (Right) Scatterplots comparing amplitudes  $|\vec{v}_{D,t}|$  to  $|\vec{v}_{ref,t}|$  at optical depths **(B)**  $\tau \approx 1$ , **(D)**  $\tau \approx 0.1$ , and **(F)**  $\tau \approx 0.01$ . The black line represents the expected solution (i.e., a coefficient of determination  $R^2 = 1$ ).

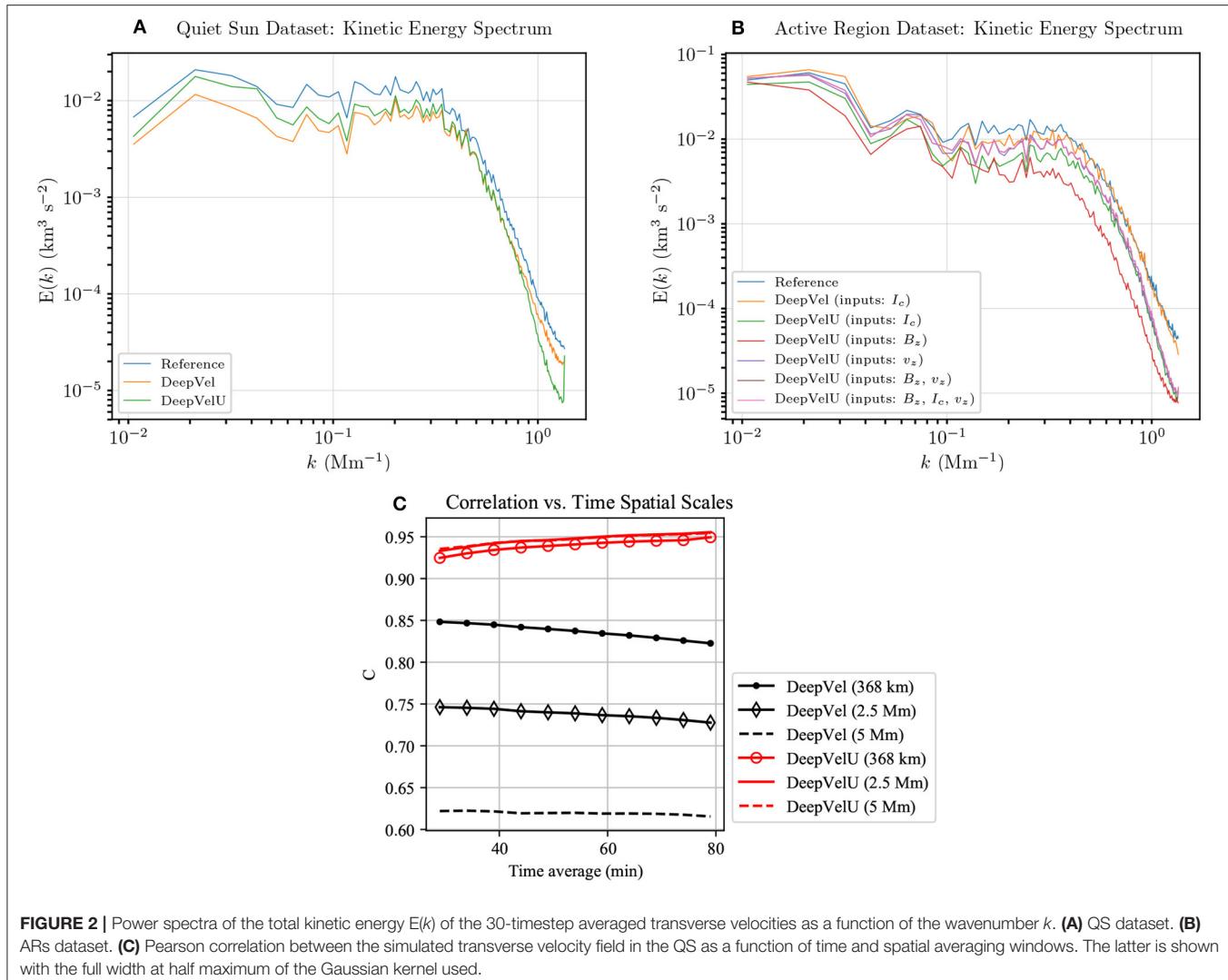
**TABLE 1 |** Comparison between the 30-timestep-averaged  $\vec{v}_{D,t}$  and  $\vec{v}_{ref,t}$  for the QS dataset and ARs dataset.

QS dataset algorithm	Quantity (units)	$\tau$	RMSE	$E_{abs}$	$E_{rel}$ (%)	C $\in [0, 1]$	A $\in [-1, 1]$	S (no units)
DeepVel	$\vec{v}_t$ (km s $^{-1}$ )	1.00 0.10 0.01	0.777 0.697 0.553	0.694 0.620 0.491	70.6 65.4 70.3	0.841 0.867 0.852	0.786 0.810 0.791	0.811 0.793 0.729
	$\vec{v}_t$ (10 $^{-3}$ s $^{-1}$ )	1.00 0.01	0.735 0.931	0.577 0.733	195.7 199.9	0.912 0.884	— —	— —
	$\vec{v}_t$ (10 $^{-3}$ s $^{-1}$ )	1.00 0.01	0.783 0.930	0.618 0.726	261.3 225.6	0.892 0.851	— —	— —
	$\vec{v}$ (km s $^{-1}$ )	1.00 0.10 0.01	0.501 0.473 0.417	0.442 0.416 0.367	43.7 43.8 52.3	0.947 0.945 0.925	0.914 0.907 0.876	0.796 0.805 0.752
	$\vec{v}_t \cdot \vec{v}_t$ (10 $^{-3}$ s $^{-1}$ )	1.00 0.01	0.618 0.864	0.490 0.682	172.4 179.9	0.941 0.913	— —	— —
	$\vec{v}_t \cdot \vec{v}_t$ (10 $^{-3}$ s $^{-1}$ )	1.00 0.01	0.662 0.852	0.525 0.674	244.5 195.6	0.920 0.913	— —	— —
	$(\vec{v} \times \vec{v})_z$ (10 $^{-3}$ s $^{-1}$ )	0.10 0.01	0.796 0.930	0.630 0.726	301.9 225.6	0.871 0.851	— —	— —
	$(\vec{v} \times \vec{v})_z$ (10 $^{-3}$ s $^{-1}$ )	0.10 0.01	0.767 0.930	0.607 0.726	281.6 225.6	0.887 0.851	— —	— —
	$(\vec{v} \times \vec{v})_z$ (10 $^{-3}$ s $^{-1}$ )	0.10 0.01	0.735 0.930	0.577 0.726	195.7 225.6	0.912 0.851	— —	— —
	$(\vec{v} \times \vec{v})_z$ (10 $^{-3}$ s $^{-1}$ )	0.10 0.01	0.783 0.930	0.618 0.726	261.3 225.6	0.892 0.851	— —	— —
ARs dataset algorithm	Quantity (units)	$ B_{ref,z} $ (G)	RMSE	$E_{abs}$	$E_{rel}$ (%)	C $\in [0, 1]$	A $\in [-1, 1]$	S (no units)
DeepVel (inputs: $I_c$ )	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.719 0.701 0.878	0.603 0.592 0.717	70.98 56.02 226.2	0.905 0.912 0.845	0.830 0.855 0.563	1.061 — —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.789 0.787 0.814	0.674 0.680 0.618	62.17 58.20 103.4	0.898 0.904 0.847	0.818 0.844 0.556	0.870 — —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	1.174 0.787 0.814	1.023 0.680 0.618	88.18 58.20 103.4	0.723 0.904 0.847	0.588 0.844 0.556	0.864 — —
DeepVelU (inputs: $B_z$ )	$\vec{v}_t$ (km s $^{-1}$ )	$< 100$ $\geq 100$	1.201 0.848	1.062 0.618	87.76 92.53	0.716 0.834	0.590 0.565	— —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.645 1.201 0.848	0.550 1.023 0.618	55.24 88.18 92.53	0.928 0.723 0.834	0.860 0.588 0.565	0.896 0.864 —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.635 0.645 0.574	0.548 0.645 0.574	50.41 55.24 105.4	0.934 0.928 0.884	0.882 0.860 0.633	— 0.896 —
DeepVelU (inputs: $v_z$ )	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.639 0.635 0.574	0.553 0.548 0.574	54.67 50.41 105.4	0.929 0.934 0.884	0.857 0.882 0.633	0.970 — —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.602 0.574 0.574	0.519 0.564 0.440	52.52 52.46 77.60	0.937 0.929 0.927	0.863 0.874 0.678	0.872 — —
	$\vec{v}_t$ (km s $^{-1}$ )	$\geq 0$ $< 100$ $\geq 100$	0.604 0.574 0.582	0.526 0.564 0.449	50.28 52.46 75.72	0.938 0.929 0.925	0.882 0.874 0.661	— — —
DeepVelU (inputs: $B_z, I_c, v_z$ )	$\vec{v}_t$ (km s $^{-1}$ )	$< 100$ $\geq 100$	0.604 0.582	0.526 0.449	50.28 75.72	0.938 0.925	0.882 0.661	— —

RMSE is the root mean squared error,  $E_{abs}$  is the mean absolute error (Equation 1),  $E_{rel}$  is the mean relative error (Equation 2), C is the Pearson correlation coefficient, A is the spatially-averaged normalized dot product (Equation 3) and S  $\equiv \int |S_{D,z}|dM / \int |S_{ref,z}|dM$  is the ratio between the integrals of the unsigned shear components of the Poynting fluxes  $S_z(\vec{v}_{D,t})$  and  $S_z(\vec{v}_{ref,t})$  (Equation 5) over the full field-of-view M.

of flows in both QS and ARs (Table 1 and Figure 3F). The granulation pattern from intensitygrams is clearly outlined in Dopplergrams by the cold sinking plasma in the intergranular lanes and the hot rising plasma at the center of granules. Meanwhile convective motions are inhibited inside Sunspots. However, flows in the penumbra are predicted by DeepVelU

to be almost purely radial with respect to the center of the Sunspot (Figure 3E), whereas the simulation penumbral flows resemble more closely those seen in Figures 3A,C. Although the metric A was slightly improved in strong field regions, vector orientations remain much less accurately reproduced than in the QS (Table 1).



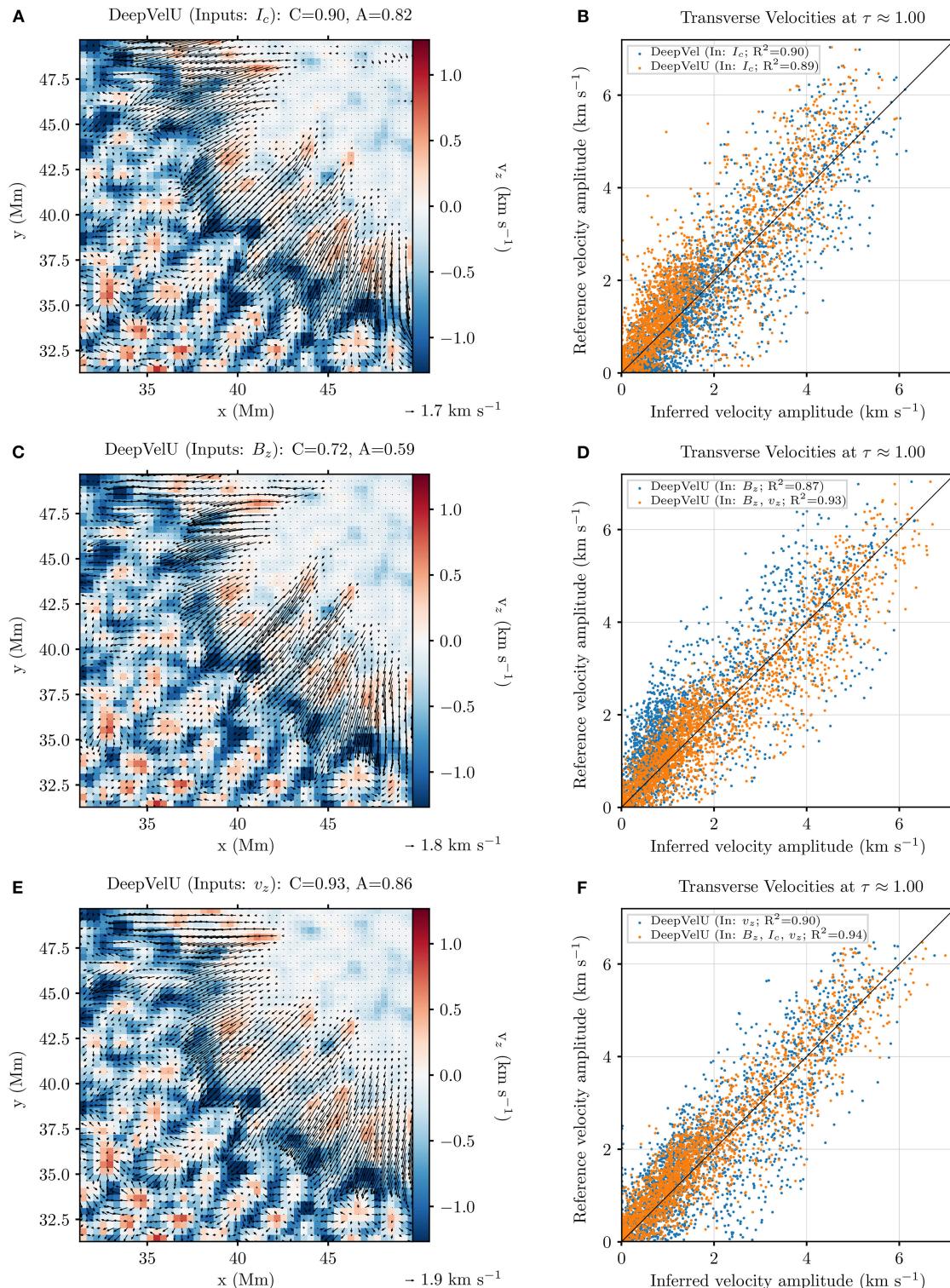
**FIGURE 2 |** Power spectra of the total kinetic energy  $E(k)$  of the 30-timestep averaged transverse velocities as a function of the wavenumber  $k$ . **(A)** QS dataset. **(B)** ARs dataset. **(C)** Pearson correlation between the simulated transverse velocity field in the QS as a function of time and spatial averaging windows. The latter is shown with the full width at half maximum of the Gaussian kernel used.

Combining  $B_{\text{ref},z}$  and  $v_{\text{ref},z}$ , which yielded the best reconstructions for ARs and QS, respectively, significantly improves the performance for ARs and the integrated Poynting flux (**Table 1**). Magnetograms provide to Dopplergrams additional signal inside Sunspots. Both quantities are coupled physically to  $\vec{v}_t$  through the magnetic induction equation, which could further explain the increase in performance. The addition of intensitygrams provides more context for the QS and improves the metrics where  $|B_{\text{ref},z}| < 100$  G, with very little to no decrease in performance where  $|B_{\text{ref},z}| \geq 100$  G (**Table 1**).

The use of Dopplergrams or combinations of inputs in DeepVelU improves the response in the power spectrum of the kinetic energy for spatial scales larger than  $k^{-1} \approx 3$  Mm (**Figure 2B**), which could again be related to the spatial scales probed in the U-net's bottleneck. Although the power is generally underestimated, its variations as a function of  $k$  are matched more consistently by DeepVelU, with DeepVel generating more signal than the simulation at supergranular scales.

## 4. CONCLUSION

We trained DeepVelU, a U-net-inspired architecture for the DeepVel neural network, using simulations of the QS and ARs and evaluated the method's response as a function of spatial and temporal scales. DeepVelU shows significant improvement over DeepVel for the QS test set. The correlations for the latter falls close to 0.6 at spatial scales of 5 Mm whereas it stays consistently above 0.9 for DeepVelU; increasing at greater time averages and plateauing at spatial scales above 3 Mm. Thus DeepVelU's QS model appears more effective than the other tracking methods tested in Tremblay et al. (2018) against the same dataset, with increased correlations and lower errors being achieved over DeepVel at granular scales and over FLCT at supergranular scales. The results for the ARs dataset are not as conclusive, but may be further improved by training for granulation, penumbra and sunspots separately. This approach could, however, introduce discontinuities at the edges of the different structures.



**FIGURE 3 |** (Left) Patches of 50 by 50 pixels<sup>2</sup> extracted from the 30-timestep-averaged  $\vec{v}_{D,t}(\tau \approx 1)$  inferred by DeepVelU from (A)  $I_c(t = t_i, t_i + \Delta t)$ , (C)  $B_z(t = t_i, t_i + \Delta t)$ , and (E)  $v_z(t = t_i, t_i + \Delta t)$ . The 30-timestep-averaged  $v_{z,ref}$  computed by the MURaM simulation and resampled to the SDO/HMI resolution is displayed as background (colorscale). (Right) Scatterplots comparing  $|\vec{v}_{ref,t}|$  to  $|\vec{v}_{D,t}|$  inferred from a combination of intensitygrams  $I_c(t = t_i, t_i + \Delta t)$ , magnetograms  $B_z(t = t_i, t_i + \Delta t)$  and Dopplergrams  $v_z(t = t_i, t_i + \Delta t)$ . The black line represents the expected solution (i.e.,  $R^2 = 1$ ). (Right) Scatterplots comparing  $|\vec{v}_{ref,t}|$  to  $|\vec{v}_{ref,t}|$  inferred from a combination of (B) consecutive intensitygrams  $I_c(t = \{t_i, t_{i+\Delta t}\})$ , (D) magnetograms  $B_z(t = \{t_i, t_{i+\Delta t}\})$  and Dopplergrams  $v_z(t = \{t_i, t_{i+\Delta t}\})$ , and (F) Dopplergrams, intensitygrams and magnetograms.

Current efforts are meant to be a proof of concept. Limitation of the method include the input image dimensions in each direction which must be a factor of  $2^n$  where  $n$  is the number of downsampling or upsampling layers in the network architecture (here  $n = 3$ ). For example, the dimensions of the sub-images ( $48 \times 48 \text{ px}^2$ ) presented to the network during training limited the number of downsampling/upsampling layers in the architecture to  $n = 3$ , corresponding to a spatial of about 3 Mm. This could explain the plateauing of the correlation above this spatial scale. In this context, development is underway for a deeper version of DeepVelU (i.e., with more downsampling/upsampling layers) that will be trained on (almost) full field-of-view images ( $256 \times 256 \text{ px}^2$  or  $94.192 \times 94.192 \text{ Mm}^2$ ) of the QS dataset and that will also double (resp. halve) the number of filters after each downsampling (resp. upsampling) operation. In addition, the simulations do not model actual supergranulation which is known to advect granules over spatial scales greater than 3 Mm. Therefore it will be worth exploring further inferences on actual photospheric observations. For example we plan to compare the inferred supergranular (QS) and moat flow patterns (ARs) with those of Attie et al. (2018) which uses a new implementation of the “Balltracking” method that is more accurate than the one of Potts et al. (2004) and that had not been tested by Tremblay et al. (2018). Similarly, the MURaM simulation includes a deep seated flow system with velocities in the 200–500 m/s range which extends about 10 Mm past the sunspot boundary, which seems in line with many observations of moat flows (Rempel, 2015). Using the data of both experiments (observations and simulation), we will compare the moat flow patterns revealed by DeepVel, DeepVelU and Balltracking.

The inclusion of Dopplergrams and line-of-sight magnetograms as inputs in the neural network architecture, both of which provide more signal than intensitygrams inside sunspots and are tied to the transverse plasma motions through the magnetic induction equation, have improved reconstructions in the context of ARs and response at supergranular scales. Evaluating the impact of additional inputs such as the transverse magnetic field vector  $\vec{B}_t$  which also appears in the magnetic induction equation or its strength  $|\vec{B}_t|$  which is not subjected to ambiguities is left as future work.

Despite the changes in architecture and the success of the neural network at capturing the spatial distribution of flows, velocity amplitudes are generally underestimated in the QS and overestimated in ARs by both DeepVel and DeepVelU. Future efforts will be dedicated to improving the inference of amplitudes.

All versions of the neural networks were trained to generate synthetic data that is consistent with a given simulation and the SDO/HMI cadence (if the velocity is multiplied by a factor of 45/60 for the QS dataset) and spatial resolution near disk center. SDO/HMI level-2 products can thus be used as input, however one should first assess if the physics and preprocessing

of the training set is consistent with the observations presented as input. For instance, SDO/HMI Dopplergrams measure the superposition of the line-of-sight components of the satellite motion, meridional flows, differential rotation, p-mode oscillations, and plasma motions. Simulation Dopplergrams only feature the latter two. Additional preprocessing steps are thus required to correct SDO/HMI Dopplergrams for the aforementioned effects (e.g., Welsch et al., 2013) or to project the simulation data in the observations space prior to training. The resulting velocity fields may then serve as synthetic observations or first estimates when performing data assimilation in an MHD model of the photosphere, or as boundary conditions driving a simulation. The method may also be used to estimate and evolve a Poynting flux vector that is representative of a given epoch of the Sun.

The velocity vector that the neural networks are trying to recover is the same vector as physics-based velocity inversion methods, i.e.,  $\vec{v}_t$  such that the magnetic induction equation is satisfied. Furthermore, DeepVelU best performed when using Dopplergrams and magnetograms as input, with both quantities appearing in the magnetic induction equation alongside  $\vec{v}$ . The training process may be revisited in the future to incorporate more effectively the physics when estimating the plasma motions, e.g., through the loss function or a physics-informed network (Raissi et al., 2019).

## DATA AVAILABILITY STATEMENT

The QS and ARs weights and biases as well as the DeepVel and DeepVelU outputs generated and analyzed for this study can be found in the DeepVelU\_Frontiers repository ([https://github.com/tremblaybenoit/DeepVelU\\_Frontiers](https://github.com/tremblaybenoit/DeepVelU_Frontiers)).

## AUTHOR CONTRIBUTIONS

BT generated the QS and ARs datasets from simulation data. BT also trained and tested all versions of the DeepVel and DeepVelU neural networks. RA acquired the ARs dataset prior to BT downsampling the data to the SDO/HMI spatial resolution. RA also provided the analysis of the QS dataset.

## ACKNOWLEDGMENTS

The authors would like to thank Prof. Laurence Perreault-Levasseur (MILA; Université de Montréal) for her advice regarding the neural network architecture. The authors would also like to thank Prof. Maria Kazachenko for her insight and her advice during the writing of this manuscript. The MURaM simulation data was kindly provided by Dr. Matthias Rempel (NCAR).

## REFERENCES

- Abbott, W. P., and Fisher, G. H. (2010). Improving large-scale convection-zone-to-corona models. *Memorie della Societa Astronomica Italiana* 81:721. Available online at: <https://www.adsabs.harvard.edu/abs/2010MmSAI..81..721A>
- Asensio Ramos, A., Requerey, I. S., and Vitas, N. (2017). DeepVel: deep learning for the estimation of horizontal velocities at the solar surface. *Astron. Astrophys.* 604:A11. doi: 10.1051/0004-6361/201730783
- Attie, R., Innes, D. E., and Potts, H. E. (2009). Evidence of photospheric vortex flows at supergranular junctions observed by FG/SOT

- (Hinode). *Astron. Astrophys.* 493, L13–L16. doi: 10.1051/0004-6361:200811258
- Attie, R., Innes, D. E., Solanki, S. K., and Glassmeier, K. H. (2016). Relationship between supergranulation flows, magnetic cancellation and network flares. *Astron. Astrophys.* 596:A15. doi: 10.1051/0004-6361/201527798
- Attie, R., Kirk, M. S., Thompson, B. J., Muglach, K., and Norton, A. A. (2018). Precursors of magnetic flux emergence in the moat flows of active region ar12673. *Space Weather* 16, 1143–1155. doi: 10.1029/2018SW001939
- Cheung, M. C. M., Schüssler, M., and Moreno-Insertis, F. (2007). The origin of the reversed granulation in the solar photosphere. *Astron. Astrophys.* 461, 1163–1171. doi: 10.1051/0004-6361:20066390
- Fisher, G. H., Abbott, W. P., Bercik, D. J., Kazachenko, M. D., Lynch, B. J., Welsch, B. T., et al. (2015). The coronal global evolutionary model: using HMI vector magnetogram and doppler data to model the buildup of free magnetic energy in the solar corona. *Space Weather* 13:369. doi: 10.1002/2015SW001191
- Fisher, G. H., and Welsch, B. T. (2008). “FLCT: a fast, efficient method for performing local correlation tracking,” in *Subsurface and Atmospheric Influences on Solar Activity, Volume 383 of Astronomical Society of the Pacific Conference Series*, eds R. Howe, R. W. Komm, K. S. Balasubramanian, and G. J. D. Petrie (San Francisco, CA), 373.
- Hagenaar, M. (2005). “Photospheric surface flows and sunspot moats,” in *Large-Scale Structures and Their Role in Solar Activity, Volume 346 of Astronomical Society of the Pacific Conference Series*, eds K. Sankarasubramanian, M. Penn, and A. Pevtsov (San Francisco, CA), 41.
- Hoeksema, J. T., Liu, Y., Hayashi, K., Sun, X., Schou, J., Couvidat, S., et al. (2014). The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: overview and performance. *Sol. Phys.* 289, 3483–3530. doi: 10.1007/s11207-014-0516-8
- Il'iarionov, E. A., and Tlatov, A. G. (2018). Segmentation of coronal holes in solar disc images with a convolutional neural network. *Monthly Notices R. Astron. Soc.* 481, 5014–5021. doi: 10.1093/mnras/sty2628
- Ioffe, S., and Szegedy, C. (2015). “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on International Conference on Machine Learning, Vol. 37* (Lille: JMLR), 448–56. doi: 10.5555/3045118.3045167
- Kazachenko, M. D., Fisher, G. H., and Welsch, B. T. (2014). A comprehensive method of estimating electric fields from vector magnetic field and Doppler measurements. *Astrophys. J.* 795:17. doi: 10.1088/0004-637X/795/1/17
- Kazachenko, M. D., Fisher, G. H., Welsch, B. T., Liu, Y., and Sun, X. (2015). Photospheric electric fields and energy fluxes in the eruptive active region NOAA 1158. *Astrophys. J.* 811:16. doi: 10.1088/0004-637X/811/1/16
- Liu, Y., and Schuck, P. W. (2012). Magnetic energy and helicity in two emerging active regions in the sun. *Astrophys. J.* 761:105. doi: 10.1088/0004-637X/761/2/105
- Longcope, D. W. (2004). Inferring a photospheric velocity field from a sequence of vector magnetograms: the minimum energy fit. *Astrophys. J.* 612, 1181–1192. doi: 10.1086/422579
- November, L. J., and Simon, G. W. (1988). Precise proper-motion measurement of solar granulation. *Astrophys. J.* 333, 427–442. doi: 10.1086/166758
- Potts, H. E., Barrett, R. K., and Diver, D. A. (2004). Balltracking: an highly efficient method for tracking flow fields. *Astron. Astrophys.* 424, 253–262. doi: 10.1051/0004-6361:20035891
- Potts, H. E., and Diver, D. A. (2008). Automatic recognition and characterisation of supergranular cells from photospheric velocity fields. *Sol. Phys.* 248, 263–275. doi: 10.1007/s11207-007-9068-5
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 680–707. doi: 10.1016/j.jcp.2018.10.045
- Rempel, M. (2015). Numerical simulations of sunspot decay: on the penumbra-evershed flow-moat flow connection. *Astrophys. J.* 814:125. doi: 10.1088/0004-637X/814/2/125
- Rempel, M., and Cheung, M. C. M. (2014). Numerical simulations of active region scale flux emergence: from spot formation to decay. *Astrophys. J.* 785:90. doi: 10.1088/0004-637X/785/2/90
- Rieutord, M., and Rincon, F. (2010). The sun's supergranulation. *Living Rev. Solar Phys.* 7, 16–17. doi: 10.12942/lrsp-2010-2
- Rieutord, M., Roudier, T., Ludwig, H.-G., Nordlund, Å., and Stein, R. (2001). Are granules good tracers of solar surface velocity fields? *Astron. Astrophys.* 377, L14–L17. doi: 10.1051/0004-6361:20011160
- Rieutord, M., Roudier, T., Rincon, F., Malherbe, J.-M., Meunier, N., Berger, T., et al. (2010). On the power spectrum of solar surface flows. *Astron. Astrophys.* 512:A4. doi: 10.1051/0004-6361/200913303
- Rieutord, M., Roudier, T., Roques, S., and Ducrot, C. (2007). Tracking granules on the Sun's surface and reconstructing velocity fields. I. The CST algorithm. *Astron. Astrophys.* 471, 687–694. doi: 10.1051/0004-6361:20066491
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *arXiv preprints arXiv:1505.04597*. doi: 10.1007/978-3-319-24574-4\_28
- Schou, J., Scherrer, P. H., Bush, R. I., Wachter, R., Couvidat, S., Raballo-Soares, M. C., et al. (2012). Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the solar dynamics observatory (SDO). *Sol. Phys.* 275, 229–259. doi: 10.1007/s11207-011-9842-2
- Schuck, P. W. (2005). Local correlation tracking and the magnetic induction equation. *Astrophys. J. Lett.* 632, L53–L56. doi: 10.1086/497633
- Schuck, P. W. (2006). Tracking magnetic footpoints with the magnetic induction equation. *Astrophys. J.* 646, 1358–1391. doi: 10.1086/505015
- Schuck, P. W. (2008). Tracking vector magnetograms with the magnetic induction equation. *Astrophys. J.* 683, 1134–1152. doi: 10.1086/589434
- Sheeley, N. R., Jr. (1969). The evolution of the photospheric network. *Sol. Phys.* 9, 347–357. doi: 10.1007/BF02391657
- Stein, R. F. (2012). Solar surface magneto-convection. *Living Rev. Sol. Phys.* 9:5. doi: 10.12942/lrsp-2012-4
- Stein, R. F., and Nordlund, Å. (2012). On the formation of active regions. *Astrophys. J. Lett.* 753:L13. doi: 10.1088/2041-8205/753/1/L13
- Tremblay, B., Roudier, T., Rieutord, M., and Vincent, A. (2018). Reconstruction of horizontal plasma motions at the photosphere from intensitygrams: a comparison between DeepVel, LCT, FLCT, and CST. *Sol. Phys.* 293:57. doi: 10.1007/s11207-018-1276-7
- Wachter, R., Schou, J., Raballo-Soares, M. C., Miles, J. W., Duvall, T. L., and Bush, R. I. (2012). Image quality of the helioseismic and magnetic imager (HMI) onboard the solar dynamics observatory (SDO). *Sol. Phys.* 275, 261–284. doi: 10.1007/s11207-011-9709-6
- Welsch, B. T. (2015). The photospheric Poynting flux and coronal heating. *Publ. Astron. Soc. Japan* 67:18. doi: 10.1093/pasj/psu151
- Welsch, B. T., Fisher, G. H., and Sun, X. (2013). A magnetic calibration of photospheric Doppler velocities. *Astrophys. J.* 765:98. doi: 10.1088/0004-637X/765/2/98

**Conflict of Interest:** 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.

Copyright © 2020 Tremblay and Attie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Classification of Magnetosheath Jets Using Neural Networks and High Resolution OMNI (HRO) Data

Savvas Raptis<sup>1\*</sup>, Sigiava Aminalragia-Giamini<sup>2</sup>, Tomas Karlsson<sup>1</sup> and Martin Lindberg<sup>1</sup>

<sup>1</sup> Division of Space and Plasma Physics, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, <sup>2</sup> Space Applications & Research Consultancy (SPARC), Athens, Greece

## OPEN ACCESS

**Edited by:**

Enrico Camporeale,  
University of Colorado Boulder,  
United States

**Reviewed by:**

Hui Li,  
National Space Science Center (CAS),  
China  
Alexei V. Dmitriev,  
Lomonosov Moscow State University,  
Russia

**\*Correspondence:**

Savvas Raptis  
savvra@kth.se

**Specialty section:**

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 05 March 2020

**Accepted:** 01 May 2020

**Published:** 05 June 2020

**Citation:**

Raptis S, Aminalragia-Giamini S, Karlsson T and Lindberg M (2020)  
Classification of Magnetosheath Jets  
Using Neural Networks and High  
Resolution OMNI (HRO) Data.  
*Front. Astron. Space Sci.* 7:24.  
doi: 10.3389/fspas.2020.00024

Magnetosheath jets are transient, localized dynamic pressure enhancements found downstream of the Earth's bow shock in the magnetosheath region. Using a pre-existing database of magnetosheath jets we train a neural network to distinguish between jets found downstream of a quasi-parallel bow shock ( $\theta_{Bn} < 45^\circ$ ) and jets downstream of a quasi-perpendicular bow shock ( $\theta_{Bn} > 45^\circ$ ). The initial database was compiled using MMS measurements in the magnetosheath (downstream) to identify and classify them as "quasi-parallel" or "quasi-perpendicular," while the neural network uses only solar wind (upstream) measurements from the OMNIweb database. To evaluate the results, a comparison with three physics-based modeling approaches is done. It is shown that neural networks are systematically outperforming the other methods by achieving a  $\sim 93\%$  agreement with the initial dataset, while the rest of the methods achieve around 80%. The better performance of the neural networks likely is due to the fact that they use information from more solar wind quantities than the physics-based models. As a result, even in the absence of certain upstream properties, such as the IMF direction, they are capable of accurately determining the jet class.

**Keywords:** magnetosheath jets, neural networks, solar wind, machine learning, bow shock

## 1. INTRODUCTION

### 1.1. Magnetosheath Jets

The magnetosphere, surrounding the Earth, offers protection from plasma flows originating from the Sun traveling at supersonic speeds. Initially, the solar wind particles interact with the Earth's bow shock and are decelerated into subsonic velocities, moving into the magnetosheath region. The interaction between the solar wind and the Earth's bow shock can in principle be modeled through the Rankine–Hugoniot relations, assuming an 1D, time stationary shock (Baumjohann and Treumann, 2012). However, there are phenomena too complex to be precisely described by the current theoretical framework. This complexity arises mainly from the geometry of the bow shock and the rapid changes in the Interplanetary Magnetic Field (IMF). A phenomenon that is generated in the interaction of the solar wind with the bow shock is the so called "magnetosheath jet." These jets are usually described as localized enhancements of dynamic pressure in the magnetosheath plasma and are attributed to a velocity or a density increase or in most cases an increase of both (e.g., Amata et al., 2011; Archer et al., 2012; Plaschke et al., 2018).

For magnetosheath jets, several terms and definitions are used in the literature (Plaschke et al., 2018). In this work, we use the term "magnetosheath jet" or simply "jet" to describe an enhancement of the dynamic pressure above the background magnetosheath level, using a time-moving average

window of  $\pm 10$  min for the dynamic pressure (e.g., Archer and Horbury, 2013; Gunell et al., 2014; Gutynska et al., 2015; Karlsson et al., 2015; Raptis et al., 2019). When an enhancement higher than two times the background level is observed, a jet is registered to a list of events.

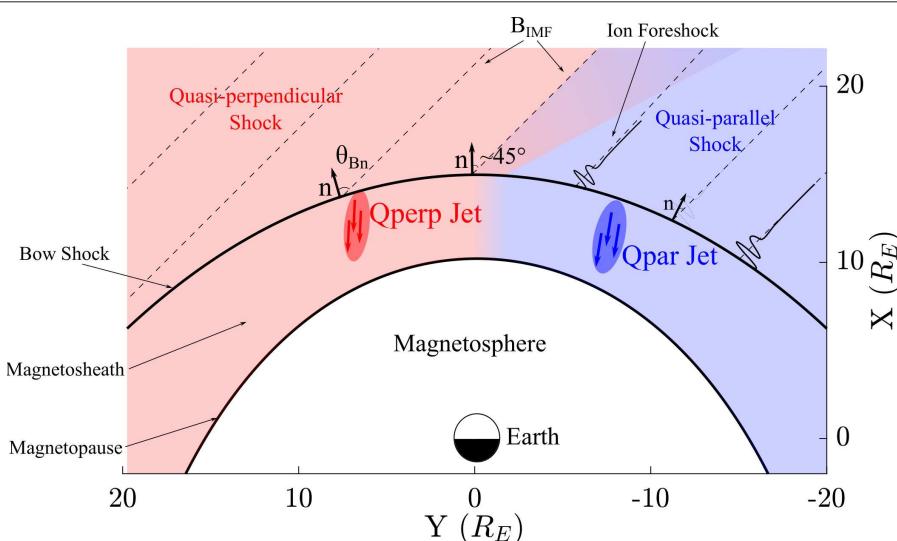
While jets have been observed since 1998 (Němeček et al., 1998), there are still several open questions regarding their origin, their morphology, and their exact generation mechanism (Plaschke et al., 2018). The predominant generation mechanism connects jets to bow shock ripples found at the quasi-parallel bow shock (Hietala et al., 2009; Hietala and Plaschke, 2013). Other phenomena that are possibly connected to jet generation may be the so called SLAMS (Short Large Amplitude Structure) that are foreshock phenomena characterized by very large magnetic field amplitudes and plasma density enhancements (Schwartz et al., 1992). It has been hypothesized that SLAMS can pass through the bow shock ripples and contribute to a density enhancement that would result in an overall increase of the dynamic pressure (Karlsson et al., 2015).

Jets are of great interest for the field of space physics and space weather. It has been suggested that they are connected to various phenomena such as the radiation belts (Turner et al., 2012; Xiang et al., 2016) and throat aurora (Han et al., 2017). Recently it has been shown that by interacting with the magnetopause, jets can trigger magnetopause reconnection (Hietala et al., 2018), which may excite surface eigenmodes (Archer et al., 2019) or even contribute to direct plasma penetration through the magnetopause (Karlsson et al., 2012). Furthermore, they appear to be occurring in other planets of our solar system and in astrophysical shocks (Giacalone and Jokipii, 2007; Plaschke et al., 2018).

An important factor that creates an intrinsic classification to shock transitions and therefore to both the magnetosheath region and the jets, is the angle ( $\theta_{Bn}$ ) between the bow shock normal vector ( $\hat{n}$ ) and the IMF ( $\mathbf{B}$ ), as depicted in **Figure 1**. Due to the differences in the bow shock formation and in particle dynamics explained below, quasi-perpendicular (Qperp) shocks ( $\theta_{Bn} > 45^\circ$ ) exhibit a sharp transition between the upstream flow and the downstream plasma, followed by a less turbulent magnetosheath region (Fuselier, 2013; Wilson, 2016). On the other hand, for quasi-parallel (Qpar) shocks ( $\theta_{Bn} < 45^\circ$ ), the transition is harder to define and the downstream plasma is irregular and strongly turbulent. The source of the different properties of each region is the dynamic behavior of solar wind particles going through the shock transition. In the case of the Qpar shock, reflected ions can travel far upstream, interact with the incoming solar wind flow and cause a number of instabilities leading to wave growth. This, in turn, creates a foreshock region which is absent in the case of Qperp shocks where the reflected particles, due to their gyration around the magnetic field, are quickly returned back to the shock and hence do not travel as far back upstream. This results in a less turbulent environment both upstream and downstream of the Qperp bow shock (Schwartz and Burgess, 1991; Balogh and Treumann, 2013).

For the generation of **Figure 1**, the bow shock and magnetopause model by Chao et al. (2002) are used. The parameters used are  $B_z = -0.22$  (nT),  $P_{dyn} = 2.15$  (nPa),  $M_{ms} = 6.09$  and  $\beta = 2.20$ . These values correspond to the average conditions of the solar wind for the periods that a Qpar or a Qperp jet was found.

Classifying jets into different categories is vital to investigate the possibility of different generation mechanisms. As discussed



**FIGURE 1 |** Sketch of the bow shock and its different configuration in the Earth's environment. An Interplanetary Magnetic Field (IMF) with an angle, approximately  $45^\circ$  with the normal at the nose of the bow shock is assumed. As a result, a quasi-perpendicular ( $\theta_{Bn} > 45^\circ$ ) shock takes place on the left part of the image and a quasi-parallel ( $\theta_{Bn} < 45^\circ$ ) on the right. The instabilities caused by the reflected ions in the Qpar case create the so called ion foreshock which changes drastically the properties between the Qpar shock and the Qperp one.

**TABLE 1 |** Solar wind quantities used as input to the neural network.

Name	Unit
Alfven Mach Number ( $\mathcal{M}_A$ )	–
Magnetosonic Mach Number ( $\mathcal{M}_{ms}$ )	–
Absolute Electric Field ( $ \mathbf{E} $ )	[mV/m]
Beta Plasma Parameter ( $\beta$ )	–
Kinetic Energy density ( $E_{kin}$ )	[nJm $^{-3}$ ]
Proton Temperature ( $T$ )	[K]
Proton Density ( $n$ )	[cm $^{-3}$ ]
Proton Absolute Velocity ( $ \mathbf{V} $ )	[km/s]
Absolute Magnetic Field ( $ \mathbf{B} $ )	[nT]
Magnetic Field X-component ( $B_x$ )	[nT]
Magnetic Field Y-component ( $B_y$ )	[nT]
Magnetic Field Z-component ( $B_z$ )	[nT]

above, there is no consensus regarding the generation of jets. By classifying jets to different bow shock configurations one can investigate both the jet properties and the associated solar wind to determine if any of the suggested mechanism apply to these subset of jets or even indicate new class-specific generation mechanisms.

## 1.2. Neural Networks

Neural Networks (NN) are widely used machine learning (ML) tools that are often employed to perform classification and regression tasks. Neural Networks were first introduced in 1943 (McCulloch and Pitts, 1943) and have been used for such tasks since at least 1958 (Rosenblatt, 1958). The basic principle behind NNs is that when provided with enough data, they are capable of adjusting their internal parameters in an optimal way to perform a specific task related to the data given. By iteratively parameterizing, and thus training, a neural network with error minimization techniques, it has been shown that when presented with unknown data the network is capable of accurately performing its trained task (Bishop, 1995). Lately, multi-layered (deep) neural networks have been used in various applications due to their ability to accurately model complex, and potentially unknown, relationships (Goodfellow et al., 2016; Samarasinghe, 2016).

In the last years, machine learning techniques, including neural networks, have been employed in heliospheric physics and space weather (Camporeale et al., 2018a,b). Their spectrum of application is quite broad, tackling many problems that traditional statistics or physics-based modeling techniques struggle with. Many applications of neural networks focus on predictive tasks such as the forecasting of solar flares (Florios et al., 2018; Jonas et al., 2018), Coronal Mass Ejections (CMEs) (Bobra and Ilonidis, 2016), CME arrival time (Liu et al., 2018), and geomagnetic indices (Bobberg et al., 2000; Wintoft et al., 2017; Chandorkar and Camporeale, 2018). Other applications focus on space environment characterization (Shin et al., 2016; Aminalragia-Giamini et al., 2018), wave recognition (Balasis et al., 2019), and the classification of the solar wind (Camporeale et al., 2017).

In this work, we apply neural networks for a supervised learning classification task. In particular, to classify magnetosheath jets using solar wind measurements and compare the results with physics-based models. The main goal of this study is to classify jets between those originating from quasi-parallel shock transitions and those originating from quasi-perpendicular ones. By doing so, we determine whether machine learning techniques can outperform physics-based models in this task and investigate the potential connection between solar wind conditions and each jet class. Finally, as detailed below, we use parts of a pre-classified dataset of jets (Raptis et al., 2019) which we evaluate by investigating the agreement of each method with the initial classification.

## 2. DATA

### 2.1. OMNIweb—Solar Wind (Upstream) Data

For the upstream conditions, which correspond to the input of the neural network, data from the OMNI database are used, available at [https://omniweb.gsfc.nasa.gov/form/omni\\_min\\_def.html](https://omniweb.gsfc.nasa.gov/form/omni_min_def.html). The OMNI data mainly originate from the ACE spacecraft that resides in the Sun-Earth L1 point (Stone et al., 1998) and are automatically time-shifted to the Earth's bow shock nose. The time-shifted data have an 1-min resolution and take into account the bow shock location and shape (King and Papitashvili, 2005). The solar wind measurements are associated with every jet as later described, resulting in a dataset of equal length to the number of jets. This dataset is then used as input to the neural networks, consisting of the 12 physical quantities shown in Table 1.

### 2.2. MMS—Magnetosheath (Downstream) Data and Jet Database

In this work, we use a list of jets initially presented in Raptis et al. (2019). The dataset is created using *in-situ* measurements from the Magnetospheric Multiscale (MMS) mission during 11/2015–03/2019. For the downstream conditions and for the initial creation of the jet dataset various plasma moment and magnetic field parameters are used. The magnetic field data are taken from the fluxgate magnetometer (FGM) (Russell et al., 2016) and ion data are taken from the fast plasma investigation (FPI) (Pollock et al., 2016). Finally, the position of MMS during each jet is registered in GSE coordinates, using as unit the Earth radius ( $R_E = 6,371\text{ km}$ ). This dataset provides a list of well-characterized and pre-classified jets. These are used for the training and the evaluation of the NN system, where the class of the jets serves as the desired classification output.

All the jets are required to satisfy a criterion of minimum dynamic pressure compared to the background magnetosheath plasma:

$$P_{dyn} = m_p n_i V_i^2 \geq \langle P_{MSH} \rangle_{20\text{ min}} \quad (1)$$

where the angular brackets indicate an average using a 20 min time moving window.  $m_p$  is the proton mass,  $n_i$  the ion number density, and  $V_i$  the ion velocity.

After each jet has been registered to the database, a classification algorithm is applied to determine its class. The main feature of the initial classification is that it uses *in-situ* MMS measurements to determine whether the jet originated from a quasi-parallel or a quasi-perpendicular bow shock configuration. This methodology was preferred over using solar wind data to calculate  $\theta_{Bn}$  for several reasons. Firstly, to avoid the errors that are generated in time-lagging procedures such as the one taking place in OMNIweb database (Mailyan et al., 2008; Case and Wild, 2012). Furthermore, due to the 1-min resolution of the database, short time scale variations of the IMF are consequently undetectable. Finally, the jets are detected in the whole magnetosheath region. As a result, a time-shift on the

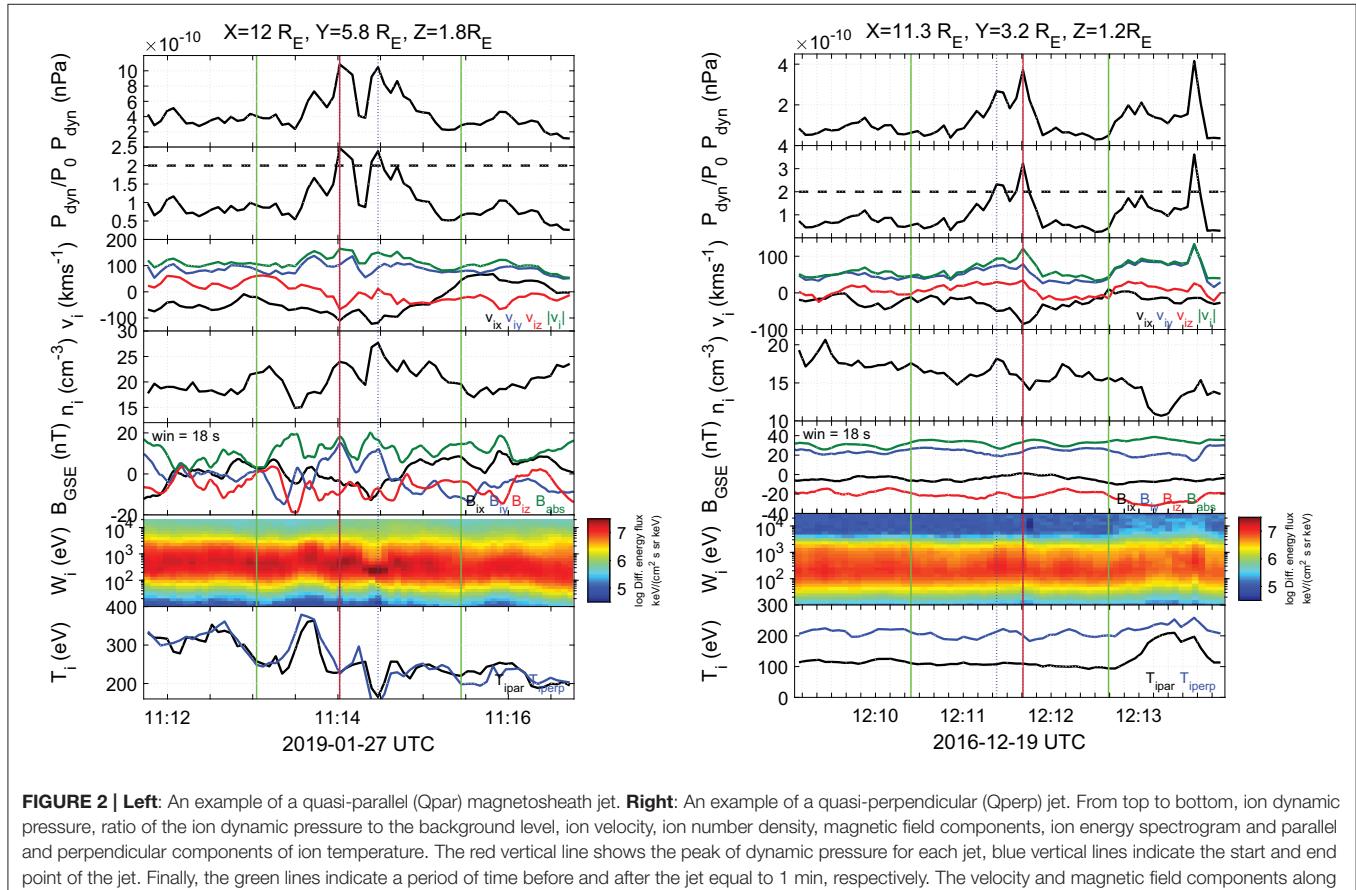
associated solar wind values is required for every jet in order to take into account the time it took for every jet to travel inside the magnetosheath. This procedure itself is difficult to be accurately implemented and it would further increase the uncertainty of the method.

The initial dataset, therefore, relies on properties found in the magnetosheath plasma regions. In particular, the algorithm uses thresholds on ion temperature anisotropy that is found to be lower in Qpar plasma than in Qperp (Anderson et al., 1994; Fuselier et al., 1994). It also takes advantage of the fact that the magnetic field's standard deviation is observed to be higher in the Qpar plasma than in the Qperp (Formisano and Hedgecock, 1973; Luhmann et al., 1986). Finally, the main difference between the Qpar and Qperp plasma regions is the high energy ion population in the ion foreshock which only exists in the Qpar bow shock (Gosling et al., 1978; Fuselier, 2013). As a result, *in-situ* measurements of temperature anisotropy, magnetic field standard deviation, and high energy ion flux were used. A summary of the basic characteristics of each class is shown in **Table 2**. From these classes, the only ones used in this work are the Qpar ( $N = 860$ ) and Qperp ( $N = 211$ ) jets.

In **Figure 2**, an example of MMS measurements for a quasi-parallel and a quasi-perpendicular jet is shown.

**TABLE 2 |** Main properties of the classes of magnetosheath jets.

Jet's class	Characteristics
Qpar	High energy ion flux, low temperature anisotropy, high magnetic field standard deviation
Qperp	Low energy ion flux, high temperature anisotropy, low magnetic field standard deviation
Boundary	Change between Qpar properties to Qperp or Vice Versa
Encapsulated	Change from Qperp properties to Qpar and back to Qperp



### 3. METHODOLOGY

#### 3.1. Input Determination

For both the physics-based modeling and the neural networks, it is necessary to provide an input that corresponds to the same time intervals in order to have a consistent comparison. The choice of the input is nontrivial, since, as discussed in the previous section, the availability of measurements and their association to jets contains multiple errors and uncertainties. Several possible inputs were examined, by either taking average or maximum values of the conditions found within a 5, 10, or 15 min period of the jet. It was found that taking average solar wind conditions starting from 5 min before the jet up to the jet observation time provided the highest agreement to the initial classification in all presented methods. As a result, the solar wind measurements ( $X_u$ ) used for each jet are defined as:

$$X_u = \frac{1}{6} \sum_{i=0}^5 (X_{t_{\text{jet}} - i\Delta t}) \quad (2)$$

where,  $t_{\text{jet}}$  is the time the jet was observed by MMS,  $\Delta t$  is equal to 60 s and subscript  $u$  refers to  $X$  being an upstream quantity.

It should be noted that while the input described in Equation (2) provided the highest agreement with the initial database, it still has its limitations. Specifically, this input choice means that jets found very far away from the bow shock nose or with extremely high or low velocities will potentially not be characterized correctly.

#### 3.2. Evaluating Jet Class With Physics-Based Modeling

In order to provide a baseline to compare the results from the neural networks, we use three different physics-based models to estimate the  $\theta_{Bn}$  angle and distinguish between Qpar and Qperp magnetosheath jets.

##### 3.2.1. Cone Angle Approximation

A simple approach to estimate  $\theta_{Bn}$  is through the cone angle:

$$\theta_{\text{cone}} = \arccos \left( \frac{|\mathbf{B}_{u,x}|}{|\mathbf{B}_u|} \right) \quad (3)$$

where,  $\mathbf{B}_{u,x}$  is the  $x$  component of the upstream magnetic field and  $\mathbf{B}_u$  is the IMF vector.  $\theta_{\text{cone}}$  is identical to  $\theta_{Bn}$  at the subsolar point of the bow shock.

By calculating the cone angle, we classify the available jets. For  $\theta_{\text{cone}} < 45$  a jet is classified as Qpar while for  $\theta_{\text{cone}} > 45$  it is classified as Qperp. This method should in principle work for the majority of the jets that are found close to the subsolar point. However, the number of the jets that are in very close proximity to the subsolar point ( $|Y, Z_{\text{GSE}}| < 2R_E$ ) is quite small (Qpar: 151/ Qperp: 108). As a result, we expect this method to perform poorly for jets found close to the flanks of the magnetosheath.

##### 3.2.2. Coplanarity Method

Another set of methods used is the so called coplanarity methods. There is a variety of methods based on the coplanarity theorem

(e.g., Paschmann and Daly, 1998). In our case, the simplest version of magnetic field coplanarity method provides the highest agreement with the initial dataset and is the one shown in this work.

Starting from Rankine–Hugoniot relations we can derive the normal vector of the bow shock as:

$$\hat{\mathbf{n}} = \pm \frac{(\mathbf{B}_d \times \mathbf{B}_u) \times \Delta \mathbf{B}}{|(\mathbf{B}_d \times \mathbf{B}_u) \times \Delta \mathbf{B}|} \quad (4)$$

In our case, the upstream (IMF) magnetic field was taken as the average value from 5 min before the observation of the jet to the time the jet was observed (Equation 2). On the other hand, the downstream (magnetosheath) magnetic field was taken as the average value of  $\pm 2.5$  min before and after the jet measurement by MMS.

This approximation should in principle be less accurate for jets found very far away from the bow shock since the jump conditions refer to points close to the shock. Furthermore, jets found at the flanks are also prone to errors since the upstream solar wind measurements are time-lagged to the bow shock nose and therefore characterize the subsolar region.

##### 3.2.3. Bow Shock Modeling

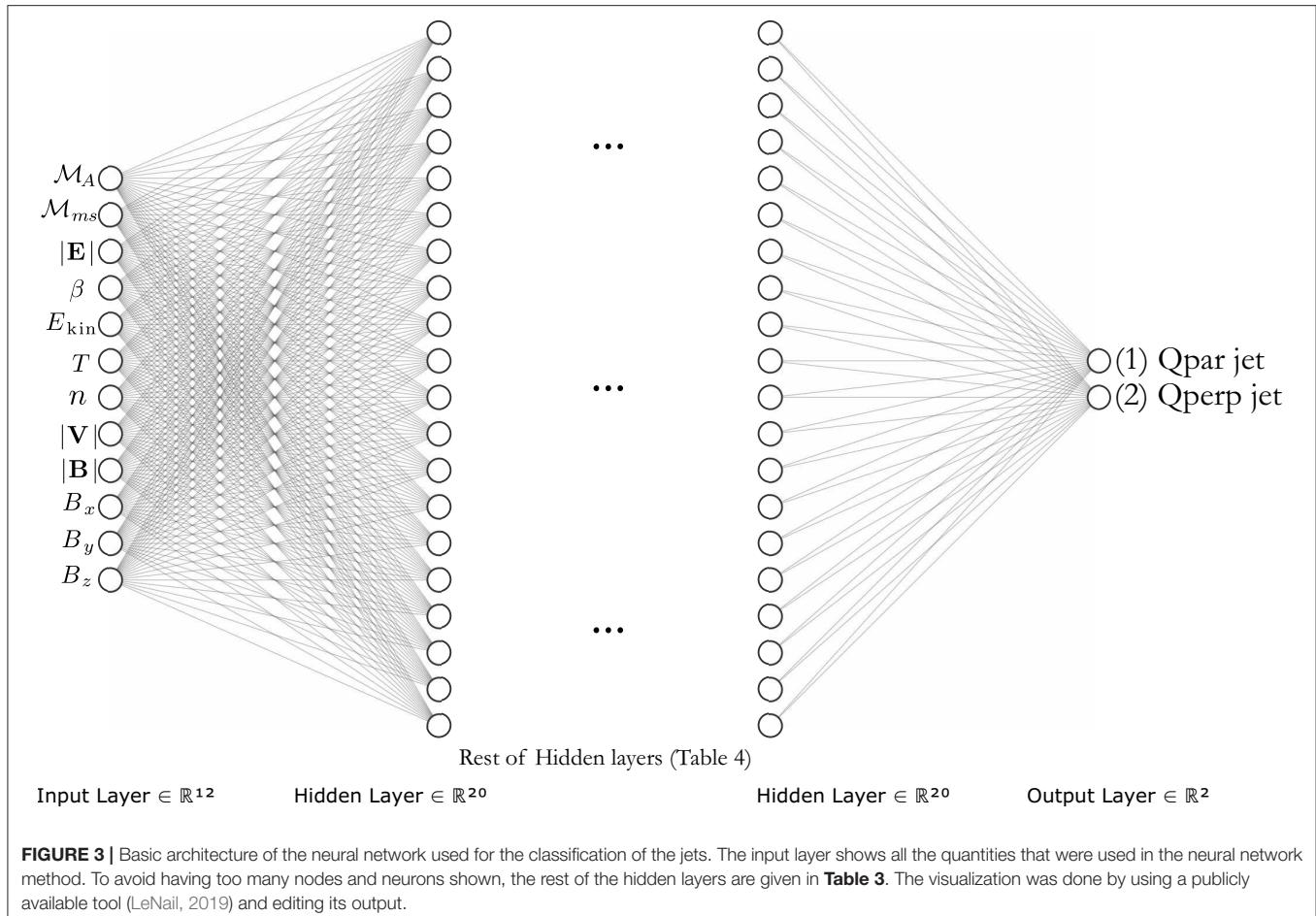
Another method to calculate  $\theta_{Bn}$  requires a model of the bow shock and an approximation of the origin of each jet.

Assuming that the jet does not get significantly accelerated or decelerated during its lifetime in the magnetosheath, one can use the maximum velocity vector ( $\mathbf{V}$ ) to propagate the jet backwards in time and find its point of origin at the bow shock. For the modeling of the bow shock, the model described by Chao et al. (2002) was used. It should be noted that this procedure is prone to several errors. To begin with, the position of the modeled bow shock may have a significant error compared to the real position (Merka et al., 2003; Turc et al., 2013). Furthermore, the assumption that the velocity is constant may introduce more errors. To derive a realistic bow shock model, we use the average associated solar wind conditions starting from 10 min before the jet up to 5 min after its observation by MMS. After we approximated a point of origin for each jet, the angle between the normal vector of that point and the IMF was calculated.

### 3.3. Evaluating Jet Class With Neural Networks

For the input of the neural networks, several inputs associated to each jet were tested. For every jet 12 solar wind measurements were used (Table 1) and were associated to it (Equation 2). From the initial number of jets (860 Qpar/211 Qperp) we exclude jets that contain corrupted data in any of the input that was used in the neural network (Table 1). As a result, the final dataset consists of 759 Qpar jets and 196 Qperp jets.

The neural network architecture, algorithm, and back-end training procedure were implemented in Python by using TensorFlow library version 2.0.0 (Abadi et al., 2015). One of the main problems of the neural network application is the treatment of the class imbalance. We are dealing with a problem where the majority class (Qpar jets) is roughly  $\sim 80\%$  of the whole dataset. Class imbalance is a non-trivial problem to optimize in



**FIGURE 3 |** Basic architecture of the neural network used for the classification of the jets. The input layer shows all the quantities that were used in the neural network method. To avoid having too many nodes and neurons shown, the rest of the hidden layers are given in **Table 3**. The visualization was done by using a publicly available tool (LeNail, 2019) and editing its output.

machine learning and there is no ideal solution for it (Goodfellow et al., 2016; Brownlee, 2020). In order to tackle this problem we utilized the imbalanced-learning library (Lemaître et al., 2017), and used under-sampling and up-sampling techniques. In particular, we used ClusterCentroid under-sampling and SMOTE up-sampling methods on the majority of the trials (Chawla et al., 2002). It should be noted, that both under-sampling and up-sampling techniques did not increase the average accuracy of the neural network significantly. Another direct way to tackle the class imbalance problem in the training procedure is to implement a weight factor in the update of the weights and biases to compensate for the differences in the classes found in the training samples. As a result, a weighted factor was used for updating the parameters of the network when the minority class (quasi-perpendicular jets) is introduced to the network.

For the optimizer of the gradient descent algorithm, we used the Adam optimizer as implemented in the Keras library (Chollet, 2015) with a smaller learning rate of  $\alpha = 0.0001$  while the rest of the parameters were left at their default status. Since the neural network is tackling a binary classification problem, the error function used for the back-propagation is the binary cross-entropy loss function. In all trials and in the final architecture we also used the batch normalization technique (Ioffe and Szegedy, 2015) and parametric rectified linear units (PReLU) (He et al.,

**TABLE 3 |** Basic architecture of the neural network used for the classification task.

Layer	Neurons
Input	12
Fully connected	20
Batch normalization	—
Fully connected	40
Batch normalization	—
Fully connected	60
Batch normalization	—
Fully connected	40
Batch normalization	—
Fully connected	20
Batch normalization	—
Output	2

2015) in the hidden layers of the network. The final architecture of the network is shown in **Figure 3** and described in **Table 3**.

### 3.4. Validation Method

For the training of a neural network we have chosen to use 80% of the jets, leaving 20% available to test the accuracy of the network.

To ensure that the results are not biased from a specific division of the data into training and test sets, the process is repeated multiple times. At each iteration, the training/testing division is made randomly resulting in different subsets. Using this method, the architecture remains the same, while allowing a reasonable sampling of the, otherwise immense, space of divisions of the data into training and testing. This method is then used to evaluate the stability of the NN classifier. From each iteration, the classifying score can be calculated and the overall mean accuracy, as well as the standard deviation of the results, are direct evidence of the system's performance. In the following section results from this iterative process are shown where 100 iterations have been used.

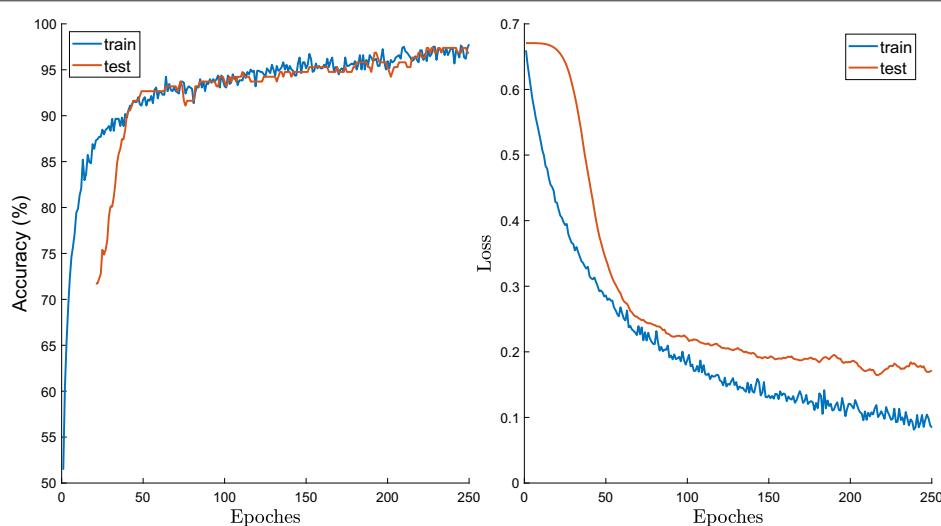
An additional method used for the validation of the classification results is the "leave-one-out" method which produces independent classification results for each and every sample. This method is very useful when dealing with datasets of small size as it can give a good measure of the system's performance in terms of correctly classified and miss-classified cases (e.g., Aminalragia-Giamini et al., 2020). With this procedure, a neural network is trained with all the available data apart from one sample which is "left out." After training, this one sample is input as a test set of size 1 and its class is evaluated by the network. This process iteratively evaluates all samples, setting a different one apart each time, and thus requires the training of a neural network as many times as the total number of jets, here  $N = 955$ . After its completion, this process produces 955 classification results where in each case only the sample tested did not participate in the training and every other available sample was used to produce the results. Finally, a single accuracy score can be calculated for each class which is indicative of the overall performance. The presented results of the "leave-1-out" method are the average results of three independent runs using this validation technique.

## 4. RESULTS

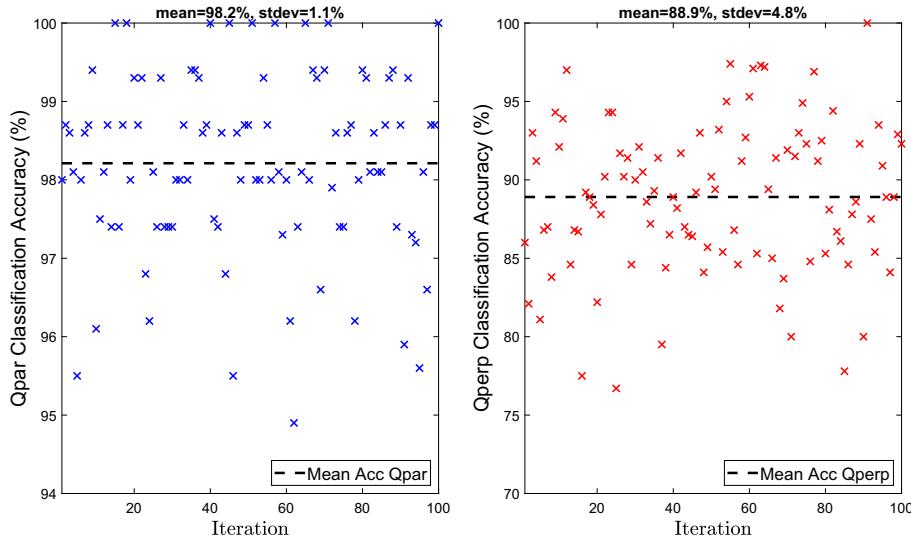
### 4.1. Neural Network

An example plot of the neural network training is shown in **Figure 4**. The test accuracy (left) increases until  $\sim 200$  epoches and as expected the test loss (right) decreases in a respective trend. Beyond the  $\sim 200$  epoch point, no significant changes were observed. After using a validation set to determine the best number of epoches and batch size, we decided to use 250 epoches with a batch size equal to 100 training samples per iteration. Finally, the 80/20 training/testing division of the data results in absolute numbers in 607 Qpar jets and 157 Qperp jets for training and 152 Qpar jets and 39 Qperp jets for testing.

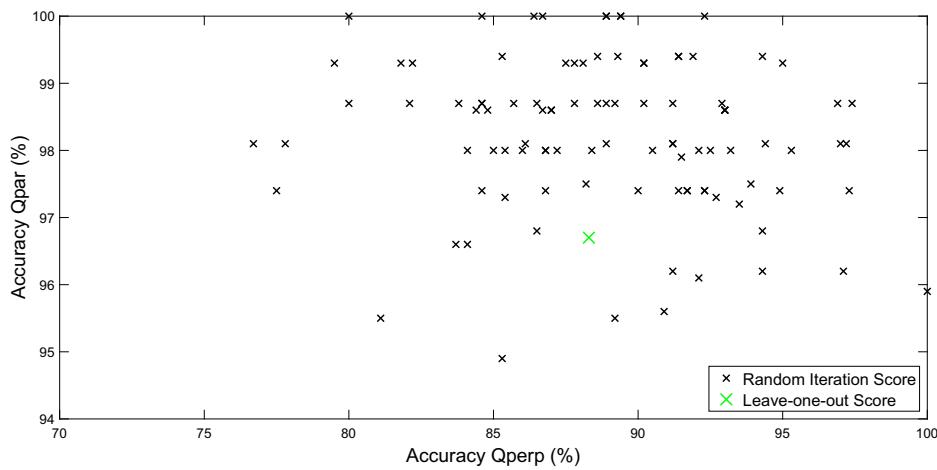
The classification results from 100 iterations with the random training/testing division for neural networks are shown in **Figures 5, 6**. **Figure 5** shows the individual classification scores for Qpar and Qperp jets where it is seen that in both cases high scores with a mean accuracy of  $\sim 98.2$  and  $\sim 88.9\%$  are achieved, respectively. Specifically for the Qpar jets, the scores are tightly clustered in the  $\sim 95 - 100\%$  range showing minimal standard deviation between iterations. On the other hand, the results of Qperp jets have a higher standard deviation, with accuracy scores ranging from  $\sim 75 - 100\%$ . This could be due to the number of jets per class not being balanced, which provides more available information for the neural network regarding the majority class (Qpar). As a result, the testing set used for each iteration contains fewer samples which makes it possible to get worse training and test splits on each iteration. Another possible explanation is that on top of the class imbalance, due to the way Qpar and Qperp shocks form, most of Qperp jets occur much closer to the subsolar region compared to Qpar jets. As discussed previously,  $\sim 50\%$  of Qperp jets are close to the subsolar region and the other half are further away toward the flanks. As a result, a split of data can make the result vary heavily if the training and test sample are not equally distributed.



**FIGURE 4 |** Example of the training procedure for a neural network run. **Left:** Accuracy vs. number of epoches. **Right:** cross-entropy loss vs. number of epoches. The results were stabilized for an epoch number of  $\sim 250$ . Blue lines show the behavior of the training subset while orange lines show that of the testing subset.



**FIGURE 5** | Accuracy of the neural network for 100 different iterations. **Left:** Results of the quasi-parallel class. **Right:** Results of the quasi-perpendicular class. The training of the shown neural networks includes the IMF vector.



**FIGURE 6** | Accuracy of the neural network for 100 random initialization of training/test set, with every point being one iteration. The x-axis represents the accuracy in quasi-perpendicular class and ranges from 70 to 100%. On the other hand, the y-axis shows the accuracy in the quasi-parallel class and ranges from 94 to 100%. Special indication of the leave one out result is marked in green color. The training of the shown neural networks includes the IMF vector.

**Figure 6** shows the resulting pairs of Qperp accuracy vs. the Qpar accuracy for each of the 100 iterations. It is seen that the distributions of the independent iterations does not exhibit any specific structure but appears random. This is important and gives confidence in the NN results since it is common in binary classification problems, such as this one, to have a competing classification problem; i.e., the higher the score for one class the lower the score for the other. This is not the case here and the differences in the respective scores appear to stem only from the random division of training/testing subsets.

Finally, in **Figure 6** the classification score from the leave-one-out process can be seen with Qpar and Qperp scores equal to ~

96.7 and ~ 88.2%, respectively. The leave-one-out process scores lie close to the mean of both the Qpar and the Qperp scores, while being slightly lower in both cases. We can speculate that the small differences may originate from the differences between the two types of validation methods. First of all, the class imbalance and the fact that resampling methods utilized in this case had to increase the dataset to a larger number of samples, could enhance the effect of class imbalance, decreasing the accuracy of the network. Furthermore, for the leave-one-out method, a validation set of 10% was used to check at which point the epoch training should be stopped. This was necessary since optimizing the epoch number or the batch size, while training with all but

**TABLE 4 |** Accuracy of each method used per class.

Class	NN - age (%)	NN - leave-1-out (%)	$\theta_{\text{cone}}$ (%)	Coplanarity (%)	Bow shock model (%)
Qpar	98	97	61	81	74
Qperp	88	88	94	79	86
Mean	93	93	77.5	80	80

The neural network accuracy is taken as the average performance of the 100 random iterations as shown in **Figures 5, 6**.

one jet is time consuming and outside the scope of this work. As a result, the difficulty in optimizing the hyper-parameters of the neural network for the leave-one-out case may have had a slight impact on the overall accuracy of the method.

## 4.2. Comparison of Neural Networks to Physics-Based Models

We now compare the classification scores obtained from the neural networks with the respective scores from the physics-based models. As seen in **Table 4**, neural networks outperform the physics-based modeling methods in reproducing the classification of the jets in the dataset.

The simple approximation of the cone angle (Equation 3) performs very well for the Qperp class, achieving the highest score of all the methods with an accuracy of  $\sim 94\%$  surpassing that of the neural networks. On the other hand, the cone angle method clearly underperforms in the Qpar class having the lowest score of all methods with an accuracy of only  $\sim 61\%$ . These results show that the cone angle approximation is not well-suited for this problem. This is expected since a large portion of jets were not found in close proximity to the subsolar region.

The coplanarity method (Equation 4) performs almost equally well for the two classes with accuracy scores of  $\sim 81$  and  $\sim 79\%$  for the Qpar and Qperp classes, respectively. The accuracy scores are quite high demonstrating it is an appropriate and accurate method for this problem. However, in both classes, it is heavily outperformed by the neural networks.

Finally, using a bow shock model to calculate the  $\theta_{Bn}$  presents an in-between case of the other two physical methods. The accuracy scores of  $\sim 74$  and  $\sim 86\%$  for the Qpar and Qperp classes, respectively, are good classification scores with a high tendency to misclassify jets as Qperp, though not as strong as in the cone-angle method. However, this method is also outperformed in both classes by the neural networks. It is interesting to note that the three physics-based methods all have mean scores, for both classes, close to  $\sim 80\%$  implying possibly an inherent limitation due to the fact that for upstream information they only use the IMF vector. This is not true for the neural networks which can accept all the available information.

## 4.3. Neural Networks Without IMF ( $\mathbf{B}_u$ ) Input

After establishing the advantage of the neural network method when providing the upstream magnetic field vector ( $\mathbf{B}_u$ ), we investigate the performance of NNs even in the absence of the, in

**TABLE 5 |** Accuracy of each method used per class.

Class	NN - age (%)	NN - leave-1-out (%)	$\theta_{\text{cone}}$ (%)	Coplanarity (%)	Bow shock model (%)
Qpar	95	95	61	81	74
Qperp	87	86	94	79	86
Mean	91	91	77.5	80	80

In this case, the magnetic field components were not included in the training procedure of the neural network method. The neural network accuracy is taken as the average performance of 100 individually trained networks as shown in **Figures 7, 8**.

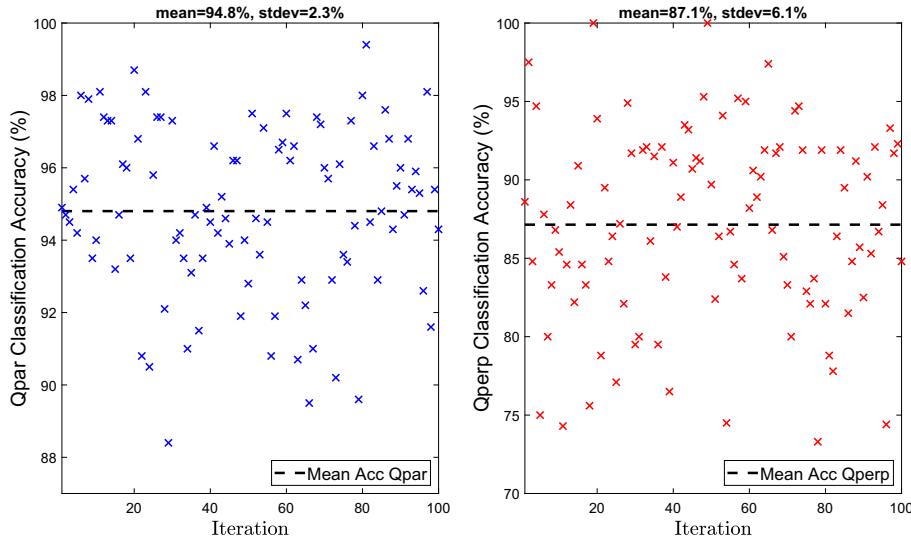
principle, vital information of the magnetic field orientation. This was done to see whether the neural network can still perform well without the IMF input that is necessary for all the other physics-based models. It should be noted, that we still used the absolute magnetic field ( $|B|$ ) as input to the NN. As a result, the input in these runs consists of all the parameters shown in **Table 1**, except the last three which are the components of the magnetic field vector.

Similarly to the previous subsections, the results of the random train/test splits are shown in **Figures 7, 8**. In **Figure 7**, it can be seen that the average classification accuracy of both Qpar ( $\sim 94.8\%$ ) and Qperp ( $\sim 87.1\%$ ) class remain high even when the NN's input does not contain directional information of the IMF. Once more, the accuracy in the class of Qpar jets has a lower standard deviation than the Qperp, while in both cases the standard deviation has increased compared to the results shown in **Figure 5**. Again, as shown in **Figure 8**, there is no specific structure regarding the scores of each iteration, similarly to the previous case reported in section 4.1. The leave-one-out result is slightly lower than the average result of Qperp jets ( $\sim 85.7\%$ ), while being the same as the average for the Qpar ones ( $\sim 94.7\%$ ).

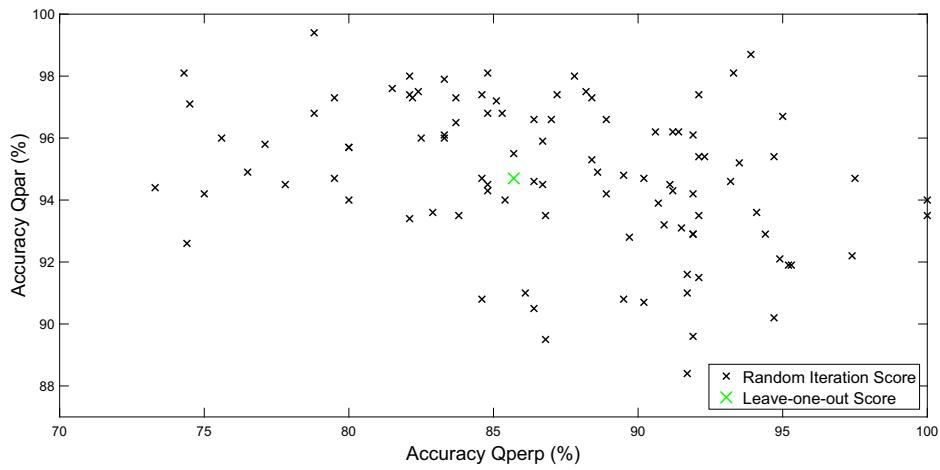
## 4.4. Dependence of Solar Zenith Angle

After deriving a classification using all the presented methods, a possible link between the misclassified jets and their position in the magnetosheath is investigated. As discussed previously, the solar wind measurements provided as input for all methods, characterize the subsolar region and are not ideal for the characterization of the flanks. As a result, we investigate if this effect is shown in the classification results of each method.

In **Figure 9**, we present the probability of jets appearing for different solar zenith angles. There are 4 histograms per jet class that correspond to the 4 presented methods of classifications. On each histogram, there are 3 plots representing the total number of jets (blue), the misclassified jets (red) and the normalized misclassified cases (black). **Figure 9A** shows the results of the Qpar jets while **Figure 9B** shows the Qperp jets. It can be seen that most cases of misclassification occur for angles close to the subsolar region ( $\leq 30^\circ$ ). However, when looking at the normalized misclassified rates (black line), it appears that when taking into account the overall number of jets, the relative misclassification rate is much higher close to the flanks ( $\geq 30^\circ$ ).



**FIGURE 7 |** Accuracy of the neural network for 100 different iterations. **Left:** Results of the quasi-parallel class. **Right:** Results of the quasi-perpendicular class. The training of the shown neural networks does not include the IMF vector.



**FIGURE 8 |** Accuracy of the neural network for 100 random initialization of training/test set, with every point being one iteration. The x-axis represents the accuracy in quasi-perpendicular class and ranges from 70 to 100%. On the other hand, the y-axis shows the accuracy in the quasi-parallel class and ranges from 88 to 100%. Special indication of the leave one out result is marked in green color. The training of the shown neural networks does not include the IMF vector.

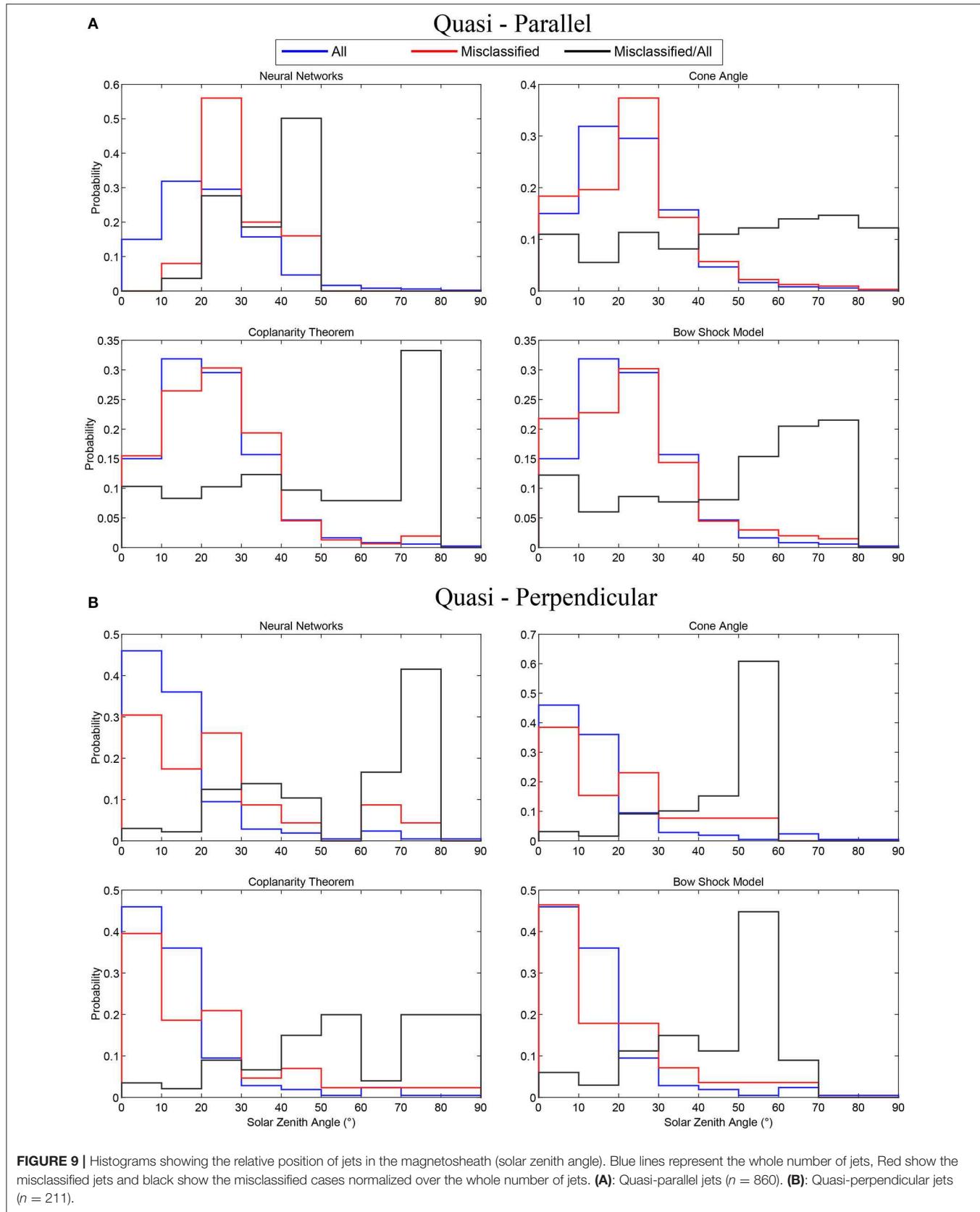
## 5. DISCUSSION AND CONCLUSIONS

From the results of the neural network application (Figures 5, 6, and Table 4), it is clearly shown that neural networks are capable of reproducing the results of the initial database with greater accuracy than the alternative physics-based models.

The different results for each method likely originates from the different properties the associated solar wind values between the Qpar and Qperp jets (Raptis et al., 2019). As shown in Figure 4 of Raptis et al. (2019), the velocity of Qperp jets is much lower than that of the Qpar jets, with the first having an average absolute ion velocity of  $\sim 100$  km/s, while Qpar have  $\sim 230$  km/s.

Furthermore, it was shown that the average solar wind velocity under which Qpar jets were found was  $\langle V_{SW,||,Jets} \rangle \approx 495$  km/s with a standard deviation of  $\sigma_{||,Jets} = 96$  km/s. On the other hand for the Qperp jet,  $\langle V_{SW,\perp,Jets} \rangle \approx 400$  km/s with  $\sigma_{\perp,Jets} = 46$  km/s. These differences in the velocity of the jets can have a large effect on the final results since they not only affect the input parameter space (solar wind velocity) but also the association timing (5-min average) differently for every class.

To begin with, the cone angle approximation is working effectively only for the Quasi-perpendicular jets. This is most likely because the majority of them were found close to the subsolar region. Another reason could be that most Qperp jets



**TABLE 6 |** Number of misclassifications grouped in different solar zenith angles for every classification method.

Method	[0 – 30]°	[30 – 50]°	[50 – 90]°
Neural networks	33	12	3
Cone angle	248	147	16
Coplanarity	146	88	10
Bow shock model	174	97	15

have a lower average velocity (Raptis et al., 2019), that along with the proximity to the subsolar region may make the 5-min average values used for the modeling more accurate. The better overall results of the bow shock modeling method originate from the estimation of the bow shock normal vector  $\hat{n}$ . By finding a point of origin for each jet, many cases that were found closer to the flanks of the magnetosheath region were correctly classified. Finally, the coplanarity method, while still producing overall worse results than the neural network, showed that a significant part of the jets can be classified using this approach. As shown in **Figure 9**, the majority of the jets, and especially Qperp jets occur close to the subsolar region ( $\leq 30^\circ$ ). However, in all the presented methods the misclassified cases occur mainly close to the flanks. This could originate from the poor characterization of OMNIweb data regarding phenomena that occur close to the flanks. In particular, OMNIweb has a poor capability of catching intervals of quasi-radial IMF (Bier et al., 2014; Suvorova and Dmitriev, 2016). Under quasi-radial IMF there is a quasi-perpendicular shock forming close to the flanks. As a result, Qperp jets found close to the flanks are probably not characterized well enough by the available solar wind measurements that were given as input to all the methods.

The superiority of the neural network provides indirect support to the initial dataset. This is achieved because the input of the neural network was independent of the one used in the initial classification. In particular, the initial database was classified using *only* magnetosheath (downstream) data, while the neural network input contained only solar wind (upstream) information. According to **Figure 9**, the highest evaluation accuracy was obtained for jets found in close vicinity of the subsolar region. As already stated previously, the OMNIweb database provides measurements that correspond to the subsolar region and therefore are not ideal to characterize the flank regions. Furthermore, the jets found very far from the bow shock could have taken a longer time to propagate in the magnetosheath region, making the choice of 5-min averaging dubious. A better estimation of the jet travel time from the bow shock to its observation point could possibly increase the accuracy of all the presented methods. These results show that the choice of using *in-situ* measurements for the determination of jets' class may indeed provide more accurate results while not limiting the classification procedure to periods of times that upstream data are available (Raptis et al., 2019). There are many jets that were found far away from the subsolar region ( $Y_{GSE} > 5R_E$ ) that were systematically misclassified by the physics-based methods. However, a good portion of them

was correctly classified by the neural network approach. As previously discussed, the main problem with jets that are found at the flanks of the magnetosheath is that the measurements taken from the solar wind do not accurately characterize this region. Furthermore, the time propagation error along with the error of the origin position is greatly enhanced the further away a jet is found from the bow shock. Nevertheless, due to the availability of such cases, the neural network was able to recognize peculiar cases, “train” for them, and correctly identify a significant portion of them. **Table 6**, shows the number of misclassifications done per method grouped in three ranges, these close to the subsolar region ( $[0 – 30]^\circ$ ) these further away from the subsolar ( $[30 – 50]^\circ$ ) and the ones far toward the flanks of the magnetosheath ( $[50 – 90]^\circ$ ). It is clear, that neural networks not only outperform the rest of the methods in the regions where jets are found more frequently but also in the not so common cases of flank jets.

To increase the accuracy of the neural network, one could in principle train two different NNs to tackle the different characterization of subsolar jets and flank jets. Then by utilizing ensemble learning methods, each network could work on its appropriate dataset possibly providing superior combined results. This task is not trivial since the boundary of where a subsolar region starts and ends is not sharp. Furthermore, depending on the properties of the jet, the association of solar wind measurements is an extremely complicated task. In this work, we used 5-min average values that while characterizing the majority of the events, they may fail to do so if a jet has very high velocities or if a significant part of its velocity lies in the  $yz$  plane, which could mean that it traveled in the magnetosheath for a longer period of time. All the above, along with the determination of the rest of the classes, shown in the presented database (**Table 2**) are planned to be done in future studies.

From a physical point of view, the most interesting result is perhaps the fact that neural networks maintained a very high accuracy even in the absence of the directional information of the IMF (**Figures 7, 8** and **Table 5**). This could be interpreted in several different ways. The most direct one is that the neural networks take advantage of the fact that in the initial database, the jets found in quasi-perpendicular plasma have on average a lower velocity and density than the jets found in the quasi-parallel magnetosheath. It is, however, not yet fully understood if this is the result of an observational bias or of a real physical mechanism (Raptis et al., 2019). The observational bias here would be that for conditions of low velocity and density, the threshold of finding a jet is easier to be satisfied (Equation 1). This would in principle allow jets that are found in the quasi-perpendicular plasma to occur primarily under low velocity and density solar wind. On the other hand, such a bias is not likely to fully explain the shown results. The conditions under which Qperp jets are found could originate from a physical process that makes Qperp jets more likely to occur under specific solar wind conditions, regardless of the IMF direction. If the latter is true, it means that an investigation of solar wind classes (e.g., Habbal et al., 1997; Camporeale et al., 2017) could give insight as to whether each

jet class (**Table 2**) occurs under different conditions or if specific conditions simply favor the formation of one class over the other. A final possible explanation is that directional information of the IMF is “hidden” in various quantities that were used for the training of the neural network. This, in turn, would allow the non-linear relationship generated by the network to accurately find the correct class of the jets by utilizing such previously undetectable information.

Neural Networks were shown to be a powerful method for the classification of magnetosheath jets. They outperformed the physics-based methods used in distinguishing between quasi-parallel and quasi-perpendicular jets. The results here, also indicate that upstream solar wind properties are sufficient to predict the class of the jets even without including the magnetic field vector. Last but not least, machine learning approaches, such as this one, can be generalized and applied to several satellite missions and space environments.

## DATA AVAILABILITY STATEMENT

MMS data are available at the MMS Science Data Center (<https://lasp.colorado.edu/mms/sdc>). OMNIweb Data are available at [https://omniweb.gsfc.nasa.gov/form/omni\\_min.html](https://omniweb.gsfc.nasa.gov/form/omni_min.html). Finally, the time of observation, the initial classification and the classification done by the neural networks and the three physical methods can be found in the **Supplementary Material** or in the associated data repository (Raptis et al., 2020).

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available online at: tensorflow.org
- Amata, E., Savin, S., Ambrosino, D., Bogdanova, Y., Marcucci, M., Romanov, S., et al. (2011). High kinetic energy density jets in the Earth's magnetosheath: a case study. *Planet. Space Sci.* 59, 482–494. doi: 10.1016/j.pss.2010.07.021
- Aminalragia-Giamini, S., Jiggens, P., Anastasiadis, A., Sandberg, I., Aran, A., Vainio, R., et al. (2020). Prediction of solar proton event fluence spectra from their peak flux spectra. *J. Space Weather Space Clim.* 10:1. doi: 10.1051/swsc/2019043
- Aminalragia-Giamini, S., Papadimitriou, C., Sandberg, I., Tsigkanos, A., Jiggens, P., Evans, H., et al. (2018). Artificial intelligence unfolding for space radiation monitor data. *J. Space Weather Space Clim.* 8:A50. doi: 10.1051/swsc/2018041
- Anderson, B. J., Fuselier, S. A., Gary, S. P., and Denton, R. E. (1994). Magnetic spectral signatures in the earth's magnetosheath and plasma depletion layer. *J. Geophys. Res. Space Phys.* 99, 5877–5891. doi: 10.1029/93JA02827
- Archer, M., Hietala, H., Hartinger, M., Plaschke, F., and Angelopoulos, V. (2019). Direct observations of a surface eigenmode of the dayside magnetopause. *Nat. Commun.* 10:615. doi: 10.1038/s41467-018-08134-5
- Archer, M., Horbury, T., and Eastwood, J. (2012). Magnetosheath pressure pulses: Generation downstream of the bow shock from solar wind discontinuities. *J. Geophys. Res. Space Phys.* 117:A5. doi: 10.1029/2011JA017468
- Archer, M. O., and Horbury, T. S. (2013). Magnetosheath dynamic pressure enhancements: occurrence and typical properties. *Ann. Geophys.* 31, 319–331. doi: 10.5194/angeo-31-319-2013
- Balasis, G., Aminalragia-Giamini, S., Papadimitriou, C., Daglis, I. A., Anastasiadis, A., and Haagmans, R. (2019). A machine learning approach for automated ulf wave recognition. *J. Space Weather Space Clim.* 9:A13. doi: 10.1051/swsc/2019010
- Balogh, A., and Treumann, R. A. (2013). *Physics of Collisionless Shocks: Space Plasma Shock Waves*. New York, NY: Springer. doi: 10.1007/978-1-4614-6099-2
- Baumjohann, W., and Treumann, R. A. (2012). *Basic Space Plasma Physics*. London: World Scientific Publishing Company. doi: 10.1142/p850
- Bier, E. A., Owusu, N., Engebretson, M. J., Posch, J. L., Lessard, M. R., and Pilipenko, V. A. (2014). Investigating the IMF cone angle control of pc3-4 pulsations observed on the ground. *J. Geophys. Res. Space Phys.* 119, 1797–1813. doi: 10.1002/2013JA019637
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press. doi: 10.1201/9781420050646.ptb6
- Boberg, F., Wintoft, P., and Lundstedt, H. (2000). Real time KP predictions from solar wind data using neural networks. *Phys. Chem. Earth C Solar Terres. Planet. Sci.* 25, 275–280. doi: 10.1016/S1464-1917(00)00016-7
- Bobra, M. G., and Ilonidis, S. (2016). Predicting coronal mass ejections using machine learning methods. *Astrophys. J.* 821:127. doi: 10.3847/0004-637X/821/2/127
- Brownlee, J. (2020). *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery. Available online at: <https://books.google.be/books?id=jaXJDwAAQBAJ>
- Camporeale, E., Carè, A., and Borovsky, J. E. (2017). Classification of solar wind with machine learning. *J. Geophys. Res. Space Phys.* 122, 10,910–10,920. doi: 10.1002/2017JA024383
- Camporeale, E., Wing, S., and Johnson, J. (2018b). Space weather in the machine learning era. *EOS* 99. doi: 10.1029/2018EO101897
- Camporeale, E., Wing, S., and Johnson, J. (eds.) (2018a). *Machine Learning Techniques for Space Weather*. Elsevier, p. 454. Available online at: <https://www.elsevier.com/books/machine-learning-techniques-for-space-weather/camporeale/978-0-12-811788-0>
- Case, N., and Wild, J. (2012). A statistical comparison of solar wind propagation delays derived from multispacecraft techniques. *J. Geophys. Res. Space Phys.* 117:A2. doi: 10.1029/2011JA016946

## AUTHOR CONTRIBUTIONS

SR and SA-G performed the data analysis. SR wrote the paper with contributions from SA-G, TK, and ML. All the authors contributed to drafting and revising the paper. All the authors contributed to the interpretation of the data.

## FUNDING

SR and TK acknowledge support from the Swedish National Space Board (SNSA grant 90/17).

## ACKNOWLEDGMENTS

We thank the MMS team for providing data and support <https://lasp.colorado.edu/mms/sdc/public/>. Furthermore, we acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service, and OMNI data. OMNI High-resolution data are available through [https://omniweb.gsfc.nasa.gov/form/omni\\_min.html](https://omniweb.gsfc.nasa.gov/form/omni_min.html).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspas.2020.00024/full#supplementary-material>

- Chandorkar, M., and Camporeale, E. (2018). "Probabilistic forecasting of geomagnetic indices using Gaussian process models." in *Machine Learning Techniques for Space Weather*, eds E. Camporeale, S. Wing, and J. R. Johnson (Elsevier), 237–258. doi: 10.1016/B978-0-12-811788-0-00009-3
- Chao, J., Wu, D., Lin, C.-H., Yang, Y.-H., Wang, X., Kessel, M., et al. (2002). "Models for the size and shape of the earth's magnetopause and bow shock," in *Cospar Colloquia Series*, Vol. 12, ed L.-H. Lyu (Taipei: Elsevier), 127–135. doi: 10.1016/S0964-2749(02)80212-8
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Chollet, F. (2015). *Keras*. Available online at: <https://github.com/fchollet/keras>
- Florios, K., Kontogiannis, I., Park, S.-H., Guerra, J. A., Benvenuto, F., Bloomfield, D. S., et al. (2018). Forecasting solar flares using magnetogram-based predictors and machine learning. *Solar Phys.* 293:28. doi: 10.1007/s11207-018-1250-4
- Formisano, V., and Hedgecock, P. (1973). Solar wind interaction with the earth's magnetic field: 3. On the earth's bow shock structure. *J. Geophys. Res.* 78, 3745–3760. doi: 10.1029/JA078i019p03745
- Fuselier, S. A. (2013). *Suprathermal Ions Upstream and Downstream from the Earth's Bow Shock*. Washington, DC: American Geophysical Union (AGU). doi: 10.1029/GM081p0107
- Fuselier, S. A., Anderson, B. J., Gary, S. P., and Denton, R. E. (1994). Inverse correlations between the ion temperature anisotropy and plasma beta in the earth's quasi-parallel magnetosheath. *J. Geophys. Res. Space Phys.* 99, 14931–14936. doi: 10.1029/94JA00865
- Giacalone, J., and Jokipii, J. R. (2007). Magnetic field amplification by shocks in turbulent fluids. *Astrophys. J. Lett.* 663:L41. doi: 10.1086/519994
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT press.
- Gosling, J., Asbridge, J., Bame, S., Paschmann, G., and Sckopke, N. (1978). Observations of two distinct populations of bow shock ions in the upstream solar wind. *Geophys. Res. Lett.* 5, 957–960. doi: 10.1029/GL005i011p00957
- Gunell, H., Wieser, G. S., Mella, M., Maggiolo, R., Nilsson, H., Darrouzet, F., et al. (2014). Waves in high-speed plasmoids in the magnetosheath and at the magnetopause. *Ann. Geophys.* 32, 991–1009. doi: 10.5194/angeo-32-991-2014
- Gutynska, O., Sibeck, D., and Omidi, N. (2015). Magnetosheath plasma structures and their relation to foreshock processes. *J. Geophys. Res. Space Phys.* 120, 7687–7697. doi: 10.1002/2014JA020880
- Habbal, S., Woo, R., Fineschi, S., O'Neal, R., Kohl, J., Noci, G., et al. (1997). Origins of the slow and the ubiquitous fast solar wind. *Astrophys. J. Lett.* 489:L103. doi: 10.1086/310970
- Han, D.-S., Hietala, H., Chen, X.-C., Nishimura, Y., Lyons, L. R., Liu, J.-J., et al. (2017). Observational properties of dayside throat aurora and implications on the possible generation mechanisms. *J. Geophys. Res. Space Phys.* 122, 1853–1870. doi: 10.1002/2016JA023394
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034. doi: 10.1109/ICCV.2015.123
- Hietala, H., Laitinen, T. V., Andréová, K., Vainio, R., Vaivads, A., Palmroth, M., et al. (2009). Supermagnetosonic jets behind a collisionless quasiparallel shock. *Phys. Rev. Lett.* 103:245001. doi: 10.1103/PhysRevLett.103.245001
- Hietala, H., Phan, T., Angelopoulos, V., Oieroset, M., Archer, M., Karlsson, T., et al. (2018). In situ observations of a magnetosheath high-speed jet triggering magnetopause reconnection. *Geophys. Res. Lett.* 45, 1732–1740. doi: 10.1002/2017GL076525
- Hietala, H., and Plaschke, F. (2013). On the generation of magnetosheath high-speed jets by bow shock ripples. *J. Geophys. Res. Space Phys.* 118, 7237–7245. doi: 10.1002/2013JA019172
- Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in: *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Vol. 37 (Lille: JMLR.org), 448–456. doi: 10.5555/3045118.3045167
- Jonas, E., Bobra, M., Shankar, V., Hoeksema, J. T., and Recht, B. (2018). Flare prediction using photospheric and coronal image data. *Solar Phys.* 293:48. doi: 10.1007/s11207-018-1258-9
- Karlsson, T., Brenning, N., Nilsson, H., Trotignon, J.-G., Vallières, X., and Fascko, G. (2012). Localized density enhancements in the magnetosheath: three-dimensional morphology and possible importance for impulsive penetration. *J. Geophys. Res. Space Phys.* 117:A3. doi: 10.1029/2011JA017059
- Karlsson, T., Kullen, A., Liljeblad, E., Brenning, N., Nilsson, H., Gunnell, H., et al. (2015). On the origin of magnetosheath plasmoids and their relation to magnetosheath jets. *J. Geophys. Res. Space Phys.* 120, 7390–7403. doi: 10.1002/2015JA021487
- King, J., and Papitashvili, N. (2005). Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *J. Geophys. Res. Space Phys.* 110:A2. doi: 10.1029/2004JA010649
- Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* 18, 559–563. Available online at: <http://jmlr.csail.mit.edu/papers/v18/16-365.html>
- LeNail, A. (2019). NN-SVG: Publication-ready neural network architecture schematics. *J. Open Source Softw.* 4:747. doi: 10.21105/joss.00747
- Liu, J., Ye, Y., Shen, C., Wang, Y., and Erdélyi, R. (2018). A new tool for CME arrival time prediction using machine learning algorithms: cat-puma. *Astrophys. J.* 855:109. doi: 10.3847/1538-4357/aaae69
- Luhmann, J., Russell, C., and Elphic, R. (1986). Spatial distributions of magnetic field fluctuations in the dayside magnetosheath. *J. Geophys. Res. Space Phys.* 91, 1711–1715. doi: 10.1029/JA091iA02p01711
- Mailyan, B., Munteanu, C., and Haaland, S. (2008). What is the best method to calculate the solar wind propagation delay? *Ann. Geophys.* 26, 2383–2394. doi: 10.5194/angeo-26-2383-2008
- McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. doi: 10.1007/BF02478259
- Merka, J., Szabo, A., Narock, T., King, J., Paularena, K., and Richardson, J. (2003). A comparison of IMP 8 observed bow shock positions with model predictions. *J. Geophys. Res. Space Phys.* 108:A2. doi: 10.1029/2002JA009384
- Němeček, Z., Šafránková, J., Přech, L., Sibeck, D., Kokubun, S., and Mukai, T. (1998). Transient flux enhancements in the magnetosheath. *Geophys. Res. Lett.* 25, 1273–1276. doi: 10.1029/98GL50873
- Paschmann, G., and Daly, P. W. (1998). *Analysis Methods for Multi-spacecraft Data*. Bern.
- Plaschke, F., Hietala, H., Archer, M., Blanco-Cano, X., Kajdič, P., Karlsson, T., et al. (2018). Jets downstream of collisionless shocks. *Space Sci. Rev.* 214:81. doi: 10.1007/s11214-018-0516-3
- Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., et al. (2016). Fast plasma investigation for magnetospheric multiscale. *Space Sci. Rev.* 199, 331–406. doi: 10.1007/s11214-016-0245-4
- Raptis, S., Aminalragia-Giamini, S., Karlsson, T., and Lindberg, M. (2020). Magnetosheath Jets (Qpar-Qperp) Classification [Data set]. *Front. Astron. Space Sci.* doi: 10.5281/zenodo.3746592
- Raptis, S., Karlsson, T., Plaschke, F., Kullen, A., and Lindqvist, P.-A. (2019). Classifying magnetosheath jets using mms - statistical properties. *Earth Space Sci. Open Arch.* 41. doi: 10.1002/essoar.10501493.2
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65:386. doi: 10.1037/h0042519
- Russell, C., Anderson, B., Baumjohann, W., Bromund, K., Dearborn, D., Fischer, D., et al. (2016). The magnetospheric multiscale magnetometers. *Space Sci. Rev.* 199, 189–256. doi: 10.1007/s11214-014-0057-3
- Samarasinghe, S. (2016). *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. New York, NY: Auerbach publications.
- Schwartz, S. J., and Burgess, D. (1991). Quasi-parallel shocks: a patchwork of three-dimensional structures. *Geophys. Res. Lett.* 18, 373–376. doi: 10.1029/91GL00138
- Schwartz, S. J., Burgess, D., Wilkinson, W. P., Kessel, R. L., Dunlop, M., and Lühr, H. (1992). Observations of short large-amplitude magnetic structures at a quasi-parallel shock. *J. Geophys. Res. Space Phys.* 97, 4209–4227. doi: 10.1029/91JA02581

- Shin, D.-K., Lee, D.-Y., Kim, K.-C., Hwang, J., and Kim, J. (2016). Artificial neural network prediction model for geosynchronous electron fluxes: dependence on satellite position and particle energy. *Space Weather* 14, 313–321. doi: 10.1002/2015SW001359
- Stone, C. E., Frandsen, A. M., Mewaldt, R., Christian, E., Margolies, D., Ormes, J., et al. (1998). The advanced composition explorer mission. *Space Sci. Rev.* 86, 1–22. doi: 10.1007/978-94-011-4762-0\_1
- Suvorova, A., and Dmitriev, A. (2016). On magnetopause inflation under radial IMF. *Adv. Space Res.* 58, 249–256. doi: 10.1016/j.asr.2015.07.044
- Turc, L., Fontaine, D., Savoini, P., Hietala, H., and Kilpua, E. K. J. (2013). A comparison of bow shock models with cluster observations during low alfén mach number magnetic clouds. *Ann. Geophys.* 31, 1011–1019. doi: 10.5194/angeo-31-1011-2013
- Turner, D. L., Shprits, Y., Hartinger, M., and Angelopoulos, V. (2012). Explaining sudden losses of outer radiation belt electrons during geomagnetic storms. *Nat. Phys.* 8:208. doi: 10.1038/nphys2185
- Wilson, L. III. (2016). “Low frequency waves at and upstream of collisionless shocks,” in *Low-Frequency Waves in Space Plasmas* eds A. Keiling, D. H. Lee, and V. Nakariakov (Hoboken, NJ: John Wiley & Sons, Inc.), 269–291. doi: 10.1002/9781119055006.ch16
- Wintoft, P., Wik, M., Matzka, J., and Shprits, Y. (2017). Forecasting KP from solar wind data: input parameter study using 3-hour averages and 3-hour range values. *J. Space Weather Space Clim.* 7:A29. doi: 10.1051/swsc/2017027
- Xiang, Z., Ni, B., Zhou, C., Zou, Z., Gu, X., Zhao, Z., et al. (2016). Multi-satellite simultaneous observations of magnetopause and atmospheric losses of radiation belt electrons during an intense solar wind dynamic pressure pulse. *Ann. Geophys.* 34, 493–509. doi: 10.5194/angeo-34-493-2016

**Conflict of Interest:** 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.

Copyright © 2020 Raptis, Aminalragia-Giamini, Karlsson and Lindberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Spectral Deconvolution With Deep Learning: Removing the Effects of Spectral PSF Broadening

Momchil Molnar<sup>1,2\*</sup>, Kevin P. Reardon<sup>1,2</sup>, Christopher Osborne<sup>3</sup> and Ivan Milić<sup>1,4</sup>

<sup>1</sup> National Solar Observatory, Boulder, CO, United States, <sup>2</sup> Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO, United States, <sup>3</sup> SUPA School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom, <sup>4</sup> Department of Physics, University of Colorado, Boulder, CO, United States

## OPEN ACCESS

### Edited by:

Sophie A. Murray,  
Trinity College Dublin, Ireland

### Reviewed by:

Hector Socas-Navarro,  
Instituto de Astrofísica de Canarias,  
Spain

Andrés Asensio Ramos,  
Instituto de Astrofísica de Canarias,  
Spain

### \*Correspondence:

Momchil Molnar  
momo@nso.edu

<sup>†</sup>DKIST Ambassador

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

Received: 03 April 2020

Accepted: 11 May 2020

Published: 23 June 2020

### Citation:

Molnar M, Reardon KP, Osborne C and Milić I (2020) Spectral Deconvolution With Deep Learning: Removing the Effects of Spectral PSF Broadening. *Front. Astron. Space Sci.* 7:29.  
doi: 10.3389/fspas.2020.00029

We explore novel methods of recovering the original spectral line profiles from data obtained by instruments that sample those profiles with an extended or multipeaked spectral transmission profile. The techniques are tested on data obtained at high spatial resolution from the Fast Imaging Solar Spectrograph (FISS) grating spectrograph at the Big Bear Solar Observatory and from the Interferometric Bidimensional Spectrometer (IBIS) instrument at the Dunn Solar Telescope. The method robustly deconvolves wide spectral transmission profiles for fields of view sampling a variety of solar structures (granulation, plage, and pores) with a photometrical precision of <1%. The results and fidelity of the method are tested on data from IBIS obtained using several different spectral resolution modes. The method, based on convolutional neural networks (CNN), is extremely fast, performing about  $10^5$  deconvolutions per second on a CPU and  $10^6$  deconvolutions per second on NVIDIA TITAN RTX GPU for a spectrum with 40 wavelength samples. This approach is applicable for deconvolving large amounts of data from instruments with wide spectral transmission profiles, such as the Visible Tunable Filter (VTF) on the DKI Solar Telescope (DKIST). We also investigate its application to future instruments by recovering spectral line profiles obtained with a theoretical multi-peaked spectral transmission profile. We further discuss the limitations of this deconvolutional approach through the analysis of the dimensionality of the original and multiplexed data.

**Keywords:** convolutional neural networks, astronomical instrumentation, spectroscopy, deep learning, deconvolution algorithm

## 1. INTRODUCTION

The finite spectral resolution of real instruments affects the inferred signal by blending the intensities at different wavelengths. This phenomenon is problematic for (solar) spectral lines, since the shape of a line encodes essential information about a range of heights in the solar atmosphere. However, some instruments use a lower spectral resolution (broader spectral transmission profile) to increase instrument throughput and reduce integration times. Such a broad spectral point-spread function (sPSF) results fundamentally in a multiplexed sampling of the line profile, with the information from a given portion of the original spectral profile sampled multiple times at various positions in the sPSF (i.e., with varying relative attenuation) as the transmission function is tuned through the line. This means that it should be possible to recover much of the underlying spectral

information from this linear combination of samplings. The concept of exploiting this multiplexing to recover spectral information was originally developed by Caccin and Roberti (1979) and later Baranyi and Ludmány (1983) in order to reconstruct spectral profiles sampled by the relatively broad (0.15–0.5 Å FWHM) sPSF of the tunable Universal Birefringent Filter (UBF) (Beckers et al., 1975). The method developed, which relied on analytical descriptions of the sPSF, was employed by Caccin et al. (1983) and Baranyi (1986) to reconstruct H-alpha and Na D line profiles recorded through a UBF. However, the data at the time were recorded on photographic film and the method was sensitive to noise and computationally demanding. The current observational demands for high-resolution imaging have resulted in instruments based on Fabry-Pérot-interferometers that have sPSF's that are again suitable for this method.

In this work, we seek to evaluate machine-learning techniques that can retrieve higher-resolution spectra from instrumentally broadened spectral profiles. The effect of spectral smearing on the line shape is shown in the left panel of **Figure 1** with an example spectrum of Ca II 8,542 Å from the FISS/BBSO (Chae et al., 2013) spectrograph. The orange line is the spectrum as observed with the full spectral resolution of FISS of about  $R \sim 150,000$  (where spectral resolution  $R$  is defined as the wavelength of observation divided by the FWHM of the profile). Instead, the blue line shows the same spectrum convolved with a Lorentzian-shaped sPSF with  $R \sim 45,000$ . Given the typical shape of an absorption line, the convolution with a broad sPSF raises the intensity around the line core and broadens the wings of the profile.

This smearing tends to increase the similarity among different spectral profiles, also reducing the spatial contrast and the ability to identify small scale structures in images of the solar atmosphere. An example of this is presented in **Figure 1** (central and right panels), with observations in the core of the Ca II 8,542 Å line from the Interferometric Bidimensional Spectrometer (IBIS) (Cavallini, 2006) instrument at the Dunn Solar Telescope. The same FOV was observed at the instrument's normal high spectral resolution ( $R \sim 200,000$ ), but also at a much lower spectral resolution ( $R \sim 50,000$ ), which was achieved by removing the “narrow passband” Fabry-Pérot interferometer (FPI) from the optical path (Reardon and Cavallini, 2008). We can see the reduced contrast in the FOV with lower spectral resolution which deteriorates the identification of the chromospheric features. Hence, mitigation of the degraded spectral purity of our observations is essential for furthering our understanding of the Sun.

Furthermore, the compressible nature of spectral lines as suggested by Asensio Ramos et al. (2007) could allow the sampling and subsequent recovery of the full spectral profiles with a lesser number of measurements by using a suitably adopted measurement basis. This approach could improve instrumental performance by increasing the sampling cadence through a reduction in the number of instrumental tuning steps needed to sample the line.

In this paper we perform experiments to test the applicability of Convolutional Neural Networks (CNNs) to perform the de-multiplexing of spectral line profiles in different scenarios. We examine the photometric accuracy than can be achieved

with these techniques. Finally, we discuss the limitations on the precision of the recovered profiles based on the dimensionality of the data derived from maximum-likelihood intrinsic-dimensionality estimate (Levina and Bickel, 2004).

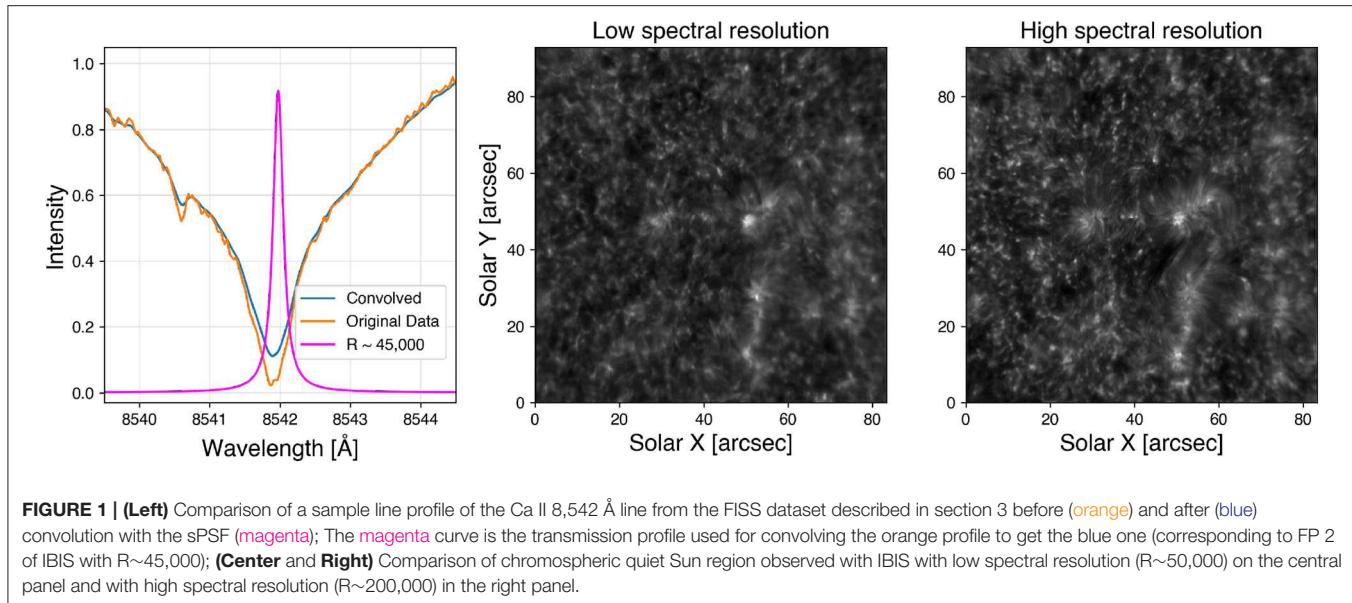
## 2. A DEEP LEARNING APPROACH

We utilize deep Convolutional Neural Networks (CNNs) for the deconvolution process as they are powerful function approximators which are widely used for pattern recognition and image processing (Goodfellow et al., 2016). We used an encoder-decoder architecture because it can extract the relevant features from noisy data (encoder) and then recreate the signal of interest from the latent space (decoder). The architecture of the network consists of three convolutional layers, followed by three symmetric upsampling (“deconvolutional”) layers, followed by two dense layers with dimensions of the output data. The three consecutive convolutional layers (and their corresponding upscaling layers) have [5, 10, 20] filters and used a three-pixel kernel. Rectified Linear unit (RELU) activation function was used for all layers with the exception of the last one where we have used a linear activation function (He et al., 2016). Furthermore, in section 3 we add the input layer to the last dense layer of the network to improve the performance of the network. This is due to the fact that in this architecture, the network has to estimate only the corrections to the convoluted signal instead of recreating the whole spectral line profile. However, this will cause the core of the spectral profile to be poorly fit since the most significant corrections are needed there (as can be seen in left panel of **Figure 1**). To alleviate this issue, we introduce a custom loss function which is a weighted mean square error. The weights are chosen inversely proportional to the intensity of the line profile so as to emphasize the precision of the recovered profiles in the line core. We trained our network with the Adam optimizer (Kingma and Ba, 2014) for about one thousand epochs before satisfactory convergence was achieved. The network was implemented with Keras under Tensorflow (Abadi et al., 2015) and can be found in the public repository of the project.

## 3. SPECTRAL DECONVOLUTION WITH CNN

### 3.1. Deconvolution of Synthetic Data From FISS

To test the CNN approach for sPSF deconvolution, we utilized Ca II 8,542 Å data from the FISS/BBSO (Chae et al., 2013) instrument ( $R \sim 150,000$ ) obtained on June 22, 2016. We created a training set by convolving each spectral profile with a Lorentzian sPSF with an effective  $R \sim 45,000$  [corresponding roughly to the FWHM of IBIS's FPI #2 transmission profile (Reardon and Cavallini, 2008)]. The bottom left panel of **Figure 2** shows a sample profile from the FISS instrument in blue and convolved with the FPI #2 profile in green. The CNN was trained with spectra from a single raster scan  $100 \times 250$  spatial pixels corresponding to  $16 \times 20$  arcseconds on the Sun centered on a pore near disk center which took 16.5 s. Satisfactory convergence



was accomplished in about 1,000 epochs with the relative RMS error at the last epoch of the training reaching about  $1.5 \times 10^{-4}$ .

The performance of the CNN was tested on a different raster (data not seen by the network previously) from FISS of the same region of the Sun acquired 5 minutes after the scan used for training. The line core intensity value and position was determined as the minimum of a parabola fitted to the 7 points around the pixel position with the lowest intensity. The core width of the line profile was measured (following Cauzzi et al., 2009) as the bisector width at the intensity equal to half the difference between intensity of the line core and the intensity at a fixed offset of 0.4 Å from the wavelength position of the line core.

The algorithm achieved 0.76% precision photometry of the line core intensity and 2.5% precision retrieving the line core width. These results are illustrated in the right column of **Figure 3**. This example shows the robustness of the ML approach for retrieving spectral line profiles. The algorithm takes about  $7 \times 10^{-6}$  s for a single inversion of 40 point spectrum on an Intel i7-4780HQ CPU and only  $0.3 \times 10^{-6}$  s on a NVIDIA TITAN RTX GPU. We take into account the I/O overhead for the GPU inversions, as we used a dataset of 16 million spectra with 40 wavelengths points (similar to the VTF full CCD readout) which amounts to about 20% of the memory of the GPU. This method is slightly faster than the `scipy.signal.deconvolve` algorithm which uses a digital filter, but the latter cannot reproduce the wings of the line well due to boundary effects. Compared to more computationally intensive algorithms such as the Richardson-Lucy (Richardson, 1972) deconvolution algorithm, we found that our algorithm is about 100 times faster. Furthermore, it does not require fine tuning of parameters once a suitable training set is provided.

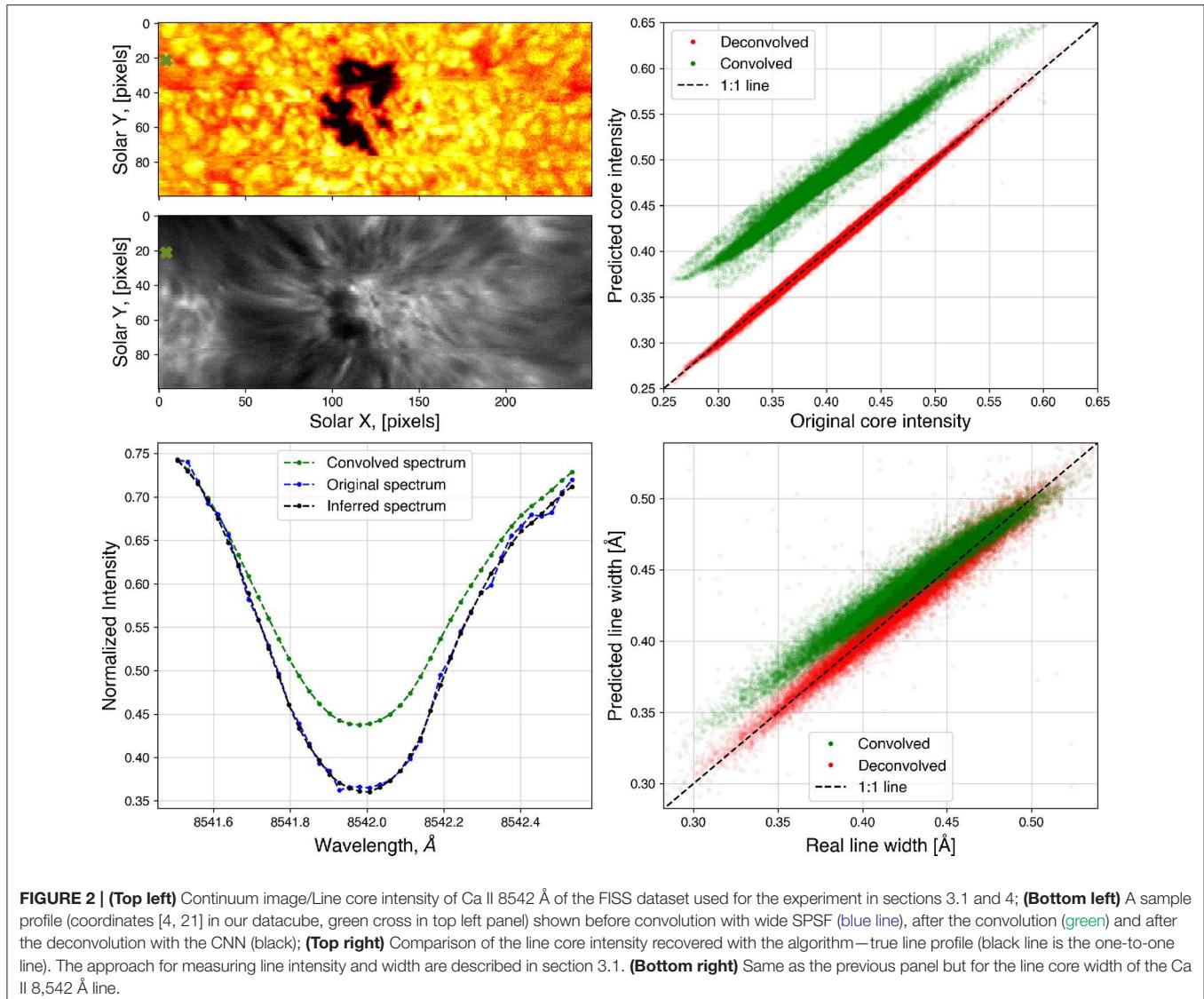
### 3.2. Deconvolution of Real Spectral Data From IBIS

To test the method on real Fabry-Pérot data we obtained a dataset with the IBIS instrument at the DST with high ( $R \sim$

200,000) and low ( $R \sim 50,000$ , similar to the FISS tests above) spectral resolution of the same region of the Sun. We achieved the different spectral resolutions by utilizing the fact that the IBIS instrument consists of two Fabry-Pérot (FP) interferometers in series, one of which has a profile three times narrower than the other (the components of the IBIS instrument are presented in the left panel of **Figure 3**). Hence, if we take the narrower FP (FP #1) out of the optical path, we obtain observations with a lower spectral resolution. We imaged a region near disk center of the Sun in the Ca II 8,542 Å line, where we scanned a spectral region of 4.4 Å centered around the line core with a spacing of 50 mÅ. We acquired five separate exposures at each wavelength [for post-processing MOMFBD (van Noort et al., 2005) reconstruction to minimize seeing effects] which resulted in two datasets of the same solar structures with different spectral resolution taken 4 minutes apart. We applied the deconvolution algorithm to the these datasets using as the input the lower spectral resolution data obtained with a single FP and as the expected output the higher spectral resolution dataset obtained with both FPs. Images from the two datasets are presented in the central and right panels of **Figure 1**.

We had limited success with deconvolving this dataset as the spectral profiles had changed significantly even over the 4 minute interval between the datasets. To illustrate this, the central panel of **Figure 3** shows the density plot of the quasi-continuum in the wings of the two datasets. The lack of obvious correlation (confirmed by visual inspection of the data) shows that the structures on the solar surface have significantly changed between the two datasets were obtained (consistent with granular lifetimes of  $\sim 8$  minutes).

To explore the validity of this deconvolution approach, we chose a subset of spatial pixels from the IBIS scans based on certain criteria to identify spectral profiles that did not change significantly between the two samplings. This step allows the CNN to train primarily on the effects from the spectral smearing, not the evolution of the solar atmosphere. The imposed criteria



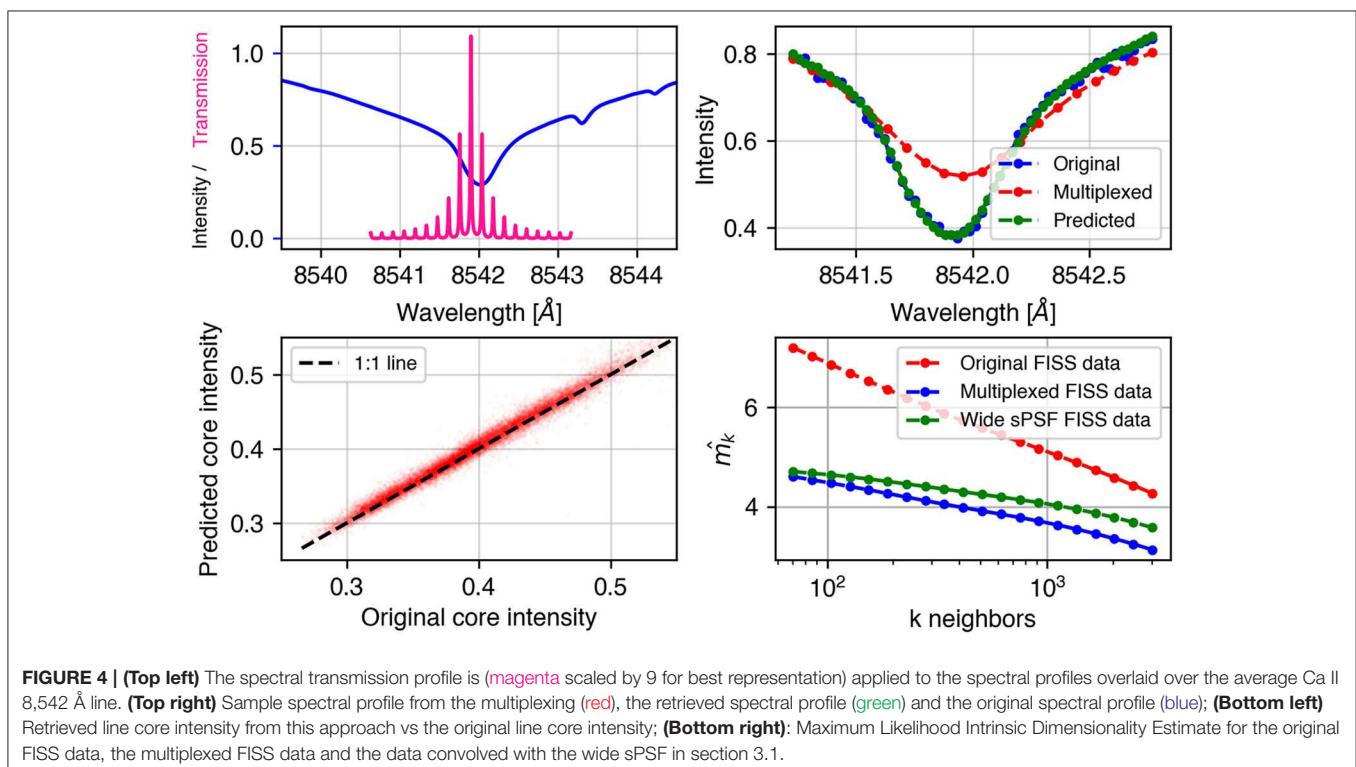
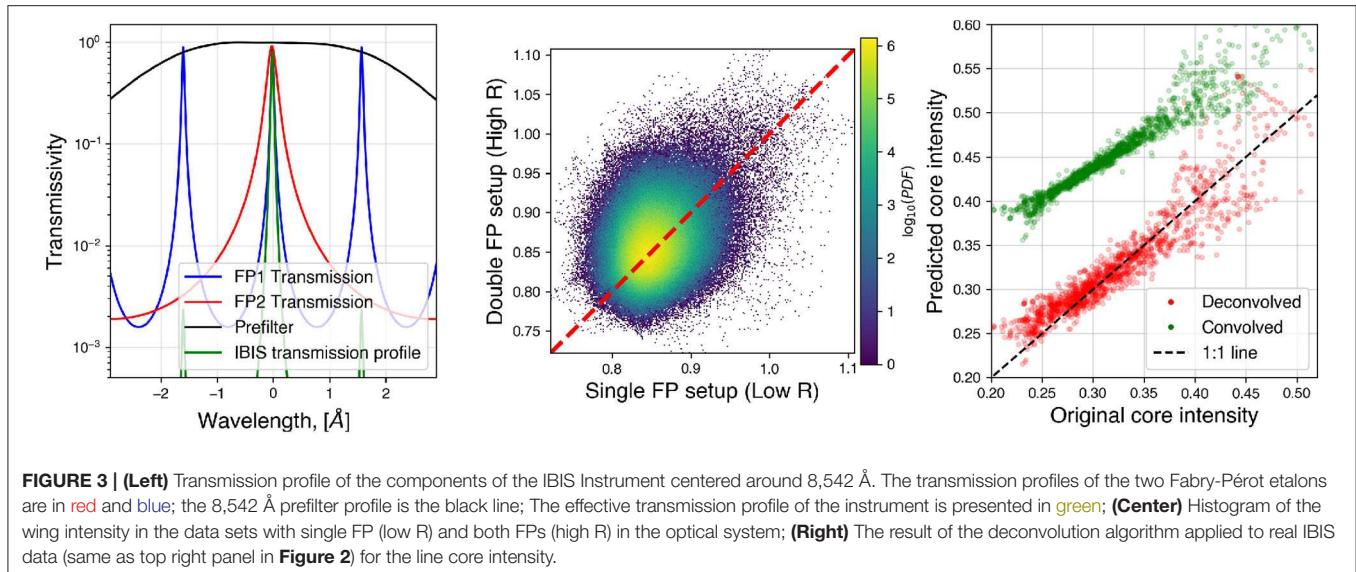
are that the measured Doppler velocity change<sup>1</sup> between the two consecutive samplings is no greater than half a resolution element (0.6 km/s) and that the location of the pixel in the cumulative distribution of intensity and width (relative to the other pixels sampled at the same spectral resolution) does not change by more than 5 percentiles. The expected vs. deconvolved core intensities are presented in the right panel of Figure 3 (compare to Figure 2). The scatter is larger than the FISS synthetic data due to the effects of solar surface evolution occurring between the acquisition of the two components of the training set, whose correction is beyond the scope of this project. Future tests of this method could emphasize obtaining a more nearly cotemporal training dataset by restricting the range of the spectral scan around line core, reducing the time separation between the different spectral resolution scans significantly. We note that distinct training datasets, derived at different times or even

using a separate instrument (e.g., a slit spectrograph), could be applied to multiple datasets obtained with a low-spectral-resolution FP instrument [e.g., VTF (Kentischer et al., 2012) at the DKIST (Tritschler et al., 2016)], under the assumption that: (a) the sPSF and straylight of each instrument is well characterized; and (b) the same general types of solar spectral profiles are sampled in both cases.

#### 4. RECOVERING UNDERSAMPLED SPECTRAL PROFILES WITH MULTI-PEAKED SPSF

It was suggested by Asensio Ramos et al. (2007) that not all wavelength points in a spectral line are linearly independent and that recovery of a full spectral profile with lesser number of measurements (in a suitably chosen basis) is possible. This presents us with the opportunity to extract useful information

<sup>1</sup>For symmetric line profiles, Doppler velocity does not depend on R.



from undersampled spectral line profiles which can result in more efficient spectral sampling or better compression techniques for space-based missions.

To test the ability of CNNs to recover spectral profiles multiplexed with multipeaked transmission profiles as suggested in Asensio Ramos and López Ariste (2010) we created a transmission profile of a hypothetical dual Fabry-Pérot interferometer with spacings of the etalons of 2.6 and 0.058 cm with 0.99 reflectance coatings. The resulting transmission profile of this hypothetical instrument is presented in the top

left panel of Figure 4 overplotted on the Ca II average line profile. The transmission profile was designed such that the higher-spectral-resolution FP generates multiple peaks within the chromospheric core of the solar spectral line while the lower-spectral-resolution FP selects a limited range such that 80% of the transmitted light is coming from the three central peaks. The properties of the FPs were chosen to optimize the precision of the deconvolutions. If the peaks of the transmission profile are too close or too far apart, the neural network's performance drops. Further optimization of the FP setup can be achieved through

exploration of the dimensionality of the data as described in the following paragraphs.

We applied the transmission profile to the FISS data used in section 3.1 where we downsampled the number of spectral samples by 3 for this particular example. A sample deconvolution is presented in the top right panel of **Figure 4**, which shows a good agreement between the original and deconvolved spectral profiles. The bottom left of **Figure 4** shows the scatter of the derived line core intensity of the multiplexed line vs. the original line core intensity. We achieve a RMS of the retrieved line core intensity of about 2% for this numerical setup. This is about three times worse than the previous experiment with FISS data in section 3.1. Our result is close to the precision obtained by Asensio Ramos (2010) where the author uses a single FP etalon with a prefilter.

To explain the lower precision of this multipeaked-multiplexing deconvolution approach compared to the deconvolution of the wide sPSF in section 3, we evaluate the dimensionality of the data. The dimensionality quantifies how much information is contained in the observations and can be used to evaluate the losses due to the spectral multiplexing. We computed the maximum likelihood intrinsic estimated dimensionality [MLIED, introduced by Levina and Bickel (2004) and suggested for spectroscopic use by Asensio Ramos et al. (2007)], which is an estimate of the dimensionality of the data based on phase density distribution. The bottom right panel of **Figure 4** shows the dimensionality estimate for the original data, the multiplexed data, and the *data convolved* with a wide sPSF vs. the number of neighboring spectra used for the computation of the dimensionality. We find that the dimensionality of the multipeaked-multiplexed data is lower than the data convolved with a wide sPSF, while the original data has the highest dimensionality. It is expected since the convolution process introduces a loss of information. This greater loss of information is why the multipeak approach (as modeled in this section) results in a lower precision compared to the results those for a single, broad sPSF.

This approach, evaluating spectral dimensionality, could be used in future design studies of instruments as a way toward building more efficient instruments, optimizing throughput and preservation of spectral information. Further work is needed to connect the dimensionality analysis (and resulting choice of instrumental sPSF) with the accuracy and precision of the retrieval of physical information from the spectral profiles via the optimal choice of parameters for the FP system.

## 5. CONCLUSIONS AND FUTURE WORK

We have presented a novel way to perform deconvolution of spectral data with deep learning. Our method is robust and reliable if the sPSF of the instrument is well known *a priori* and we have a reliable training set. Our method can deconvolve a single, 40-wavelength spectrum in 0.3 microsecond on a NVIDIA TITAN RTX GPU with a photometric precision of the line core intensity of <1%. The speed of the proposed algorithm makes it very effective for processing large numbers of spectra, with

further improvements possible if the deconvolution is performed on batches of data on a GPU. With the next generation of solar instruments (such as the VTF at the DKIST), which will produce terabytes of spectral data per day, the speed of deconvolutional techniques will become increasingly important.

The technique was demonstrated here only for non-polarized spectroscopic measurements, but full spectropolarimetric measurements (including also the spectral dependence of the circularly and linearly polarized components of the signal) are a key aspect of observational solar science. There is no conceptual reason why this method could not be extended to the measurement of the Stokes profiles, given suitable training sets. However, since the polarized components of the signal tend to be just a small fraction of the overall signal (a few percent or less), any systematic errors introduced into the deconvolved profiles might bias the recovery of the information about the magnetic field. Future work will evaluate the application of this method to this common usage scenario.

We successfully recovered spectral line profiles observed with a multipeaked spectral transmission profile, as suggested before in Asensio Ramos and López Ariste (2010), using a theoretical dual Fabry-Pérot etalon instrument with optically realistic parameters. Our numerical experiments showed that a careful choice in the separation of the peaks of the transmission profile allows the retrieval of the spectral line profiles with a photometric precision of about  $\sim 2\%$  while requiring 3 times fewer spectral samples. This could be used in the design of future Fabry-Pérot based instruments that would require fewer measurements (higher cadence) and potentially have higher transmission (shorter exposure times).

Future work will include obtaining a more suitable dataset for improving the results from the experiment with IBIS data in section 3.2. In order to apply this deconvolutional approach to real observations in a routine manner we will need training sets consisting of low and high resolution data of a variety of regions on the Sun. One approach would be to obtain nearly simultaneous observations of the same region of the Sun with low and high spectral resolution instruments at comparable spatial resolution. Another feasible way to create the training dataset is by numerically convolving data from a high-spectral-resolution instrument with the known sPSF of the low-spectral-resolution instrument to generate simulated observations. Both approaches have advantages and disadvantages but provide alternative approaches to real world applications of this method. We therefore hope that future instruments will consider the approaches described here and in Asensio Ramos and López Ariste (2010) to leverage the advantages of machine learning and compressive sensing to more efficiently retrieve information from the solar spectrum and further our understanding of the Sun.

## DATA AVAILABILITY STATEMENT

The datasets and the code for this study can be found in the github repository of the author, [https://github.com/momomolnar/SPSF\\_remove](https://github.com/momomolnar/SPSF_remove).

## AUTHOR CONTRIBUTIONS

KR proposed the setup of the two problems and suggested the approach to solving it and obtained the FISS data. KR and MM acquired the IBIS data at the DST. MM reduced the IBIS data and constructed the CNNs with help from CO and IM. MM performed the tests of the accuracy of the methods. All authors contributed to the manuscript.

## FUNDING

MM was supported for the work on this article from the GEH fellowship provided from the University of Colorado, Boulder. Funding for the DKIST Ambassadors program was provided by the National Solar Observatory, a facility of the National Science Foundation, operated under Cooperative Support Agreement Number AST-1400405. CO acknowledges support from the UK's Science and Technology Facilities Council (STFC) doctoral training grant ST/R504750/1.

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Available online at: tensorflow.org.
- Asensio Ramos, A. (2010). Compressed sensing for next generation instruments. *Astron. Nachrich.* 331:652. doi: 10.1002/asna.201011394
- Asensio Ramos, A. and López Ariste, A. (2010). Compressive sensing for spectroscopy and polarimetry. *Astron. Astrophys.* 509:A49. doi: 10.1051/0004-6361/200913019
- Asensio Ramos, A., Socas Navarro, H., Lopez Ariste, A., and Martinez Gonzalez, M. J. (2007). The intrinsic dimensionality of spectropolarimetric data. *Astrophys. J.* 660, 1690–1699. doi: 10.1086/513069
- Baranyi, T. (1986). Study of solar  $\text{H}\alpha$  line profiles by means of filtergrams. *Publ. Debrecen Heliophys. Observ.* 6:25.
- Baranyi, T., and Ludmány, A. (1983). Synthesis of  $\text{H}\alpha$ -profiles from filter transmission functions. *Publ. Debrecen Heliophys. Observ.* 5, 595–602.
- Beckers, J. M., Dickson, L., and Joyce, R. S. (1975). Observing the sun with a fully tunable Lyot-Öhman filter. *Appl. Opt.* 14, 2061–2066. doi: 10.1364/AO.14.002061
- Caccin, B., Falciari, R., Roberti, G., Sambuco, A. M., and Smaldone, L. A. (1983). Bidimensional analysis of solar active regions and flares - part one - imaging spectroscopy with universal birefringent filters. *Solar Phys.* 89, 323–339. doi: 10.1007/BF00217254
- Caccin, B., and Roberti, G. (1979). A method for a rough reconstruction of line profiles from a series of narrow-band filtergrams. *Mem. Soc. Astron. Ital.* 50, 393–404.
- Cauzzi, G., Reardon, K., Rutten, R. J., Tritschler, A., and Uitenbroek, H. (2009). The solar chromosphere at high resolution with IBIS. IV. Dual-line evidence of heating in chromospheric network. *Astron. Astrophys.* 503, 577–587. doi: 10.1051/0004-6361/200811595
- Cavallini, F. (2006). IBIS: A new post-focus instrument for solar imaging spectroscopy. *Solar Phys.* 236, 415–439. doi: 10.1007/s11207-006-0103-8
- Chae, J., Park, H.-M., Ahn, K., Yang, H., Park, Y.-D., Nah, J., et al. (2013). Fast imaging solar spectrograph of the 1.6 meter new solar telescope at big bear solar observatory. *Solar Phys.* 288, 1–22. doi: 10.1007/s11207-012-0147-x
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. Available online at: <http://www.deeplearningbook.org>.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90
- Kentischer, T. J., Schmidt, W., von der Lühe, O., Sigwarth, M., Bell, A., Halbgewachs, C., et al. (2012). "The visible tunable filtergraph for the ATST," in *Proceedings Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 844677 (Amsterdam).
- Kingma, D. P., and Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. arXiv[Preprint]. arXiv:1412.6980. Available online at: <https://arxiv.org/abs/1412.6980>.
- Levina, E., and Bickel, P. J. (2004). "Maximum likelihood estimation of intrinsic dimension," in *Advances in Neural Information Processing Systems 17*. (Vancouver, BC: MIT Press), 777–784.
- Reardon, K. P., and Cavallini, F. (2008). Characterization of Fabry-Perot interferometers and multi-etalon transmission profiles. The IBIS instrumental profile. *Astron. Astrophys.* 481, 897–912. doi: 10.1051/0004-6361:20078473
- Richardson, W. H. (1972). Bayesian-based iterative method of image restoration\*. *J. Opt. Soc. Am.* 62, 55–59. doi: 10.1364/JOSA.62.000055
- Tritschler, A., Rimmele, T. R., Berukoff, S., Casini, R., Kuhn, J. R., Lin, H., et al. (2016). Daniel K. Inouye Solar Telescope: high-resolution observing of the dynamic Sun. *Astron. Nachrich.* 337:1064. doi: 10.1002/asna.201612434
- van Noort, M., Rouppe van der Voort, L., and Löfdahl, M. G. (2005). Solar image restoration by use Of multi-frame blind de-convolution with multiple objects and phase diversity. *Solar Phys.* 228, 191–215. doi: 10.1007/s11207-005-782-z

## ACKNOWLEDGMENTS

MM would like to thank the organizers of the ML in Helio conference for the support to attend this great workshop. CO is grateful to the members of the National Solar Observatory for many scintillating discussions. The authors would like to thank Kyeore Lee, Jongchul Chae, and Kwangsu Ahn for generously providing the FISS spectra. Furthermore, the authors would like to thank the referees for their comments which helped improve the manuscript and Andres Asensio Ramos for the useful discussions which helped with the CNN approach and the Maximum Likelihood Dimensionality Estimate. The National Solar Observatory (NSO) is operated by the Association of Universities for Research in Astronomy, Inc. (AURA), under cooperative agreement with the NSF. IBIS has been designed and constructed by the INAF/Osservatorio Astrofisico di Arcetri with contributions from the Università di Firenze, the Università di Roma Tor Vergata, and upgraded with further contributions from NSO and Queens University Belfast.

**Conflict of Interest:** 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.

Copyright © 2020 Molnar, Reardon, Osborne and Milić. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Analyzing AIA Flare Observations Using Convolutional Neural Networks

Teri Love\*, Thomas Neukirch and Clare E. Parnell

School of Mathematics and Statistics, University of St Andrews, St. Andrews, United Kingdom

## OPEN ACCESS

### Edited by:

Sophie A. Murray,  
Trinity College Dublin, Ireland

### Reviewed by:

Michael S. Kirk,  
The Catholic University of America,  
United States

James Paul Mason,  
University of Colorado Boulder,  
United States

Mark Cheung,  
Lockheed Martin Solar and  
Astrophysics Laboratory (LMSAL),  
United States

### \*Correspondence:

Teri Love  
tl28@st-andrews.ac.uk

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 02 April 2020

**Accepted:** 26 May 2020

**Published:** 26 June 2020

### Citation:

Love T, Neukirch T and Parnell CE  
(2020) Analyzing AIA Flare  
Observations Using Convolutional  
Neural Networks.  
*Front. Astron. Space Sci.* 7:34.  
doi: 10.3389/fspas.2020.00034

In order to efficiently analyse the vast amount of data generated by solar space missions and ground-based instruments, modern machine learning techniques such as decision trees, support vector machines (SVMs) and neural networks can be very useful. In this paper we present initial results from using a *convolutional neural network* (CNN) to analyse observations from the *Atmospheric Imaging Assembly* (AIA) in the 1,600 Å wavelength. The data is pre-processed to locate flaring regions where flare ribbons are visible in the observations. The CNN is created and trained to automatically analyse the shape and position of the flare ribbons, by identifying whether each image belongs into one of four classes: two-ribbon flare, compact/circular ribbon flare, limb flare, or quiet Sun, with the final class acting as a control for any data included in the training or test sets where flaring regions are not present. The network created can classify flare ribbon observations into any of the four classes with a final accuracy of 94%. Initial results show that most of the images are correctly classified with the compact flare class being the only class where accuracy drops below 90% and some observations are wrongly classified as belonging to the limb class.

**Keywords:** solar flares, ribbons, machine learning, classification, CNNs

## 1. INTRODUCTION

The steady improvement of technology and instrumentation applied to solar observations has led to the generation of vast amounts of data, for example the *Solar Dynamics Observatory* (SDO) collects approximately 1.5 terabytes of data everyday (Pesnell et al., 2012). The analysis of these data products can be made much more efficient by the use of modern machine learning techniques such as decision trees, support vector machines (SVMs) and neural networks. In this paper we describe some initial results we obtain using a convolutional neural network (CNN) to analyse SDO data. Basic applications of CNNs to solar physics data classification is shown in e.g., Kucuk et al., 2017; Armstrong and Fletcher, 2019, however CNNs have also started being applied to the prediction of solar events, in particular flares and CMEs, that can affect space weather as considered, for example, by Bobra and Couvidat (2015), Nagem et al. (2018), and Fang et al. (2019).

In this paper we focus on solar flares and in particular on the classification of the morphology of flares displaying visible flare ribbons (e.g., Kurokawa, 1989; Fletcher and Hudson, 2001). Throughout this paper, flare observations from the *Atmospheric Imaging Assembly* (AIA) Lemen et al. (2012) onboard SDO were used, specifically AIA 1,600 Å. These observations clearly show the flare ribbons as they appear on the solar surface.

The locations and shapes of flare ribbons are thought to be closely linked to the geometry and topology of the solar magnetic field in the flaring region. For example, the ribbon shapes and lengths have been connected to the presence of separatrix surfaces and quasi-separatrix layers (QSLs) (e.g.,

Aulanier et al., 2000; Savcheva et al., 2015; Janvier et al., 2016; Hou et al., 2019). The ribbon shapes found and analyzed throughout these studies are mostly two-ribbon flares with two “J” shaped ribbons, however it is known that other ribbon shapes can also occur with circular or compact flare ribbons also being observed. One motivation of the work presented in this paper is to create a tool that allows the classification of large data sets to generate a catalog of flares associated with their ribbons, which could automatically be detected and classified. The catalog could then, for example, be used in connection with magnetic field models to obtain better statistics on the possible correlation of ribbon geometry and magnetic field structure.

This paper considers all C, M, and X class flares (see e.g., Fletcher et al., 2011, for a definition of GOES classes) that occurred between November 2012 and December 2014 and attempts to classify the shape of all observable flare ribbons. To do this a CNN consisting of two hidden layers was created and trained to predict four classes of ribbons and flares. These four classes are two-ribbon flares, limb flares and circular/compact ribbon flares, with the fourth class acting as a control class to process quiet Sun images that may also be processed through the CNN. The network was trained on a dataset containing 540 images (including validation images), and was tested using an unseen dataset containing 430 images.

The paper is structured as follows. In section 2, we describe the design and training of our CNN. The preparation of the data used in the paper is discussed in section 3, our results are presented in section 4 and we conclude with a discussion of our findings in section 5.

## 2. METHODS

Convolutional neural networks (CNNs) are a type of machine learning technique commonly used to find patterns in data and classify them. Instead of being given explicit instructions or mathematical functions to work they use patterns and trends in the data, initially found through a “training data” set. This data set should be the set of inputs for the CNN—usually a subset of the data that one would initially want to classify or detect. This allows the network to “learn” the patterns and trends such that it can independently classify unknown data.

### 2.1. CNN Design

To create a basic CNN there must be at least 3 layers; an input layer, a hidden layer and an output layer (e.g., Cun et al., 1990; Hinton et al., 2012; LeCun et al., 2015; Szegedy et al., 2015; Krizhevsky et al., 2017). The input layer is the first initial network layer which accesses the data arrays inputted into the model which are to be trained upon. The data input has usually been through some pre-processing before being used by the network, the pre-processing used on the AIA data is discussed in section 3.

The hidden layer is a convolutional layer where instead of applying a layer using matrix multiplication, as in general neural networks, a layer using a mathematical convolution is used instead. Although this is the basic set-up for a CNN, most CNNs have multiple hidden layers before having a fully connected output layer. The different types of hidden layers that can be used

are: convolutional, pooling, dropout (Hinton et al., 2012) and fully connected layers. The final output layers are usually built from fully connected (*dense*) layers. These layers take the output from the hidden layers and process it such that for each data file a pre-defined class is predicted by the network.

A convolutional layer performs an image convolution of the original image data using a kernel to produce a *feature map*. These kernels can be any size but are commonly chosen to be of size  $3 \times 3$ . The stride of the kernel can also be set in the convolutional layers indicating how many pixels it should skip before applying the kernel to the input—this has been set as 1 for the CNN here such that the kernel has been applied to every pixel in the input. If larger features were to be classified larger strides could be used.

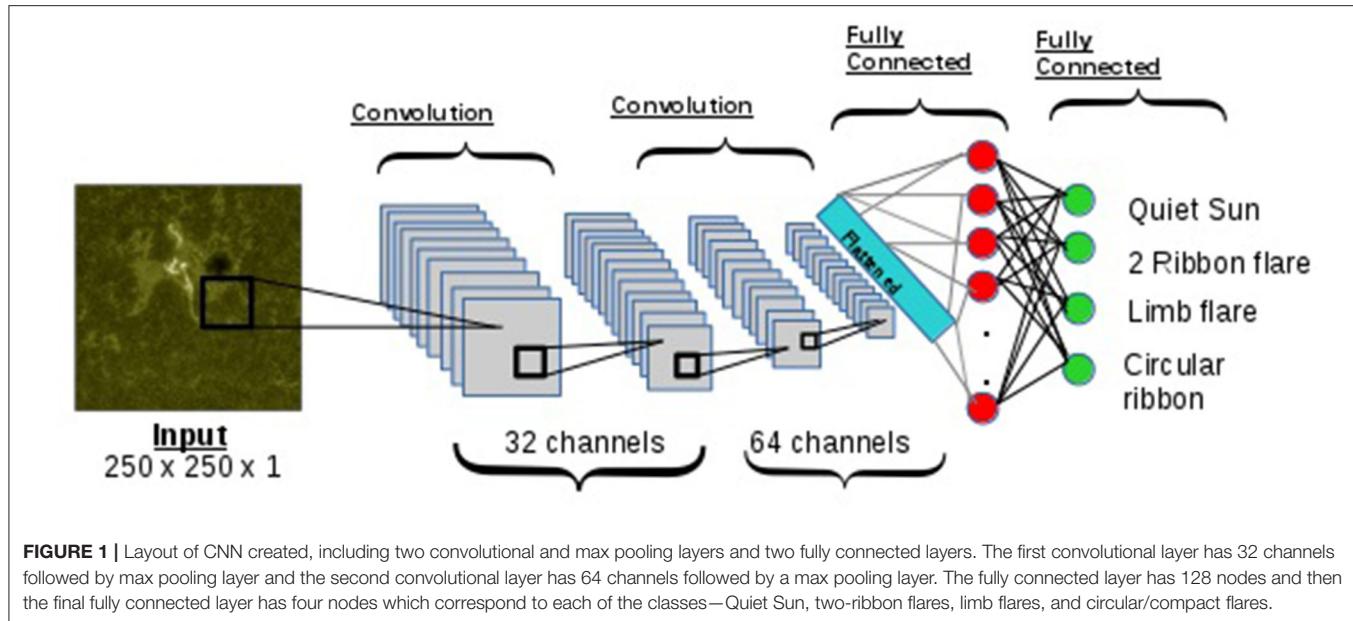
The kernel moves over every point in the input data, producing a single value for each  $3 \times 3$  region by summing the result of a matrix multiplication. The value produced is then placed into a feature map which is passed onto the next layer. As the size of the feature map will be smaller than the input, the feature map is padded with zeros to ensure the resulting data is the same size as the original input. After the feature map is produced the convolutional layer has an associated *activation function* which produces non-linear outputs for each layer and determines what signal will be passed onto the next layer in the network. A common activation function used is the *rectified linear unit* (ReLU, Nair and Hinton, 2010), which is defined by;

$$f(x) = \max(0, x). \quad (1)$$

Other activation functions such as *linear*, *exponential* or *softmax* (see Equation 3) can also be implemented, however for the convolutional layers in our model only ReLU is used, as the function can be calculated and updated quickly with gradients remaining high (close to 1), with ReLU also avoiding the vanishing gradient problem.

Although convolutional layers make up the majority of the hidden layers within a CNN, other hidden layers are also important to avoid over-training of the network. Implemented after convolutional layers, *pooling* layers are commonly used to deal with this. Pooling layers help to reduce over-fitting of data and reduce the number of parameters produced throughout training—which causes the CNN to learn faster. The most common type of pooling is *max pooling* which takes the maximum value in each window and places them in a new and smaller feature map. The window size for the max pooling layers can be varied similarly to the convolutional kernels, however throughout this paper all max pooling layers had a kernel of size  $2 \times 2$ . Although the feature map size is being reduced, the max pooling layers will keep the most important data and pass it onto the next training steps.

For the CNN created to analyse the flare ribbons, two convolutional layers were implemented after the input layer. These layers were both followed by max-pooling layers with a stride of 2. Both layers were implemented using ReLU activation functions, however the first convolutional layer had 32 nodes



**FIGURE 1 |** Layout of CNN created, including two convolutional and max pooling layers and two fully connected layers. The first convolutional layer has 32 channels followed by max pooling layer and the second convolutional layer has 64 channels followed by a max pooling layer. The fully connected layer has 128 nodes and then the final fully connected layer has four nodes which correspond to each of the classes—Quiet Sun, two-ribbon flares, limb flares, and circular/compact flares.

**TABLE 1 |** Details pf each CNN layer with the number of filters, size of kernels, and activation functions used shown.

Layer	Number of nodes	Kernel size (weights)	Stride	Activation function
Convolution	32	$3 \times 3$	1	ReLU
Max Pooling	/	$2 \times 2$	2	/
Convolution	64	$3 \times 3$	1	ReLU
Max Pooling	/	$2 \times 2$	2	/
Fully Connected	128	$61 * 61 * 64 * 128$	/	ReLU
Fully Connected (Output)	4	$128 \times 4$	/	Softmax

whereas the second layer was implemented with 64 nodes before being passed onto fully connected layers.

Once the convolutional and pooling layers have been implemented as hidden layers, the final feature map output is passed onto output layers which allows the data to be classified. These classification layers are made up of fully connected (FC) layers—similar to those in a normal neural network. FC layers only accept one-dimensional data and so the data must be flattened before being passed into them. The neurons in the FC layers have access to all activations in previous layers—this allows them to classify the inputs. The final fully connected layer should have the number of classes as its units, with each output node representing a class.

An additional output layer that can be implemented before a FC layer is a *dropout layer*. This layer is implemented before a FC layer to indicate that random neurons should be ignored in the next layer i.e., they have *dropped out* of the training algorithm for the current propagation. Hence if a FC layer is indicated to have 10 neurons, a random set of these will be ignored when training (see e.g., Hinton et al., 2012, for further information on dropout layers).

The CNN was created and trained using Keras Chollet et al. (2015), with the network layout shown in **Figure 1**. This shows

the two convolutional and pooling layers previously discussed, with a dropout layer implemented before the data is passed onto two FC layers, with 128 and 4 nodes, respectively. A breakdown of all parameters used in each layer are shown in **Table 1**.

## 2.2. Model Training

The previous section described the basic design of the CNN used throughout this paper. Here we will describe the training process carried out on the model.

When data is passed through the network, at each layer a loss function is used to update the model weights. This loss function carries out the process known as *back-propagation* (Hecht-Nielson, 1989), where differentiation takes place and the network learns the optimal classifications for each training image. The loss function chosen for our model is known as *categorical cross entropy*. This cross entropy loss is calculated as follows;

$$CEP = - \sum_{i=1}^M y(x_i) \log(p(x_i)), \quad (2)$$

where  $M$  is the number of classes (here  $M = 4$ ) and  $y$  is the binary indicator (0 or 1) such that if  $y = 1$  the observation belongs to the

class and  $y = 0$  if it does not. Finally  $p$  is the probability that the observation belongs to a class,  $c$ .

The probability,  $p(x_i)$ , of each class is calculated using a softmax distribution such that;

$$p(x_i) = \frac{e^{x_i}}{\sum_k e^{x_i}}. \quad (3)$$

This function should tend toward 1 if an observation belongs to a single class and tends to 0 for the other 3 classes to indicate that the network does not recognize it as belonging to those classes. The resultant classification is selected by choosing the largest probability that lies above  $p(x_i) = 0.5$ .

The network is trained on 540 1,600 Å AIA images. The data processing is discussed in section 3, with each image used containing a single flare, unless it belongs to the quiet Sun class. The four classes are as follows:

#### 1. Quiet Sun

No brightenings present on the surface, hence should give an indication of general background values (It should be noted that none of these observations are taken on the limb).

#### 2. Two-ribbon Flare

Two flare ribbons must be clearly defined in the observations. However, the shape does not matter here e.g., if there are 2 semi circular ribbons the flare is classified as a two-ribbon flare and not a circular flare.

#### 3. Limb Flares

The solar limb must be clearly observed in this snapshot observation with a flare brightening being visible. The limb class was chosen to start at a specific distance from the solar limb to reduce confusion with other classes. This will be discussed further in section 4.

#### 4. Circular Flare Ribbons

Here a circular ribbon shape of any size must be observed. It should be ideally a singular ribbon so as not to be confused with the two-ribbon flare class. Compact flares were also included here, they appear in the data as round “dot” like shapes.

Classes were divided almost evenly to stop observational bias from entering the model during training and although there is a slight class imbalance it is not large enough to affect the accuracy of the model. From the training set used, 40% of the data was used as a validation data set with the remaining 60% used to train the model. The learning rate chosen was  $10^{-4}$  with a batch size of 32 selected for both training and validation to allow the use of mini-batch gradient descent throughout training. Although larger batch sizes would speed up the training process, to get better generalization of the model a smaller batch size was picked to improve the model accuracy.

**Figure 2** shows the results from training and testing the model. **Figures 2A,B** show the results from training, with the training and validation accuracy plotted in **Figure 2A**. It is shown that the network was trained only for 10 epochs to prevent over-fitting. The training accuracy was 98% and the validation accuracy was slightly lower at 94%, these are excellent accuracies for the number of epochs used. **Figure 2B** shows the

training and validation loss for the same number of epochs. Both losses fall quite sharply and then start to level off, these could be improved with a larger data set which could be run for more epochs. The loss leveling out indicates that training should be stopped to prevent over-fitting and further improvements can be made from creating larger data sets. To further validate the training process and its outputs, k-fold cross validation was implemented, similar to that implemented by Bobra and Couvidat (2015). The loss and accuracy values from five-fold cross validation are shown in **Figure 2E**, with the mean accuracy across the five-folds being approximately  $92.9 \pm 2.98\%$ .

## 3. DATA PREPARATION

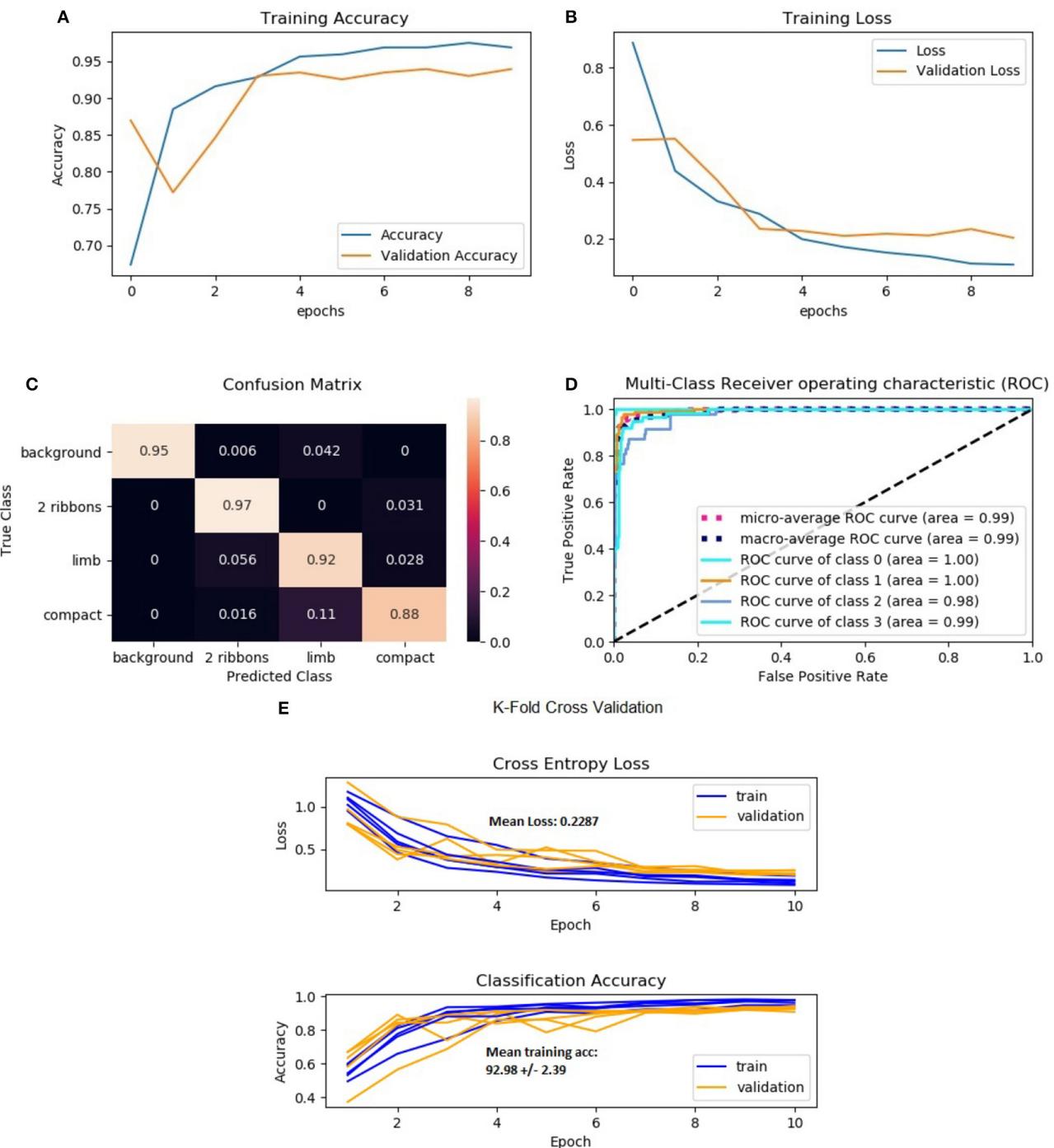
To create a neural network that can analyse the flare ribbons observed, a robust data set of flaring regions and their ribbons was created. The data set must be created from observations from the same wavelength and instrument to ensure the CNN will not train on varying parameters such as wavelength or smaller features that would perhaps only be found by using a certain instrument. Due to this the data has been collected from the AIA on board the *Solar Dynamics Observatory* (SDO) at the 1,600 Å wavelength. This wavelength has been chosen as it observes the upper photosphere and transition region allowing for a clearer view of the flare ribbons than those observed in the EUV wavelengths.

To find dates where flares were observed on the solar disk, the flare ribbon database created by Kazachenko et al. (2017) was used. From this database all flares that occurred between November 2012 and December 2014 were included in the training set, this included all C, M, and X class flares. To create a training set all of the flares included must be labeled as belonging to a class that is defined for the CNN. Flares where ribbons were not well-defined were removed from the data set. This resulted in a training set containing 540 image samples with 160 quiet Sun regions, 160 two-ribbon flares, 95 limb flares, and 125 circular flares.

When creating the training and test sets, flares have been chosen such that they should clearly fall into a particular class. To be able to classify each image the following process was implemented.

For each flare, the observation was chosen at peak flare time according to the *Heliosysics Event Knowledgebase* (HEK) (Hurlburt et al., 2012). It should be noted that this means the CNN does not take into account the evolution of the flare ribbons from the start to the end of the flare, although this is something that could possibly be included in further work. For some observations there is more than one flare present and in this case both regions are processed and classified separately, although they occurred on the solar disc simultaneously.

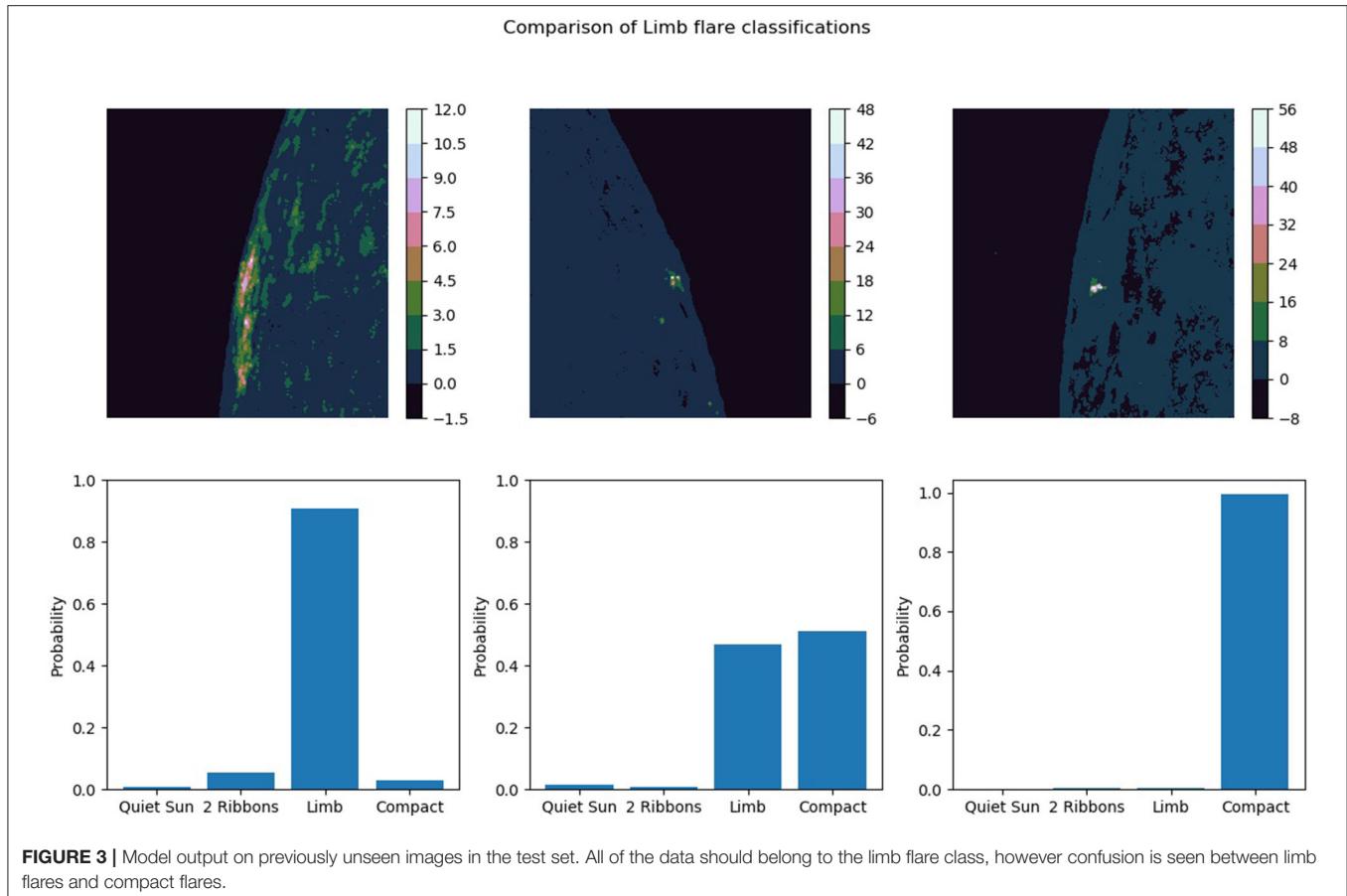
Once the flare position has been located, a bounding box is created around the central flare position. For each flare this creates a bounding box of size  $500 \times 500 \times 1$  pixels. This step was included to reduce the size of the data the neural network would



**FIGURE 2 | (A)** Training accuracies with both validation and training accuracies shown over 10 epochs; **(B)** Training and validation loss shown over 10 epochs; panel **(C)** shows the confusion matrix created on the test set, with the diagonal showing the correctly identified ribbon types; panel **(D)** shows the receiver operating characteristic (ROC) curve which has been modified to include a curve for each class and the micro and macro average curves; panel **(E)** shows the results for loss and accuracy whilst using k-fold cross validation, where  $k = 5$ .

have to process due to large data sets increasing the number of training parameters quickly. The original AIA level 1 data files are  $4,096 \times 4,096 \times 1$  in size, hence this step allows the data

input size to be drastically reduced. This code works in a similar way to that of an object detector creating bounding boxes around objects to be classified.



Once located each image is labeled manually according to the classes previously discussed; the quiet sun, two-ribbon flares, circular/compact ribbon flares, or limb flares. Once one of these has been chosen, the label is entered into an array ready for training the CNN.

Once each image has been classified the final steps of the data preparation is to ensure all ROIs were of a suitable size for the CNN to process, hence the data was down-sampled so each image was of size  $250 \times 250 \times 1$ . Hence the final set of input data would be of size  $n \times 250 \times 250 \times 1$ , where  $n$  is the total number of ROI samples contained within the training data.

The final step for the data preparation was to normalize the data slightly before training, this will ensure the best results when training the CNN and so all of the ROIs were normalized using their z-scores as follows:

$$\text{normalized} = \frac{\text{data} - \text{mean}(\text{data})}{\text{standard deviation}(\text{data})}.$$

Once all of the above processes had been carried out on the observations the CNN could begin training as discussed in section 2.2.

## 4. RESULTS

Once training was completed the network was tested using a previously unseen data set. This test set contained 430 images consisting of 160 quiet sun images, 160 two-ribbon flares, 47 limb flares, and 63 circular ribbon. Note that some flares included in the test data may have occurred in the same active regions as images included in the training data set. The test outputs are shown using a confusion matrix and ROC curves as shown in Figures 2C,D.

A confusion matrix is a good way to visualize model performance on test data that has already been labeled. It summarizes the number of correct and incorrect classifications and shows them by plotting the predicted classes against the true classes of the data. The confusion matrix shown in Figure 2C indicates the percentage of data correctly classified by the diagonal. It shows that for the quiet Sun, two-ribbons and limb classes approximately 95% of all test data was correctly classified, however for compact flares only 88% of the data is being correctly classified with approximately 11% being incorrectly classified as limb flares. This may be due to the distortion of ribbons on the limb, making them look almost compact or circular in shape. The 11% being incorrectly classified could possibly be corrected by training the model further on a larger data set.

**Figure 2D** shows multiple *receiver operating characteristic* (ROC) curves. A ROC curve is plotted as the true positive rate (TPR) against the false positive rate (FPR) at various thresholds. The area under the ROC curve (AUC) indicates the performance of the model as a classifier. The closer to 1 the AUC is indicates how well the model works, with 0 indicating that the model is not classifying anything correctly. hence the further to the left of the diagonal the ROC curve lies the better the classifier. The ROC curves in **Figure 2D** show how well the model works for each class, with high AUC values found—all approximately 99%.

To further investigate the model outputs for the limb class, three different images from the test set were considered. **Figure 3** shows these three flares and their probabilities of belonging to each class. The first flare is clearly identified as a limb flare with the flaring region sitting just away from the limb. For the second flare it is shown that the model is confused, with very little difference in the confidence that the flare is either a compact or limb flare, both with approximately 50% probability that the flare could belong to either class. For the final limb flare considered, the model is almost 100% confident that the flare belongs to the compact/circular ribbon class. This may be due to the flare being slightly further from the limb and so instead of picking up the limb region and the flare, the network has only identified the flare which looks to belong to the circular ribbon class. To rectify this problem in further work some changes to the network and its input could be applied, this could include the inclusion of spatial co-ordinates as one of the inputs which could help with the confusion about which images belong to the limb class.

## 5. DISCUSSION

In this paper, we have demonstrated a basic application of CNNs to solar image data. In particular, the model classifies the shapes of solar flare ribbons that are visible in 1,600 Å AIA observations. The four classes chosen (Quiet Sun, two-ribbons, Limb flares, Compact/Circular ribbons) were picked due to there being obvious differences between each class, hence more complicated classes could have been chosen but may have effected the overall performance of the CNN. Each of the classes chosen when tested were all found to be well-defined with most of the images being correctly classified by the network, with an overall accuracy of approximately 94%.

The network created is a shallow CNN with only two convolutional layers, unlike deeper networks used on solar image data (Kucuk et al., 2017; Armstrong and Fletcher, 2019). Both of these papers tried to classify solar events such as flares, coronal holes and sunspots, with varying instruments used. However, even with such a shallow CNN as used here, the accuracy of the overall model is still good at approximately 96%. Our model currently focuses on flare ribbon data and Analyzing their positions and shapes. This model and data could be compared to a similar setup used to analyse the MNIST dataset containing variations of the numbers 0 to 9 (e.g., Lecun et al., 1998).

However, to generalize the model further training could be carried out on features such as sunspots or prominences which can also be viewed in the current wavelength, although to do this a deeper network would be needed to extract finer features in the data. Varying the image wavelengths for the AIA data or using a different instrument such as SECCHI EUVI observations from STEREO (*Solar Terrestrial Relations Observatory*) or EIS EUV observations from Hinode could also make the model more robust.

If it was chosen to implement more layers in the network, a CNN such as the VGG network could be used (Simonyan and Zisserman, 2014). These networks would take longer to train, particularly on larger data sets containing more images and classes and would require more epochs to properly train the network. As well as increasing the number of convolutional layers used, other layers or parameters could also be modified to alter the model speed and performance. The parameters discussed in **Table 1** could all be altered to affect the model speed and accuracy.

The main result from this paper shows that even with a shallow CNN we can get excellent accuracy in the dataset that we considered here. Such a result is encouraging and shows basic CNNs can be very useful tools in analyzing large datasets. The model created in this paper can be applied to other data pipelines and can be used to locate many more features from Solar observations obtained from both space and ground-based instruments.

## DATA AVAILABILITY STATEMENT

The research data supporting this publication can be accessed at <https://doi.org/10.17630/fa62b9e5-4bd5-4c35-82db-4910d3df62f5>. The code can be found at <https://github.com/Terilove/AIARibbonCNN.git>.

## AUTHOR CONTRIBUTIONS

TL created the neural network and carried out the data analysis. TN and CP regularly contributed to the project intellectually by providing ideas and guidance. All authors contributed to the writing of the paper.

## FUNDING

TL acknowledges support by the UK's Science and Technology Facilities Council (STFC) Doctoral Training Centre Grant ST/P006809/1 (ScotDIST). TN and CP both acknowledge support by the STFC Consolidated Grant ST/S000402/1.

## ACKNOWLEDGMENTS

The authors would like to thank the continued support from STFC. The AIA data used are provided courtesy of NASA/SDO and the AIA science team.

## REFERENCES

- Armstrong, J., and Fletcher, L. (2019). Fast solar image classification using deep learning and its importance for automation in solar physics. *Sol. Phys.* 294:80. doi: 10.1007/s11207-019-1473-z
- Aulanier, G., DeLuca, E. E., Antiochos, S. K., McMullen, R. A., and Golub, L. (2000). The topology and evolution of the bastille day flare. *Astrophys. J.* 540, 1126–1142. doi: 10.1086/309376
- Bobra, M. G., and Couvidat, S. (2015). Solar flare prediction using SDO/HMI Vector magnetic field data with a machine-learning algorithm. *Astrophys. J.* 798:135. doi: 10.1088/0004-637X/798/2/135
- Chollet, F. et al. (2015). *Keras*. Available online at: <https://github.com/fchollet/keras>
- Cun, Y. L., Boser, B., Denker, J. S., Howard, R. E., Hubbard, W., Jackel, L. D., et al. (1990). *Handwritten Digit Recognition with a Back-Propagation Network*. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Fang, Y., Cui, Y., and Ao, X. (2019). Deep learning for automatic recognition of magnetic type in sunspot groups. *Adv. Astron.* 2019:9196234. doi: 10.1155/2019/9196234
- Fletcher, L., Dennis, B. R., Hudson, H. S., Krucker, S., Phillips, K., Veronig, A., et al. (2011). An observational overview of solar flares. *Space Sci. Rev.* 159:19. doi: 10.1007/s11214-010-9701-8
- Fletcher, L., and Hudson, H. (2001). The magnetic structure and generation of EUV flare ribbons. *Solar Phys.* 204, 69–89. doi: 10.1023/A:1014275821318
- Hecht-Nielson, R. (1989). “Theory of the backpropagation neural network,” in *International 1989 Joint Conference on Neural Networks* (Washington, DC), 593–605. doi: 10.1109/IJCNN.1989.118638
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. Available online at: <https://arxiv.org/pdf/1207.0580v1.pdf>
- Hou, Y., Li, T., Yang, S., and Zhang, J. (2019). A secondary fan-spine magnetic structure in active region 11897. *Astrophys. J.* 871:4. doi: 10.3847/1538-4357/aaef4f
- Hurlburt, N., Cheung, M., Schrijver, C., Chang, L., Freeland, S., Green, S., et al. (2012). Heliophysics event knowledgebase for the solar dynamics observatory (SDO) and beyond. *Sol. Phys.* 275, 67–78. doi: 10.1007/s11207-010-9624-2
- Janvier, M., Savcheva, A., Pariat, E., Tassev, S., Millholland, S., Bommier, V., et al. (2016). Evolution of flare ribbons, electric currents, and quasi-separatrix layers during an X-class flare. *Astron. Astrophys.* 591:A141. doi: 10.1051/0004-6361/201628406
- Kazachenko, M., Lynch, B. J., Welsch, B. T., and Sun, X. (2017). A database of flare ribbon properties from the solar dynamics observatory. I. reconnection flux. *Astrophys. J.* 845:49. doi: 10.3847/1538-4357/aa7ed6
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386
- Kucuk, A., Banda, J., and Angryk, R. (2017). “Solar event classification using deep convolutional neural networks,” in *Artificial Intelligence and Soft Computing, ICAISC 2017* (Cham: Springer), 118–130. doi: 10.1007/978-3-319-59063-9\_11
- Kurokawa, H. (1989). High-resolution observations of H $\alpha$  flare regions. *Space Sci. Rev.* 51, 49–84. doi: 10.1007/BF00226268
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791
- Lemen, J. R., Title, A. M., Akin, D. J., Boerner, P. F., Chou, C., Drake, J. F., et al. (2012). The atmospheric imaging assembly (AIA) on the solar dynamics observatory (SDO). *Sol. Phys.* 275, 17–40. doi: 10.1007/978-1-4614-3673-7\_3
- Nagem, T., Qahwaji, R., Ipson, S., Wang, Z., and Al-Waisy, A. (2018). Deep learning technology for predicting solar flares from (geostationary operational environmental satellite) data. *Int. J. Adv. Comput. Sci. Appl.* 9, 492–498. doi: 10.14569/IJACSA.2018.090168
- Nair, V., and Hinton, G. E. (2010). “Rectified linear units improve restricted Boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML-10* (Madison, WI: Omnipress), 807–814.
- Pesnell, W. D., Thompson, B. J., and Chamberlin, P. C. (2012). The solar dynamics observatory (SDO). *Sol. Phys.* 275, 3–15. doi: 10.1007/s11207-011-9841-3
- Savcheva, A., Pariat, E., McKillop, S., McCauley, P., Hanson, E., Su, Y., et al. (2015). The relation between solar eruption topologies and observed flare features. I. Flare ribbons. *Astrophys. J.* 810:96. doi: 10.1088/0004-637X/810/2/96
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. Available online at: <https://arxiv.org/pdf/1409.1556.pdf>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). “Going deeper with convolutions,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA), doi: 10.1109/CVPR.2015.7298594

**Conflict of Interest:** 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.

Copyright © 2020 Love, Neukirch and Parnell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Incorporating Physical Knowledge Into Machine Learning for Planetary Space Physics

**Abigail R. Azari<sup>1\*†</sup>, Jeffrey W. Lockhart<sup>2</sup>, Michael W. Liemohn<sup>1</sup> and Xianzhe Jia<sup>1</sup>**

## OPEN ACCESS

### Edited by:

Bala Poduval,  
University of New Hampshire,  
United States

### Reviewed by:

Simon Wing,  
Johns Hopkins University,  
United States

Michael S. Kirk,

The Catholic University of America,  
United States

Ryan McGranaghan,  
ASTRA, LLC, Spain

### \*Correspondence:

Abigail R. Azari  
azari@umich.edu

### †Present address:

Abigail R. Azari,  
Space Sciences Laboratory, University  
of California, Berkeley, Berkeley, CA,  
United States

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 05 March 2020

**Accepted:** 28 May 2020

**Published:** 08 July 2020

### Citation:

Azari AR, Lockhart JW, Liemohn MW  
and Jia X (2020) Incorporating  
Physical Knowledge Into Machine  
Learning for Planetary Space Physics.  
*Front. Astron. Space Sci.* 7:36.  
doi: 10.3389/fspas.2020.00036

<sup>1</sup> Climate and Space Sciences and Engineering Department, University of Michigan, Ann Arbor, MI, United States, <sup>2</sup> Sociology Department, University of Michigan, Ann Arbor, MI, United States

Recent improvements in data collection volume from planetary and space physics missions have allowed the application of novel data science techniques. The Cassini mission for example collected over 600 gigabytes of scientific data from 2004 to 2017. This represents a surge of data on the Saturn system. In comparison, the previous mission to Saturn, Voyager over 20 years earlier, had onboard a ~70 kB 8-track storage ability. Machine learning can help scientists work with data on this larger scale. Unlike many applications of machine learning, a primary use in planetary space physics applications is to infer behavior about the system itself. This raises three concerns: first, the performance of the machine learning model, second, the need for interpretable applications to answer scientific questions, and third, how characteristics of spacecraft data change these applications. In comparison to these concerns, uses of "black box" or un-interpretable machine learning methods tend toward evaluations of performance only either ignoring the underlying physical process or, less often, providing misleading explanations for it. The present work uses Cassini data as a case study as these data are similar to space physics and planetary missions at Earth and other solar system objects. We build off a previous effort applying a semi-supervised physics-based classification of plasma instabilities in Saturn's magnetic environment, or magnetosphere. We then use this previous effort in comparison to other machine learning classifiers with varying data size access, and physical information access. We show that incorporating knowledge of these orbiting spacecraft data characteristics improves the performance and interpretability of machine learning methods, which is essential for deriving scientific meaning. Building on these findings, we present a framework on incorporating physics knowledge into machine learning problems targeting semi-supervised classification for space physics data in planetary environments. These findings present a path forward for incorporating physical knowledge into space physics and planetary mission data analyses for scientific discovery.

**Keywords:** planetary science, automated event detection, space physics, Saturn, physics-informed machine learning, feature engineering, domain knowledge, interpretable machine learning

## 1. INTRODUCTION

Planetary space physics is a young field for large-scale data collection. At Saturn for example, it was only in 2004 that the first Earth launched object orbited this planet (Cassini) and landed on Titan (Huygens). After arriving Cassini collected data about Saturn and its near-space environment for 13 years, resulting in 635 GB of scientific data (NASA Jet Propulsion Laboratory, 2017a). To put this into perspective, the Voyager I mission which flew by Saturn in 1980 had onboard  $\sim$ 70 kB of memory total (NASA Headquarters, 1980). The Cassini mission represents the first large-scale data collection of Saturn. This enabled the field of planetary science to apply statistics to large-scale data sizes, including machine learning, to the most detailed spatio-temporally resolved dataset of the planet and its environment.

This surge of data is not unique to Saturn science. In planetary science broadly, Mars in 2020 has eight active missions roving along the surface and orbiting (Planetary Society, 2020). The Mars Reconnaissance Orbiter alone has already collected over 300 TB of data (NASA Jet Propulsion Laboratory, 2017b). It is commonly accepted that upcoming missions will face similar drastic advances in the collection of scientific data. Traditionally planetary science has employed core scientific methods such as remote observation and theoretical modeling. With the new availability of sampled environments provided by these missions, methods in machine learning offer significant potential advantages. Applying machine learning in planetary space physics differs from other common applications. Cassini's data are characteristic of other planetary and space physics missions like the Magnetospheric Multiscale Mission (MMS) at Earth and the Juno mission to Jupiter. The plasma and magnetic field data collected by these missions are from orbiting spacecraft. This conflates spatial and temporal phenomena. This is a shared characteristic with the broader field of geoscience which often represents complex systems undergoing significant spatio-temporal changes with limitations on quality and resolution (Karpatne et al., 2019).

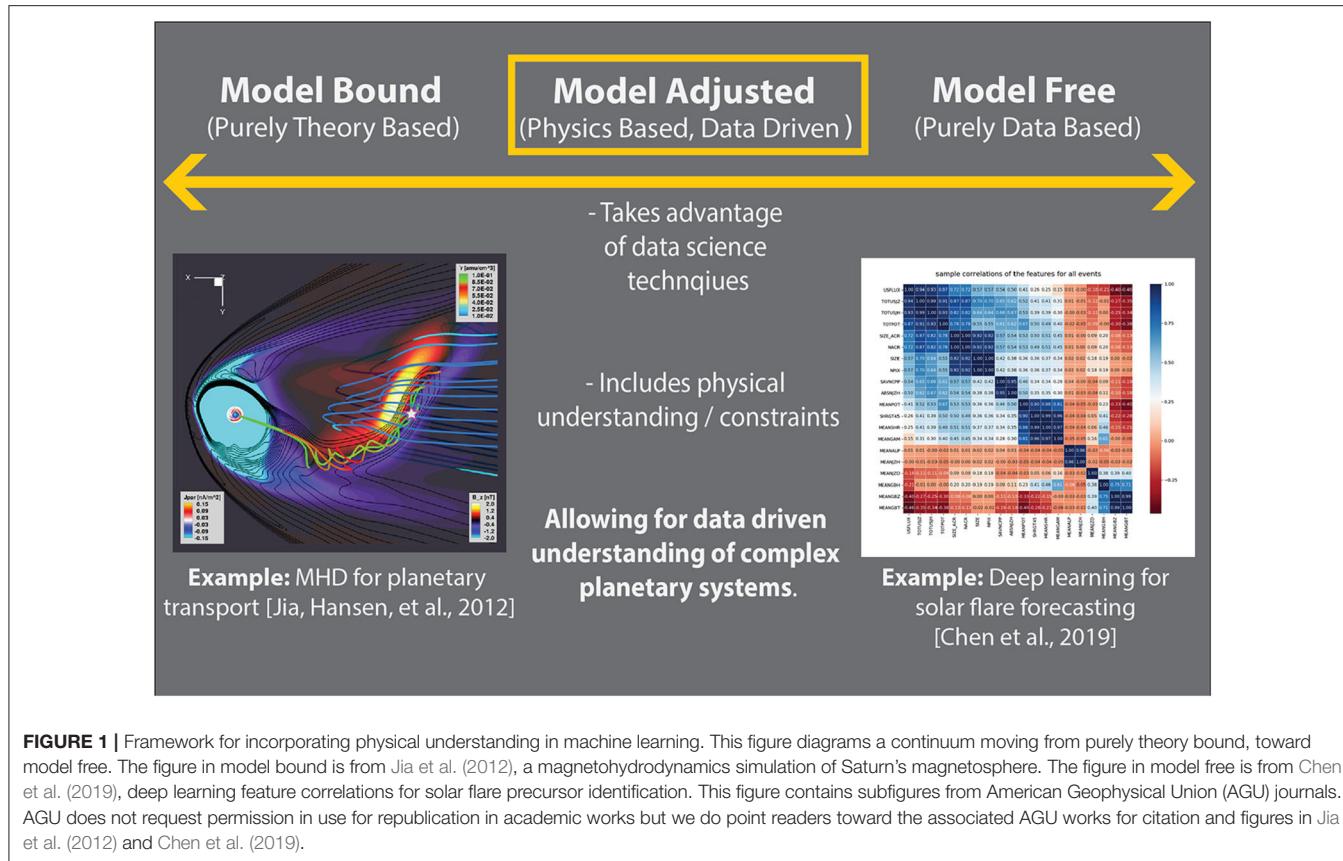
The main use of these data in planetary science is to advance fundamental scientific theories. This requires the ability to infer meaning from applications of statistical methods. Unlike similar missions at Earth, machine learning for space physics data at Saturn has limited direct application to the prediction of space weather. A central interest in space weather prediction is to give lead-time information for operational purposes. As a result, the prediction accuracy in machine learning applications in space weather prediction is seen as paramount. In comparison, at Saturn, machine learning applications require highly interpretable and explainable techniques to investigate scientific questions (Ebert-Uphoff et al., 2019). How to improve machine learning generally from an interpretability standpoint is itself an active research area in domain applications of machine learning (e.g., Molnar, 2019). Within this work we specifically focus on evaluating and implementing interpretable machine learning. Interpretable machine learning usually relies on domain knowledge and is therefore domain specific, but it can be extended to generally refer to models with functional forms simple enough for humans to understand how they make

predictions, such as logical rules or additive factors (Rudin, 2019). Complexity depends in part on what constitutes common knowledge within a domain. Scientists are trained to interpret different models depending on their field. As a result models will range in perceived interpretability across fields. While the final models must be relatively simple in order for humans to understand their decision process, the algorithms which produce optimal interpretable models often require solving computationally hard problems. Importantly, despite widespread myths about performance, interpretable models can often be designed to perform as well as un-interpretable or "black box" models (Rudin, 2019).

In planetary science it's important to discern the workings of a model in order to understand the implications for the workings of physical systems. Interpretability is not the same as explainability: explainability refers to any attempt to explain how a model makes decisions, typically this is done afterwards and without reference to the model's internal workings. Interpretability, however, refers to whether the inner workings of the model, its actual decision process, can be observed and understood (Rudin, 2019). Within this work we are concerned with interpretability in order to gain scientifically actionable results from applied machine learning. The dual challenges of spatio-temporal data and interpretability are compounded for planetary orbiting spacecraft. Complications for orbiting spacecraft can range from rare opportunities for observation, and engineering constraints on spacecraft data transmission. A main interest in this work is to begin to ask: how can machine learning be used within these constraints to answer fundamental scientific questions?

Scientists have approached interpretable machine learning for physics in two ways. First, they have added known physical constraints and relationships into modeling. Within the space weather prediction community, such integration has shown promise in improving the performance of deep learning models over models that do not account for the physics of systems (Swiger et al., 2020). Several fields including biology have argued for an equal value of domain knowledge and machine learning techniques for that reason (see discussion within Coveney et al., 2016). These discussions have culminated in several reviews for scientific fields on the integration of machine learning for data rich discovery (Butler et al., 2018; Bergen et al., 2019). Second, scientists have long tried to use machine learning for the discovery of physical laws from machine learning (e.g., Kokar, 1986). Recently, this work has turned to deep learning tools (e.g., Ren et al., 2018; Raissi et al., 2019; Iten et al., 2020). However, as Rudin (2019) points out, explanations for the patterns deep learning tools find are often inaccurate and at worst, totally unrelated to both the model and the world it models. These two approaches lie on a continuum between valuing increasing data and model freedom, or incorporating physical insight and model constraint.

In **Figure 1**, we present a diagram for considering physical theory and machine learning within the context of theoretical constraints. The examples at one end of the continuum represent applications of traditional space physics from global theory driven modeling, while those at the other end of the continuum



**FIGURE 1 |** Framework for incorporating physical understanding in machine learning. This figure diagrams a continuum moving from purely theory bound, toward model free. The figure in model bound is from Jia et al. (2012), a magnetohydrodynamics simulation of Saturn's magnetosphere. The figure in model free is from Chen et al. (2019), deep learning feature correlations for solar flare precursor identification. This figure contains subfigures from American Geophysical Union (AGU) journals. AGU does not request permission in use for republication in academic works but we do point readers toward the associated AGU works for citation and figures in Jia et al. (2012) and Chen et al. (2019).

focus on data driven approaches to space weather and solar flare prediction. The model adjusted center presented below takes advantage of data, but limits or constrains the application by merging with domain understanding. Our work is in the middle of the continuum. We leverage domain knowledge about space physics, while also aiming to learn more about the physical system we study. Importantly, we use an interpretable machine learning approach so that we can be more confident in drawing physical insights from the model.

We present comparisons between a range of data sizes and physics incorporation to classify unique plasma transport events around Saturn using the Cassini dataset. As a characteristic data set of space physics and planetary environments, this provides valuable insights toward future implementation of automated detection methods for space physics and machine learning. We focus on three primary guiding axes in this work to address implementations of machine learning. First, we address the performance and accuracy of the application. Second, we consider how to increase interpretability of machine learning applications for planetary space physics. Third, we tackle how characteristics of spacecraft data change considerations of machine learning applications. All of these issues are essential to consider in applications of machine learning to planetary and space physics data for scientific interpretation.

To investigate these questions and provide a path toward application of machine learning to planetary space physics

datasets, we compare and contrast physics-based and non-physics based machine learning applications. In section 2, we discuss the previous development of a physics-based semi-supervised classification from Azari et al. (2018) for the Saturn system within the context of common characteristics of orbiting spacecraft data. We then provide an outline for general physics-informed machine learning for automated detection with space physics datasets in section 3. Section 4 describes the machine learning model set up and datasets that we use to compare and contrast physics-based and non-physics based event detection. Section 5 details the implementation of logistic regression and random forest classification models as compared to this physics-based algorithm with the context of physics-informed or model adjusted machine learning. Section 6 then concludes with paths forward in applications of machine learning for scientific insight in planetary space physics.

## 2. BACKGROUND: SATURN'S SPACE ENVIRONMENT AND DATA

Saturn's near space environment where the magnetic field exerts influence on particles, or magnetosphere, ranges from the planet's upper atmosphere to far from the planet itself. On the dayside the magnetosphere stretches to an average distance of 25 Saturn radii ( $R_S$ ) with a dynamic range between 17 and 29  $R_S$  (Arridge

et al., 2011) ( $1 R_S = 60,268$  km). This distance is dependent on a balance between the internal dynamics of the Saturn system and the Sun's influence from the solar wind. Within this environment a complex system of interaction between a dense disk of neutrals and plasma sourced from a moon of Saturn, Enceladus, interacts with high-energy, less dense plasma from the outer reaches of the magnetosphere (see **Figure 2**).

This interaction, called interchange, is most similar to Rayleigh-Taylor instabilities and results in the injection of high-energy plasma toward the planet. In **Figure 2**, a system of interchange is detailed with a characteristic Cassini orbit cutting through the interchanging region. The red box in this figure is presented as an illustrative slice through the type of data obtained to characterize interchange. One of the major questions in magnetospheric studies is how mass, plasma, and magnetic flux moves around planets. At the gas giant planets of Saturn and Jupiter, interchange is thought to be playing a fundamental role in system-wide transport by bringing in energetic material to subsequently form the energetic populations of the inner magnetosphere, and to transport plasma outwards from the moons. Until Cassini, Saturn never had a spacecraft able to develop statistics based on large-scale data sizes to study this mass transport system.

The major scientific question surrounding studying these interchange injections is what role these injections are playing in the magnetosphere for transport, energization, and loss of

plasma. To answer this question, it's essential to understand where these events are occurring and the dependency of these events on other factors in the system, such as influence from other plasma transport processes and spatio-temporal location. From Cassini's data, several surveys of interchange had been pursued by manual classification, but these surveys disagreed on both the identification of events and resulting conclusions (Chen and Hill, 2008; Chen et al., 2010; DeJong et al., 2010; Müller et al., 2010; Kennelly et al., 2013; Lai et al., 2016). The main science relevant goal was to create a standardized, and automated, method to identify interchange injections. This list needed to be physically justified to allow for subsequent conclusions and comparisons.

In section 2.1, we provide background on the Cassini dataset and summarize the previous development of a physics-based detection method in section 2.2. We then provide a generalized framework in the following section 3 for incorporating physical understanding into machine learning with the development of this previous physics-based method as an example. Subsequent sections investigate comparisons of this previous physics-based effort to other automated identification methods.

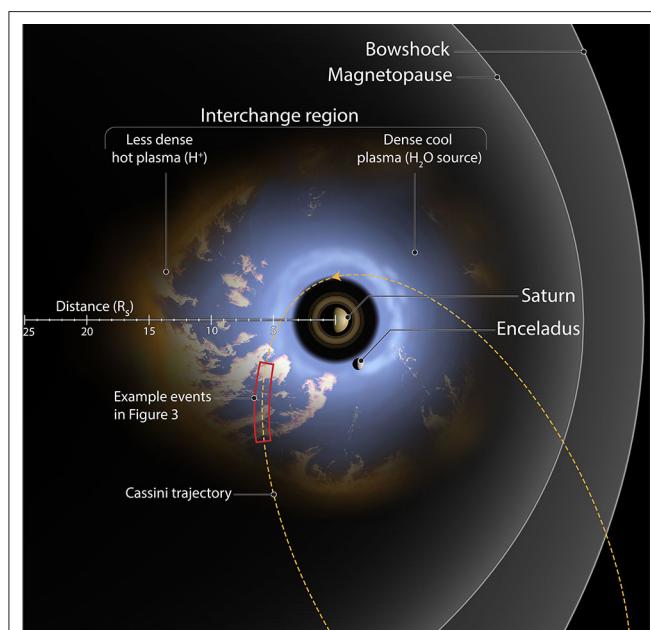
## 2.1. Cassini High-Energy Ion Dataset

Cassini has onboard multiple plasma and wave sensors which are in various ways sensitive to interchange injections. However, none of the previous surveys focused on high-energy ions, which are the primary particle species transported inwards during injections. In **Figure 3**, a series of injections are shown in high-energy (3–220 keV) ions ( $H^+$ ) and magnetic field datasets. This figure shows three large injections between 0400 and 0600 UTC followed by a smaller injection after 0700 most noticeable in the magnetic field data. It is evident from these examples that using different sensors onboard Cassini will result in different identification methods for interchange injections. This was a primary driver in a standardized identification method for these events. The top two panels detail the Cassini Magnetospheric Imaging Instrument: Charge Energy Mass Spectrometer (CHEMS) dataset while the last contains the Cassini magnetometer magnetic field data (Dougherty et al., 2004; Krimigis et al., 2004).

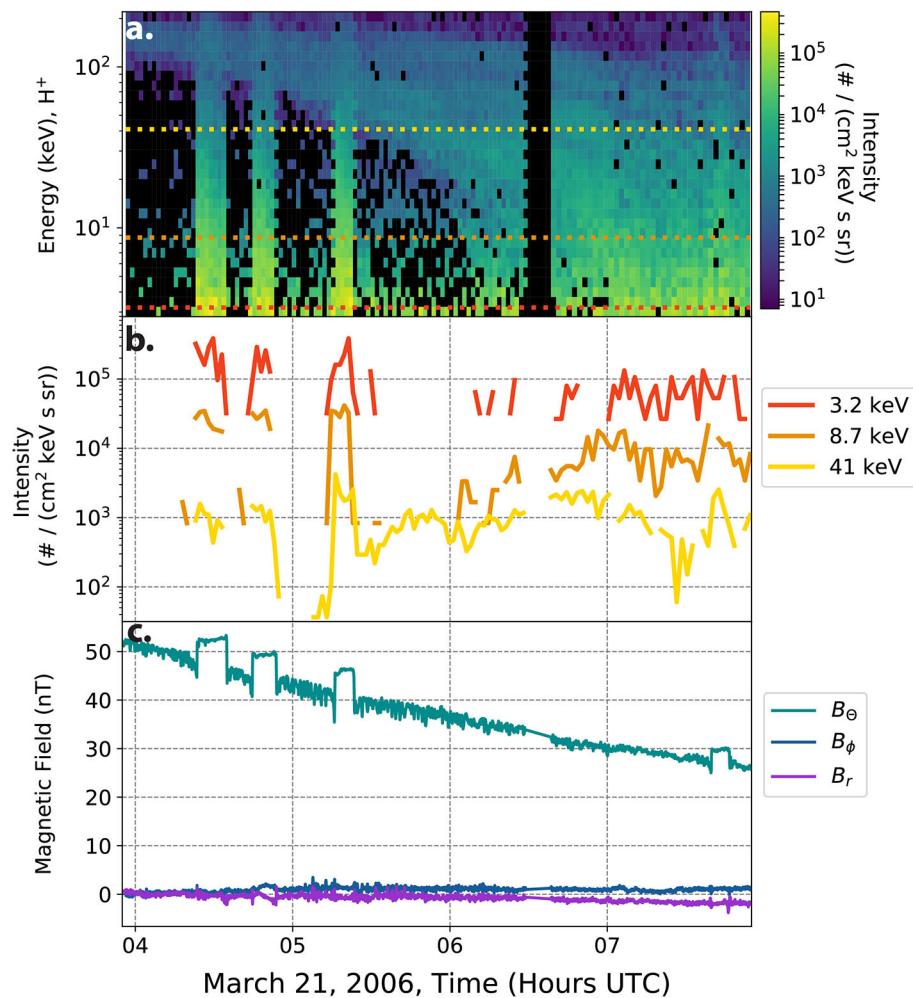
The CHEMS instrument onboard Cassini collected multiple species of ion data and finds the intensity of incoming particles in the keV range of data. This datastream can be thought of as unique energy channels, each with a spacecraft position and time dependence. In **Figure 3b**, three unique energy channels are shown from the overall data in the top panel, to illustrate the nature of these high-energy data. This type of spatio-temporal data is often a characteristic of space physics missions (see Baker et al., 2016, for a review of MMS' data products).

## 2.2. Development of Physics-Based Detection Method

When applying automated or machine learning methods, such data discussed above provides unique challenges and characteristics including: rare events (class imbalances), spatio-temporal sampling, heterogeneity in space and time, extreme high-dimensionality, and missing or uncertain data (Karpatne et al., 2019). These challenges are in addition to desired



**FIGURE 2 |** Diagram of interchange injection in the Saturn system. The illustrated orbit is an equatorial Cassini orbit from 2005. Injections are denoted by the pale orange material interspersed with the water sourced plasma from Enceladus. Along the example orbit the red box denotes a hypothetical segment of Cassini data discussed in **Figure 3**. The purpose of developing an automated event detection is to identify the pale orange material traveling toward the planet. This figure is produced in consultation with, and copyright permissions from Falconieri Visuals.



**FIGURE 3 |** Series of interchange injections characterized by high-energy ions. **(a)** Details an energy time spectrogram of the intensity from the Cassini CHEMS sensor. The color black denotes flux either below the colorbar limit or missing data. The three lines are placed at the energy channels for the plot in **(b)**. **(b)** Shows the same CHEMS data, but split out into three characteristic energies over the entire CHEMS range. **(c)** Shows the magnetic field data in KRTP (Kronocentric body-fixed, J2000 spherical coordinates).

interpretability. It's essential that an interpretable model is used to learn substantive information about this application. One common use of machine learning is to input a large number of variables and/or highly granular raw data (e.g., individual sensor readings or image pixel values) into a model, letting the algorithm sort out relationships among them. Such models are inherently "black boxes" because the number and granularity of variables, not to mention complicated recursive relationships among them, makes it difficult or impossible for humans to interpret (Rudin, 2019). One solution to this issue is to reduce dimensionality to fewer, more meaningful-to-humans inputs. But at the same time, the model needs to be informative, and the inputs need to be meaningful. Incorporating domain knowledge and then letting the model determine their effectiveness in the system of study is a potential framework to consider.

For this reason, when developing a detection method to standardize, characterize, and subsequently build off the detected

list, a physics-based method was chosen to address these unique challenges. This previous effort is discussed in Azari et al. (2018) and the resultant dataset is located on the University of Michigan's Deep Blue Data hub (Azari, 2018). We build on this effort in the present work to provide a new evaluation of alternative solutions for data-driven methods.

To develop this physics-based method, the common problems in space physics data described in Karpatne et al. (2019) were considered and addressed to develop a single dimension array ( $S$ ).  $S$  was then used in a style most similar to a single dimensional logistic regression to find the optimum value for detecting interchange events. This classification was standardized in terms of event severity, as well as physically bounded in definition of events. As a result, it was able to be used to build up a physical understanding of the high-energy dynamics around Saturn's magnetosphere including: to estimate scale sizes (Azari et al., 2018) and to demonstrate the influence of tail

injections as compared to the ionosphere (Azari et al., 2019). Following machine learning practices,  $S$  was designed through cross validation. It was created to perform best at detecting events in a training dataset and then evaluated on a separate test dataset. These sets contain manually identified events and were developed from 10% of the dataset (representing 7,375/68,090 time samples). Training and test dataset selection and limiting spatial selection is of critical importance in spatio-temporal varying datasets. Our particular selection considerations are discussed in following sections. The training set was used to optimize the final form of  $S$ . The test dataset was used to compare performance and prevent over fitting. The same test and training datasets are used in the following sections.

$S$  was developed in Azari et al. (2018) to provide a single-dimension parameter which separated out the multiple dependencies of energy range and space while dealing with common challenges in space physics and planetary datasets.  $S$  is calculated from  $S_r$  by removing the radial dependence through normalization. In mathematical form,  $S_r$  can be written as:

$$S_r = \sum_{e=0}^{14} w^{(Z_{e,r} - C)} \quad (1)$$

$S$  can be thought of as a single number which describes the intensification of particle flux over a normalized background. In other words,  $S$  can be calculated as:  $S = (S_r - \bar{S}_r)/\sigma_{S_r}$ . In which  $\bar{S}_r$  is the average radially dependent average and  $\sigma_{S_r}$  the radially dependent standard deviation. These calculations allow for  $S$  to be used across the entire radial and energy range for optimization in units of standard deviation. The variables  $w$  and  $C$  represent weighting values which are optimized for and discussed in the following section. The notations of  $e$  and  $r$  represent energy channel and radial value.  $Z_{e,r}$  represents a normalized intensity value observed by CHEMS. This is similar to the calculation of  $S$  from  $S_r$ .

Additional details on the development, and rationale behind,  $S$  are described in section 3 as a specific example for a general framework for inclusion of physical information into machine learning.

The final form of  $S$  depends non-linearly on the intensity values of the CHEMS sensor and radial distance. In **Figure 4**, we show the dependence of the finalized  $S$  value over the test dataset for the intensity at a single energy value of 8.7 keV and over all radial distances. Within this figure the events in the test dataset are denoted with dark pink dots. From **Figures 4d,e**, it's evident that  $S$  disambiguates events from underlying distributions, for example in **Figure 4b**. By creating  $S$  it was possible to create a single summary statistic which separated events from a background population.

The strategies pursued in developing  $S$  are most applicable for semi-supervised event detection with space physics data. They can, however, prove a useful guide in starting to incorporate physical knowledge into other applications in heliophysics and space physics. Within the previous effort we used the model optimization process from machine learning to guide a physics incorporated human effort. This was a solution to

incorporating the computational methods employed in machine learning optimization to a human-built model. The end result was optimized in a similar fashion as machine learning models but through manual effort to ensure physical information preservation. Moving from this effort, we now present a framework for expanding the style of integrating human effort and physical-information into other applications for space physics data.

Below we provide a framework for incorporating physical-understanding into machine learning. In each strategy we discuss common issues in space physics data, using a similar phraseology as Karpatne et al. (2019). In addition to characteristics in the structure of geoscience data, we also add interpretability as a necessary condition. For space physics and planetary data, the challenges within Karpatne et al. (2019) are often compounded and where appropriate we note potential overlap. After each strategy, we provide a walk-through of the development of  $S$  employed in Azari et al. (2018).

### 3. FRAMEWORK FOR PHYSICS INCORPORATION INTO MACHINE LEARNING

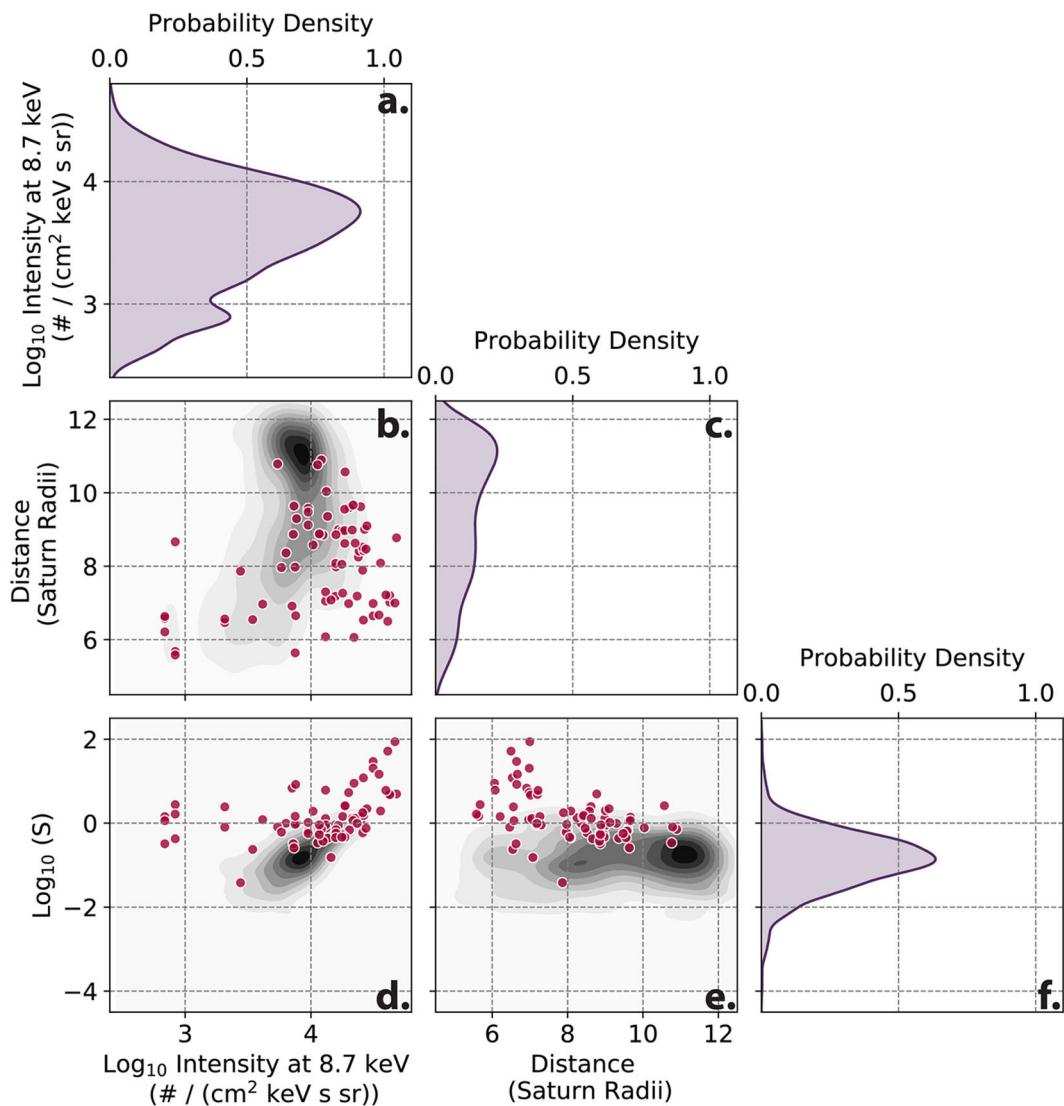
This framework focuses on interpretable semi-supervised event detection with space physics data from orbiters for the end goal of scientific analysis. Depending on the problem posed certain solutions could be undesirable. For a similarly detailed discussion on creating a machine learning workflow applied to problems in space weather (see Camporeale, 2019). The framework presented here can be thought of as a directed application of feature engineering for space physics problems, mostly for requiring interpretability. In general the strategies below provide a context for careful consideration of the nature of domain application which is essential for applications of machine learning models to gather scientific insights.

**1. Limit to region of interest.** Orbiting missions often range over many environments and limiting focus to regions of interest can assist in automated detection by increasing the likelihood of detection of events.

**Issues:** heterogeneity in space and time, rare events (class imbalance)

**Example:** The Cassini dataset represents a wide range of sampled environments, the majority of which do not exhibit interchange. In addition, the system itself undergoes seasonal cycles, changing in time, presenting a challenge to any long-ranging spatial or temporal automated detection. The original work targeted a specific radial region between 5 and 12  $R_S$  in the equatorial plane. This region is known to be sensitive to interchange from previous studies. Similarly, each season of Saturn was treated to a separate calculation of  $S$ , allowing for potential temporal changes to the detection of interchange.

**2. Careful consideration of training and test datasets.** Due to the orbiting nature of spacecraft, ensuring randomness in training and test datasets is usually not sufficient to create a representative set of data across space and time. For event studies, considerations of independence for training and test



**FIGURE 4 |** Distributions of S parameter developed in Azari et al. (2018). This figure represents a subset of multiple dependencies of S from a kernel density estimation (kde). The data used in this figure is from the test dataset of the data. **(a,c,f)** Represent a single dimension kde of a CHEMS energy channel intensity, spacecraft location in radial distance, and of S. **(b,d,e)** Represent two dimensional distributions. This figure was developed using the Seaborn statistics package's kde function (Waskom et al., 2020).

dataset while containing prior and post-event data (at times critical for event identification) are important. This is similar to recent strides in activity recognition studies with spatio-temporal data, in which training set considerations drastically affect the accuracy of activity classification (e.g., Lockhart and Weiss, 2014a,b).

**Issues:** heterogeneity in space and time, spatio-temporal data, rare events, small sample sizes

**Example:** While the test and training set represent 10% of the data for the worked example, the 10% was taken such that it covered the widest range of azimuthal and radial values, while still being continuous in time and containing a range of events.

3. **Normalize and/or transform.** Many space environments have a spatio-temporal dependent background. Normalizing separately to spatial or other variables will address these dependencies and can prove advantageous if these are not critical to the problem.

**Issues:** heterogeneity in space and time, spatio-temporal sampling, multi-dimensional data

**Example:** As seen **Figure 4b** flux values depend on radial distance and energy value. Similarly, flux exhibits log scaling, where values can range over multiple powers of 10 in the span of minutes to hours as seen in **Figure 3**. To handle the wide range of values from the CHEMS sensor, each separate energy channel's intensity was first converted into logarithmic

space before then being normalized by subtracting off the mean and dividing by its standard deviation. Effectively, this transforms the range of intensities to a near-normal distribution dependent on radial distance and energy value (see  $Z_{e,r}$  in Equation 1). A similar treatment is performed on creating the final  $S$  from  $S_r$ . This is important due to the commonality of normalcy assumptions in which models can assume normally distributed data on the same scale across inputs.

- 4. Incorporate physical calculations.** Space physics data can come with hundreds if not thousands of features. While many machine learning techniques are designed for just this kind of data, they do not typically yield results that are amenable to human interpretation and scientific insight into the processes of physical systems. They express a complex array of relationships among raw measurements that do little to help humans build theory or understanding. Summary statistics like summing over multiple variables, or taking integrals, can preserve a large amount of information from the raw data for the algorithm while leaving scientists with smaller sets of relationships between more meaningful variables to interpret. For other fields rich in noisy and incomplete time-series data with a longer history of automated detection methods, summary statistic transformations have been a valuable way of handling this type of data for improved performance (e.g., Lockhart and Weiss, 2014a).

**Issues:** interpretability, multi-dimensional data, missing data

**Example:** To address missing values . building up summary statistics, for example through summing over multiple energy channels can help. This creates an particle pressure like calculation (see sum in Equation 1). Particle pressure itself is not used to identify events, as the ability to tune the exact parameters was desired in the identification of injections and developing  $S$  proved more reliable. This allows for the lower 14 energy channels to contribute without removing entire timepoints from the calculation where partial data is missing and also increasing interpretability of the end result . Only the lower 14 channels are used as the higher energy channels also show long duration background from earlier events drifting in the Saturn environment (see **Figure 3**).

- 5. Compare with alternate metrics.** Dependent on your use case, the trade-off costs between false positives and false negatives could be different from the default settings in standard machine learning tools. Investigating alternate metrics of model performance and accuracy are useful toward increasing interpretability.

**Issues:** interpretability, rare events (class imbalance)

**Example:** In the training and test datasets only 2.4% of the data exist in an event state. This proves to be challenging for then finding optimum detection due to the amount of false positives and usage for later analysis. In Equation (1) scaling factors of  $w$  and  $C$  are introduced. These scale factors are chosen by optimizing for the best performance of the Heidke Skill Score (HSS) (Heidke, 1926). HSS is more commonly used in weather forecasting than in machine learning penalty calculations but has shown potential for handling rare events (see Manzato, 2005, for a discussion of HSS). In section

5, we evaluate how HSS performs as compared to other regularization schemes (final values:  $w = 10$ ,  $C = 2$ ).

- 6. Compare definitions of events, consider grounding in physical calculations.** Much of the purpose of developing an automated detection is to standardize event definitions. Developing a list of events then can become tricky.

**Issues:** lack of ground truth, interpretability, rare events (class imbalance)

**Example:** At this point in the calculation of  $S$ , there is a single number, in units of standard deviations, for each time point. This calculation so far, takes in the flux of the lowest 14 energy channels of CHEMS before normalizing and combining these values to return a single value at each time. This number is higher (in the useful units of standard deviation) for higher flux intensifications and lower for flux drop outs. The final question becomes at which  $S$  value should an event be considered real or false.

Based on the training dataset, 0.9 standard deviations above the mean of  $S$  is the optimum parameter for peak HSS performance. As discussed in section 2.2 0.9 was determined through optimizing against the training set. Since  $S$  is in terms of standard deviations, additional higher thresholds can be implemented to sub-classify events into more or less severe cases with a physical meaning (ranking). This allowed for the application as a definition task with a physical justification.

- 7. Investigate a range of machine learning models and datasets.** Incorporating a range of machine learning models, from the most simple to the most complex in addition to varying datasets, can offer insights in the nature of the underlying physical data.

**Issues:** interpretability

**Example:** In developing  $S$ , alternative feature inclusions were considered.  $S$  was settled on for its grounding in physical meaning. A secondary major consideration was its accuracy compared to other machine learning applications. In the following sections we discuss additional models.

As similarly discussed within Camporeale (2019), the desire to incorporate physical calculations comes from an interest in using machine learning for knowledge discovery. In the use cases of interest here, both the needs for accuracy and interpretability are essential. These presented strategies are designed to improve the potential performance for semi-supervised classification problems and the interpretability for subsequent physical understanding. Creating the final form of  $S$  was a labor intensive process to create and then optimize. Due to  $S$ 's non-linear dependence on the features shown in **Figure 4**, this was a non-trivial task. Similarly expanding  $S$  into additional dimensions is challenging. This is where the machine learning infrastructure offers significant advantages as compared to the previous effort. In the following sections 4 and 5 we discuss alternative solutions to identification of interchange.

## 4. METHODS: MODELS AND EXPERIMENTAL SETUP

In the previous physics-based approach, events were defined through intensifications of  $H^+$  only, allowing for comparisons

to other surveys and advancement of the understanding of events. This was a non-intuitive approach as common logic in application of machine learning algorithms suggests that greater data sizes will result in additional accuracy given a well-posed problem. To explore both the potential for higher accuracy as well as interpretability of the application, we compare the performance of two distinct machine learning models with access to varying data set sizes. Below we discuss models we use in this comparison effort.

## 4.1. Models

Two commonly used machine learning models for supervised classification are logistic regression and random forest classification. Both are considered standard classification models when applying machine learning and performing comparative studies (Couronné et al., 2018). While both models can be interpreted by humans, the additive functional form of logistic regression and the broad literature on interpreting it make it highly interpretable. Random Forest models consist of easy to interpret logical rules, but the large numbers and weighted combinations of those rules mean it is less interpretable (Rudin, 2019). The original physics-based algorithm was designed with a logistic regression method in mind, but with significant adjustment. Comparisons to this model are directly informative as a result. Logistic regression categorizes for binary decisions by fitting a logistic form, or a sigmoid. Logistic regression is a simple, but powerful, method toward predicting categorical outcomes from complex datasets. The basis of logistic regression is associated with progress made in the nineteenth century in studying chemical reactions, before becoming popularized in the 1940s by Berkson (1944) (see Cramer, 2002, for a review). When implemented and optimized using domain knowledge, highly interpretable models, like logistic regression, generally perform as well as less interpretable models and even deep learning approaches (Rudin, 2019).

Random forest in comparison classifies by building up collection of decision trees trained on random subsets of the input variables. The predictions of all trees are then combined in an ensemble to develop the final prediction. Similar to logistic regression, the method of random forest has been built over time with the most modern development associated with Breiman (2001). While logistic regression requires researchers to specify the functional form of relationships among variables, random forests add complexity toward classification decisions, by allowing for arbitrary, unspecified non-linear dependencies between features, also known as model inputs.

The models used within this chapter are from the scikit-learn machine learning package in Python (Pedregosa et al., 2011). Within the logistic regression the L2 (least squares) regularization penalty is applied. Within the random forest a grid search with 5-fold cross-validation is used to find the optimum depth between 2 and 5, while the number of trees is kept at 50. These search parameters are chosen to constrain the random forest within the perspective of the noisy nature of the CHEMS dataset and to prevent over fitting. Alterations to this tuning parameter scheme are not seen to alter the results in the following section. Events are relatively rare in the data (2.4% of the data in the training

and test datasets corresponds with an event), and this can bias the fit of models. As such, unless otherwise noted, we use class weighting to adjust the importance of data from each class (event and non-event) inversely proportional to its frequency so that the classes exert balanced influence during model fitting. This results in events weighted higher more important than non-events due to their rarity. Performance is shown in section 5 against the test dataset defined above.

## 4.2. Dataset Definitions and Sizes

To explore the performance of logistic regression and random forest, four distinct subsets of the Cassini plasma and magnetic field data are utilized ranging in data complexity and size as follows:

1. S\C (Spacecraft) Location and Magnetic Field  
6 features, 68,090 time samples
2. S\C Location, Magnetic Field, and H<sup>+</sup> flux (3–220 keV)  
38 features, 68,090 time samples
3. Low Energy H<sup>+</sup> flux (3–22 keV)  
14 features, 68,090 time samples
4. Azari et al. (2018) (S Value)  
1 feature, 68,090 time samples

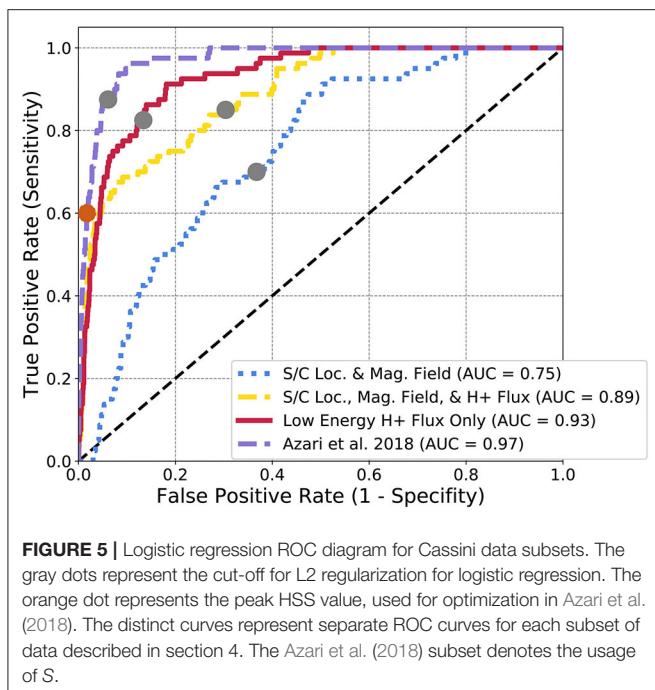
These subsets are chosen to represent additional features, complexity, and physics inclusion. All of these subsets should be sensitive in varying amounts toward identification of interchange injections as evidenced in Figures 3, 4. The first two datasets are a comparison of increasing features that should assist in identification of interchange injection. The third dataset includes less features, but is the originator most similar to the derived parameter from Azari et al. (2018). The final dataset contains the single summary statistic array of the S parameter. In the following result section, these four dataset segments are used to evaluate the two models.

## 5. RESULTS AND DISCUSSION

We are interested in evaluating how the former physics-based S parameter performs with other commonly used subsets of space physics data. Our primary goal in this section is to investigate the trade off between the performance of these more traditional models and their interpretability, and therefore usage for scientific analyses. We complete this through applying supervised classification models and evaluate the ease of interpretability and their relative performance.

### 5.1. Supervised Logistic Regression Classification

In Figure 5, the ROC curve of a logistic regression for all four subsets of Cassini data is presented. ROC or receiver operating characteristics, are a common method employed for visualizing the efficacy of classification methods (see Fawcett, 2006, for a generalized review of ROC analysis). ROC curves in this particular example are created by sweeping over a series of classification thresholds. Ideally a perfect classifier will result in a curve that carves a path nearest to the upper left corner.



Area under the curve, or AUC is presented as a metric to understand the overall performance of each logistic regression evaluation. AUC has the ideal parameters of ranging between 0 and 1, with 0.5 representative of random guessing, 1 representing perfect classification, and 0 as the inverse of truth. AUC can be thought of as an average accuracy of a model and isn't sensitive to class-balance and thresholds. ROC curves present the ratios of true positive rate (y-axis) to false positive rate (x-axis). This can be thought of as the trade off for classifiers between events successfully identified (y-axis), and events unsuccessfully identified (x-axis).

The purple curve represents the logistic regression with only the derived physics-based S as an input. This is rather redundant with optimizing by hand as it's a single variable space. Instead the purple curve is provided as a benchmark against the identical performance and curves found within Azari et al. (2018). From this figure, this single summary statistic (S) outperforms all other subsets of Cassini data with an AUC approaching near 1.0 (0.97). This is evidence for the current case, that incorporating physical information, even at the expense of greater dataset size improved the performance of certain machine learning applications.

Following this it is not the largest dataset that has the second best performance. Instead, the red curve which contains only the low energy H<sup>+</sup> intensities shows the best performance of the non physics-adjusted datasets. The magnetic field is a useful parameter for the prediction of interchange as demonstrated in Figure 3, but the form of the logistic regression is unable to use this information successfully. This is possibly due to the higher time resolution needed for interchange identification from magnetic field data and any future identification work needs to focus on adjusting the magnetic field inputs and models. The current dataset is processed such that each time point

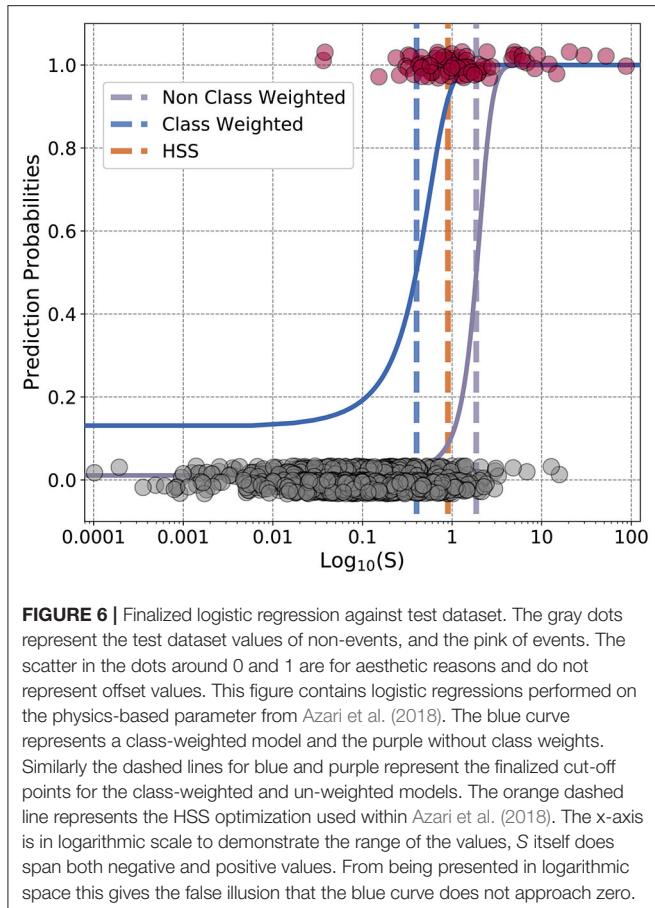
in the CHEMS set is matched with a single magnetic field vector. Normally within interchange analyses, the magnetic field information is of a much higher resolution. It is likely if a study pursued solely magnetic field data of higher time resolution and processed these data to represent pre and post event states dependent on time, the performance of the magnetic field data would be improved. It's evident from Figure 4, that S exhibits non-linear behavior from the distribution of S on intensity, distance, and energy. Similarly the magnetic field values likely range over a far range due to the background values, that the linear dependency requirements of logistic regression are unable to use this information. Without the flux data especially (the blue curve) logistic regression is unable to predict interchange as compared to the previous physics-based parameter.

The AUC doesn't capture the entire picture for our interest. While it shows the performance of the algorithm, it contains information for multiple final classifications of events. The gray dots on Figure 5 demonstrates the chosen cut-point for L2 regularization for class weighted events, or the final classification decision for an optimal trade between real events and false events. Within the previous section, the Heidke Skill Score or HSS was discussed as the final threshold separating events from non-events (denoted as the orange dot on Figure 5). Deciding the threshold of what separates an injection event from a non-event is critical for the implementation of statistical analysis on the results especially in this case, in which non-events outnumber events at a ratio of ~50:1. One solution would be to rank events, in similar style of the previous work of S with categories of events (Azari et al., 2018).

## 5.2. Rare Event Considerations

We now move to evaluating the previous HSS optimization to the logistic regression L2 regulation for both class weighted and non-class weighted models. In Figure 6, the final forms of the weighted and non-weighted logistic regression for the trivial 1 dimensional array case of the S parameter are shown. The thresholds for the final decisions and for HSS are shown as vertical lines (the orange dashed line represents HSS). Due to the extreme imbalance of non-events to events, implementing class weighting results in large shifts between what is considered an injection event or not. We suggest that the class imbalance inherent in this problem is the main rationale between the differences of HSS and other regularizations. Between the two decision points of the blue and purple vertical lines there are 46 real events, but 202 non-events. This means that if using class-weighting in logistic regression for this problem, 202 non-events would be classified as events. Non-intuitively, for this application where the final events are used to understand the Saturn system, it's advantageous to use a non-class weighted model, as it limits the non-events. However the un-class-weighted model results in removing many real events as well as can be seen in the bulk of the pink events (real events) being misclassified by the purple vertical line.

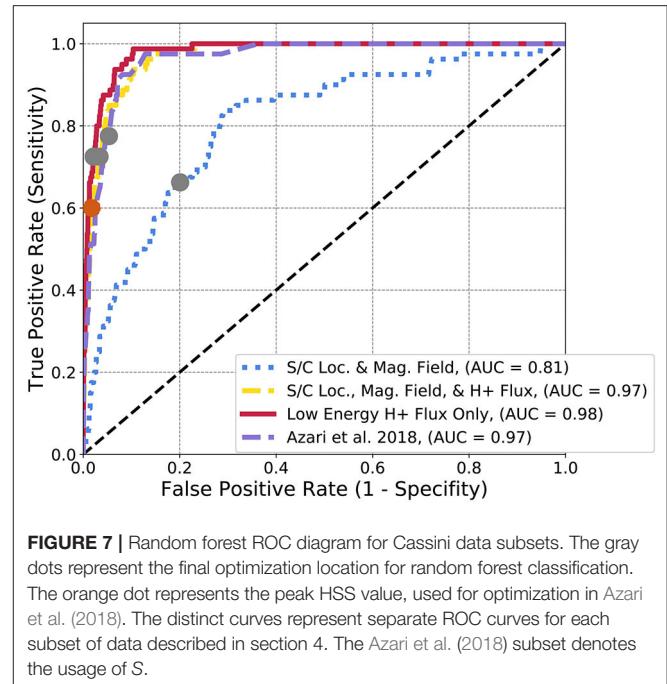
The Heidke Skill Score provides an in-between choice of these by providing a higher threshold than the class-weighted, and lower than non-class weighted. The logistic regression for the S parameter shown here is easily intuited since the X-axis



represents only one variable. The power of machine learning however is most advantageous in multiple dimensions. HSS has shown to be a more applicable metric for rare events. Other skill scores, such as the True Skill Score have also shown promise in machine learning applications to space physics (Bobra and Couvidat, 2015). Skill score metrics themselves have a long and rich history in space physics before more recent applications in machine learning with interest originating in space weather prediction (see Morley, 2020, for an overarching review of space weather prediction). We also direct the reader to discussions of metrics for physical model and machine learning prediction of space weather (Liemohn et al., 2018; Camporeale, 2019). How can these traditional metrics for space applications be integrated into the regularization schemes? Future work in machine learning applications should consider shared developments between the physical sciences communities usage of skill scores and regularization of models.

### 5.3. Supervised Random Forest Classification

In Figure 7, the ROC diagram for the same subsets of data but for a random forest model are presented. In this case, unlike the logistic regression, other subsets of data can reproduce the same performance (or AUC) as the derived parameter.



All curves, with the exception of the spacecraft location and magnetic field, quickly approach or slightly surpass the AUC of the physics-based parameter at 0.97, with small differences in the performance of the low energy  $H^+$  flux (0.98) and of the combined spacecraft location, all flux, and magnetic field (0.97). The model form of random forest allows for non-linear behavior in the intensity and magnetic field data to find injection events. Increasing the features then helps in the case of random forest whereas it did not for logistic regression. Similar to the logistic regression, HSS results in a different ratio between true positive rate and false positive rate than the random forest model cut-off point with the gray dots.

Comparing back to logistic regression, even with a relatively complex model such as random forest, the AUC of the best ROC curves are near-identical. Given that  $S$  is an array, this is not that surprising. In both cases the physics-derived parameter outperforms or is effectively equivalent to all other data subsets, including those with access to a much richer information set and therefore more complex model. For the application of interpretability for then gathering scientific conclusions, logistic regression is advantageous as it presents a much simpler model. However, random forest, has shown ability to mimic the underlying physics adjustments through selection of datasets.

Within these results, it's evident that the  $S$  parameter performs as well as simplistic machine learning models. Given that  $S$  is also grounded in a physics-based definition dependent on solely a variable flux background, this offers advantages to subsequent usage in scientific results. However, many of the adjustments in creating  $S$  can be implemented into other space physics data, and integrated into machine learning as evidenced here. In the description of the development of  $S$ , several challenges in geoscience data from the framework discussed in Karpatne

et al. (2019), and CHEMS specific solutions were presented. From the above evaluation, it is evident that applications of machine learning are useful to the task of automated event detection from flux data, but with diminishing interpretability. A potential solution to both enhancing the interpretability, similar to the  $S$  based parameter, but also incorporating the advantages of machine learning is presented in **Figure 1**. Rather than consider incorporation of physics-based information as deleterious to the implementation of machine learning, we have found that including this information simplifies the application, enhances the interpretability, and improves the overall performance.

## 6. CONCLUSION AND FUTURE DIRECTIONS

Planetary space physics has reached a data volume capacity at which implementation of statistics including machine learning is a natural extension of scientific investigation. Within this work we addressed how machine learning can be used within the constraints of common characteristics of space physics data to investigate scientific questions. Care should be taken when applying automated methods to planetary science data due to the unique challenges in spatio-temporal nature. Such challenges have been broadly discussed for geoscience data by Karpatne et al. (2019), but until now limited attention in comparison to other fields has been given toward reviews of planetary data.

Within this work we have posed three framing concerns for applications of machine learning to planetary data. First, it's important to consider the performance and accuracy of the application. Second, it's necessary to increase interpretability of machine learning applications for planetary space physics. Third, it's essential to consider how the underlying issue characteristics of spacecraft data changes applications of machine learning. We argue that by including physics-based information into machine learning models, all three concerns of these applications can be addressed.

For certain machine learning models the performance can be enhanced but importantly in this application, the interpretability improves along with handling of characteristic data challenges. To reach this conclusion we presented a framework for incorporating physical information into machine learning. This framework targeted considerations for increasing interpretability and addressing aspects of spacecraft data into machine learning with space physics data. In particular, it addresses challenges such as the spatio-temporal nature of orbiting spacecraft, and other common geoscience data challenges (see Karpatne et al., 2019). After which we then cross-compared a previous physics-based method developed using the strategies in the framework to less physics-informed but feature rich datasets.

The physics-based semi-supervised classification method was built on high-energy flux data from the Cassini spacecraft to Saturn (see Azari et al., 2018). In investigating the

accuracy of machine learning applications, we demonstrated this physics-based approach outperformed automated event detection for simple logistic regression models. It was found that traditional regularization through L2 penalties both under, and overestimated ideal cutoff points for final event classification (depending on class weighting). Instead, metrics more commonly used in weather prediction, such as the Heidke Skill Score, showed promise in class imbalance problems. This is similar to work demonstrating the applicability of True Skill Score in heliophysics applications (Bobra and Couvidat, 2015). Future work should consider building on the rich history of prediction metrics in the space physics community for shared development between the physical sciences usage of skill scores and in regularization of models.

While logistic regression is a more interpretable model, random forest proved that with the addition of more and lower level variables from the Cassini mission, the model could approximate our physics-based logistic model successfully. In this case physics-informed or model adjusted machine learning, can each the same performance but with different levels of interpretability, thus different ability to draw further conclusions about implications of the results. The logistic approach provides a coefficient and threshold for a meaningful physical quantity,  $S$ , effectively the normalized intensification of particle flux. The random forest approach can provide an "importance" score for  $S$  or show a large number of conjunction rules involving it, but neither is as useful for human analysts. A forest model using a large number of raw variables instead of a small number of more meaningful ones like  $S$  is even harder for humans to make sense of. Deep neural networks, as multi-layered webs of weighted many-to-many relationships, are even less informative for human analysts interested in understanding the workings of the model and physical system. Further, findings that the interpretable model performs as well or better than other approaches demonstrate that, despite the widespread myth to the contrary, there is no inherent tradeoff between performance and interpretability (Rudin, 2019). For example, the ability to further split and define identified events based on their flux intensity using  $S$  gives the ability to address further scientific questions as to the fundamental mechanisms behind the interchange instability itself. The simplistic model of logistic regression which results in the same performance as random forest is highly advantageous for the current task.

The framework and comparison presented here opens up avenues toward consideration of applying machine learning to answer planetary and space physics questions. In the future, cross-disciplinary work would greatly advance the state of these applications. Particularly within the context of interpretability toward scientific conclusions through physics-informed, or model adjusted machine learning. The inclusion of planetary science and space physics domain knowledge in application of data science allows for the pursuit of fundamental questions. We have found that incorporating physics-based information increases the interpretability, and improves the overall performance of machine learning applications for scientific insight.

## DATA AVAILABILITY STATEMENT

The events analyzed for this study can be found in the Deep Blue Dataset under doi: 10.7302/Z2WM1BMN (Azari, 2018). The original datasets from the CHEMS (Krimigis et al., 2004) and MAG (Dougherty et al., 2004) instruments can be found on the NASA Planetary Data System (PDS). Details on the most recent datasets for CHEMS and MAG can be found on the Cassini-Huygens Archive page at the PDS Planetary Plasma Interactions node (<https://pds-ppi.igpp.ucla.edu/mission/Cassini-Huygens>). Associated data not included in the above repositories can be obtained through contacting the corresponding author.

## AUTHOR CONTRIBUTIONS

We use the CRediT (Contributor Roles Taxonomy) categories for providing the following contribution description (see Brand et al., 2015). AA led the conceptualization and implemented the research for this manuscript including the investigation, visualization, formal analysis, and original drafting of this work. JL assisted in the conceptualization and discussions of methodology in this work along with editing the manuscript. ML provided funding acquisition, resources, supervision, and assisted in conceptualization along with editing the manuscript. XJ provided funding acquisition, resources, supervision, and assisted in conceptualization. All authors contributed to the article and approved the submitted version.

## FUNDING

This material was based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1256260 and was partially funded by the Michigan Space Grant Consortium under NNX15AJ20H. JL received funding through an NICHD training grant to the Population Studies Center at the University of Michigan (T32HD007339). ML was funded by NASA grant NNX16AQ04G.

## ACKNOWLEDGMENTS

We would like to thank Monica Bobra, Brian Swiger, Garrett Limon, Kristina Fedorenko, Dr. Nils Smit-Anseeuw, and Dr. Jacob Bortnik for relevant discussions related to this draft. We would also like to thank the conference organizers of the 2019 Machine Learning in Heliophysics conference at which this work was presented, and the American Astronomical Society Thomas Metcalf Travel Award for funding travel to this conference. This work has additionally appeared as a dissertation chapter (Azari, 2020). **Figure 2**'s copyright is held by Falconieri Visuals. It is altered here with permission. **Figure 1** contains graphics from Jia et al. (2012) and Chen et al. (2019) which can be found in journals with copyright held by AGU. We would like to thank Dr. Jon Vandegriff for assistance with the CHEMS data used within this work.

## REFERENCES

- Arridge, C. S., André, N., McAndrews, H. J., Bunce, E. J., Burger, M. H., Hansen, K. C., et al. (2011). Mapping magnetospheric equatorial regions at Saturn from Cassini Prime Mission observations. *Space Sci. Rev.* 164, 1–83. doi: 10.1007/s11214-011-9850-4
- Azari, A. R. (2018). *Event List for “Interchange Injections at Saturn: Statistical Survey of Energetic H+ Sudden Flux Intensifications”*. University of Michigan - Deep Blue. doi: 10.7302/Z2WM1BMN
- Azari, A. R. (2020). *A data-driven understanding of plasma transport in Saturn’s magnetic environment*. (Ph.D. thesis). University of Michigan, Ann Arbor, MI, United States.
- Azari, A. R., Jia, X., Liemohn, M. W., Hospodarsky, G. B., Provan, G., Ye, S.-Y., et al. (2019). Are Saturn’s interchange injections organized by rotational longitude? *J. Geophys. Res.* 124, 1806–1822. doi: 10.1029/2018JA026196
- Azari, A. R., Liemohn, M. W., Jia, X., Thomsen, M. F., Mitchell, D. G., Sergis, N., et al. (2018). Interchange injections at Saturn: Statistical survey of energetic H+ sudden flux intensifications. *J. Geophys. Res.* 123, 4692–4711. doi: 10.1029/2018JA025391
- Baker, D. N., Riesberg, L., Pankratz, C. K., Panneton, R. S., Giles, B. L., Wilder, F. D., et al. (2016). Magnetospheric Multiscale instrument suite operations and data system. *Space Sci. Rev.* 199, 545–575. doi: 10.1007/s11214-014-0128-5
- Bergen, K. J., Johnson, P. A., de Hoop, M. V., and Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363:eau0323, doi: 10.1126/science.aau0323
- Berkson, J. (1944). Application of the logistic function to bio-assay. *J. Am. Stat. Assoc.* 39, 357–365. doi: 10.1080/01621459.1944.10500699
- Bobra, M. G., and Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys. J.* 798:135. doi: 10.1088/0004-637X/798/2/135
- Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learn. Publish.* 28, 151–155. doi: 10.1087/20150211
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature* 559, 547–555. doi: 10.1038/s41586-018-0337-2
- Camporeale, E. (2019). The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weath.* 17, 1166–1207. doi: 10.1029/2018SW002061
- Chen, Y., and Hill, T. W. (2008). Statistical analysis of injection/dispersion events in Saturn’s inner magnetosphere. *J. Geophys. Res.* 113:A07215. doi: 10.1029/2008JA013166
- Chen, Y., Hill, T. W., Rymer, A. M., and Wilson, R. J. (2010). Rate of radial transport of plasma in Saturn’s inner magnetosphere. *J. Geophys. Res.* 115:A10211. doi: 10.1029/2010JA015412
- Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., DuFumier, B., Zhou, T., et al. (2019). Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters. *Space Weath.* 17, 1404–1426. doi: 10.1029/2019SW002214
- Couronné, R., Probst, P., and Boulesteix, A. L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 19:270. doi: 10.1186/s12859-018-2264-5
- Coveney, P. V., Dougherty, E. R., and Highfield, R. R. (2016). Big data need big theory too. *Philos. Trans. R. Soc. A* 374. doi: 10.1098/rsta.2016.0153
- Cramer, J. S. (2002). *The Origins of Logistic Regression*. Tinbergen Institute Working Paper No. 2002-119/4.
- DeJong, A. D., Burch, J. L., Goldstein, J., Coates, A. J., and Young, D. T. (2010). Low-energy electrons in Saturn’s inner magnetosphere

- and their role in interchange injections. *J. Geophys. Res.* 115:A10229. doi: 10.1029/2010JA015510
- Dougherty, M. K., Kellock, S., Southwood, D. J., Balogh, A., Smith, E. J., Tsurutani, B. T., et al. (2004). The Cassini magnetic field investigation. *Space Sci. Rev.* 114, 331–383. doi: 10.1007/s11214-004-1432-2
- Ebert-Uphoff, I., Samarasinghe, S. M., and Barnes, E. A. (2019). Thoughtfully using artificial intelligence in Earth science. *Eos Transactions American Geophysical Union*. 100. doi: 10.1029/2019EO135235
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognit. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010
- Heidke, P. (1926). Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst (Calculation of the success and goodness of strong wind forecasts in the storm warning service). *Geografiska Annaler Stockholm* 8, 301–349. doi: 10.1080/20014422.1926.11881138
- Iten, R., Metger, T., Wilming, H., del Rio, L., and Renner, R. (2020). Discovering physical concepts with neural networks. *Phys. Rev. Lett.* 124:10508. doi: 10.1103/PhysRevLett.124.010508
- Jia, X., Hansen, K. C., Gombosi, T. I., Kivelson, M. G., Tóth, G., Dezeeuw, D. L., et al. (2012). Magnetospheric configuration and dynamics of Saturn's magnetosphere: a global MHD simulation. *J. Geophys. Res.* 117:A05225. doi: 10.1029/2012JA017575
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2019). Machine learning for the geosciences: challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* 31, 1544–1554. doi: 10.1109/TKDE.2018.2861006
- Kennelly, T. J., Leisner, J. S., Hospodarsky, G. B., and Gurnett, D. A. (2013). Ordering of injection events within Saturnian SLS longitude and local time. *J. Geophys. Res.* 118, 832–838. doi: 10.1002/jgra.50152
- Kokar, M. M. (1986). “Coper: A methodology for learning invariant functional descriptions,” in *Machine Learning*. The Kluwer International Series in Engineering and Computer Science (Knowledge Representation, Learning and Expert Systems), eds T. M. Mitchell, J. G. Carbonell, and R. S. Michalski (Boston, MA: Springer), 151–154. doi: 10.1007/978-1-4613-2279-5\_34
- Krimigis, S. M., Mitchell, D. G., Hamilton, D. C., Livi, S., Dandouras, J., Jaskulek, S., et al. (2004). Magnetosphere Imaging Instrument (MIMI) on the Cassini mission to Saturn/Titan. *Space Sci. Rev.* 114, 233–329. doi: 10.1007/s11214-004-1410-8
- Lai, H. R., Russell, C. T., Jia, Y. D., Wei, H. Y., and Dougherty, M. K. (2016). Transport of magnetic flux and mass in Saturn's inner magnetosphere. *J. Geophys. Res.* 121, 3050–3057. doi: 10.1002/2016JA022436
- Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weath.* 16, 2079–2102. doi: 10.1029/2018SW002067
- Lockhart, J. W., and Weiss, G. M. (2014a). “Limitations with activity recognition methodology and data sets,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjust Publication* (Seattle, WA), 747. doi: 10.1145/2638728.2641306
- Lockhart, J. W., and Weiss, G. M. (2014b). “The benefits of personalized smartphone-based activity recognition models,” in *2014 SIAM International Conference on Data Mining* (Philadelphia, PA). doi: 10.1137/1.9781611973440.71
- Manzato, A. (2005). An odds ratio parameterization for ROC diagram and skill score indices. *Weath. Forecast.* 20, 918–930. doi: 10.1175/WAF899.1
- Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Available online at: <https://christophm.github.io/interpretable-ml-book/>
- Morley, S. K. (2020). Challenges and opportunities in magnetospheric space weather prediction. *Space Weath.* 18:e2018SW002108. doi: 10.1029/2018SW002108
- Müller, A. L., Saur, J., Krupp, N., Roussos, E., Mauk, B. H., Rymer, A. M., et al. (2010). Azimuthal plasma flow in the Kronian magnetosphere. *J. Geophys. Res.* 115:A08203. doi: 10.1029/2009JA015122
- NASA Headquarters (1980). *Voyager Backgrounder, Release No: 80-160*. NASA Headquarters. Available online at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19810001583.pdf>
- NASA Jet Propulsion Laboratory (2017a). *Cassini Huygens by the Numbers*. NASA Jet Propulsion Laboratory. Available online at: <https://solarsystem.nasa.gov/resources/17761/cassini-huygens-by-the-numbers/>
- NASA Jet Propulsion Laboratory (2017b). *Mars Reconnaissance Orbiter By the Numbers*. NASA Jet Propulsion Laboratory. Available online at: <https://mars.nasa.gov/resources/7741/mars-reconnaissance-orbiter-by-the-numbers/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. Available online at: <https://dl.acm.org/doi/10.5555/1953048.2078195>
- Planetary Society (2020). *Missions to Mars*. Planetary Society. Available online at: <https://www.planetary.org/explore/space-topics/space-missions/missions-to-mars.html>
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. doi: 10.1016/j.jcp.2018.10.045
- Ren, H., Stewart, R., Song, J., Kuleshov, V., and Ermon, S. (2018). Learning with weak supervision from physics and data-driven constraints. *AI Mag.* 39, 27–38. doi: 10.1609/aimag.v39i1.2776
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. doi: 10.1038/s42256-019-0048-x
- Swiger, B. M., Liemohn, M. W., and Ganushkina, N. Y. (2020). Improvement of plasma sheet neural network accuracy with inclusion of physical information. *Front. Astron. Space Sci.* doi: 10.3389/fspas.2020.00042
- Waskom, M., Botvinnik, O., Ostblom, J., Lukauskas, S., Hobson, P., Gelbart, M., et al. (2020). *mwaskom/seaborn: v0.10.0* (January 2020). doi: 10.5281/zenodo.3629446

**Conflict of Interest:** 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.

Copyright © 2020 Azari, Lockhart, Liemohn and Jia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Dynamic Time Warping as a New Evaluation for Dst Forecast With Machine Learning

Brecht Laperre\*, Jorge Amaya and Giovanni Lapenta

Department of Mathematics, Centre for Mathematical Plasma Astrophysics, KU Leuven, Leuven, Belgium

Models based on neural networks and machine learning are seeing a rise in popularity in space physics. In particular, the forecasting of geomagnetic indices with neural network models is becoming a popular field of study. These models are evaluated with metrics such as the root-mean-square error (RMSE) and Pearson correlation coefficient. However, these classical metrics sometimes fail to capture crucial behavior. To show where the classical metrics are lacking, we trained a neural network, using a long short-term memory network, to make a forecast of the disturbance storm time index at origin time  $t$  with a forecasting horizon of 1 up to 6 h, trained on OMNIWeb data. Inspection of the model's results with the correlation coefficient and RMSE indicated a performance comparable to the latest publications. However, visual inspection showed that the predictions made by the neural network were behaving similarly to the persistence model. In this work, a new method is proposed to measure whether two time series are shifted in time with respect to each other, such as the persistence model output versus the observation. The new measure, based on Dynamical Time Warping, is capable of identifying results made by the persistence model and shows promising results in confirming the visual observations of the neural network's output. Finally, different methodologies for training the neural network are explored in order to remove the persistence behavior from the results.

## OPEN ACCESS

**Edited by:**

Veronique A. Delouille,  
Royal Observatory of Belgium,  
Belgium

**Reviewed by:**

Hui Li,  
National Space Science Center (CAS),  
China  
Octav Marghitu,  
Space Science Institute, Romania

**\*Correspondence:**

Brecht Laperre  
brecht.laperre@kuleuven.be

**Specialty section:**

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 06 March 2020

**Accepted:** 04 June 2020

**Published:** 22 July 2020

**Citation:**

Laperre B, Amaya J and Lapenta G (2020) Dynamic Time Warping as a New Evaluation for Dst Forecast With Machine Learning.  
*Front. Astron. Space Sci.* 7:39.  
doi: 10.3389/fspas.2020.00039

## 1. INTRODUCTION

The disturbance storm time (Dst) index is calculated from the measurements of four ground-based magnetometer stations, located close to the equator and spread evenly across the Earth (Sugiura and Kamei, 1991). The index, introduced by Sugiura (1963), is the average of the magnetic disturbance of the Earth's magnetic field horizontal component. Most often, the Dst is used as a measure of the strength of the axi-symmetric magnetosphere currents, capturing the dynamics of the inner magnetospheric current system. The most important types of these dynamics are geomagnetic storms. The Dst index can be used to identify the three phases of these storms: the initial phase, the main phase and the recovery phase.

Geomagnetic storms are large perturbations in the Earth's magnetic field. They are caused by the coupling between solar wind and magnetosphere, in particular the southward component of the interplanetary magnetic field (IMF) (Burton et al., 1975). When magnetic reconnection happens between the IMF and the Earth's magnetosphere, an influx of energetic particles from the solar wind into the magnetosphere occurs. This increases the intensity of the Earth's ring

current, which is reflected by the Dst index (Akasofu, 1981; Gonzalez et al., 1994). In the case of very intense geomagnetic storms, satellites and power grids can be damaged, as well as extensive pipeline systems such as those used to transport gas, oil and water (Kasinskii et al., 2007). The most prominent examples of intense geomagnetic storms are the Carrington event and the 2003 Halloween storm. From these historical storms, it is clear that there will be global economical losses and damages when another intense geomagnetic storm would hit Earth (Kapperman and Albertson, 1990; Pulkkinen et al., 2005; Council, 2008). Consequently, the ability to timely forecast geomagnetic storms has been a topic of interest in geophysics for the past three decades.

Because of how the Dst index is linked to geomagnetic storms, forecasting the Dst is the most direct way of forecasting geomagnetic storms. Burton et al. (1975) was one of the first to construct a model of the Dst. Their linear data-driven model consisted of a driver and decay term. These terms were estimated using the available data of the solar wind velocity and density, and the southward component of the IMF. Others have tested and improved the initial model by modifying the driver and decay term (see e.g., Klimas et al., 1998; Temerin and Li, 2002, 2006), with Ji et al. (2012) providing an overview of these models. However, a linear coupling between solar wind and magnetosphere is inadequate for predicting large geomagnetic disturbances, and non-linear systems are required to fully capture the behavior of the Dst (Iyemori, 1990).

A popular approach to modeling this non-linearity is through neural networks. Feed-forward neural networks were used first for this application. Lundstedt and Wintoft (1994) was one of the first to use this type of neural network, using as input the  $B_z$  component of the IMF and the velocity and density of the solar wind to forecast the Dst 1 h in advance. Later, Stepanova and Pérez (2000) used the previous Dst values as input to predict the Dst 3 h in advance. Bala and Reiff (2012) was able to forecast the Dst 6 h in advance by using the Boyle index as their most important input (Boyle et al., 1997). Lazzús et al. (2017) used a feed-forward neural network, but used particle-swarm optimization instead of backpropagation (Eberhart and Kennedy, 1995) to train their model. Using the past Dst as an input, they were able to provide a forecast of the Dst up to 6 h in advance, showing the benefit of this type of training algorithm.

With the introduction of the recurrent neural network, in particular the Elman recurrent network (Elman, 1990), new forecast models where introduced. Wu and Lundstedt (1997) used an Elman recurrent network to provide a forecast of the Dst index up to 6 h in advance. Similarly, many more used an Elman recurrent network and input from the solar wind to forecast the Dst index (see e.g., Barkhatov et al., 2001; Lundstedt et al., 2002; Palloccchia et al., 2006; Watanabe et al., 2002). More recently, Gruet et al. (2018) used a long short-term memory (LSTM) neural network instead, and combined it with Gaussian processes to provide both a regressive and a probabilistic forecast of the Dst for 1–6 h in advance (Hochreiter and Schmidhuber, 1997).

However, some authors detected problems with the forecast of the Dst. Stepanova and Pérez (2000) used a feed-forward neural network and previous Dst values to provide a forecast up to 3

h in advance. More advanced forecasts had shown a time shift between the observed and predicted Dst, forecasting geomagnetic storms too late. This effect was also detected by Wintoft and Wik (2018), who evaluated forecasts of the Kp and Dst with their neural network up to 3 h in advance. Their Kp forecast had a time shift between forecast and observation for 2 and 3 h in advance, while Dst forecast of the main phase of geomagnetic storms showed time shifts at already 1 h in advance.

This paper aims to highlight this time shift problem. Section 2 sets up an experiment where a recurrent neural network is trained to forecast the Dst 6 h in advance. It compares our model with those in the literature, and concludes by highlighting the time shift observed in geomagnetic storm forecasts. Then, in section 3 a new measure is introduced that is capable of accurately measuring this time shift between observation and prediction. Finally, section 4 looks further into why this time shift behavior is observed, and potential solutions.

## 2. THE EXPERIMENT

This section concerns the details of the experiment, together with the discussion of the results that lead to the discovery of the problems. The initial problem was to train a neural network to forecast the values of the Dst-index at times  $t + l$ ,  $l \geq 1$ , while having information up to time  $t$ .

The layout is the following: first the data used to train the neural network model is explained. Second, the processing of the data is discussed. Then the method of evaluation is described. Afterwards, the neural network model is described in detail. Finally, the results of the model are analyzed and discussed.

The following terminology will be used throughout this section. Box et al. (2015) defines  $\text{Dst}(t + l)$  as a forecast at *origin*  $t$  with a *lead time*  $l$ . We will differ from this terminology, instead naming the  $\text{Dst}(t + l)$  a forecast at *forecasting horizon*  $t + l$ .

### 2.1. The Data

The data used to train and test the neural network model were obtained from the NASA/GSFC's OMNI database (King and Papitashvili, 2005). From the database, the hourly averages of the solar wind velocity  $V_{sw}$  and density  $\rho_{sw}$ , the IMF z-component  $B_z$  and magnitude  $|B|$ , together with the geomagnetic Dst index were extracted. These physical quantities will be referred to as 'features' throughout the paper. In particular, these features were measured between 00:00, 14 January 2001 and 23:00, 31 December 2016, and the full extracted data set contains a total of 139,944 entries.

### 2.2. Preprocessing the Data

The preprocessing of the data is done with the following steps. The data are first split into a training, test, and validation set, to prevent information bias through validation leakage. Next, each set is scaled and normalized using scaling parameters measured from the training set. Finally invalid measurements are removed from each of the sets by using a sliding window. Each step of this procedure is discussed in more details below.

Before assigning entries to one of the data sets, the data are split into monthly samples. The first sample corresponds to the month January of 2001, the final sample to the month December

of 2016, equating to 181 samples. The reason for this split is the high temporal correlation of the data. Hourly samples are highly correlated, which causes the model to artificially perform better on the test set (Camporeale, 2019).

The data sets are constructed as follows. The test set consists of the months April, August, and December of each year, corresponding to 25% of the total data. These months have been chosen arbitrarily, and mainly ensure a good spread of the test data over the given time period. We expect to see the same kind of results when the test set would be taken differently. An experiment to determine the effect of the chosen test set on the performance will be done in section 2. From the remaining data, 60% of the months are distributed randomly into the training set. The remaining 15% of the data are placed in the validation set. This corresponds to 77,544, 20,520, and 33,120 entries in training, validation, and test set, respectively.

The choice of training, validation and test set plays a huge role in the performance of the model. A difference in these sets makes direct comparison of forecasting models difficult, as stated by Lazzús et al. (2017). In order to measure the variance caused by this choice, 10-fold cross validation is performed and the results are reported in section 2.6.

After this step, the sets are scaled and normalized. This ensures that every feature lies within the same range of values, ensuring comparability of the different features and faster convergence of the machine learning algorithm (see e.g., Juszczak et al., 2002). The full transformation process is done in two steps: first the scaling constants are determined from the training set, then the transformation is applied on the training, validation, and test set. The features are transformed by removing the mean and scaling to unit variance:

$$\begin{aligned}\bar{X}_{train} &= \frac{X_{train} - \mu_{train}}{\sigma_{train}}, & \bar{X}_{valid} &= \frac{X_{valid} - \mu_{train}}{\sigma_{train}}, \\ \bar{X}_{test} &= \frac{X_{test} - \mu_{train}}{\sigma_{train}},\end{aligned}\quad (1)$$

where  $X$  is the data set,  $\mu_{train}$  the mean of the training data set and  $\sigma_{train}$  the standard deviation of the training set.

The final step of the preprocessing extracts valid time series window samples from each set. A sample is valid if, for every feature, there are no missing or invalid measurements. The output can be ignored, because the Dst index contains no missing values. The samples are extracted by moving a sliding window over the full data set. The size of the sliding window is equal to the size of the length of the time series used as input for the model, which in our case will have a length 6 h. We chose 6 h based on results of Lazzús et al. (2017), who evaluated their time series input from length  $t$  back to  $t - 48h$  through an exhaustive procedure, and found no significant improvement in the forecast when using data from times further in the past than  $t - 6h$ . The target of the forecast is the Dst index from times  $t + 1h$  to  $t + 6h$ . We chose 6 h to be able to compare to other results found in the literature. After the preprocessing, there are 74,117, 19,596, and 32,166 valid samples in the training, validation, and test set, respectively. This corresponds to about 96% of the initial total data.

## 2.3. Evaluation of the Model

Throughout the paper, evaluation of the model's forecast is done through three distinct methods. The first method compares the neural network model to a naive forecasting model, a so-called baseline model. The second method is evaluation by use of a set of metrics. The third method is k-fold cross-validation (k-fold CV), a statistical method to test the model's predictive capabilities. All of these methods are explained in this section.

### 2.3.1. Baseline Model

The baseline model is a simplified empirical law that can generate a zero order forecast. The most simple type of time-series forecast is done with the persistence model. This model assumes that the value at the current time step does not change, so the next time step is predicted as:

$$Dst(t + 1h) = Dst(t). \quad (2)$$

This model is easily extended to forecast multiple hours in the future:

$$Dst(t + p) = Dst(t), \quad p \in \mathbb{N}. \quad (3)$$

The work by Owens et al. (2013) has shown that the persistence model can be a reliable predictor for some solar wind parameters, comparable to numerical models when evaluated in a point-by-point assessment method. In particular, geomagnetic activity and solar wind speed show good results when evaluated with a 27-day persistence model, and can be used as a benchmark for more sophisticated models.

### 2.3.2. Metrics

Now the set of metrics used for the model evaluation are defined. The root mean square error (RMSE) and the Pearson linear correlation coefficient ( $R$ ) are often used for the evaluation of time series. In addition, the set of fit performance metrics recommended by Liemohn et al. (2018) will also be used. These are the linear fit parameters, the mean absolute error (MAE), the mean error (ME), and the prediction efficiency (PE). All these metrics and their definitions are summed up in this section. Define  $M_i$  as the forecast of the model and  $O_i$  the corresponding real observational value, with  $i = 1, \dots, N$ , and  $N$  the number of samples.

- The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2}. \quad (4)$$

This metric puts emphasis on outliers, which in the case of the Dst index corresponds to geomagnetic storms. A low RMSE thus corresponds to good accuracy of the forecast of geomagnetic storms.

- The Pearson linear correlation coefficient ( $R$ ), is given by:

$$R = \frac{\text{cov}(M_i, O_i)}{\sqrt{\text{var}(M_i)\text{var}(O_i)}}. \quad (5)$$

This correlation coefficient gives a more global view of the prediction. It indicates if the model predicts the trend of the index correctly.

- Since the model  $M$  is predicting the observations  $O$ , a linear relationship is expected. Computing the linear fit of the model  $M$  to the observations  $O$  allows the results to be evaluated independent of time. Call  $A$  the offset of the linear fit, and  $B$  the slope. Then these parameters linearly relate  $M$  to  $O$  as follows:

$$M_i = A + B \cdot O_i. \quad (6)$$

In the case of a perfect prediction,  $B$  is 1 and  $A$  is 0.

- The MAE emphasizes the “usual” state of the index, which in the case of the Dst index corresponds to quiet time, where no geomagnetic storms are happening. The MAE indicates how well the model predicts these values. This value is defined as

$$MAE = \frac{1}{N} \sum_i^N |M_i - O_i|. \quad (7)$$

- The ME indicates if the model systematically over- or under-predicts the observations, based on its positive or negative sign. If the mean error is zero, under and over predictions are balanced:

$$ME = \frac{1}{N} \sum_i^N (M_i - O_i). \quad (8)$$

- Finally, the PE is used to quantify the model’s ability at reproducing the time variation of the Dst index:

$$PE = 1 - \frac{\sum_{i=1}^N (M_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2}, \quad (9)$$

where  $\bar{O}$  is the average of the observational values. The maximum value of the PE is 1, corresponding to a perfect prediction at all times. A prediction efficiency equal to or less than zero indicates that the model is incapable of forecasting the time variation seen in the observations.

## 2.4. K-Fold Cross-Validation

K-fold CV is a training technique for the model in order to learn its capabilities, without having to make use of the test set and so prevent an information bias. When applying CV, the training and validation sets defined in section 2.1 are used, unless stated otherwise.

The technique firsts randomly splits the (monthly) data into  $k$  equal-sized partitions (or folds). One of these folds is picked as the validation set, and the other remaining folds are used as the training set. The model is trained on the training set and evaluated on the validation set, and the evaluation is stored. This is repeated until every fold has been used exactly once as the validation set. By taking the average of the result of each run, an estimate of the predictive performance is given. This allows us to evaluate the model on different parameters, both hyperparameters and features, while preventing any optimization on the test set, as this would otherwise invalidate our results.

## 2.5. The Model

A neural network model is constructed and trained to forecast the Dst index for forecasting horizons  $t + 1h$  up to  $t + 6h$ . As input, the model receives a multidimensional time series  $X_t$ , containing the data described in section 2.1, ranging from time  $t - 6h$  to time  $t$ , as displayed in Equation (10). The output then consists of a 6-dimensional vector  $Y_t$ , corresponding to the forecast Dst values for forecasting horizon  $t + 1h$  up to  $t + 6h$ .

$$\underbrace{\begin{pmatrix} V_{sw}(t-6h) & V_{sw}(t-5h) & \dots & V_{sw}(t) \\ |B|(t-6h) & |B|(t-5h) & \dots & |B|(t) \\ \vdots & \vdots & \ddots & \vdots \\ Dst(t-6h) & Dst(t-5h) & \dots & Dst(t) \end{pmatrix}}_{X_t} \rightarrow \underbrace{\begin{pmatrix} Dst(t+1h) \\ Dst(t+2h) \\ \vdots \\ Dst(t+6h) \end{pmatrix}}_{Y_t} \quad (10)$$

The programming language Python, and in particular the package PyTorch (Paszke et al., 2017), was used to implement and train the neural network. A link to the source code can be found in Appendix.

### 2.5.1. Model Description and Training

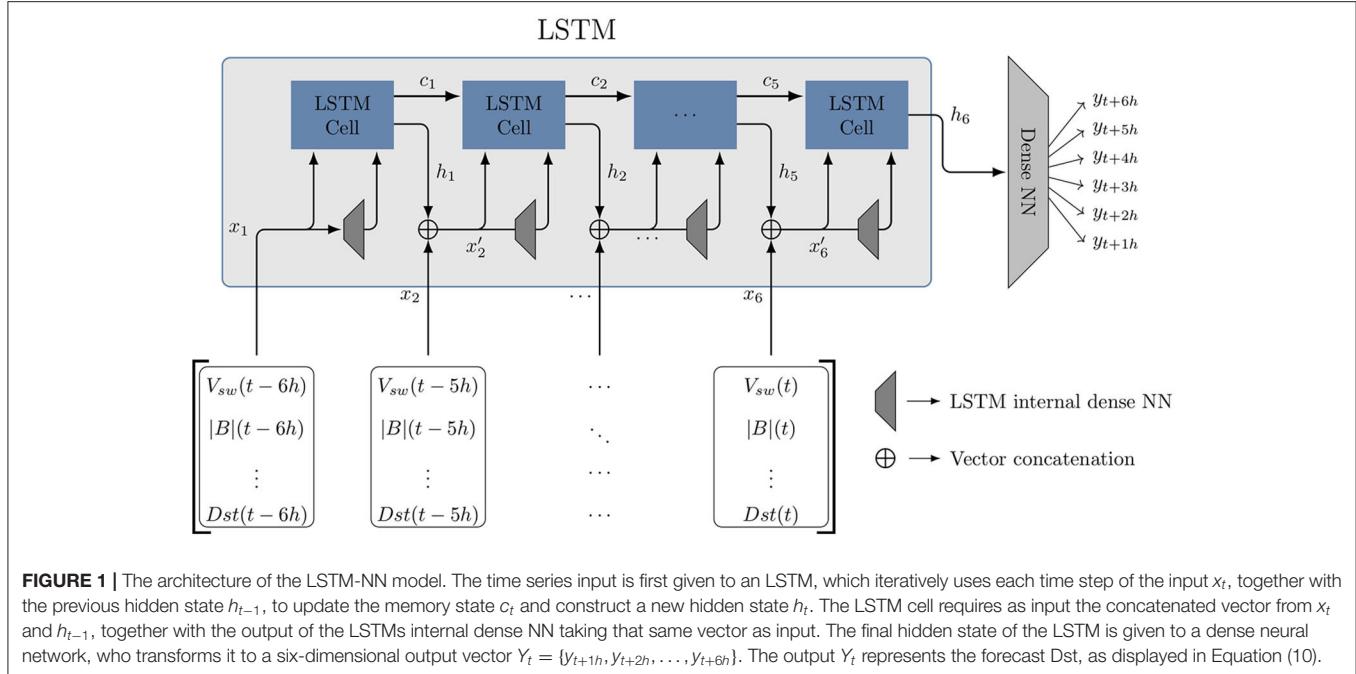
The model architecture of the neural network model consists of a LSTM, combined with a dense neural network, as shown in Figure 1.

LSTMs are a type of recurrent neural network, where the output of the previous iteration is used as an additional input. When multiple past events are used as inputs, classic recurrent neural networks loose training information in a phenomena called gradient vanishing (Hochreiter, 1998). LSTMs are designed to work better for long time series, by incorporating two internal memory states: the hidden state,  $h_t$ , and the memory state,  $c_t$ . A detailed explanation of how these memory states retain information and how they work is presented in Hochreiter and Schmidhuber (1997). We refer the reader to this publication for mode details on the exact internal workings of the LSTM.

The LSTM memory machinery is encapsulated in the LSTM cell. Figure 1 shows how the input,  $X_t$  is connected to the output,  $Y_t$ . Each time step,  $x_{t-\tau}$  of the input  $X_t$  is given to a single LSTM cell. The input is concatenated to the hidden memory output of the preceding LSTM cell,  $h_{t-1}$ , forming  $x'_t$ . The input  $x'_t$  and the internal weights of the current LSTM cell change the information  $c_{t-1}$  from the previous cell by propagating, modifying or blocking the information. Finally, a new hidden state  $h_t$  is created by first transforming  $x'_t$  with one of the LSTMs internal neural networks and then combining it with the new memory state  $c_t$ . The final output of the cell, the hidden state  $h_t$ , is obtained by the internal combination of the cell input  $x'_t$  and the memory of previous cells,  $c_t$ , using an internal neural network.

The final hidden state vector,  $h_6$ , is then fed to a classical, fully connected feed-forward neural network, where the input is transformed to a vector of the size of the target vector  $Y_t$ . Throughout the rest of the paper, we will refer to this model as the “LSTM-NN model.”

The model is trained using the RMSProp method, an unpublished adaptive learning rate method proposed by Geoff Hinton (see <http://www.cs.toronto.edu/~tijmen/csc321/slides/>



lecture\_slides\_lec6.pdf). As with any other neural network based technique, LSTMs “learn” by iteratively adjusting the weights of the inter-neuron connections. The iterative process is similar to the Newton-Raphson method, adjusting the free parameter of the model by gradual changes based on the derivative of the error between the real output and the forecast output (Hecht-Nielsen, 1992; LeCun et al., 2012). LSTMs are composed of a chain of neural networks that learn together. This requires a special optimization method called RMSprop. RMSprop ensures that the error is correctly propagated backwards through all the chain of neural networks that compose the LSTM. The method has been used successfully for training LSTMs in generating complex sequences (Graves, 2013).

The PyTorch library provides an implementation of this method. The error criterion of the model was the mean squared error loss:

$$MSE = \frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2. \quad (11)$$

When the model is trained with a training set, the error on this set will be smaller for every iteration (or epoch). In order to prevent over-fitting, the process where the model starts memorizing the training set instead of learning the training set, after every epoch, the performance of the model on the validation set is checked. When the model performance stops improving on the validation set is a good indicator that the model is starting to over-fit on the training set, and we can stop the training of the model. This is the classic early stopping method.

### 2.5.2. Parameters of the Model

When training a neural network, there are many parameters that can be chosen that have an impact on the performance.

However, finding the optimal value for these parameters is problem-dependent, the so-called No Free Lunch theorem Wolpert and Macready (1997), and finding the optimal values is a computationally exhaustive task. In our model, there are three sets of hyperparameters. The first set is intrinsic to the architecture of our model itself, the second set to the learning method, and the third set to the training method. The hyperparameters were obtained by manually tweaking their values over the course of 15–20 runs and evaluating their performance using 7-fold CV. We found that for our particular case, we observed no significant changes in the accuracy of the model caused by the tweaking. Because the model seemed robust under the tweaking, we decided to not do an exhaustive search for the optimal values of the hyperparameters. We now provide an overview of the important hyperparameters and the values we gave them. A final list of the parameters are summed up at the end of this section.

In our model itself, the LSTM has a few hyperparameters that have impact on its performance. The first is the number of neurons in the hidden layer of the LSTM. This number must be large enough to ensure it can encode the process behind the data, but not too large to prevent over-fitting and plain computational cost. This number has been determined by performing CV, and we found the best performance to be around 50 neurons. Next there is the number of layers in the LSTM. Multiple LSTM's can be stacked on top of each other, where the first LSTM gives the intermediate hidden states (see Figure 1) as input for the second LSTM, and so on. This increases the complexity and computational cost of the model. In our search we tested using multiple layers of LSTMs, but did not find any significant performance increase and set the number of layers to 1. Finally, there is the option to make the LSTM bidirectional, where the model has access to both future and past states. But this is

**TABLE 1** | Evaluation of the LSTM-NN model and the persistence model with the metrics from section 2.3.2.

Forecasting horizon	RMSE (nT)	R	A	B (nT)	MAE (nT)	ME (nT)	PE
<b>LSTM-NN model</b>							
$t + 1h$	3.731	0.980	-0.166	0.960	2.391	0.318	0.960
$t + 2h$	5.689	0.953	-0.770	0.915	3.820	0.271	0.907
$t + 3h$	7.155	0.924	-1.438	0.866	4.823	0.198	0.853
$t + 4h$	8.172	0.899	-1.963	0.826	5.479	0.161	0.808
$t + 5h$	8.926	0.878	-2.358	0.793	5.929	0.168	0.771
$t + 6h$	9.566	0.859	-2.769	0.761	6.280	0.143	0.737
<b>Persistence model</b>							
$t + 1h$	4.745	0.974	1.884	<b>0.975</b>	3.265	2.179	0.935
$t + 2h$	6.853	0.934	0.101	<b>0.935</b>	4.400	0.875	0.865
$t + 3h$	8.998	0.895	1.571	<b>0.898</b>	5.884	2.798	0.767
$t + 4h$	9.913	0.860	-0.652	<b>0.864</b>	6.300	0.979	0.717
$t + 5h$	10.937	0.829	-1.487	<b>0.834</b>	6.943	0.507	0.655
$t + 6h$	11.864	0.799	-1.592	<b>0.805</b>	7.464	0.742	0.594

The values in bold show where the persistence model performed better than the LSTM-NN model.

unnecessary in the context of the Dst forecast, thus this has been set to false.

The RMSProp method has a lot of tweakable parameters, but we will focus on the two most important parameters, the learning rate and the momentum. The learning rate is the most important parameter, and controls how strongly the model weights will be changed by the error gradient. A too large learning rate might cause unstable learning, with the performance swinging widely during training and preventing convergence. A too small learning rate might cause the model to barely change and converge too slowly or not at all. The momentum parameter will affect the learning rate parameter throughout the training process, and will accelerate the training process. Most often, the momentum parameter is chosen close to 1. We found that setting the learning rate to 0.0003 and the momentum to 0.8 gave us the best performance, with stable convergence of the training error.

Finally, when training the model, we can also set a few parameters that can affect the performance. There are two parameters that are important: the number of epochs and the batch-size of the training set. The number of epochs decide how many times we loop over the full training set for training. This number must be large enough that the model has time to converge to the ideal solution before the training is stopped by the classic early stopping method. The batch-size determines how many samples of the training data are given to the model before the error is computed and backpropagation is applied. Setting the batch-size to one would corresponds with so-called online learning, where the model is trained separately on every sample. The opposite is offline learning, i.e., setting the batch-size to the size of the training set, so the model is optimized on the accumulated error over the complete training set. Offline learning is almost never used as it fails to learn more outlying cases, and online learning is more prone to over-fitting. Using a small batch-size is typically recommended. We found that setting

the batch-size to 64 gave a fast convergence and did not have a large impact on the performance.

Finally, we sum up the parameters of the model:

### 1. LSTM hyperparameters

- Number of hidden layer neurons: 50
- Number of layers: 1
- Bidirectional: False

### 2. RMSProp hyperparameters:

- Learning rate: 0.0003
- Momentum: 0.8

### 3. Training hyperparameters

- Number of epochs: 30
- Training set batch-size: 64

## 2.6. Results and Discussion

The LSTM-NN model is now evaluated using the defined metrics and baseline model of section 2.3. The evaluation is discussed and a comparison of the model to some of the latest forecasting models is made. Finally, the forecast is visually observed.

The first analysis examines whether the LSTM-NN model performs better than the persistence model defined in section 2.3.1. **Table 1** displays the results from the metrics defined in section 2.3.2, applied on the forecast of both the LSTM-NN model and the persistence model. The LSTM-NN model was found to overall perform better than the persistence model. Only the linear relation of the persistence model is consistently better for every forecasting horizon compared to the LSTM-NN model.

Taking a closer look at the remaining metrics, it seems that the MAE of the LSTM-NN model and the persistence model are similar. It increases with the forecasting horizon, and always remains smaller than the RMSE. The correlation, linear model parameters, and prediction efficiency of the LSTM-NN model are close to that of the persistence model for forecasting horizon  $t+1h$  and  $t+2h$ ; however, its accuracy quickly disappears for later forecasting horizons. This would indicate that the persistence model could serve as a strong benchmark for nowcasts of the Dst index. Taking into account the results reported in **Table 1**, we conclude that using the more complicated LSTM-NN model will result in better forecasts. In particular, forecasts made at forecasting horizon  $t+3h$  to  $t+6h$  show significant improvement in accuracy compared to the persistence model.

Next we compare the LSTM-NN model to the models reported in the work of Gruet et al. (2018) and Lazzús et al. (2017). Both publications present a neural network trained on OMNIWeb data used to forecast the Dst index. The model by Gruet et al. (2018) also makes use of LSTM modules in their model, while the model of Lazzús et al. (2017) consists of a feed-forward neural network instead of a RNN, trained with a particle-swarm optimization method. The performance of their models were evaluated with the RMSE and Pearson correlation coefficient, and are summarized in **Table 2**, together with the results from the LSTM-NN model and the persistence model.

However, before we can do quantitative comparison of the LSTM-NN model with the two other models presented in the

**TABLE 2 |** The RMSE and Pearson linear correlation coefficient of the persistence model and LSTM-NN model compared to the models of Gruet et al. (2018) and Lazzús et al. (2017).

Forecasting horizon	Persistence	LSTM-NN	Gruet et al. (2018)	Lazzús et al. (2017)
<b>Correlation</b>				
$t + 1h$	$0.974 \pm 0.003$	$0.980 \pm 0.008$	0.966	0.982
$t + 2h$	$0.934 \pm 0.009$	$0.953 \pm 0.010$	0.946	0.949
$t + 3h$	$0.895 \pm 0.015$	$0.924 \pm 0.013$	0.923	0.918
$t + 4h$	$0.860 \pm 0.019$	$0.899 \pm 0.017$	0.902	0.887
$t + 5h$	$0.829 \pm 0.023$	$0.878 \pm 0.019$	0.882	0.858
$t + 6h$	$0.799 \pm 0.026$	$0.859 \pm 0.021$	0.865	0.826
<b>RMSE (nT)</b>				
$t + 1h$	$4.75 \pm 0.47$	$3.73 \pm 0.78$	5.34	4.24
$t + 2h$	$6.85 \pm 0.85$	$5.69 \pm 0.59$	6.65	7.05
$t + 3h$	$9.00 \pm 1.10$	$7.16 \pm 0.55$	7.86	8.87
$t + 4h$	$9.91 \pm 1.39$	$8.17 \pm 0.59$	8.86	10.44
$t + 5h$	$10.94 \pm 1.56$	$8.92 \pm 0.70$	9.59	11.65
$t + 6h$	$11.86 \pm 1.74$	$9.57 \pm 0.83$	10.24	13.09

The confidence intervals were obtained from the cross-validation experiment, detailed in section 2.6 and shown in **Figure 2**.

literature, we have to keep in mind the following problem: direct comparison of two different neural network models is considered bad practice when the sets used to train and test both models are not identical, as also stated by Lazzús et al. (2017). Because we are unable to recreate the exact same training and test set used by Gruet and Lazzús, we instead will quantify the impact that the choice of training and test set has on the performance of the LSTM-NN model. By performing 10-fold CV, we can measure the variance in performance caused by the choice of training and test set, as 10-fold CV will replicate the effect of training the model for 10 different choices of training and test set.

From this we can learn two things. The first has already been discussed in section 2.2, namely a way to quantify the effect of our final choice of training, validation and test set. By choosing an ideal training and test set, it is possible to cause an artificial improvement of the model performance. Computing the variance caused by the choice of training and test makes it possible to determine if the reported performance is an outlier, or expected. The results are shown in **Figure 2**. The average performance of the LSTM-NN and the persistence model from the CV is indicated by the blue and dark green bars, and the standard deviation on the performance is indicated by the error bars. The uncertainty interval obtained from this experiment has also been included in **Table 2**. We find that for both RMSE and correlation, the reported performance of the LSTM-NN and the persistence model lies inside the variation, indicating that there is no artificial improvement of the results by choosing an ideal training and test set.

The second reason to do this experiment was to be able to perform a qualitative comparison of the LSTM-NN with the model of Gruet and Lazzús. Their reported values have also been added to **Figure 2**. Let us first look at the results for the correlation in the left bar chart. The model of Lazzús et al. (2017)

has a performance that is comparable to that of the LSTM-NN model for forecasting horizons  $t + 1h$  to  $t + 4h$ . However, for later forecasting horizons, the LSTM-NN performs significantly better. The model by Gruet et al. (2018) is outperformed by the LSTM-NN for the forecasting horizon  $t + 1h$ , but performance equivalently for all the later forecasting horizons.

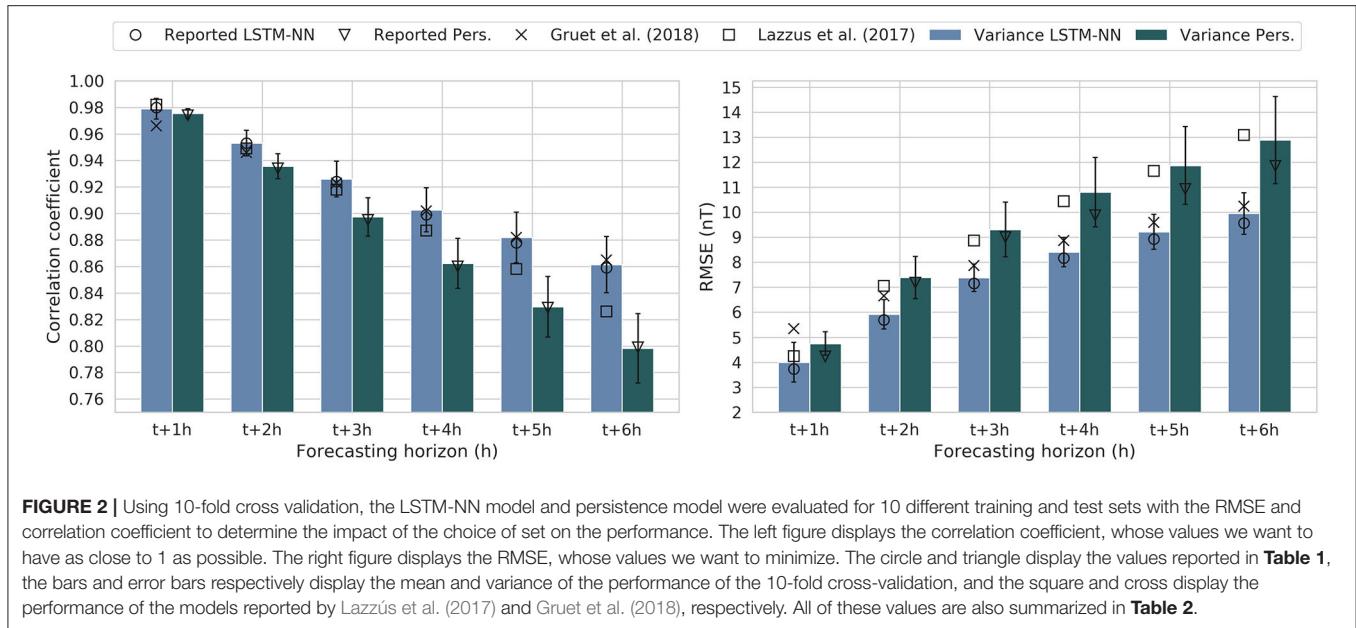
Looking at the RMSE, we see that the model by Lazzús has comparable performance for forecasting horizon  $t + 1h$ . However, for later forecasting horizons the LSTM-NN model is significantly better. This seems to agree with the observation made by Gruet et al. (2018), stating that the choice of an LSTM module in the model architecture improves the accuracy of the forecast. Looking at the performance of Gruet, the LSTM-NN shows a significant improvement for forecasting horizons  $t + 1h$  to  $t + 2h$ , but is equivalent in performance for later times.

In conclusion, our model seems to be comparable to that of Lazzús for forecasting horizons  $t + 1h$  and  $t + 2h$ , but outperforms it for later times. Our model shows some improvement over that of Gruet for forecasting horizons  $t + 1h$  and  $t + 2h$ , but is otherwise comparable to theirs.

Finally, a visual observation of the forecast is analyzed. **Figure 3** displays three geomagnetic storms contained in the test set, together with the forecast of the LSTM-NN model for forecasting horizon  $t + 1h$ ,  $t + 3h$ , and  $t + 5h$ . The first column displays the  $t + 1h$  forecast, and seems to be an almost perfect prediction of the storm. However, the forecast of the Dst-index for forecasting horizon  $t + 3h$  and  $t + 5h$ , displayed in column 2 and 3 of **Figure 3**, shows a distinct delay in the forecasting of the main phase. Take for example the prediction at forecasting horizon  $t + 5h$ : the sudden offset of the storm is predicted 5 h too late.

This brings us to the main problem of this paper. The purpose of the experiment was to create a LSTM-NN model that forecasts the Dst-index with the same accuracy and correlation as other presented architectures. We managed to create such a model, but, when visually inspecting the forecast, it was observed that there is a distinct time shift between forecast and observation. If geomagnetic storms are forecast only when they start, it means the LSTM-NN model will not give us any more information than the persistence model. While it is not possible to say that the models from Gruet et al. (2018) and Lazzús et al. (2017) also have this problem, we believe that one should pay close attention to this problem and ensure it does not happen.

An additional problem that most modern machine learning techniques have to face is that rare events can not be properly forecasted. Neural networks learning by gradient descent requires that patterns show up frequently in the data. In order to forecast dangerous super-storms, like the 2003 Halloween or the Carrington events, the networks must have to learn to identify them. From 2001 to 2016 there are only 100 entries of the Dst index recorded with values below  $-200$  nT (including consecutive hours of individual storms). Possible solutions to this issue can be of four types: (1) data augmentation by duplication, where months with high number of storms are used multiple times in a singular epoch, (2) generative data augmentation, where a second machine learning technique, like auto-encoders or generative adversarial networks, is used



to generate artificial storms, (3) augmentation by computer simulations, using 3D models of the interaction of the solar wind and the magnetosphere of the planet to artificially generate data with large storms, and (4) multi-tier machine learning architectures, where multiple models specialize in the detection of different types of inputs and storms strengths. These solutions are out of the scope of the present paper but will be studied in a future work.

### 3. THE DYNAMIC TIME WARPING METHOD

Section 2 revealed that the LSTM-NN model failed to give an accurate forecast of the Dst index, and in particular geomagnetic storms, despite the evaluation of the model indicating that the model should have a high accuracy and correlation. This problem has also been observed by Wintoft and Wik (2018) and Stepanova and Pérez (2000), but not by other similar forecasting models. It is often unclear whether or not this was overlooked or if the forecasting model did not have this problem. Wintoft and Wik were able to detect this time shift by manually shifting their forecast in time and analysing the correlation coefficient between shifted forecast and observation, while Stepanova and Pérez visually observed this time shift.

Because this phenomena seems detrimental in the evaluation of a forecasting model, we propose a new method, which we will name the “warping measure.” This measure is more capable of quantifying the time shift between model and observation, and is based on the Dynamic Time Warping (DTW) algorithm (see Berndt and Clifford, 1994), a method that measures the relative similarity between two time series. At the very least, we expect the warping measure to be able to detect the forecast made by a persistence model. What follows first is a brief overview of the DTW algorithm, followed by the modifications we made to tailor the algorithm to our specific problem.

### 3.1. Dynamic Time Warping

The DTW algorithm is a method first developed for speech recognition and is now commonly used in the fields of economics, biology, and database analysis (see e.g., Wang et al., 2012; Skutkova et al., 2013). DTW is mainly used as a measure to investigate how much a sample time series matches or is contained in a target time series. The strength of DTW is that it can compare two time series even though they might be shifted or stretched in time, which is a property that is essential to our goal. This section summarizes the algorithm developed by Berndt and Clifford (1994). A visualization of this algorithm is shown in Figure 4. Take two time series, Q and S, of length  $n$  and  $m$ , respectively.

$$Q = [q_1, q_2, \dots, q_i, \dots, q_n], \quad (12)$$

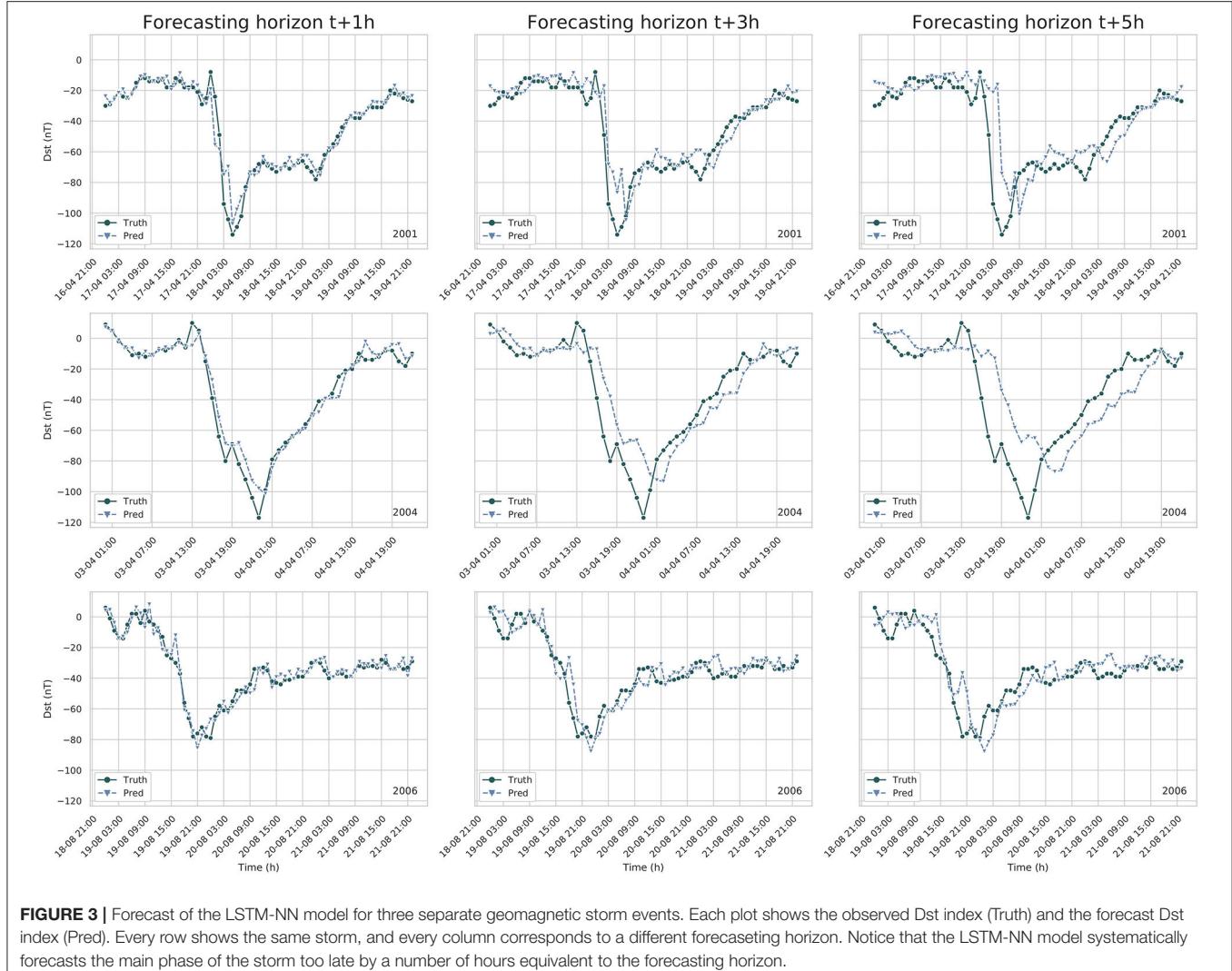
$$S = [s_1, s_2, \dots, s_j, \dots, s_m]. \quad (13)$$

The DTW algorithm first constructs a distance between these two time-series by placing them in an  $n \times m$  grid. Each grid point  $(i, j)$  then corresponds to an alignment of  $q_i$  and  $s_j$ . An alignment is given a cost by a distance function  $d(q_i, s_j)$ . The distance function can be chosen freely, and for our case the Euclidean distance function,  $d(x, y) = \sqrt{(x - y)^2}$  is used. The DTW algorithm then searches for a path (the so-called warping path)  $P$  in this grid that minimizes the sum of said distance. The warping path  $P$  can be defined as:

$$P = [p_1, p_2, \dots, p_K], \text{ with } \max(m, n) \leq K < m + n - 1, \quad (14)$$

where each point  $p_k$  corresponds to a grid point  $(i, j)_k$ . The path must then minimize the cost function, so

$$DTW(Q, S) = \min \sqrt{\sum_{i=1}^K p_k}, \quad (15)$$



**FIGURE 3 |** Forecast of the LSTM-NN model for three separate geomagnetic storm events. Each plot shows the observed Dst index (Truth) and the forecast Dst index (Pred). Every row shows the same storm, and every column corresponds to a different forecasting horizon. Notice that the LSTM-NN model systematically forecasts the main phase of the storm too late by a number of hours equivalent to the forecasting horizon.

and must hold to the following conditions:

1. Boundary conditions: the beginning and the end of the sequences are matched;
2. Continuity: there are no gaps, every point is mapped to at least one other point;
3. Monotonicity: the points are ordered in time,  $i_{k-1} \leq i_k$  and  $j_{k-1} \leq j_k$ ;
4. Warping window  $w$ : an optional constraint that sets the maximum distance in time between  $i_k$  and  $j_k$  to  $w$ :  $|i_k - j_k| \leq w$ .

In order to find this optimal path, the following dynamic programming technique can be used. Starting at point  $(1, 1)$ , the cumulative distance  $\Delta$  at each grid point is computed by the following recursive equation:

$$\Delta[i, j] = d(q_i, s_j) + \min(\Delta[i-1, j-1], \Delta[i-1, j], \Delta[i, j-1]). \quad (16)$$

Once all the cumulative distances are computed, the optimal warping path can be found by starting at the point  $(n, m)$  and

tracing backwards in the grid, taking the smallest value each time. This is displayed in **Figure 5**.

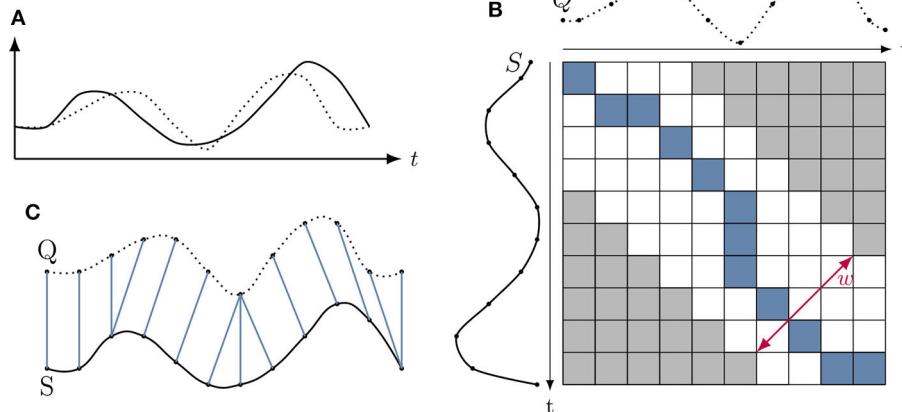
A warping window constraint can be added on the algorithm. This window will change the warping cost and warping path  $P$ . Let  $w \in \mathbb{N}$  be the warping window, then

$$\forall p_k \in P, \quad |i_k - j_k| \leq w. \quad (17)$$

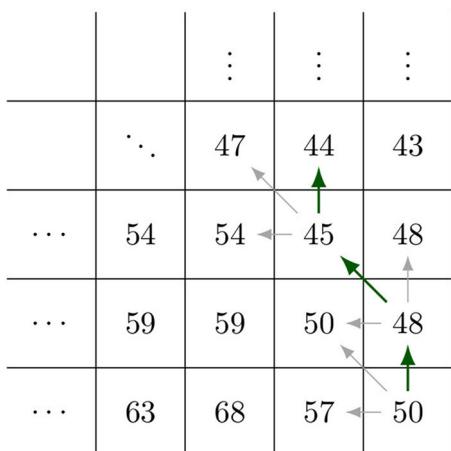
Faster and better implementations of this algorithm exist (see e.g., Keogh and Pazzani, 1999; Keogh and Ratanamahatana, 2005; Salvador and Chan, 2007; Lemire, 2009), but they are outside the scope of this text.

### 3.2. The Warping Measure

It should be mentioned that the DTW algorithm does not satisfy the necessary properties to be a metric. For example, it is easy to see that the algorithm does not satisfy the triangle inequality. Consequently, this method will be called a measure, and not a metric.



**FIGURE 4 | (A)** Shows two time series we want to compare. **(B)** Illustrates the cumulative distance matrix, together with a warping window  $w$  and the ideal warping path  $P$  in blue. **(C)** Illustrates the warping path  $P$  aligning the two time series.



**FIGURE 5 |** Illustration of how the warping path is determined from the distance matrix  $D$ .

This measure does not make use of the warping cost, and instead uses the information contained in the warping path  $W$ . The measure is then able to determine how exactly a forecast time series is shifted in time in comparison to the true or observed time series. Take the two time series  $M$  and  $O$ , where  $M$  is the predicted time series and  $O$  the true time series.

$$M = [m_1, m_2, \dots, m_i, \dots, m_n], \quad (18)$$

$$O = [o_1, o_2, \dots, o_j, \dots, o_m]. \quad (19)$$

The DTW algorithm is applied on these time series, giving a cost matrix  $D$  of dimensions  $n \times m$ . The warping constraint defined in Equation (17) is applied, and  $w$  is set equal to the forecasting horizon time. However, an additional constraint is included: the warping window is restrained such that the algorithm only compares the prediction  $M$  at  $t+p$  with the observations  $O$  from

time  $t$  to  $t+p$ , i.e., predictions are not compared to observations that are in the future. Applying this constraint can be done as a modification of the warping constraint defined in Equation (17):

$$\text{Let } m_i \text{ be the modeled value, } o_j \text{ the observation, then } w \geq i - j \geq 0. \quad (20)$$

This is also illustrated in **Figure 6**. After computing the warping path, we take each step  $p_k = (i_k, j_k)$  and compute what we define as the warp value:

$$\Delta t = |i_k - j_k|, \text{ with } \Delta t \in [0, 1, \dots, w]. \quad (21)$$

Finally, a histogram is taken from all the different values of  $\Delta t$ . The percentages reflect how time series  $M$  is shifted compared to time series  $O$ . We now present the results of this measure applied to the persistence model prediction and the LSTM-NN model prediction.

### 3.3. Results

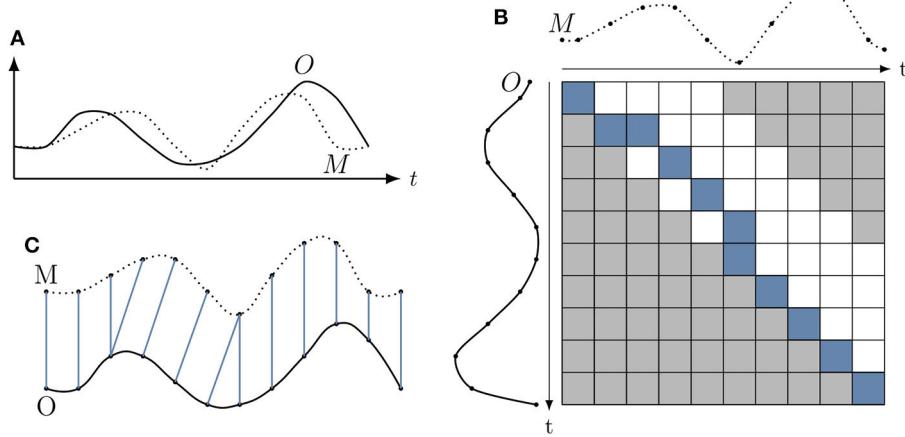
#### 3.3.1. DTW Measure Applied to the Persistence Model

The warping measure is first applied to the forecast of the persistence model. The persistence model can be seen as the textbook example for this algorithm. Assuming that the persistence model is set as follows:

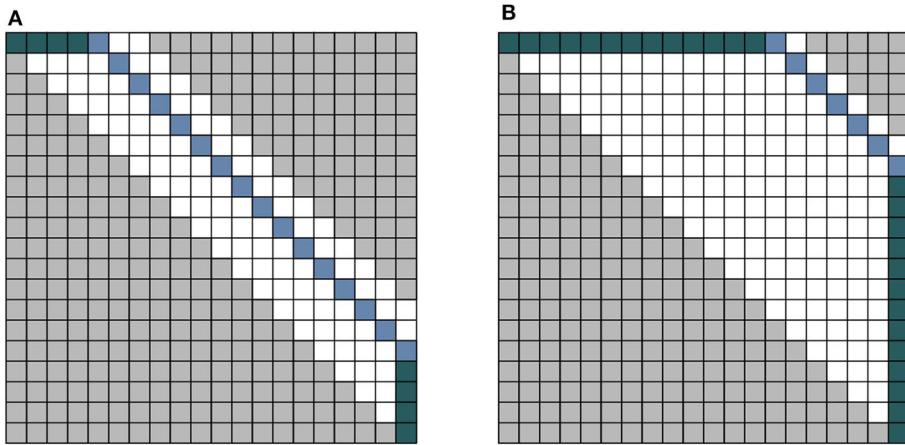
$$Dst(t+p) = Dst(t), \quad p \in \mathbb{N}, \quad (22)$$

then the algorithm should detect that almost all of the forecast values are shifted with a time  $p$  compared to the actual observation. The algorithm will not detect 100% of the values to be shifted with  $p$ , because of the constraint in the DTW algorithm that forces the beginning and the end points of the two time series to match, as discussed in section 3.

The persistence model is applied to the test set defined in section 2.1, and the resulting warp values are shown in **Table 3**.



**FIGURE 6 |** Overview of the warping measure. **(A)** Displays the two time-series  $O$  and  $M$  that are compared for alignment. However, the warping algorithm is adapted with a new window  $w$  that prevents comparing values of  $M$  with values of  $O$  that lie in the future, as seen in **(B)**. **(C)** Shows the alignment from **(B)** from which the measure is determined. Notice that there is no alignment from  $M$  to points in  $O$  that lie in the relative future of  $M$ .



**FIGURE 7 |** An illustration of a potential problem with the measure. **(A)** Illustrates the result for a persistence model with a small time-shift. **(B)** Illustrates the case of a persistence model with a large time shift. When the time series is too small, the counts will be dominated by the green-colored block, while the actual truth will appear very small due to normalization.

The results confirm our expectations, where except for a few percentile, all the values are detected to be shifted by the forecasting horizon.

One potential problem that can arise is when the time shift in the two compared time series is very large. First, the algorithm will take longer to run as the window-size  $w$  needs to be much larger. Second, because the boundary conditions require the beginning and end of both sequences to match, if the time series is too short, the algorithm might give a shifted results. Take the extreme example shown in Figure 7, showing both a persistence model with respective time shift  $s_1$  and  $s_2$ . If the time shift is very large and the time series is small, the number of values counted to have shift  $s_2$  are made insignificant due to the path also including the boundary condition. This effect is already starting to show in

the final row of Table 3, where around 2% of the shifted values are artificial due to the boundary constraint. Potential changes to the algorithm that could account for this problem is a topic for future work.

### 3.3.2. DTW Applied to the LSTM-NN Model

The normalized values of the DTW measure applied to the LSTM-NN model presented in section 2.5 can be seen in Table 4. The highest percentages are located on the offset diagonal, identical to the results of the persistence model. As discussed before, this indicates that a shift in time exists between the observations and the model predictions. This confirms that our observation of the results discussed in section 2.6 are happening throughout the whole time series. We notice that the second

**TABLE 3 |** The row-normalized fractions of the warping measure on the persistence model.

Forecast horizon	Time shift						
	0 h	1 h	2 h	3 h	4 h	5 h	6 h
$t + 1h$	0.003	<b>0.997</b>	0.0	0.0	0.0	0.0	0.0
$t + 2h$	0.003	0.003	<b>0.994</b>	0.0	0.0	0.0	0.0
$t + 3h$	0.004	0.003	0.003	<b>0.991</b>	0.0	0.0	0.0
$t + 4h$	0.003	0.003	0.003	0.003	<b>0.988</b>	0.0	0.0
$t + 5h$	0.004	0.003	0.003	0.003	0.003	<b>0.984</b>	0.0
$t + 6h$	0.004	0.003	0.003	0.003	0.003	0.003	<b>0.981</b>

The algorithm detects that the persistence model is shifted with the expected number of hours.

The largest values of each row are annotated in bold.

**TABLE 4 |** The row-normalized fractions of the DTW measure on the LSTM-NN model.

Forecast horizon	Time shift						
	0 h	1 h	2 h	3 h	4 h	5 h	6 h
$t + 1h$	0.352	<b>0.578</b>	0.045	0.015	0.006	0.003	0.002
$t + 2h$	0.115	0.334	<b>0.428</b>	0.069	0.03	0.015	0.009
$t + 3h$	0.074	0.113	0.287	<b>0.355</b>	0.09	0.047	0.034
$t + 4h$	0.068	0.066	0.115	0.249	<b>0.309</b>	0.11	0.083
$t + 5h$	0.073	0.054	0.073	0.117	0.225	<b>0.28</b>	0.178
$t + 6h$	0.079	0.052	0.06	0.079	0.117	0.215	<b>0.397</b>

The results are in agreement with the visual inspection of the model: for each forecasting horizon time, the highest percentage is located at the corresponding time shift.

The largest values of each row are annotated in bold.

highest percentage of each row is located on the diagonal, indicating that the model is actually capable of providing some accurate prediction of the Dst for one hour into the future, similar to the observations of Wintoft and Wik (2018) and Stepanova and Pérez (2000).

## 4. DISCUSSION

### 4.1. Dst Index Analysis

What follows is a statistical analysis of the Dst index itself. The autocorrelation of the Dst is shown in **Figure 8B**. Notice the very high autocorrelation of the Dst index with itself for delay times up to  $t + 7h$ . This can also be seen in the lag plot, shown in **Figure 8A**. This could explain why the persistence model has such high accuracy and correlation when evaluated with the metrics of section 2.3.2, shown in **Table 1**. We believe that this also explains why the linear fit parameters of the persistence are so high.

The partial autocorrelation is also an important value. The partial autocorrelation  $\alpha(k)$ , defined by Equation (23), behaves as the autocorrelation between  $z_t$  and  $z_{t+k}$ , adjusted from the intermediate variables  $z_{t+1}, z_{t+2}, \dots, z_{t+k-1}$  (Box et al., 2015).

$$\alpha(k) = \begin{cases} \text{cor}(z_{t+1}, z_t) \text{ for } k = 1 \\ \text{cor}(z_{t+k} - P_{t,k}(z_{t+k}), z_t - P_{t,k}(z_t)) \text{ for } k \geq 2, \end{cases} \quad (23)$$

where  $k$  is the lag between the two time series values  $z_t$  and  $z_{t+k}$ , and  $P_{t,k}(z)$  is an operator of orthogonal projection of  $z$  onto the linear subspace of the Hilbert space spanned by  $z_{t+1}, z_{t+2}, \dots, z_{t+k}$ . The partial autocorrelation of the Dst can be seen in **Figure 8C**. This shows what can actually be learned from the Dst index: after the correction applied by the autocorrelation, only the Dst at time step  $t + 2h$  still has some significant correlation to the Dst at time  $t$ . This would explain why the neural network model has difficulty accurately predicting values beyond  $t + 1$ , and instead relies on behaving as a persistence model to predict the next values.

### 4.2. Removing the Autocorrelation

The autocorrelation properties of the Dst index are most likely the causes of the problem in the forecast. Direct workarounds consist of either changing the input or the output. A first solution is not to include the Dst index in the input vectors, as done by Wu and Lundstedt (1997). This gives a forecast based purely on the solar wind parameters.

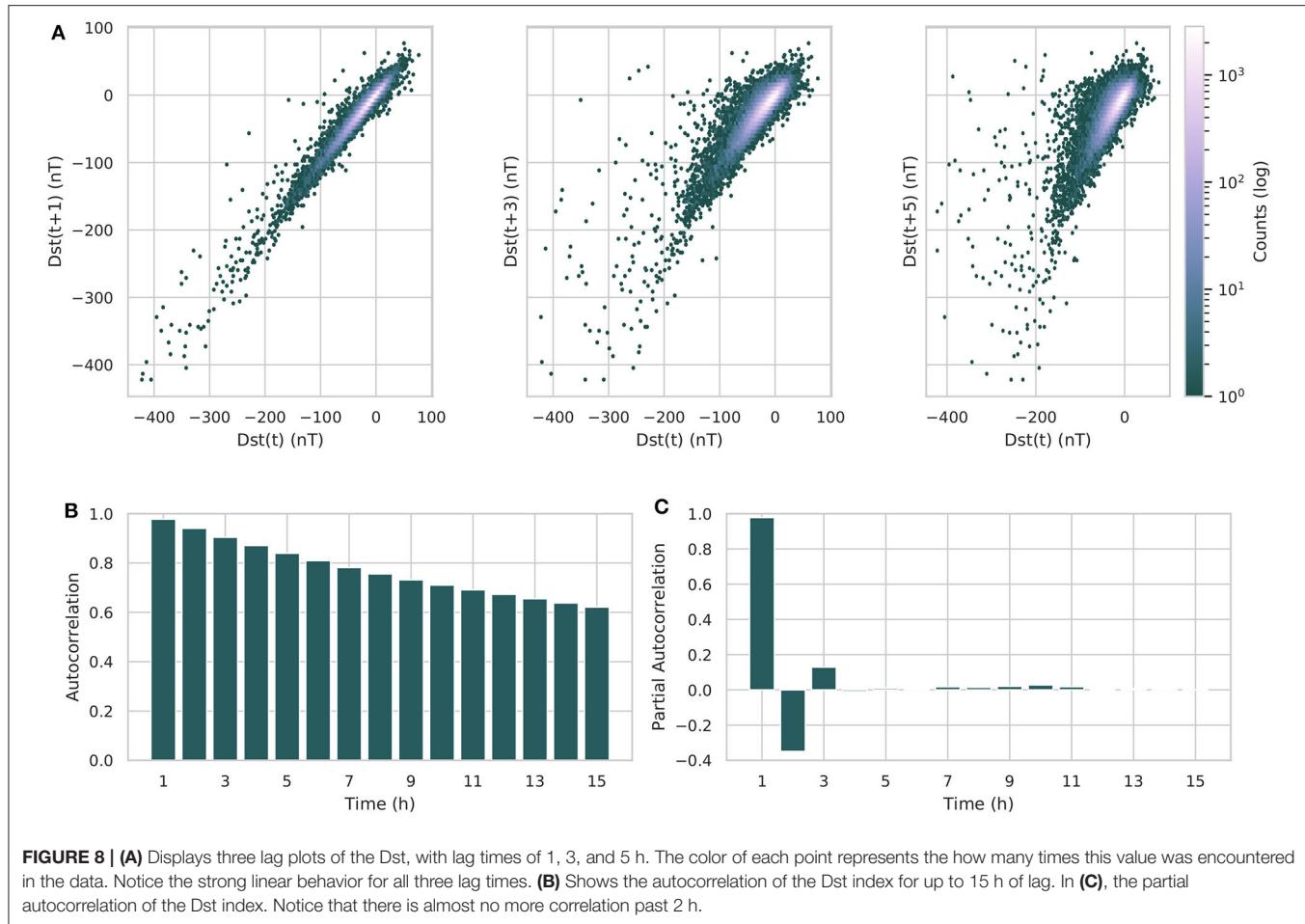
Another solution is to de-trend the Dst time series, and instead forecast the change in the Dst. Let us call  $\Delta\text{Dst}$  the difference of the Dst between two time steps:

$$\Delta\text{Dst}(t) = \text{Dst}(t) - \text{Dst}(t - 1). \quad (24)$$

This parameter has also been introduced by Wintoft and Wik (2018). However, they do not forecast the  $\Delta\text{Dst}$  directly with their model, but use it as a parameter for data selection. A lag plot of  $\Delta\text{Dst}$  shows that the correlation with the previous time step has almost completely vanished, as is visible in **Figure 9A**. Computing the autocorrelation confirms this, as seen in **Figure 9B**. Notice now that the autocorrelation behaves almost identical to the partial-autocorrelation, displayed in **Figure 9C**.

The  $\Delta\text{Dst}$  is a new parameter that we can use to train the LSTM-NN model with. The experiment described in section 2 is repeated, only this time the model will forecast the  $\Delta\text{Dst}$  model. As input we use, next to the parameters described in section 2.1, also the previous values of the  $\Delta\text{Dst}$ . **Tables 5, 6** show the results of the LSTM-NN using this data. The forecasting of the  $\Delta\text{Dst}$  seems to work well for forecasting horizons of 1–2 h. For later forecasting horizons, the correlation coefficient decreases sharply, and the prediction efficiency becomes close to zero. The RMSE does not increase substantially when the forecasting horizon increases. The results of the DTW measure are shown in **Table 6**. Notice the absence of a persistence effect, as most values are no longer on the offside diagonal. Only in the last row does there seem to be some delay, but this could be explained by taking the prediction efficiency of **Table 5** into account, which is close to 0 for forecasting horizon  $t + 6h$ . This means that the forecast most likely no longer resembles the observed time series anymore, and the evaluation of the DTW algorithm does not have much meaning anymore.

The forecast now accurately shows us that the predictive power of the LSTM is linked to the partial autocorrelation of the  $\Delta\text{Dst}$ , and demonstrates the difficulty to provide an accurate forecast beyond  $t + 2h$ . The advantage of using the  $\Delta\text{Dst}$  is that there is no longer a false sense of accuracy. The



persistence effect gave the illusion of a strong forecast, while the  $\Delta\text{Dst}$  does not. Using this as a basis, it will be much more transparent when a forecasting model provides us with an actual accurate forecast.

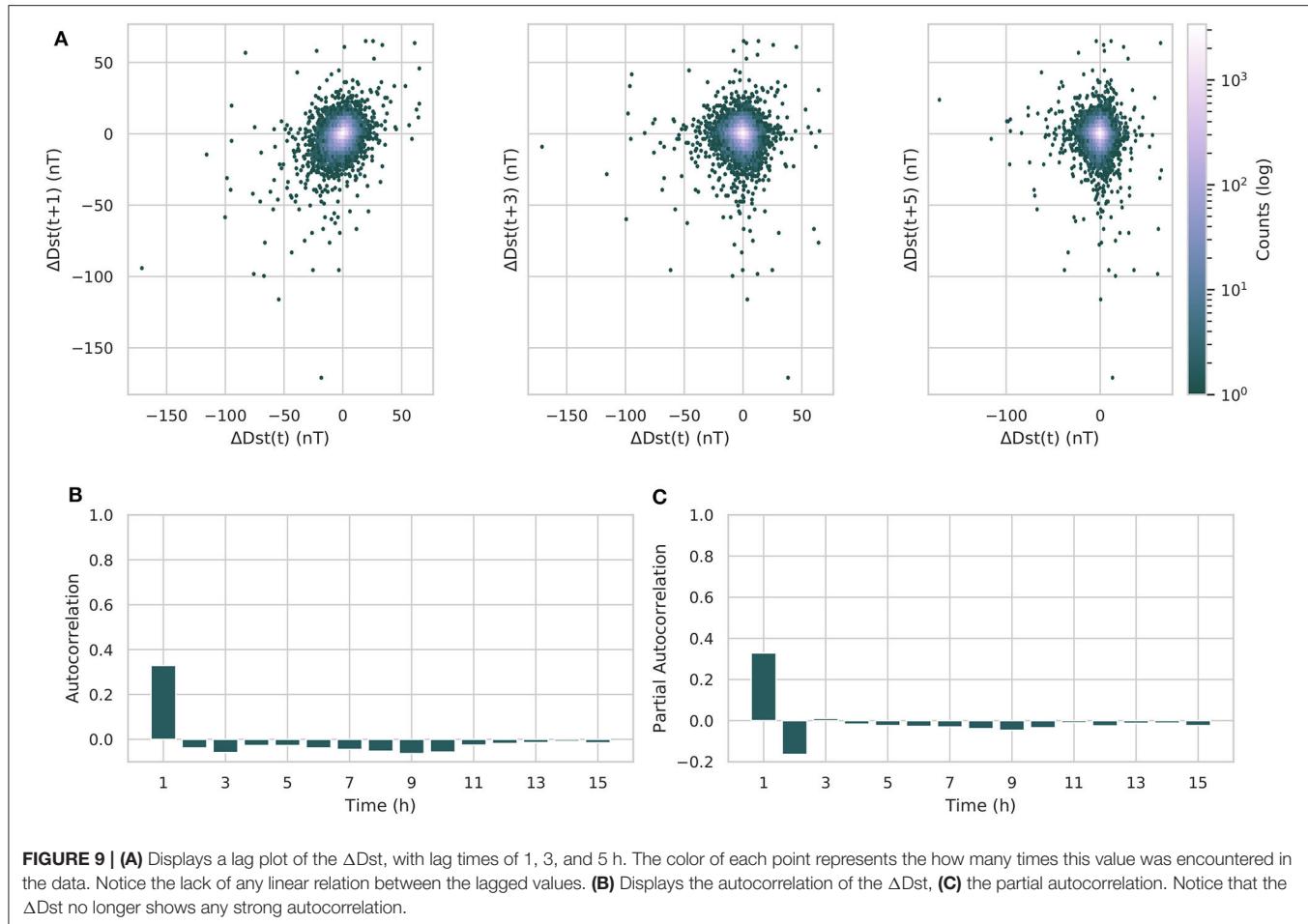
Finally we discuss the possible causes of why forecasting the Dst is so difficult. We believe that there are two problems that we have not yet taken into account. The first is the variation of the geo-effectiveness of the quiet solar wind, mainly caused by how the tilt of the Earth effects the interaction of magnetosphere with the solar wind. Together with the inclination of the equatorial plane of the Sun, this causes a yearly variation which was not taken into account in this experiment. This is called the Russell-McPherron effect (Russell and McPherron, 1973) and has been shown to effect the Dst index (Siscoe and Crooker, 1996). The second is that we believe that it is misguided to forecast the Dst index using the solar wind data measured at L1. These measurements are taken too close to the Earth, which causes an intrinsic limit on how far in the future we can give a forecast. We believe that having measurements at L5 would provide a large improvement in our abilities to provide timely forecasts, as discussed by Hapgood (2017). The effects of using measurements at L5 could be explored in future research, where simulations

such as EUFHORIA (Pomoell and Poedts, 2018) are used to provide artificial measurements.

## 5. CONCLUSIONS

An LSTM-based neural network, called the LSTM-NN model, is trained to forecast the Dst index 1–6 h in the future, using solar wind parameters and the Dst from 6 h before the prediction as an input. While the evaluation scores have indicated that the LSTM-NN model is comparable to the latest publications, visual inspection shows that the model's forecast behavior is similar to that of a persistence model, using the last known input of the Dst as its output. Although the prediction performs better than the persistence model, showing that some information can be learned from the solar wind, the LSTM-NN model effectively fails in its objective.

In order to detect this new type of error, a new test is developed based on the DTW algorithm, to measure the shift between observation and model prediction. DTW can compare two time series in a time window, instead of comparing two values on the same timestamp such as done by the RMSE and the



**TABLE 5 |** Evaluation of the LSTM-NN model and the persistence model with the metrics from section 2.3.2 when forecasting  $\Delta\text{Dst}$ .

Forecasting horizon	RMSE (nT)	R	A	B (nT)	MAE (nT)	ME (nT)	PE
<b>LSTM-NN model</b>							
$t + 1h$	3.215	0.630	-0.064	0.396	2.156	-0.064	0.397
$t + 2h$	3.807	0.393	-0.031	0.163	2.505	-0.029	0.154
$t + 3h$	3.943	0.305	-0.016	0.106	2.571	-0.016	0.091
$t + 4h$	3.995	0.260	0.059	0.082	2.601	0.057	0.064
$t + 5h$	4.057	0.197	-0.018	0.053	2.640	-0.021	0.033
$t + 6h$	4.075	0.168	0.009	0.038	2.646	0.006	0.025
<b>Persistence model</b>							
$t + 1h$	5.160	0.311	1.729	0.311	3.524	1.730	-0.553
$t + 2h$	6.170	-0.042	1.530	-0.042	4.189	1.532	-1.220
$t + 3h$	6.178	-0.080	1.068	-0.080	4.169	1.069	-1.226
$t + 4h$	6.078	-0.063	0.685	-0.063	4.065	0.686	-1.155
$t + 5h$	6.067	-0.035	1.144	-0.035	4.076	1.145	-1.147
$t + 6h$	6.088	-0.016	1.490	-0.016	4.150	1.491	-1.162

correlation coefficient, allowing the detection of temporal trends. By using the output of the DTW algorithm, first a least-distance mapping is given between the two time series, which can then

**TABLE 6 |** The DTW measure of the LSTM-NN model forecasting the  $\Delta\text{Dst}$ .

Forecast horizon	Time shift						
	0 h	1 h	2 h	3 h	4 h	5 h	6 h
$t + 1h$	0.426	<b>0.449</b>	0.053	0.03	0.019	0.014	0.009
$t + 2h$	<b>0.482</b>	0.355	0.061	0.028	0.026	0.024	0.024
$t + 3h$	<b>0.424</b>	0.319	0.152	0.021	0.026	0.03	0.028
$t + 4h$	0.287	<b>0.283</b>	0.25	0.107	0.022	0.028	0.023
$t + 5h$	0.186	0.192	0.228	<b>0.229</b>	0.118	0.024	0.022
$t + 6h$	0.147	0.139	0.165	0.188	<b>0.22</b>	0.127	0.013

Notice that the persistence-like behavior is absent.

be used to compare the timestamps of the points mapped to each other. This gives us a measure of the time warp between these two time series, from which we can infer a potential persistence effect.

When this new measure was applied to the persistence model, the results were as expected, and completely captured the temporal behavior of the persistence model. When the measure was applied to the time series forecasting of the LSTM-NN model, it detected the temporal lag in the forecast, proving its usefulness.

Finally, the possible origin of this lag was discussed by observing the autocorrelation of the time series, together with

possible different experiments that do not suffer this temporal lag. It was shown that the forecasting of the differentiated Dst did not have this temporal lag. The LSTM-NN model showed promising results for forecasting horizons of t+1h and t+2h, but later forecasts did not have a very high accuracy or correlation to the observations. Future studies focusing on forecasting the differentiated Dst could provide more transparent results. We believe that new research also has to explore the effect of the variability of the solar wind interacting with the magnetosphere in function of the Earth tilt and the inclination of the solar equatorial plane.

Finally, we believe that the observational data measured at L1 plays a big role in limiting the forecast horizon of the Dst index. Looking at the effects of having measurements at L5 should be further explored in future work, using simulations to provide the artificial measurements.

As a concluding remark, we would like to emphasize that researchers should be very prudent when reporting results of time series forecasting with the metrics defined in section 2.3.2. These metrics fail to capture behaviors that are only seen when taking into account the temporal dimension of the forecasting, and could provide misleading results.

## REFERENCES

- Akasofu, S.-I. (1981). Energy coupling between the solar wind and the magnetosphere. *Space Sci. Rev.* 28, 121–190. doi: 10.1007/BF00218810
- Bala, R., and Reiff, P. (2012). Improvements in short-term forecasting of geomagnetic activity. *Space Weath.* 10. doi: 10.1029/2012SW000779
- Barkhatov, N. A., Korolev, A. V., Ponomarev, S. M., and Sakharov, S. Y. (2001). Long-term forecasting of solar activity indices using neural networks. *Radiophys. Quant. Electron.* 44, 742–749. doi: 10.1023/A:1013019328034
- Berndt, D. J., and Clifford, J. (1994). “Using dynamic time warping to find patterns in time series,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, AAAIWS’94* (AAAI Press), 359–370.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: John Wiley & Sons.
- Boyle, C., Reiff, P., and Hairston, M. (1997). Empirical polar cap potentials. *J. Geophys. Res.* 102, 111–125. doi: 10.1029/96JA01742
- Burton, R. K., McPherron, R., and Russell, C. (1975). An empirical relationship between interplanetary conditions and DST. *J. Geophys. Res.* 80, 4204–4214. doi: 10.1029/JA080i031p04204
- Camporeale, E. (2019). The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weath.* 17, 1166–1207. doi: 10.1029/2018SW002061
- Council, N. R. (2008). *Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report*. Washington, DC: The National Academies Press.
- Eberhart, R., and Kennedy, J. (1995). “New optimizer using particle swarm theory,” in *Proceedings of the International Symposium on Micro Machine and Human Science* (Nagoya).
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402\_1
- Gonzalez, W., Joselyn, J.-A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B., et al. (1994). What is a geomagnetic storm? *J. Geophys. Res.* 99, 5771–5792. doi: 10.1029/93JA02867
- Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.
- Gruet, M. A., Chandorkar, M., Sicard, A., and Camporeale, E. (2018). Multiple-hour-ahead forecast of the Dst index using a combination of long short-term memory neural network and Gaussian process. *Space Weath.* 16, 1882–1896. doi: 10.1029/2018SW001898
- Hapgood, M. (2017). L1l5together: report of workshop on future missions to monitor space weather on the sun and in the solar wind using both the L1 and L5 Lagrange points as valuable viewpoints. *Space Weath.* 15, 654–657. doi: 10.1002/2017SW001652
- Hecht-Nielsen, R. (1992). “Theory of the backpropagation neural network,” in *Neural Networks for Perception*, ed. H. Wechsler (Cambridge, MA: Elsevier), 65–93.
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertain. Fuzz. Knowl. Based Syst.* 6, 107–116. doi: 10.1142/S0218488598000094
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *J. Geomagn. Geoelectr.* 42, 1249–1265. doi: 10.5636/jgg.42.1249
- Ji, E. Y., Moon, Y. J., Gopalswamy, N., and Lee, D. H. (2012). Comparison of Dst forecast models for intense geomagnetic storms. *J. Geophys. Res.* 117, 1–9. doi: 10.1029/2011JA016872
- Juszczak, P., Tax, D., and Duin, R. P. (2002). “Feature scaling in support vector data description,” in *Proceedings of ASCI* (Citeseer), 95–102.
- Kapperman, J. G., and Albertson, V. D. (1990). Bracing for the geomagnetic storms. *IEEE Spectr.* 27, 27–33. doi: 10.1109/6.48847
- Kasinskii, V., Ptitsyna, N., Lyahov, N., Tyasto, M., Villoresi, G., and Iucci, N. (2007). Effect of geomagnetic disturbances on the operation of railroad automated mechanisms and telemechanics. *Geomagn. Aeron.* 47, 676–680. doi: 10.1134/S0016793207050179
- Keogh, E., and Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowl. Inform. Syst.* 7, 358–386. doi: 10.1007/s10115-004-0154-9
- Keogh, E. J., and Pazzani, M. J. (1999). “Scaling up dynamic time warping to massive datasets,” in *European Conference on Principles of Data Mining and Knowledge Discovery* (Prague: Springer), 1–11.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://omniweb.gsfc.nasa.gov/>.

## AUTHOR CONTRIBUTIONS

BL performed and analyzed the experiment, developed and tested the new technique, and wrote the manuscript. JA planned the study and has provided substantial intellectual contribution and interpretation of the results. All authors took part in the manuscript revision and have read and approved the submitted version.

## FUNDING

This research has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 776262 (AIDA, <http://www.aida-space.eu/>). The OMNI data used for this research were obtained from the GSFC/SPDF OMNIWeb interface at <https://omniweb.gsfc.nasa.gov/>. The code will be made publicly available after publication on the project’s website at <http://www.aida-space.eu/>.

- King, J., and Papitashvili, N. (2005). Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *J. Geophys. Res.* 110. doi: 10.1029/2004JA010649
- Klimas, A. J., Vassiliadis, D., and Baker, D. N. (1998). DST index prediction using data-derived analogues of the magnetospheric dynamics. *J. Geophys. Res.* 103, 20435–20447. doi: 10.1029/98JA01559
- Lazzús, J. A., Vega, P., Rojas, P., and Salfate, I. (2017). Forecasting the Dst index using a swarm-optimized neural network. *Space Weath.* 15, 1068–1089. doi: 10.1002/2017SW001608
- LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. (2012). “Efficient backprop,” in *Neural Networks: Tricks of the Trade*, eds G. Montavon, G. B. Orr, and K.-R. Müller (Berlin: Springer), 9–48.
- Lemire, D. (2009). Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recogn.* 42, 2169–2180. doi: 10.1016/j.patcog.2008.11.030
- Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weath.* 16, 2079–2102. doi: 10.1029/2018SW002067
- Lundstedt, H., Gleisner, H., and Wintoft, P. (2002). Operational forecasts of the geomagnetic dst index. *Geophys. Res. Lett.* 29, 34–41. doi: 10.1029/2002GL016151
- Lundstedt, H., and Wintoft, P. (1994). Prediction of geomagnetic storms from solar wind data with the use of a neural network. *Ann. Geophys.* 12, 19–24. doi: 10.1007/s00585-994-0019-2
- Owens, M. J., Challen, R., Methven, J., Henley, E., and Jackson, D. (2013). A 27 day persistence model of near-earth solar wind conditions: a long lead-time forecast and a benchmark for dynamical models. *Space Weath.* 11, 225–236. doi: 10.1002/swe.20040
- Pallocchia, G., Amata, E., Consolini, G., Marcucci, M. F., and Bertello, I. (2006). Geomagnetic Dst index forecast based on IMF data only. *Ann. Geophys.* 24, 989–999. doi: 10.5194/angeo-24-989-2006
- Paszke, A., Gross, S., Chintala, S., and Chanan, G. (2017). Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration. *PyTorch* 6.
- Pomoell, J., and Poedts, S. (2018). Euhforia: European heliospheric forecasting information asset. *J. Space Weath. Space Clim.* 8:A35. doi: 10.1051/swsc/2018020
- Pulkkinen, A., Lindahl, S., Viljanen, A., and Pirjola, R. (2005). Geomagnetic storm of 29–31 october 2003: geomagnetically induced currents and their relation to problems in the swedish high-voltage power transmission system. *Space Weath.* 3. doi: 10.1029/2004SW000123
- Russell, C., and McPherron, R. (1973). Semiannual variation of geomagnetic activity. *J. Geophys. Res.* 78, 92–108. doi: 10.1029/JA078i001p00092
- Salvador, S., and Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* 11, 561–580. doi: 10.3233/IDA-2007-11508
- Siscoe, G., and Crooker, N. (1996). Diurnal oscillation of DST: a manifestation of the russell-mcpherron effect. *J. Geophys. Res.* 101, 24985–24989. doi: 10.1029/96JA01875
- Skutkova, H., Vitek, M., Babula, P., Kizek, R., and Provanik, I. (2013). Classification of genomic signals using dynamic time warping. *BMC Bioinformatics* 14:S1. doi: 10.1186/1471-2105-14-S10-S1
- Stepanova, M. V., and Pérez, P. (2000). “Autoprediction of Dst index using neural network techniques and relationship to the auroral geomagnetic indices,” in *Geofísica Internacional*.
- Sugiura, M. (1963). *Hourly Values for Magnetic Storm Variation for International Geophysical Year*. Technical Report/Patent No. NASA-TM-X-55238, X-611-63-131. Greenbelt, MD: NASA Goddard Space Flight Center.
- Sugiura, M., and Kamei, T. (1991). Equatorial dst index 1957–1986, iaga bull., 40. (*Int. Serv. Geomagn. Indices Publ. Off, Saint Maur, 1991*), eds A. Berthelier and M. Menville
- Temerin, M., and Li, X. (2002). A new model for the prediction of Dst on the basis of the solar wind. *J. Geophys. Res.* 107:SMP 31-1-SMP 31-8. doi: 10.1029/2001JA007532
- Temerin, M., and Li, X. (2006). Dst model for 1995–2002. *J. Geophys. Res.* 111:A4. doi: 10.1029/2005JA011257
- Wang, G.-J., Xie, C., Han, F., and Sun, B. (2012). Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Phys. A* 391, 4136–4146. doi: 10.1016/j.physa.2012.03.036
- Watanabe, S., Sagawa, E., Ohtaka, K., and Shimazu, H. (2002). Prediction of the DST index from solar wind parameters by a neural network method. *Earth Planets Space* 54, e1263–e1275. doi: 10.1186/BF03352454
- Wintoft, P., and Wik, M. (2018). Evaluation of kp and dst predictions using ace and dscovr solar wind data. *Space Weath.* 16, 1972–1983. doi: 10.1029/2018SW001994
- Wolpert, D. H., and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* 1, 67–82. doi: 10.1109/4235.585893
- Wu, J.-G., and Lundstedt, H. (1997). Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks. *J. Geophys. Res.* 102, 14255–14268. doi: 10.1029/97JA00975

**Conflict of Interest:** 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.

Copyright © 2020 Laperre, Amaya and Lapenta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

### ADDITIONAL INFORMATION

Training and evaluating the model takes on average 10 min on a machine with the following specifications:

- **OS:** Windows 10
- **Processor:** Intel(R) Core(TM) i7-8850H CPU @ 2.60 GHz, 2592 Mhz, 6 Core(s), 12 Logical Processor(s)
- **RAM:** 16 Gb

The source code and experiment can be found on the following webpage: [https://github.com/brechtlaperre/DTW\\_measure](https://github.com/brechtlaperre/DTW_measure).



# Improvement of Plasma Sheet Neural Network Accuracy With Inclusion of Physical Information

Brian M. Swiger<sup>1\*</sup>, Michael W. Liemohn<sup>1</sup> and Natalia Y. Ganushkina<sup>1,2</sup>

<sup>1</sup> Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, United States, <sup>2</sup> Space Research and Observation Technologies, Space and Earth Observation Centre, Finnish Meteorological Institute, Helsinki, Finland

## OPEN ACCESS

### Edited by:

Thomas Berger,  
University of Colorado Boulder,  
United States

### Reviewed by:

Alexei V. Dmitriev,  
Lomonosov Moscow State University,  
Russia

David Malaspina,  
University of Colorado Boulder,  
United States

### \*Correspondence:

Brian M. Swiger  
swigerbr@umich.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 03 April 2020

**Accepted:** 16 July 2020

**Published:** 30 July 2020

### Citation:

Swiger BM, Liemohn MW and Ganushkina NY (2020) Improvement of Plasma Sheet Neural Network Accuracy With Inclusion of Physical Information. *Front. Astron. Space Sci.* 7:42.  
doi: 10.3389/fspas.2020.00042

The near-Earth plasma sheet is the source for electrons in the inner magnetosphere. The coupling between the solar wind and the near-Earth plasma sheet is dominated by non-linear processes, making any relationship difficult to infer. We report on the development of a neural network to capture the non-linear behavior between solar wind variations and the response of energetic electron flux in the plasma sheet. To train the neural network algorithm, we developed a data set with inputs from solar wind monitoring spacecraft. The targets come from three probes of the Time History of Events and Macroscale Interactions during Substorms mission as the spacecraft traversed the plasma sheet from years 2008–2019. Preliminary findings during the development of the neural network model show that tuning input parameters based on previously known physical properties is conducive to improving model performance.

**Keywords:** neural network, plasma sheet, solar wind, machine learning, keV electron flux, deep learning, feature engineering, space weather

## 1. INTRODUCTION

The fluxes of <200 keV electrons in the Earth's inner magnetosphere constitute the seed population, which is critically important for radiation belt dynamics. It is through cyclotron resonance with the electrons of energies between a few and tens of keV (Kennel and Petschek, 1966; Kennel and Thorne, 1967; Li et al., 2008, 2012) that chorus waves are generated outside the plasmapause in association with the injection of Plasma Sheet (PS) electrons into the inner magnetosphere (Tsurutani and Smith, 1974; Meredith et al., 2002). Whistler mode chorus waves play an important role in accelerating the seed electron population to relativistic energies in the outer radiation belt (Horne et al., 2005; Chen et al., 2007). Moreover, low-energy electrons (electrons with energies less than about 100 keV) are responsible for hazardous space weather phenomena such as surface charging (Garrett, 1981; Davis et al., 2008). The electron flux of low energies varies significantly with geomagnetic activity and even during quiet time periods. The source of the low-energy electrons is the PS. Much of the behavior of the PS is driven by variations in the solar wind (SW) and interplanetary magnetic field (IMF) upstream of Earth's bow shock (e.g., Aubrey and McPherron, 1971; Nishida and Lyon, 1972; Tsutomu and Teruki, 1976; Terasawa et al., 1997; Wing et al., 2005; Nagata et al., 2008; Cao et al., 2013). It is therefore an important challenge to understand the distribution of energetic plasma entering the inner magnetosphere, as dependent upon SW driving.

Several studies have examined the link between SW variations and PS particles. For example, Borovsky et al. (1998) found that there are several PS properties that are highly correlated with upstream SW. Namely, the density, temperature, and total pressure of the PS are highly correlated

with density, velocity, and dynamic pressure of the SW, respectively. Tsyganenko and Mukai (2003) used Geotail and Advance Composition Explorer (ACE) (Stone et al., 1998) data to develop an empirical model for PS ions dependent upon SW driving. Luo et al. (2011) additionally used Geotail and ACE to investigate the PS electron population with an empirical model for electron fluxes with energy  $>38$  keV. Their model achieved good performance compared to observations; yet, it had limitations. Not being able to measure below 38 keV and using integrated flux made it impossible to accurately describe the behavior of electrons with lower energy. More recently, Dubyagin et al. (2019) estimated PS differential electron fluxes from Maxwellian and Kappa distribution functions derived from plasma moments obtained using an empirical model developed by Dubyagin et al. (2016). They found that for thermal and superthermal energies ( $\lesssim 1$  keV) the estimations are accurate within a factor of two. Yet, for higher energy ( $\geq 10$  keV), the estimates of electron flux diverge by more than an order of magnitude from observations. They suggest that to obtain a realistic representation of PS electrons at these energies, a flux based model should be developed.

Considering the limitations of previous empirical relationships to model plasma sheet properties from SW, an alternative method is to utilize machine learning (ML). ML is viable in the present era, yet there are challenges regarding the utility of using so-called “black- or gray-box” ML techniques. (e.g., Camporeale, 2019). We have a sufficiently large amount of observations (more than two 11-year solar cycles) of the SW and PS, and we have the necessary modern computational resources to process this amount of data. Bortnik et al. (2016) described a ML method to predict the value of some observable in the inner magnetosphere dependent upon a set of inputs and their time history. Using the methodology described by Bortnik et al. (2016), Chu et al. (2017) developed a neural network model of electron density in the inner magnetosphere with inputs of spacecraft location and time history of several geomagnetic indices. In a similar study, Zhelavskaya et al. (2017) used geomagnetic indices but included SW parameters as inputs to neural networks. They found their neural networks that included a combination of SW parameters and geomagnetic indices performed the best. Yue et al. (2015) used Support Vector Regression ML to develop an inner PS pressure model, during substorm growth phases only, based on inputs of SW dynamic pressure, sunspot number, Cross Polar Cap Potential, and the Auroral Electrojet Index. Their ML model was able to predict the observed pressure in the near-Earth PS during the substorm growth phase. In the present work, we used ML to develop an electron flux-based empirical model of the near-Earth PS during all times, using only the time history of upstream SW plasma and IMF parameters. The purpose of this brief report is to (1) establish that a machine learned model can, with some skill, predict the electron flux in the PS from SW input drivers only, and (2) demonstrate that using large amounts of data in a machine learning model is not as useful as using a limited dataset while applying established physical knowledge as inputs.

## 2. METHODS

We used a feed-forward neural network (NN) to investigate the response of 1–200 keV energy electrons in the near-Earth PS to variations in the SW upstream of Earth’s bow shock. There are two versions of a NN, which we label as Version 1 and Version 2, that we will describe. In both versions, there is an input layer, two hidden layers, and an output layer. The differences in both versions involve the number and types of inputs, the number of nodes in each hidden layer, and the amount of underlying physical information included as input.

### 2.1. Data Description

The data that were used in this study come from OMNI (King and Papitashvili, 2005) and Time History of Events and Macroscale Interactions during Substorms (THEMIS) (Angelopoulos, 2008). OMNI combines upstream SW measurements and calculated derivations for several plasma parameters. Measurements from multiple Lagrange L<sub>1</sub> spacecraft have been combined and propagated to the assumed Earth’s bow shock at approximately 15 Earth radii (R<sub>E</sub>). An advantage of OMNI data, rather than data directly from the source spacecraft, is its continuity over several decades and multiple spacecraft. Each THEMIS satellite carries an Electrostatic Analyzer (ESA) (McFadden et al., 2008) and Solid State Telescope (SST) (Angelopoulos et al., 2008), which combined measure electrons in the energy range from a few eV to a few MeV. All OMNI and THEMIS data were obtained via the NASA Goddard Space Physics Data Facility.

#### 2.1.1. Version 1 Target Data

The target data are PS electron flux of energies between approximately 1–200 keV. The THEMIS Science Team has combined measurements from the ESA and SST instruments into a single data product called GMOM (Ground combined MoMents). Altogether, there are 46 energy channels in the GMOM data set ranging from about 5 eV to  $\sim 300$  keV. We chose 17 energy channels of electron flux between 1 and 200 keV because this energy range is most correlated with the generation of chorus waves and with spacecraft surface charging. The approximate energy of each channel is shown in Table 1. The log<sub>10</sub> of the energetic flux values make up the target vector,  $\vec{y}$  (Equation 1a), which has 17 entries, one for each energy channel. Although there are several methods for filtering spacecraft observations to the PS (e.g., Roziers et al., 2009; Dubyagin et al., 2016), we adopt the method used by Ruan et al. (2005) that uses only a single criterion of plasma  $\beta \geq 1$ . The  $\beta \geq 1$  criterion follows from average properties of the central PS described by Baumjohann et al. (1989).

All of the data that we use for training the neural networks are from three probes, THEMIS-A, -D, and -E. The spacecraft have a nominal spin rate of 3 s, and thus have flux data with nominal time cadence of the same. However, electron flux enhancements resulting from magnetotail processes occur on minute time scales (e.g., Bame et al., 1967). We down-sampled the GMOM flux data to 1 min by taking the mean of intervals closed on the left and open on the right. From all observations marked by the THEMIS mission team with a good data quality flag from 1 February 2008

**TABLE 1 |** Comparison of architecture and test metrics for Version 1 and Version 2 neural networks.

Neural network models comparison		
Characteristic	Version 1	Version 2
<i>Physical Considerations</i>		
Plasma sheet location	$\beta \geq 1$	$\beta \geq 1$ , MLT
SW time resolution	30 min	5 min
SW time history	8 h	4 h
SW parameters	$V_x, V_y, V_z,  V $ $B_x, B_y, B_z,  B $ $P_{dyn}, n, T, E, \beta$	$B_y, B_z,  V , n$
Solar cycle period	Solar minimum	Declining phase
Number of inputs	208	174
Number of examples	~650K	~280K
<i>1st hidden layer</i>		
Number of nodes	624 (208 × 3)	522 (174 × 3)
Activation fn	ReLU	ReLU
<i>2nd hidden layer</i>		
Number of nodes	1,248 (624 × 2)	1,044 (522 × 2)
Activation fn	ReLU	ReLU
<i>Output layer</i>		
Number of outputs	17	17
Activation	linear	linear
Train target std. dev.	1.31	1.29
Test target std. dev.	1.41	1.39
Loss function	Mean squared error	
Optimization	Adam; batch size: 50	
Apprx. energies	1.3, 2.2, 3.9, 6.7, 11.7, 20.2	
of channels used (in keV)	27, 28, 29, 30, 31, 41 52, 65.5, 93, 139, 203	
<i>Test Metrics</i>		
Bias	-0.05	0.15
Extremes	0.46	0.36
PE (skill)	0.76	0.80
MSE (loss)	0.47	0.38
MAE	0.51	0.44
Association	0.43	0.69
sMAPE	90%	80%
SSPB	-30%	20%
MSA	151%	110%

Orange cells highlight characteristics that were expected to be advantageous toward modal performance. Blue cells highlight the metrics with the better test score of the two model versions. PE, prediction efficiency; MSE, mean squared error; MAE, mean absolute error; SSPB, symmetric signed percentage bias; sMAPE, symmetric mean absolute percentage error; MSA, median symmetric accuracy.

through 31 July 2019, we selected those that occurred when the spacecraft were between  $-9R_E \geq XY_{GSM} \geq -11R_E$ , had a measured plasma  $\beta \geq 1$ , and were on the night side [between magnetic local times (MLTs) 18–06]. The spatial region chosen does not relate to any static structure in the magnetosphere. Rather, we presume that varying characteristics in the PS at these locations will be captured by the model since they are dependent

upon SW driving. Combining observations that fit these criteria from all three spacecraft yielded around 830,000 one-minute observations. Note that not all of these were used in training due to missing data in the input data set, which is described next.

### 2.1.2. Version 1 Input Data

The inputs to the Version 1 NN are OMNI data from  $-0.5$  to  $-8$  h of each event identified in the target data set. Following evidence that the magnetosphere acts as a low-pass filter of the SW (Ilie et al., 2010), we used 30-min averaged OMNI data. Creating the input vector for each event is shown in Equations (1b) and (1c). We assumed a time delay of  $\tau = 30$  min to account for the time that it would take variations in the upstream SW to have an effect in the magnetotail. Thirteen OMNI parameters were used: SW proton number density, three velocity components and flow speed as well as IMF geocentric magnetic  $B_x$ ,  $B_y$ ,  $B_z$ , and  $|B|$ . Derived parameters included are SW proton temperature, electric field, dynamic pressure, and plasma beta. A full investigation quantifying the importance of these SW input drivers to PS electron flux is underway, however, such an investigation is beyond the scope of this brief report. As a preconditioning step, the values of each OMNI parameter were scaled to the range  $[-1, 1]$  by dividing all observations by the observed absolute maximum value between 2008 and 2019 for that particular parameter. With inclusion of these parameters and their time history, each input vector had 208 features. For each input vector  $\vec{x}_i$ , if there were any missing data, that input vector and its associated 1-min output vector  $\vec{y}_i$  were discarded from the database. Approximately 26% of training examples were removed due to missing data, reducing the total number from about 830,000 to 613,952. We intentionally did not randomize our training and testing sets due to the time series nature of the observations in the PS. Rather, we selected February 2008 to February 2018 as the training data and March 2018 to July 2019 as the test data.

$$\vec{y}_i = \log_{10} [\text{eflux}_1, \dots, \text{eflux}_{17}] \quad (1a)$$

$$\xi = [B_x, B_y, B_z, |B|, V_x, V_y, V_z, |V|, E, n, T, P_{dyn}, \beta] \quad (1b)$$

$$\vec{x}_i = [\bar{\xi}_{t_{y_i}-\tau}, \bar{\xi}_{t_{y_i}-\tau-\Delta t}, \dots, \bar{\xi}_{t_{y_i}-\Delta T}] \quad (1c)$$

In Equations (1b) and (1c),  $\bar{\xi}$  is the 30-min averaged  $\xi$ ,  $\tau = 30\text{min}$  is a time delay,  $t_{y_i}$  is observation time of  $\vec{y}_i$ ,  $\Delta t = 30\text{min}$ , and  $\Delta T = 8\text{h}$ .

### 2.1.3. Version 2 Training Data

Based on physical understanding of the behavior of PS electrons to SW driving, we made changes to the input dataset. The Version 1 NN uses 30-min averaged OMNI data. However, 1–200 keV energy electron flux in the near-Earth PS can vary on timescales of minutes. By averaging out the smaller scale variations using the 30-min averaged SW, we had neglected to include information that could potentially increase the accuracy of the training. Moreover, many previous studies (e.g., Newell et al., 2007), have identified SW and IMF parameters that tend to influence the response of the magnetosphere, typically in some

functional form. These studies have shown the most important contribution to magnetosphere response to be a combination of  $n$ ,  $V$ ,  $B_Y$ , and  $B_Z$  (e.g., Newell et al., 2007; Balikhin et al., 2010).

In the Version 2 NN, we restrict our input to include only these four parameters. The PS responds to solar wind variations through an increase in dayside reconnection and dynamic pressure, allowing plasma and energy to enter Earth's magnetosphere where it is stored in the magnetotail. The release of this stored energy both increases Earthward plasma flow and magnetic flux in the near-Earth PS, leading to increased energetic electron flux there. We chose these four parameters as indicators of how much reconnection and dynamic pressure will be impacting the dayside magnetosphere. We note that three of these parameters— $V$ ,  $B_Y$ , and  $B_Z$ —have long been shown to be an accurate predictor of energy input from the SW to the magnetosphere (e.g., Perreault and Akasofu, 1978).

### 2.1.3.1. Version 2 input data

For Version 2, we alter the input data to include only the four parameters from section 2.1.3. These parameters are averaged to 5 min and restricted to a time history of  $-0.5$  to  $-4$  h. To them, we add two inputs related to the spacecraft position. In a similar manner described by Bortnik et al. (2016), we encode spacecraft location by including the magnetic local time (MLT). For the purposes of this study only, we made a simplification that the characteristics of the PS within  $\pm 1 R_E$  in radial distance at  $10 R_E$  are approximately consistent. Similar to the Version 1 input vector, the OMNI parameters that are included start at  $t_0 - 30\text{min}$ . The Version 2 input data is shown in Equation 2. Each input vector has 174 features; there are 172 OMNI inputs ( $4 \text{ parameters} \cdot (240 - 25) \text{ min}/5 \text{ min}$ ) and 2 position inputs. The creation of the target vectors was unchanged.

$$\phi = \frac{\text{MLT}}{24} 2\pi \quad \xi = [B_Y, B_Z, |V|, n] \quad (2a)$$

$$\vec{x}_i = [\bar{\xi}_{t_{y_i}-\tau}, \bar{\xi}_{t_{y_i}-\tau-\Delta t}, \dots, \bar{\xi}_{t_{y_i}-\Delta T}, \cos \phi_i, \sin \phi_i] \quad (2b)$$

In Equation 2,  $\bar{\xi}$  is the 5-min averaged  $\xi$ ,  $\tau = 30\text{ min}$  is a time delay,  $t_{y_i}$  is observation time of  $\vec{y}_i$ ,  $\Delta t = 5\text{ min}$ ,  $\Delta T = 4\text{ h}$ , and MLT is the magnetic local time of the spacecraft when each of the  $i$  observations were recorded. Unlike the Version 1 OMNI inputs, data in Version 2 were not scaled to the range  $[-1, 1]$ . Similar to the Version 1 input data, if there were any missing OMNI data in an input vector, then we discarded that  $(\vec{x}_i, \vec{y}_i)$  example from the dataset. Since it is more likely to have missing data when averaging over 5 min than when averaging over 30 min, a much larger number, about 66%, of training examples were excluded in the Version 2 data. After removing examples with missing data, there were 282,294 total training examples remaining. As with Version 1, 10% of the examples were reserved for testing. The date ranges for Version 2 data are training: February 2008–August 2015 and testing: September 2015–May 2017. We note that these are not the same training/testing periods that were used for the Version 1 model. We discuss this discrepancy and its consequences in section 4.

## 2.2. Neural Network Description

A NN has the proven ability to fit any non-linear function between two sets of variables (Hornik et al., 1989). While we can be confident that some ambient SW plasma eventually finds its way to the PS, all of the non-linear methods involved for how it arrives there are not completely understood (e.g., Wing et al., 2014). To capture the unknown non-linear processes, we used the machine learning tool of a NN as a statistical mapping between upstream SW and PS observations.

### 2.2.1. Version 1 Neural Network

The Version 1 NN used an input layer, two hidden layers, and an output layer. The number of nodes in each hidden layer is based on a multiple of the number of inputs to that layer. The first hidden layer has 624 nodes, which is the number of inputs times three, and the second hidden layer has 1,248 nodes, which is 624 times two. All neurons in both layers are activated using the rectified linear unit function (ReLU, defined in Equation 3). In the output layer, a linear activation function is used to render the log of the flux values.

$$\text{ReLU : } f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (3)$$

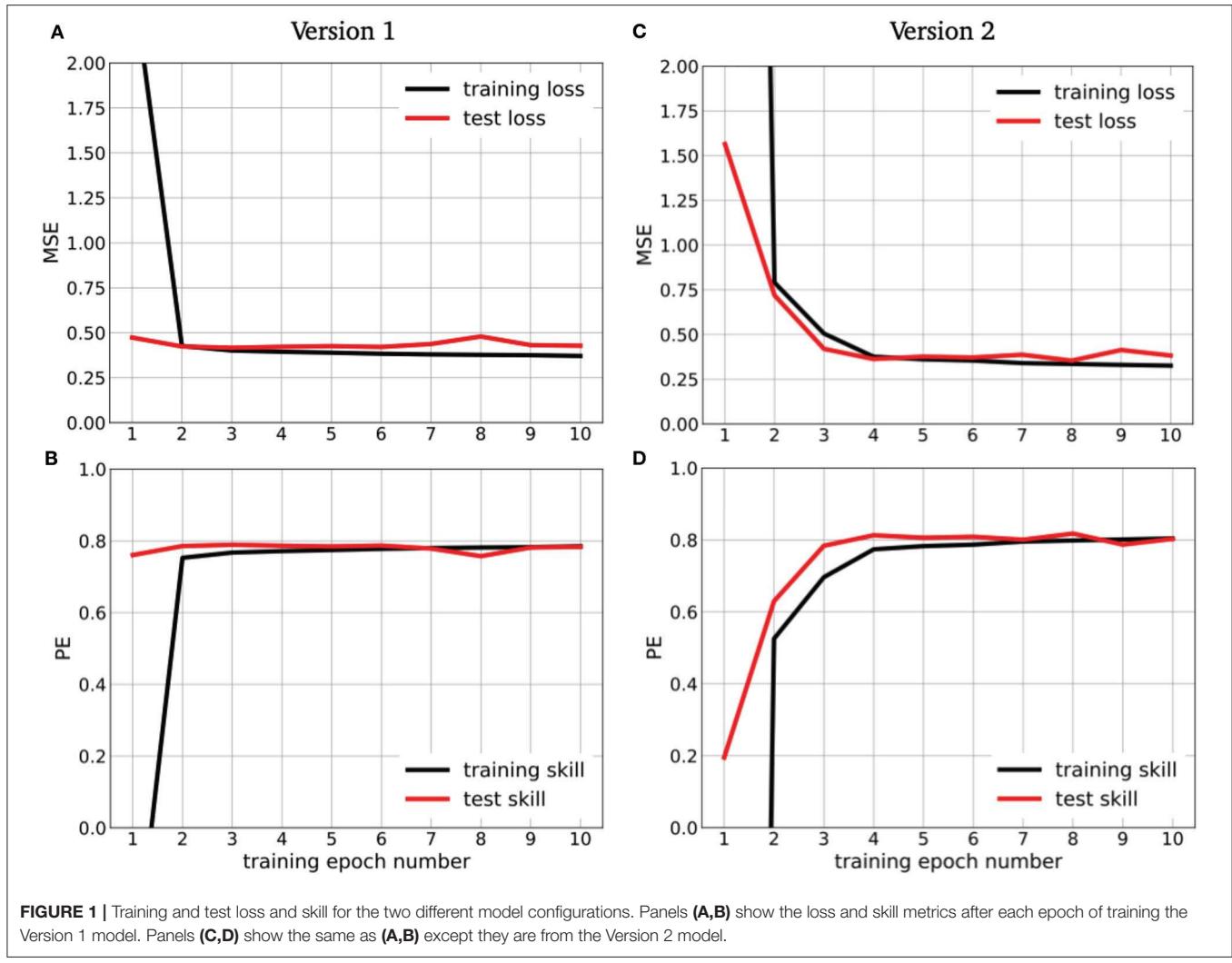
### 2.2.2. Version 2 Neural Network

We made modifications to the NN by modifying both the inputs, targets, and NN architecture. See section 2.1.3.1 for descriptions of how the inputs were modified from the Version 1 model. The NN architecture modifications from Version 1 to Version 2 are as follows. The number of inputs to the Version 2 model is 174. The first hidden layer has 522 nodes, which is three times the number of inputs. The second hidden layer has 1,044 nodes which is twice the number of nodes in the first hidden layer. As in the Version 1 model, all nodes in both hidden layers are activated using ReLU, and the output layer is activated using a linear function.

## 2.3. Neural Network Training

We utilized Keras with Tensorflow (Abadi et al., 2015) software for our NN training. Weights and biases were updated using a loss function of mean squared error (MSE) and the Adam optimization algorithm (Kingma and Ba, 2014) with hyperparameters set to  $\alpha$  (learning rate) = 0.001,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-7}$ . Although MSE was calculated using all 17 energy channels in the  $\vec{y}$ - and  $\hat{\vec{y}}$ -vectors, the weights and biases for each channel were updated independently. For both versions, we stopped training when it was detected that the test loss had stopped decreasing after three consecutive epochs. This occurred after ten epochs for both versions. We trained both NNs in batches of 50 training examples, resulting in several thousand updates per epoch.

**Figure 1** shows the training loss and test skill for both Version 1 and Version 2 neural networks. In **Figures 1A–D**, the black line was calculated using training data and the red line was calculated using the test data. **Figure 1A** is the MSE of predicted output vector vs. observed electron flux channels calculated after each epoch of training for model Version 1. We define a single



epoch of training to be a complete pass through all training data. The curves do not look “smooth” because the weights were updated after each batch of 50 training examples and the loss was only recorded after each complete epoch (11,052 batches per epoch in Version 1 and 5,082 batches per epoch in Version 2). **Figure 1B** shows the skill of the model calculated after each epoch. Model skill was determined using the prediction efficiency (PE) metric, which was calculated as unity minus the ratio of the MSE to the observed variance. **Figures 1C,D** show the loss and skill of the Version 2 model for both train and test data after each epoch of training. In Version 1, the final training loss was 0.34, and the final training skill was 0.76. In Version 2, the training loss and skill are similar to Version 1, at 0.33 and 0.80, respectively. The test loss for both versions have final values of 0.47 for Version 1, and 0.38 for Version 2. The final test skill between the two versions are Version 1 PE is 0.76 and Version 2 PE is 0.80.

### 3. RESULTS

We have calculated several model-observation metrics for the two neural networks in order to evaluate their performance.

Each metric was calculated using the data designated as test for both versions. Observations include the full set of  $\vec{y}$ , and model output,  $\hat{\vec{y}}$ , is obtained by applying the trained weights and biases to the test inputs,  $\vec{x}$ . The bottom section of **Table 1**, labeled *Test section*, shows all of the test metrics calculated for both versions of the NN. We use several different metrics for a more comprehensive model comparison (e.g., Liemohn et al., 2018). The first five (“Bias” through “MAE”) are calculated using the log of flux values and the last four (“Association” through “MSA”) are calculated using actual flux values. We highlighted in blue the metric between the two versions that more closely represents the observations. For all metrics calculated except Bias and Extremes, the Version 2 NN outperforms the Version 1 NN.

Bias is calculated as  $\text{mean}(\hat{\vec{y}}) - \text{mean}(\vec{y})$ . Version 1 has a slight negative bias of  $-0.05$  and Version 2 has a larger, positive bias of  $0.15$ . The Extremes are the ratio of the range of model flux to the range of observed flux. An Extreme score of 1 would indicate that the model output perfectly captures the observed range of flux values. Since the Version 1 score of  $0.46$  is closer to unity than the Version 2 score of  $0.36$ , we

can infer that the Version 1 model is better at capturing the range of observed flux values than Version 2. PE and MSE are defined in section 2.3, and for both, Version 2 outperforms Version 1. In training, the algorithm was attempting to minimize MSE on the training data, and PE was monitored to impede overfitting (see section 2.3). The MSE of Version 1 is 0.47 and the MSE of Version 2 is 0.38. The Version 2 PE of 0.80 is an improvement over the Version 1 PE of 0.76. We use Mean Absolute Error (MAE) as a second measure of the spread of the deviation between observed and modeled values. The Version 1 MAE is 0.51 and the Version 2 MAE is 0.44. If we take the square root of MSE to obtain the Root Mean Square Error (RMSE), then RMSE and MAE have the same units, in this case  $\log_{10}(\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1})$ . RMSE is larger than MAE for both models, which reveals that it is likely that there is a substantial spread of modeled flux values compared to the observed values.

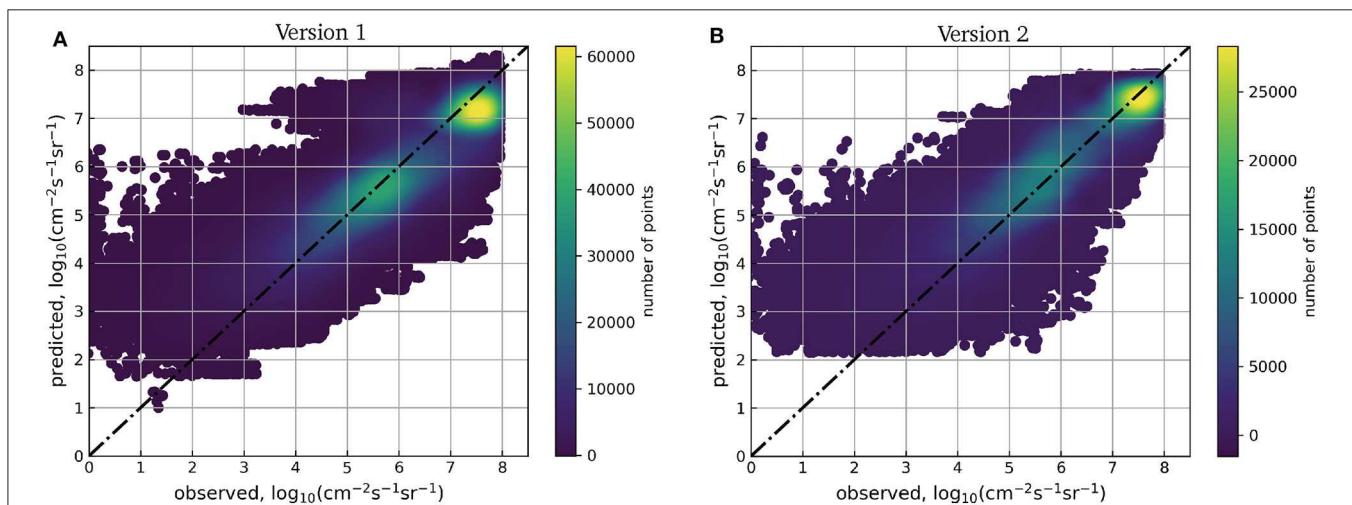
The remaining metrics that are described were calculated using actual flux values. With Association, we use the standard textbook  $r^2$  value commonly used for regression analysis. Our interpretation of Association is that the Version 1 model captures only 43% of the variance in the observed fluxes, while the Version 2 model is capturing nearly 70% of the variance in observed fluxes. The symmetric mean absolute percent error (sMAPE) ranges from 0 to 200 percent. The Version 1 model has a sMAPE of 90% and the Version 2 model has a sMAPE of 80%. The signed symmetric percent bias (SSPB) and the median symmetric accuracy (MSA) are two metrics described by Morley et al. (2018) that provide a more robust comparison of flux values that vary by orders of magnitude. SSPB is calculated using the median value of the log of flux, rather than the mean of the log of flux. Consistent with the Bias calculated using the mean of the log of flux values, The SSPB for Version 1 is negative, at  $-30\%$  and the SSPB for Version 2 is positive at  $20\%$ . The fact that the absolute SSPB is lower for Version 2 than Version 1 implies that there is larger spread in the Version 1 modeled values than in the

Version 2 modeled values compared to the observed values. This is consistent with the comparison of MAE and RMSE described in the previous paragraph. Despite the name, the MSA is a measure of error, and the values of MSA for Version 1 and Version 2 models indicates that there is less error in the Version 2 model. A percent error of 100% would imply that on average, there is a factor of two in the discrepancy between the observed flux and the modeled flux. The Version 2 NN achieves a MSA of 110%, which is better than the Version 1 MSA of 151%.

The comparison of observed to modeled electron flux using observations from the test data and modeled output from both versions of the NN is shown in Figure 2. Figure 2A shows the scatter for the Version 1 NN and Figure 2B shows the scatter for the Version 2 NN. The Version 1 scatter diagram shows a higher number of points overall than the Version 2 scatter diagram, because there was a larger amount of data in the Version 1 dataset (see section 2.1.2). The black diagonal dash-dotted line is a hypothetical ideal perfect correlation between observation and model. For both model versions, the scatter shows a clustering of the densest points close to the black line. The Version 2 model shows a larger portion of the points closer to the black line. Figure 2 is a general picture of model output and observational comparisons, and we are hesitant to draw conclusions regarding the behavior of plasma sheet electrons from it.

## 4. DISCUSSION

The Version 1 model assumes no a priori knowledge about which quantities in the SW are important contributions to near-Earth PS variations. We made this choice in order to allow the NN to appropriately weight any parameters or combinations thereof that might have been overlooked by previous SW-magnetosphere coupling studies. There is a longer time history of the SW that is used in Version 1 than Version 2. The Version 1 NN is also trained on more than twice as many training examples than Version 2. These factors might suggest that the Version 1 NN



would produce more accurate output than the Version 2 NN. However, despite the assumed advantages of the Version 1 model, the Version 2 model outperforms the Version 1 model in most model-data comparison metrics calculated (see **Table 1**). We propound that Version 2 outperforms Version 1 because of two modifications that were made based on physical information.

The first modification toward incorporating physical information relates to the time resolution of the inputs. When deciding to use 30-min averaged SW inputs, we used evidence that the magnetosphere acts as a low-pass filter of SW variations (e.g., Ilie et al., 2010). However, there is evidence that some higher time resolved information contained in the SW effects the PS (Lyons et al., 2009; Wang et al., 2017). Additionally, much of the behavior in the PS occurs on cycles of a few hours, i.e., substorm activity (e.g., Hones, 1972). By reducing the SW time history to 4 h, we neglect information that is likely relevant to predicting the electron flux in the PS. Others have found time delays of 6 h (Nagata et al., 2008) and 8 h (Borovsky et al., 1998) between SW/IMF variations and PS response. Even though the Version 2 model does not consider delays longer than 4 h, it still outperforms Version 1.

The second item of physical information that we introduced to the input set is the spatial distribution of observed electron flux. It is widely observed that there is a dawn-dusk asymmetry of electron fluxes in the PS (e.g., Walker and Farley, 1972; Lui and Rostoker, 1991; Sarafopoulos et al., 2001; Imada et al., 2008). We would expect both a higher flux and a larger number flux enhancements in the post local midnight, dawn section of the PS. Moreover, Wang et al. (2007) demonstrated that the spatial distribution of electrons within the PS is correlated with varying SW parameters. Wang et al. (2011) additionally show that the distribution of electron flux can be characterized by MLT as electrons drift closer toward the inner magnetosphere. Therefore, treating the PS as uniform in electron flux at a single radial distance, as modeled by the Version 1 NN, is physically inappropriate. By including the spacecraft location as an input for the Version 2 model, we are encoding the physical knowledge that the variation of electron flux is dependent upon spatial location within the PS.

Both model versions were trained using periods of the solar cycle that include quiet and active periods: Version 1, solar minimum through the declining phase and Version 2, solar minimum through solar maximum of solar cycle 24. However, Version 1 was tested on a period of solar quiet (solar minimum) and Version 2 was tested with data during a solar active period (declining phase). While we might expect the model to perform better during quiet SW conditions, this is not what we see when comparing Version 1 to Version 2. Moreover, there is not a substantial difference in the variance of training and testing target data between Versions 1 and 2. The Version 1 target data has standard deviations of 1.31 and 1.41  $\log_{10} (cm^{-2}s^{-1}sr^{-1})$  for train and test sets, respectively, while the Version 2 target data has standard deviations of 1.29 and 1.39  $\log_{10} (cm^{-2}s^{-1}sr^{-1})$ , respectively. This further indicates that including physical information is more important than using a larger amount of data when training these neural networks.

## 5. CONCLUSION

In summary, this study showed that including additional physical understanding, even while reducing the data set and inputs in other ways, improved the quality of the NN predictive capability. With neural networks, tracing the contribution from inputs to outputs is difficult, hindering interpretability of results, i.e., determining which inputs contributed to which output, or finding a functional mapping between inputs and outputs. Azari et al. (2020) showed that incorporating physical knowledge into ML additionally improves scientific interpretability along with performance for certain models. Development of a robust NN model of PS electron flux from SW input using additional physical understanding shows promise for improving performance.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://spdf.gsfc.nasa.gov/pub/data/> (under identifiers OMNI and THEMIS). Model output data and specific observed data are available at the following link: <https://doi.org/10.7302/559r-t639>.

## AUTHOR CONTRIBUTIONS

We use the CRediT (Contributor Roles Taxonomy) categories (Brand et al., 2015) for providing the following contribution description. NG led the conceptualization and provided resources and supervision. ML assisted in conceptualization and formal analysis, provided the resources, funding acquisition, supervision, and aided in project administration. BS designed the methodology, conducted the investigation, performed data visualization and formal analysis, and wrote the original draft. All authors have contributed toward the revision and editing of the manuscript.

## FUNDING

This work was funded in part by NASA Grant #NNX17AB87G. Work of NG was supported by NASA grants #NNX17AI48G (ROSES 2016), #80NSSC20K0353 (ROSES 2018), and Heliophysics Phase I DRIVE Science Center SOLSTICE (Solar Storms and Terrestrial Impacts Center) #80NSSC20K0600. BS was partially funded by Michigan Space Grant Consortium, NASA grant #NNX15AJ20H.

## ACKNOWLEDGMENTS

The authors thank the THEMIS mission team and OMNIWeb team for data availability. Matplotlib ([matplotlib.org](https://matplotlib.org)) plotting software was used to generate figures shown in this report. Neural network training was performed using Keras (<https://keras.io>) with Tensorflow (<https://tensorflow.org>). BS would like to thank useful discussions with A. R. Azari during manuscript preparation.

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Technical report. Available online at: <https://tensorflow.org>
- Angelopoulos, V. (2008). The THEMIS mission. *Space Sci. Rev.* 141, 5–34. doi: 10.1007/s11214-008-9336-1
- Angelopoulos, V., Sibeck, D., Carlson, C. W., McFadden, J. P., Larson, D., Lin, R. P., et al. (2008). First results from the THEMIS mission. *Space Sci. Rev.* 141, 453–476. doi: 10.1007/s11214-008-9378-4
- Aubry, M. P., and McPherron, R. L. (1971). Magnetotail changes in relation to the solar wind magnetic field and magnetospheric substorms. *J. Geophys. Res.* 76, 4381–4401. doi: 10.1029/JA076i019p04381
- Azari, A. R., Lockhart, J. W., Liemohn, M. W., and Jia, X. (2020). Incorporating physical knowledge into machine learning for planetary space physics. *Front. Astron. Space Sci.* 7:36. doi: 10.3389/fspas.2020.00036
- Balikhin, M. A., Boynton, R. J., Billings, S. A., Gedalin, M., Ganushkina, N., Coca, D., et al. (2010). Data based quest for solar wind-magnetosphere coupling function. *Geophys. Res. Lett.* 37:L24107. doi: 10.1029/2010GL045733
- Bame, S. J., Asbridge, J. R., Felthauer, H. E., Hones, E. W., and Strong, I. B. (1967). Characteristics of the plasma sheet in the Earth's magnetotail. *J. Geophys. Res.* 72, 113–129. doi: 10.1029/JZ072i001p00013
- Baumjohann, W., Paschmann, G., and Cattell, C. A. (1989). Average plasma properties in the central plasma sheet. *J. Geophys. Res. Space Phys.* 94, 6597–6606. doi: 10.1029/JA094iA06p06597
- Borovsky, J. E., Thomsen, M. F., and Elphic, R. C. (1998). The driving of the plasma sheet by the solar wind. *J. Geophys. Res. Space Phys.* 103, 17617–17639. doi: 10.1029/97JA02986
- Bortnik, J., Li, W., Thorne, R. M., and Angelopoulos, V. (2016). A unified approach to inner magnetospheric state prediction. *J. Geophys. Res. Space Phys.* 121, 2423–2430. doi: 10.1002/2015JA021733
- Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learn. Publ.* 28, 151–155. doi: 10.1087/20150211
- Camporeale, E. (2019). The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather* 17, 1166–1207. doi: 10.1029/2018SW002061
- Cao, J., Duan, A., Reme, H., and Dandouras, I. (2013). Relations of the energetic proton fluxes in the central plasma sheet with solar wind and geomagnetic activities. *J. Geophys. Res. Space Phys.* 118, 7226–7236. doi: 10.1002/2013JA019289
- Chen, Y., Reeves, G., and Friedel, R. (2007). The energization of relativistic electrons in the outer van allen radiation belt. *Nat. Phys.* 3, 614–617. doi: 10.1038/nphys655
- Chu, X., Bortnik, J., Li, W., Ma, Q., Denton, R., Yue, C., et al. (2017). A neural network model of three-dimensional dynamic electron density in the inner magnetosphere. *J. Geophys. Res. Space Phys.* 122, 9183–9197. doi: 10.1002/2017JA024464
- Davis, V. A., Mandell, M. J., and Thomsen, M. F. (2008). Representation of the measured geosynchronous plasma environment in spacecraft charging calculations. *J. Geophys. Res. Space Phys.* 113:A10204 doi: 10.1029/2008JA013116
- Dubyagin, S., Ganushkina, N., and Liemohn, M. (2019). On the accuracy of reconstructing plasma sheet electron fluxes from temperature and density models. *Space Weather* 17, 1704–1719. doi: 10.1029/2019SW002285
- Dubyagin, S., Ganushkina, N. Y., Sillanpää, I., and Runov, A. (2016). Solar wind-driven variations of electron plasma sheet densities and temperatures beyond geostationary orbit during storm times. *J. Geophys. Res. Space Phys.* 121, 8343–8360. doi: 10.1002/2016JA022947
- Garrett, H. B. (1981). The charging of spacecraft surfaces. *Rev. Geophys.* 19, 577–616. doi: 10.1029/RG019i004p00577
- Hones, E. W. (1972). Plasma sheet variations during substorms. *Planet. Space Sci.* 20, 1409–1431. doi: 10.1016/0032-0633(72)90048-7
- Horne, R. B., Thorne, R. M., Glauert, S. A., Albert, J. M., Meredith, N. P., and Anderson, R. R. (2005). Timescale for radiation belt electron acceleration by whistler mode chorus waves. *J. Geophys. Res. Space Phys.* 110:A03225. doi: 10.1029/2004JA010811
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* 2, 359–366. doi: 10.1016/0893-6080(89)90020-8
- Ilie, R., Liemohn, M. W., and Ridley, A. (2010). The effect of smoothed solar wind inputs on global modeling results. *J. Geophys. Res. Space Phys.* 115:A01213. doi: 10.1029/2009JA014443
- Imada, S., Hoshino, M., and Mukai, T. (2008). The dawn-dusk asymmetry of energetic electron in the earth's magnetotail: observation and transport models. *J. Geophys. Res. Space Phys.* 113:A11201. doi: 10.1029/2008JA013610
- Kennel, C. F., and Petschek, H. E. (1966). Limit on stably trapped particle fluxes. *J. Geophys. Res.* 71, 1–28. doi: 10.1029/JZ071i001p00001
- Kennel, C. F., and Thorne, R. M. (1967). Unstable growth of unducted whistlers propagating at an angle to the geomagnetic field. *J. Geophys. Res.* 72, 871–878. doi: 10.1029/JZ072i003p00871
- King, J. H., and Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *J. Geophys. Res.* 110:A02104. doi: 10.1029/2004JA010649
- Kingma, D. P., and Ba, J. (2014). “Adam: a method for stochastic optimization,” in *3rd International Conference on Learning Representations*. Technical report.
- Li, W., Thorne, R., Bortnik, J., McPherron, R., Nishimura, Y., Angelopoulos, V., et al. (2012). Evolution of chorus waves and their source electrons during storms driven by corotating interaction regions. *J. Geophys. Res. Space Phys.* 117:A08209. doi: 10.1029/2012JA017797
- Li, W., Thorne, R. M., Meredith, N. P., Horne, R. B., Bortnik, J., Shprits, Y. Y., et al. (2008). Evaluation of whistler mode chorus amplification during an injection event observed on CRRES. *J. Geophys. Res. Space Phys.* 113:A09210. doi: 10.1029/2008JA013129
- Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather* 16, 2079–2102. doi: 10.1029/2018SW002067
- Lui, W. W., and Rostoker, G. (1991). Effects of dawn-dusk pressure asymmetry on convection in the central plasma sheet. *J. Geophys. Res. Space Phys.* 96, 11501–11512. doi: 10.1029/91JA01173
- Luo, B., Tu, W., Li, X., Gong, J., Liu, S., Burin des Roziers, E., et al. (2011). On energetic electrons (>38 kev) in the central plasma sheet: data analysis and modeling. *J. Geophys. Res. Space Phys.* 116:A09220. doi: 10.1029/2011JA016562
- Lyons, L. R., Kim, H.-J., Xing, X., Zou, S., Lee, D.-Y., Heinselman, C., et al. (2009). Evidence that solar wind fluctuations substantially affect global convection and substorm occurrence. *J. Geophys. Res. Space Phys.* 114:A11306. doi: 10.1029/2009JA014281
- McFadden, J. P., Carlson, C. W., Larson, D., Ludlam, M., Abiad, R., Elliott, B., et al. (2008). The THEMIS ESA plasma instrument and in-flight calibration. *Space Sci. Rev.* 141, 277–302. doi: 10.1007/978-0-387-89820-9\_13
- Meredith, N. P., Horne, R. B., Iles, R. H. A., Thorne, R. M., Heynderickx, D., and Anderson, R. R. (2002). Outer zone relativistic electron acceleration associated with substorm-enhanced whistler mode chorus. *J. Geophys. Res. Space Phys.* 107, SMP 29-1-SMP 29-14. doi: 10.1029/2001JA000146
- Morley, S. K., Brito, T. V., and Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather* 16, 69–88. doi: 10.1002/2017SW001669
- Nagata, D., Machida, S., Ohtani, S., Saito, Y., and Mukai, T. (2008). Solar wind control of plasma number density in the near-earth plasma sheet: three-dimensional structure. *Ann. Geophys.* 26, 4031–4049. doi: 10.5194/angeo-26-4031-2008
- Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., and Rich, F. J. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *J. Geophys. Res.* 112:A01206. doi: 10.1029/2006JA012015
- Nishida, A., and Lyon, E. F. (1972). Plasma sheet at lunar distance: structure and solar-wind dependence. *J. Geophys. Res.* 77, 4086–4099. doi: 10.1029/JA077i022p04086
- Perreault, P., and Akasofu, S. I. (1978). A study of geomagnetic storms. *Geophys. J. Int.* 54, 547–573. doi: 10.1111/j.1365-246X.1978.tb05494.x
- Roziers, E. B. D., Li, X., Baker, D. N., Fritz, T. A., Friedel, R., Onsager, T. G., et al. (2009). Energetic plasma sheet electrons and their relationship with the solar wind: a cluster and geotail study. *J. Geophys. Res. Space Phys.* 114:A02220. doi: 10.1029/2008JA013696

- Ruan, P., Fu, S. Y., Zong, Q.-G., Pu, Z. Y., Cao, X., Liu, W. L., et al. (2005). Ion composition variations in the plasma sheet observed by cluster/rapid. *Geophys. Res. Lett.* 32, doi: 10.1029/2004GL021266
- Sarafopoulos, D. V., Sidiropoulos, N. F., Sarris, E. T., Lutsenko, V., and Kudela, K. (2001). The dawn-dusk plasma sheet asymmetry of energetic particles: an interball perspective. *J. Geophys. Res. Space Phys.* 106, 13053–13065. doi: 10.1029/2000JA900157
- Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., et al. (1998). The advanced composition explorer. *Space Sci. Rev.* 86, 1–22. doi: 10.1007/978-94-011-4762-0\_1
- Terasawa, T., Fujimoto, M., Mukai, T., Shinohara, I., Saito, Y., Yamamoto, T., et al. (1997). Solar wind control of density and temperature in the near-earth plasma sheet: Wind/geotail collaboration. *Geophys. Res. Lett.* 24, 935–938. doi: 10.1029/96GL04018
- Tsurutani, B. T., and Smith, E. J. (1974). Postmidnight chorus: a substorm phenomenon. *J. Geophys. Res.* 79, 118–127. doi: 10.1029/JA079i001p00118
- Tsutomu, T., and Teruki, M. (1976). Flapping motions of the tail plasma sheet induced by the interplanetary magnetic field variations. *Planet. Space Sci.* 24, 147–159. doi: 10.1016/0032-0633(76)90102-1
- Tsyganenko, N. A., and Mukai, T. (2003). Tail plasma sheet models derived from geotail particle data. *J. Geophys. Res. Space Phys.* 108:1136. doi: 10.1029/2002JA009707
- Walker, R. J., and Farley, T. A. (1972). Spatial distribution of energetic plasma sheet electrons. *J. Geophys. Res.* 77, 4650–4660. doi: 10.1029/JA077i025p04650
- Wang, C.-P., Gkioulidou, M., Lyons, L. R., Wolf, R. A., Angelopoulos, V., Nagai, T., et al. (2011). Spatial distributions of ions and electrons from the plasma sheet to the inner magnetosphere: comparisons between themis-geotail statistical results and the rice convection model. *J. Geophys. Res. Space Phys.* 116:A11216. doi: 10.1029/2011JA016809
- Wang, C.-P., Kim, C. Y., Weygand, J. M., Hsu, T.-S., and Chu, X. (2017). Effects of solar wind ultralow frequency fluctuations on plasma sheet electron temperature: regression analysis with support vector machine. *J. Geophys. Res. Space Phys.* 122, 4210–4227. doi: 10.1002/2016JA023746
- Wang, C.-P., Lyons, L. R., Nagai, T., Weygand, J. M., and McEntire, R. W. (2007). Sources, transport, and distributions of plasma sheet ions and electrons and dependences on interplanetary parameters under northward interplanetary magnetic field. *J. Geophys. Res. Space Phys.* 112:A10224. doi: 10.1029/2007JA012522
- Wing, S., Johnson, J. R., Chaston, C. C., Echim, M., Escoubet, C. P., Lavraud, B., et al. (2014). Review of solar wind entry into and transport within the plasma sheet. *Space Sci. Rev.* 184, 33–86. doi: 10.1007/s11214-014-0108-9
- Wing, S., Johnson, J. R., Newell, P. T., and Meng, C.-I. (2005). Dawn-dusk asymmetries, ion spectra, and sources in the northward interplanetary magnetic field plasma sheet. *J. Geophys. Res. Space Phys.* 110:A08205. doi: 10.1029/2005JA011086
- Yue, C., Wang, C.-P., Lyons, L., Wang, Y., Hsu, T.-S., Henderson, M., et al. (2015). A 2-d empirical plasma sheet pressure model for substorm growth phase using the support vector regression machine. *J. Geophys. Res. Space Phys.* 120, 1957–1973. doi: 10.1002/2014JA020787
- Zhelavskaya, I. S., Shprits, Y. Y., and Spasovska, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *J. Geophys. Res. Space Phys.* 122, 11227–11244. doi: 10.1002/2017JA024406

**Conflict of Interest:** 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.

Copyright © 2020 Swiger, Liemohn and Ganushkina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Neural Network Based Identification of Energy Conversion Regions and Bursty Bulk Flows in Cluster Data

Vlad Constantinescu\* and Octav Marghitu

*Institute of Space Science, Bucharest, Romania*

## OPEN ACCESS

**Edited by:**

Enrico Camporeale,  
University of Colorado Boulder,  
United States

**Reviewed by:**

Andrei Runov,  
University of California, Los Angeles,  
United States  
Arnaud Masson,  
European Space Astronomy Centre  
(ESAC), Spain

**\*Correspondence:**

Vlad Constantinescu  
vlad@spacescience.ro

**Specialty section:**

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 03 April 2020

**Accepted:** 06 July 2020

**Published:** 14 August 2020

**Citation:**

Constantinescu V and Marghitu O  
(2020) Neural Network Based  
Identification of Energy Conversion  
Regions and Bursty Bulk Flows in  
Cluster Data.  
*Front. Astron. Space Sci.* 7:51.  
doi: 10.3389/fspas.2020.00051

Neural networks (NN) provide a powerful pattern recognition tool, that can be used to search large amounts of data for certain types of “events”. Our specific goal is to make use of NN in order to identify events in time series, in particular energy conversion regions (ECRs) and bursty bulk flows (BBFs) observed by the Cluster spacecraft in the magnetospheric tail. ECRs are regions where  $\mathbf{E} \cdot \mathbf{J} \neq 0$  is rather well-defined and observed on time scales from a few minutes to a few tens of minutes ( $\mathbf{E}$  is the electric field and  $\mathbf{J}$  the current density). BBFs are high speed plasma jets, known to make a significant contribution to magnetospheric dynamics. Not surprisingly, ECRs are often associated with BBFs. The manual examination of the Cluster plasma sheet data from the summer of 2001 provided start-up sets of several ECRs and, respectively, BBFs, used to train feed-forward back-propagation NNs. Subsequently, larger volumes of Cluster data were searched for ECRs and BBFs by the trained NNs. We present the results obtained and discuss the impact of the signal-to-noise ratio on these results.

**Keywords:** neural networks, feedforward backpropagation, cluster, energy conversion region, bursty bulk flow

## INTRODUCTION

As sensor resolutions and sampling frequencies increase, data available from space missions is steadily increasing. For example, the SMILE mission will generate 30 Gbits/orbit of data (Raab et al., 2016). Processing these data requires large amounts of time spent by researchers in order to identify interesting events. For example, Paschmann et al. (2018) assembled a database with thousands of manually selected events from the MMS mission.

Pattern recognition tasks can be handled very well by the human brain which has a highly complex, non-linear and parallel structure (Haykin, 2009). Such an information processing system is able to perform these tasks much faster than today’s computers. An artificial neural network is based on a simplified model of the biological neural network and brings the pattern recognition power of the brain into the world of computers.

As another example, Wing et al. (2003) showed the application of a multilayer feedforward neural network in the classification of radar signals from ionospheric irregularities. The neural network implementation correctly classified 98% of the signals.

In this paper we show results obtained by using a feedforward neural network for automatically locating regions of interest in time series. We searched for events in data from the Cluster mission, consisting of four identical spacecraft launched in pairs in July and August 2000, with a perigee of  $4 R_E$  and an apogee of  $19.6 R_E$  (Escoubet et al., 2001). The mission, whose operational phase started in February 2001, allows for *in situ* exploration of particle and field data, with emphasis on investigations that require multi-point data and techniques—for example, analysis of vector

fields, like deriving current density from magnetic field data, or examination of magnetospheric boundary layers, like, e.g., the magnetopause. While in the following we concentrate on Cluster data, the broader goal of the paper is to investigate and illustrate the potential benefits of applying neural networks to time series of space physics interest. We aim to provide examples of using NN algorithms to quickly identify specific events, and by doing so to enable automated build-up of event databases. This can help a better exploitation of big volumes of data, much of which are time series, and by that a better view over the phenomena under study and a better insight to the relevant physics.

In the next section we present the type of neural networks architecture used in this paper and detail the software implementation of the algorithm. In section Neural Network Identification of ECRs: Key Questions, we describe several difficulties encountered during the application of the algorithm and the solutions to overcome those. In sections Neural Network Identification of ECRs: Comparison With an Algorithmic Approach and Neural Network Identification of BBFs, we present the application of this algorithm on two types of time series data provided by instruments onboard Cluster spacecrafts, namely energy conversion regions (ECRs) and bursty bulk flows (BBFs). These two examples illustrate the cases of low and high signal to noise ratio, for ECRs and BBFs, respectively, and provide qualitative information on the impact of this parameter, as well as a test bed for an upcoming quantitative assessment. In the last section we present our conclusions.

## NEURAL NETWORKS

### Introduction: Various NN Types

Depending on the interconnection between the artificial neurons, neural networks can have different architectures. We further present three main classes:

- **Single-layer feedforward networks:** contain just an input and an output layer. The information is passed from the input directly to the output. This network architecture contains only one layer of processing neurons, as the input layer does not perform any computations.
- **Multi-layer feedforward networks:** contain more than one processing layer. Information is passed from the input layer to the output layer via one or more hidden layers.
- **Recurrent networks:** these networks contain at least one feedback loop connecting the output to the input.

Neural networks can learn either helped by a “teacher” or without one. The first case is called **supervised learning** and it involves a previously known set of input-output pairs that can be presented to the network. When learning without a “teacher,” the method is called **unsupervised learning**. In this case, the network attempts to split the input data in different classes. A combination of the two methods is **reinforcement learning**, where the network is just presented with the consequences of its actions. In this case, the network attempts to minimize a given criterion by modifying the decisions it makes.

### Feedforward Back Propagation NN

In order to identify the regions of interest in time series, we make use of a feedforward back propagation neural network, which is of single or multi-layer feedforward network type. The feed-forward direction refers to the traveling of the processed input toward the output when the trained network is used. The back propagation refers to the traveling of the error backwards through the network as part of the training of the network.

During the training phase, the network learns to map a series of input vectors to the corresponding, known, output vectors. The learning process modifies the values of the inter-neuron connection weights.

The first layer of the network, the input layer, has the size equal to the size of the input vector and does not perform any computations. Its function is to feed the input information to the network. The next layers, one or more hidden layers or directly the output layer, process the information.

As data travel through the network, it is adjusted via the inter-neuron connection weights. Based on these inputs, each neuron computes its activation function. The usual types of activation functions are sigmoid and linear, and both of them are used in this paper.

The forward mechanism involves the processing of the input data by the neurons, through use of weights and transfer functions. While the latter are fixed, the weights of the network are adjusted during the training process. Data are multiplied with the corresponding weights and the input of each neuron is computed as the sum of the weighted input values:

$$I_j^{(1)} = \sum_{i=1}^m \omega_{ji} O_i^{(0)}$$

Where

- $I_j^{(1)}$  is the input of neuron  $j$  from layer 1
- $\omega_{ji}$  is the weight of the connection from input  $i$  to neuron  $j$
- $O_i^{(0)}$  is the output of the  $i$ -th element of the previous layer (input value in the case of the first layer of neurons)
- $m$  is the number of elements of the previous layer (input in the case of the first layer of neurons).

The computed weighted sum  $I_j^{(1)}$  is further passed to the activation function of the neuron  $j$  and the corresponding output computed:

$$O_j^{(1)} = f(I_j^{(1)})$$

Where

- $O_j^{(1)}$  is the output of neuron  $j$  from layer 1
- $f$  is the activation function of the neuron

In case of the last layer, the output ( $O_j^{(1)}$ ) of the neuron is an element in the output vector of the network.

During training, the values of the corresponding output vector for a given input vector are known. The error vector ( $e$ ) is

computed as the difference between the desired output and the actual output:

$$e = D - O$$

Where

- $D$ —desired output vector
- $O$ —actual output vector.

The aim of the algorithm is to minimize the error, so the weights are adjusted backward from the output, in an iterative process:

$$\omega_{ji}(n+1) = \omega_{ji}(n) + \alpha [\Delta\omega_{ji}(n-1)] + \eta\delta_j(n)I_i(n)$$

Where

- $\omega_{ji}(n+1)$  is the adjusted weight (step  $n+1$ )
- $\omega_{ji}(n)$  is the unadjusted weight (step  $n$ )
- $\alpha$  is the momentum constant
- $\eta$  is the learning rate
- $\delta_j$  is the local gradient—takes into account the error and the derivative of the activation function.

Essentially, the training phase represents the non-linear optimization of the weights, by using specific functional forms for the activation functions and interconnection between neurons. The learning rate and momentum constant are fixed parameters for a given instance of the network, that can be adjusted during the training phase. Together with the number of layers and the number of neurons in each layer, they represent the global parameters of the network. The user has to tune these parameters during the training phase in order to improve the results of the network. As detailed under Section Neural Network Identification of ECRs: Key Questions, the key tasks of the training phase are both the training *per se* and identifying the network configuration that optimizes the results.

The training stops when one of these conditions is met:

- The error drops below a certain value
- The specified number of training epochs was reached
- The error decrease rate is slow enough.

## Key Features/Parameters

The sizes of the input and output layers are determined by the sizes of the input and corresponding output, respectively. The hidden layer size is determined experimentally by the user. There is no fixed rule that specifies the size or even the presence of the hidden layer.

In general, an increase in the size of the hidden layer can help the network to better learn the features of the training set. On the other hand, if the network size is increased too much, it might lose its ability to generalize and therefore to address data sets which are different from the training set—which is the actual goal of the training. A correctly configured and trained NN must be able to accurately evaluate new data, based on key features learned from the training set.

The neural network architecture presented in this paper relies either on none or one hidden layer and can be tuned to better perform the task at hand by adjusting the size of

this hidden layer, the learning rate ( $\eta$ ) and the momentum constant ( $\alpha$ ). Further degrees of freedom that need to be handled are the intrinsic variability of the network results and setting the training stop condition—all detailed under Section Neural Network Identification of ECRs: Key Questions.

## Software Implementation

We implemented a feedforward neural network algorithm (private communication by Simon Wing) in C. The C programming language gives more flexibility in choosing the platform where to run the software (Linux or Windows). For switching between platforms, the program requires only a recompilation on the target platform.

The program configuration parameters are read from a file, whose name is given at runtime as a command line parameter. Inside the parameter file, the user must specify:

- Number of layers of the network
- Transfer function for each layer
- Number of neurons for each layer
- Number of training pairs
- Name of the file containing training data (input-output pairs)
- Number of testing pairs
- Name of the file containing testing data (input data)
- Name of the file containing the evolution of the error rate
- Name of the file containing the response of the network to the testing data.

The parameters that control the evolution of the network during training, the learning rate and the momentum, are specified inside the C code and can be modified if needed. This setup gives the user the possibility to explore in parallel multiple network configurations. A simple script can start the predefined configurations, each with its separate output file. A batch system allows the execution of multiple instances in parallel, each with different parameters and input data. This allows the user to search the parameter space more efficiently.

## NEURAL NETWORK IDENTIFICATION OF ECRs: KEY QUESTIONS

The manual examination of the Cluster plasma sheet data from the summer of 2001 provided a first set of energy conversion regions (ECRs; Marghitu et al., 2010), where  $E \cdot J \neq 0$ , with  $E$  the electric field and  $J$  the current density. Estimates of  $E$  are typically available on Cluster from more than one instrument, providing a necessary redundancy when the electric field is low, while  $J$  is inferred from the magnetic field,  $B$ , measured by the four Cluster satellites, as a direct application of Ampère's law.

More specifically, the electric field, was derived as  $E = -\mathbf{V} \times \mathbf{B}$ , with plasma velocity,  $\mathbf{V}$ , inferred from the Hot Ion Analyzer (HIA) and Composition and Distribution Function (CODIF) sensors of the Cluster Ion Spectrometer (CIS) experiment. At each time, the electric field was obtained as an average value, by using ion data from Cluster 1, Cluster 3, and Cluster 4, where one or both sensors were operational (no CIS sensor was operational on Cluster 2). Data from the Electric Field and Wave (EFW)

experiment were only used to cross-check the CIS estimates. EFW provides just the spin plane projection of the electric field and the assumption  $\mathbf{E} \cdot \mathbf{B} = 0$  is needed to derive the full electric field vector. In the tail plasma sheet, where we searched for ECR events, the angle between the magnetic field vector and the Cluster spin plane was in general too small for inferring the missing electric field component (perpendicular to the spin plane) from the condition  $\mathbf{E} \cdot \mathbf{B} = 0$ . Therefore, EFW data were just used to cross-check the spin plane electric field, in particular the dawn-dusk,  $E_y$  component ( $E_x$ , typically small and affected by a Sun offset, is less reliable). The current density,  $J$ , was computed from magnetic field data of the FluxGate Magnetometer (FGM) experiment, by using the Curlometer method (Dunlop et al., 2002), taking care of the spacecraft separation, as well as planarity and elongation of the Cluster tetrahedron (see below). For further details on the electric field and current density estimates, used to derive the power density  $\mathbf{E} \cdot \mathbf{J}$ , the reader is referred to Marghitu et al. (2006).

Among the observed ECRs, some were concentrated generator regions (CGRs),  $\mathbf{E} \cdot \mathbf{J} < 0$ , where mechanical energy is converted into electromagnetic energy, while others, more numerous in the geomagnetic tail, were concentrated load regions (CLRs),  $\mathbf{E} \cdot \mathbf{J} > 0$ , where the sense of energy conversion is reversed. As illustrated in **Figure 1**, in both cases energy conversion is rather well-defined and observed for a relatively short time (a few minutes to a few tens of minutes). Since energy conversion is associated with interesting signatures in the plasma parameters (notably plasma velocity and related BBFs), and is known to be an important ingredient of key plasma processes, it appeared as useful to replace the time-consuming manual search with an automated procedure.

An algorithmic procedure has been developed by Hamrin et al. (2009a) which led to the identification of 151 ECR events, of which 116 CLRs and 35 CGRs, in the Cluster crossings of the plasma sheet from 2001. The adjusted and refined procedure was applied later on to Cluster plasma sheet crossings from 2001, 2002, and 2004 (Hamrin et al., 2010), resulting in a total of 555 ECRs, of which 428 CLRs and 127 CGRs. A broader set of events, extended to cover also 2003 and 2005 (with due care to the small Cluster tetrahedron size in 2003 and multi-scale configuration in 2005), was used to select the NN training base, consisting of 81 CLRs and the testing set consisting of 326 CLRs (see section Selection of the Training Set).

The 81 CLRs selected for training had the following distribution over 2001–2005: 11 of 2001, 28 of 2002, 19 of 2003, 21 of 2004, and 2 of 2005. The testing set consisted of 326 events, distributed as follows: 59 of 2001, 107 of 2002, 46 of 2003, 108 of 2004, and 6 of 2005. As indicated by Hamrin et al. (2009a), for all selected ECRs the tetrahedron was reasonably regular (elongation and planarity  $<0.4$ ), which applied also for the events of 2003 and (few) of 2005. Moreover, ECR events that were too short ( $<100$  s), too weak (absolute value of average  $\mathbf{E} \cdot \mathbf{J} < 0.4$  pW/m<sup>3</sup>, absolute value of integrated power density  $<200$  pJ/m<sup>3</sup>), or too close to the kinetic regime (duration multiplied by plasma velocity  $<5$  proton gyro-radii) were not selected. Further details of the event selection algorithm are provided by Hamrin et al. (2009a). While specific thresholds of this procedure were tuned

manually, its application provided a fair selection of ECR events, whose features could be examined subsequently in a consistent manner (Hamrin et al., 2009a,b, 2010).

In a first stage, the trained NNs were used to identify both CLRs and CGRs. Accordingly, the NN output was a vector of the same size as the input data, populated with one of three values: 1 for CLRs, -1 for CGRs, and 0 in rest. Later on, as described below (sections Sliding Window Approach, Selection of the Training Set), we concentrated just on CLRs.

## Size of the Input/Output Layer

Our first approach to identify energy conversion events with neural networks consisted of dividing the data into fixed size intervals (of about 100 elements). For each input interval, we had a corresponding output interval of the same size. This approach resulted in a highly complex network architecture, with 100 input neurons, a hidden layer of various sizes, and 100 output neurons. Since the feed-forward backpropagation NN we used was fully connected (every neuron in a layer was connected to all the neurons in the next layer), the training algorithm had to compute a large number of weights, which required significant training time and computer resources (memory, cpu). The trained network was also not very good in identifying the events. The data used to train the network consisted of concatenated intervals of satellite readings, containing both event (1 or -1 desired output) and non-event data (0 desired output), of which non-event data were by far dominant, therefore the trained network regarded the non-event data as the “right” ones and mostly ignored the event data.

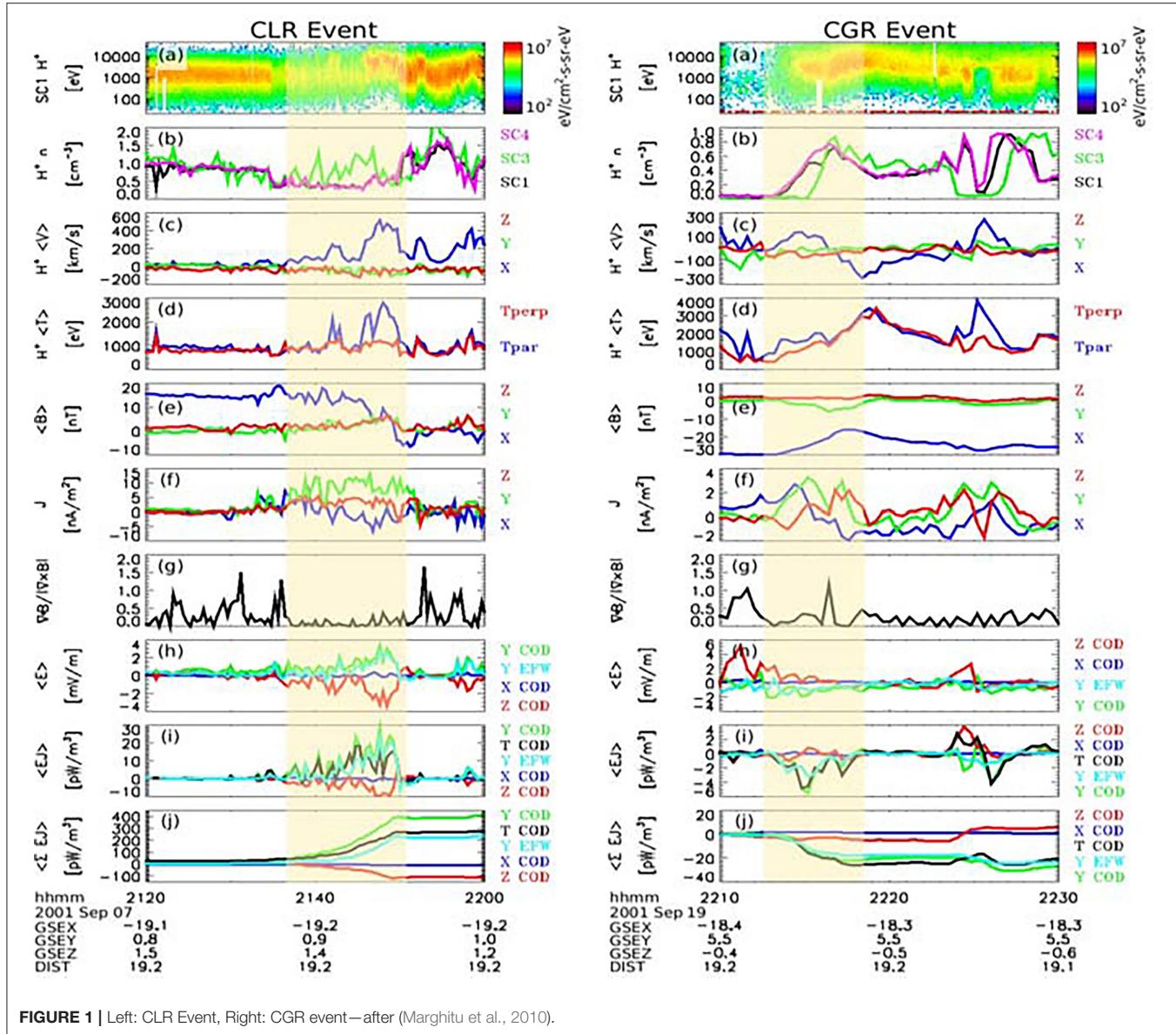
## Use of Synthetic Data for Training and Testing

Using real  $\mathbf{E} \cdot \mathbf{J}$  data raised additional problems: the training set was limited and the data used for training could not be explored later—in a consistent manner—for the presence of CLRs and CGRs. In order to overcome these problems, we tried to use synthetic data for training and testing the NN. When building the synthetic data sets, we randomized the intensity, duration, and sign. During the tests with synthetic data, we noticed an improvement in detection for less complex network configurations. The network worked better for a smaller number of neurons on the hidden layer, but after training with more input data.

For simplicity, we further considered only the search for CLR events ( $\mathbf{E} \cdot \mathbf{J} > 0$ ), as presented in section Selection of the Training Set. This search is also similar to the case of BBFs (section Neural Network Identification of BBFs), where events are as well-positive and NN output is accordingly just 1 (event) and 0 (non-event). Moreover, in the tail plasma sheet CLRs dominate over CGRs, consistent with the large-scale load character of this region.

## Sliding Window Approach

In order to decrease the training time and the complexity of the network, we opted eventually for another structure of the NN by implementing a sliding window algorithm. The sliding window consists of a certain number of input neurons, typically



**FIGURE 1 |** Left: CLR Event, Right: CGR event—after (Marghitu et al., 2010).

more than one, and just one output neuron; which implies a great decrease in the complexity of the NN and accordingly of the training time (from days to tens of minutes). As the window moves across the time series, the output is 1 when the window slides over the CLR event.

## Selection of the Training Set

During the tests with real and synthetic data we noticed the importance of a heterogeneous and representative selection of input data for training the network. By including many non-event data points (associated with an output value of 0), the network is biased toward false negatives (more values of 0 in the output), therefore a balanced input selection is required (see also Section Size of the Input/Output Layer).

Starting from the dataset of ECRs observed in 2001, 2002, and 2004, analyzed by Hamrin et al. (2010), extended to cover also

2003 and 2005 (as described above), we constructed training sets with equally distributed events, in duration or intensity. In the data presented to the network during training, one must maintain a balance between event and non-event intervals. Given that the quality of the network training is judged by the mean error of the output (difference between desired and actual output), if one uses mostly non-event intervals (i.e., mostly 0 desired output) a network trained to supply mostly 0 will be wrongly considered well-trained. By using a better tuned training set, we managed to improve the detection accuracy of the NN.

A key question for this better tuning was the uniform selection of training data. For this aim, we sorted the events considering their duration, as well as their intensity, quantified by the median  $E \cdot J$  value (similar to Hamrin et al., 2009a). The median was computed over the length of each event, consisting of at least 25 points (minimum duration 100 s, with 4 s per point—see also

the brief description of the event selection algorithm above). After sorting the events, we tested two selection methods: a linear sampling, 1 out of  $n$ , and a logarithmic sampling, to take into account that most events are weak, i.e., of short duration and low intensity. While the linear sampling selects mostly weak events, and thus is essentially biased comparable to the case of dominant non-event intervals, the logarithmic sampling emphasizes longer/stronger events, which turns out to compensate this bias and to provide better results. Hamrin et al. (2009a) found that intensity is indeed distributed close to logarithmic, while Hamrin et al. (2009b) found the same for duration, in particular of CLRs, therefore logarithmic sampling appears to be a better choice.

While the linear selection 1 out of  $n$  is self-explanatory (every  $n$ th event), in the case of logarithmic sampling we used the natural logarithms of the respective values and made the selection according to the following formula:

$$\text{Index} = 1 + (N - 1) \cdot \frac{\ln\left(\frac{y_1 + j \frac{y_N - y_1}{M}}{y_1}\right)}{\ln\left(\frac{y_N}{y_1}\right)}$$

Where

- $\text{Index}$  is the index of the element to be selected from the input set, comprising in our case 407 CLR events.
- $y_{1, N}$  are the elements of the input set (of size  $N$ , in our case 407), from which we select elements of the training set;  $y_1$  is the first element of the input set and  $y_N$  the last one.
- $j$  is the index in the destination (selected) training set, containing selected elements, and runs from 0 to  $M$ ; in our case  $M$  is 80 and the training set comprises 81 events.

In **Figure 2** we present the duration of the events in the training set, selected by linear sampling 1 out of 5 (top panel-red line) and logarithmic sampling (top panel-blue line), as well as the duration of the events used for testing (bottom panel). The full set of events includes 407 CLRs, the training sets 81 events, and the testing set 326 events. For logarithmic sampling, 81 was the maximum number that ensured distinct indices after rounding at the low  $\text{index}$  end. For the linear selection, 1 out of 5 provides a training set of similar size with the logarithmic selection. As expected, the logarithmic sampling provides a more uniform selection as compared to the linear sampling, that is, a somewhat better representation of the longer events. The plots are similar (not shown) when duration is replaced by intensity, in agreement with the distributions observed by Cluster (Hamrin et al., 2009a,b).

### Other Parameters: Momentum Constant, Learning Rate, Stop Condition, Initial Conditions

One difficulty encountered in the tests with ECR data was the unstable NN behavior, with fast growth of the weights (defining the connections between neurons) and error (i.e., the difference between the actual response of the NN and the target response) sometimes leading to numerical overflow.

The momentum constant ( $\alpha$ ) controls the adjustment of the network's weights, based on the previous evolution of the weight. The value of  $\alpha$  must be kept  $>-1$  and smaller than 1. A value of 0 means no momentum influence during training the network. The momentum constant can prevent the network from stopping in a local minimum during training. In our tests, we found that a momentum constant of 0.5 favors the stability of the network.

A smaller learning rate parameter ( $\eta$ ) determines a smaller change in the synaptic weights of the network from one iteration to another. If the learning rate parameter is too large, in order to speed up the training, the resulting large changes of the weights may render the evolution of the network unstable during the training phase. The value of the learning rate parameter should be kept above 0 and less or equal to 1. In practice, we found appropriate a small  $\eta$  value, of 0.0001.

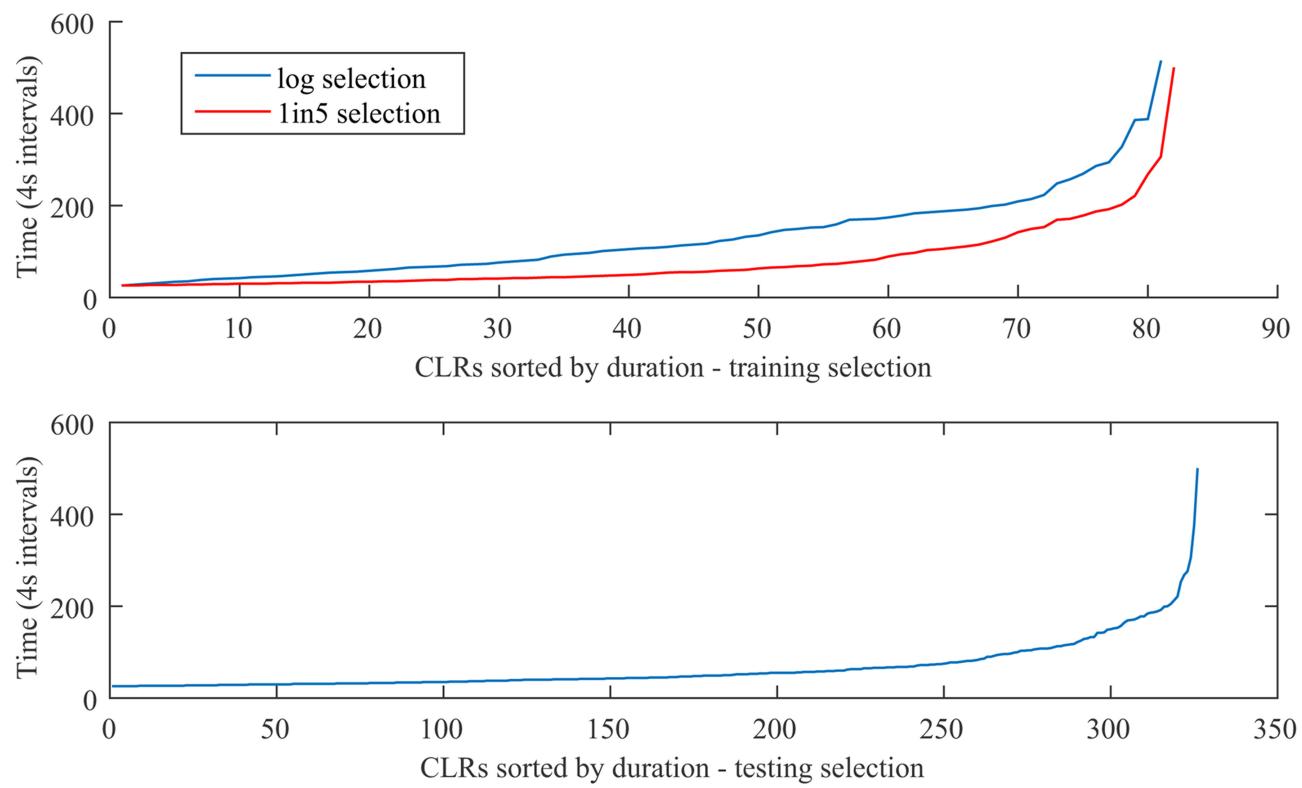
Two more important features proved to be the stop condition and the intrinsic variability of the network. Thus, the stop condition had to be formulated in terms of relative decrease in the error, as opposed to a fixed number of iterations. The intrinsic variability of the network was related to the randomly selected initial weights. In practice, the network was trained several times, and the best instance selected for further operation.

## NEURAL NETWORK IDENTIFICATION OF ECRs: COMPARISON WITH AN ALGORITHMIC APPROACH

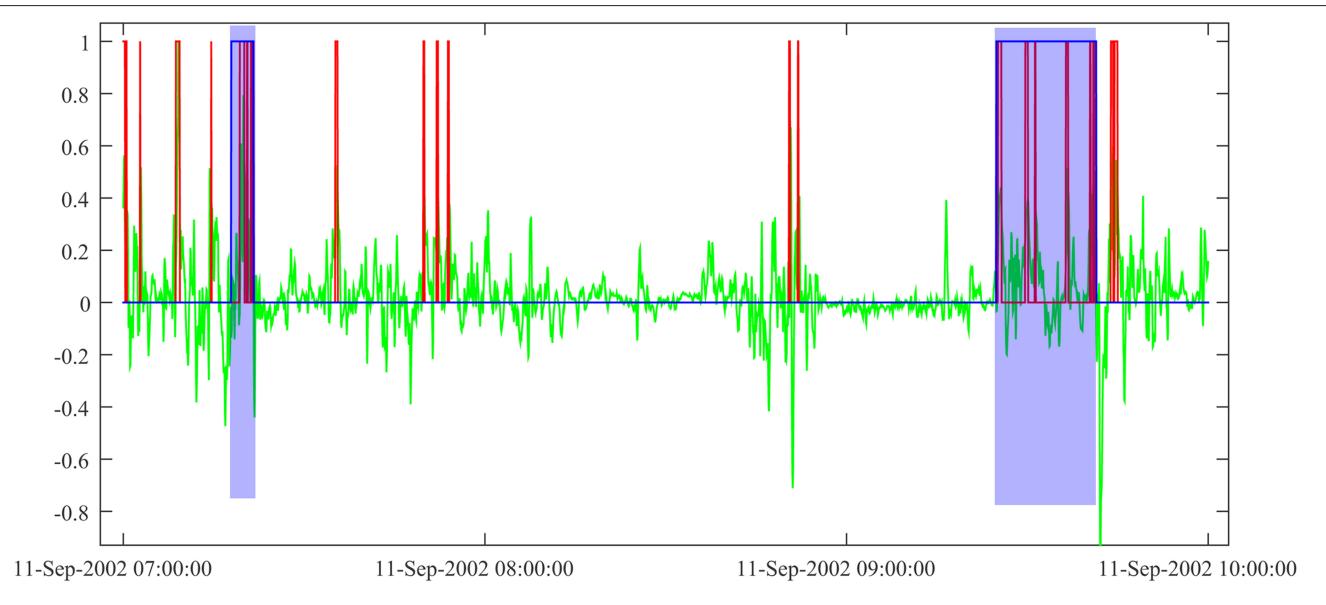
The detailed examination of the results on real E-J data (not concatenated individual events) shows that the NN identifies neighboring events as separate ones, while the semi-automatic algorithmic procedure set up by Hamrin et al. (2009a) includes a post-processing that merges such neighboring events into a single one. As detailed below, since post-processing introduces further degrees of freedom, we decided to skip it for the time being and to compare the results in a way that is less sensitive to this step, by using the cumulative sum of E-J.

In **Figure 3** we present a sample output of the NN search for ECR events over a time interval of  $\sim 3$  h. The NN configuration in this case included no hidden layer and the sliding window (Section Sliding Window Approach) had five points. During our tests, we explored the use of different window sizes, from 5 to 101 points, with various hidden layer sizes, from 0 (no hidden layer) to 71. Following the evaluation detailed below, the NN configuration behind **Figure 3** provided the results that we regarded as the best match to the algorithmic approach, used for reference. This means that most events identified by the algorithmic approach were also detected by the NN (which does not exclude additional events detected by the NN).

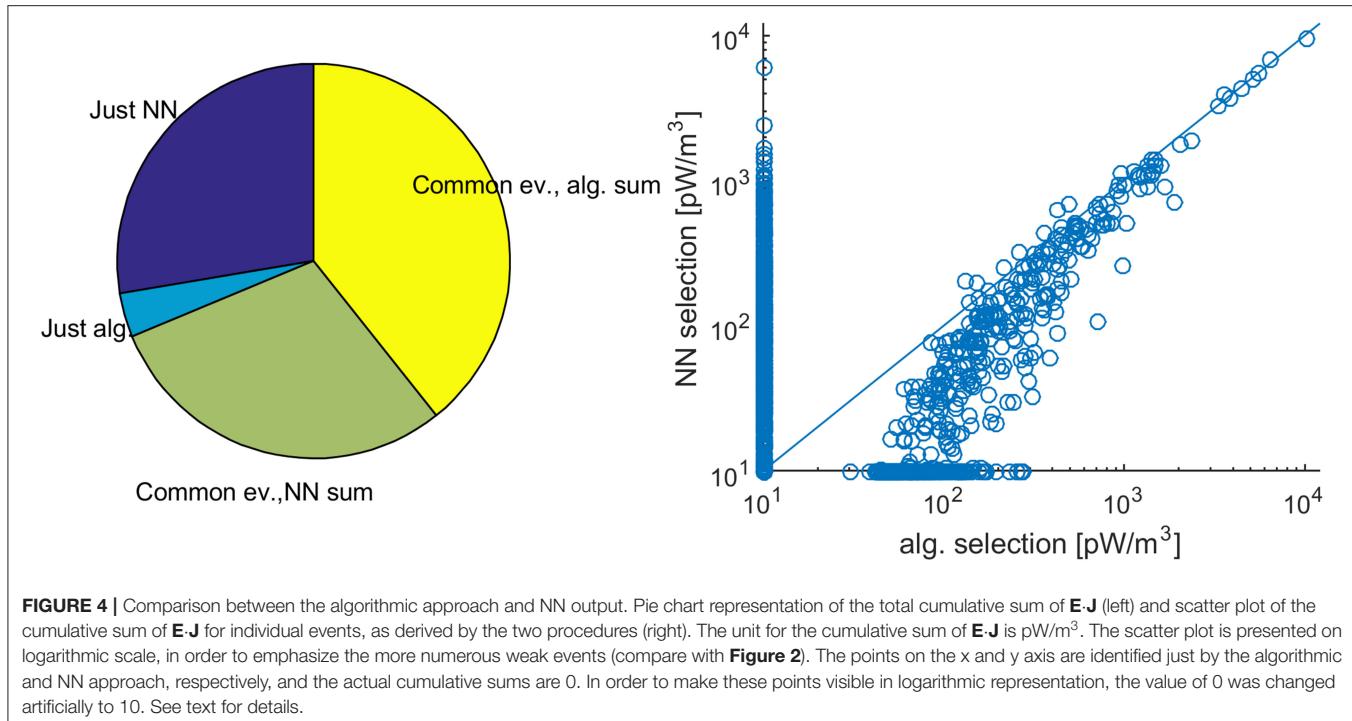
In order to evaluate the multiple network configurations, we compared the NN output with the results of the algorithmic approach by building pie charts and scatter plots for all network configurations, as illustrated in **Figure 4**. This allowed a quick and overall visual comparison of the events selected by the NN with those selected by the algorithmic approach. **Figure 4** corresponds to a NN with no hidden layer and an input sliding window of five elements. The events used for training were



**FIGURE 2 |** CLRs sorted by duration—training selection (top), testing selection (bottom). The top panel shows both the logarithmic selection (blue) of 81 events and the linear selection (red) 1 out 5, which provides a similar number of events with the logarithmic one. The x axis shows the event index in the respective set, while the y axis shows the event duration in Cluster spin periods of 4 s (the number on y axis should be multiplied by four to get the duration in seconds). See text for details.



**FIGURE 3 |** NN output on **E-J** data: NN output (rounded) in red, **E-J** in green (normalized to the maximum value of the time interval, here 17.4 pW/m<sup>3</sup>). Events found by the algorithmic approach are shaded in blue.



**FIGURE 4 |** Comparison between the algorithmic approach and NN output. Pie chart representation of the total cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  (left) and scatter plot of the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  for individual events, as derived by the two procedures (right). The unit for the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  is  $\text{pW}/\text{m}^3$ . The scatter plot is presented on logarithmic scale, in order to emphasize the more numerous weak events (compare with **Figure 2**). The points on the x and y axis are identified just by the algorithmic and NN approach, respectively, and the actual cumulative sums are 0. In order to make these points visible in logarithmic representation, the value of 0 was changed artificially to 10. See text for details.

logarithmically distributed in duration and selected from Cluster data between 30th May 2001 and 30th December 2004.

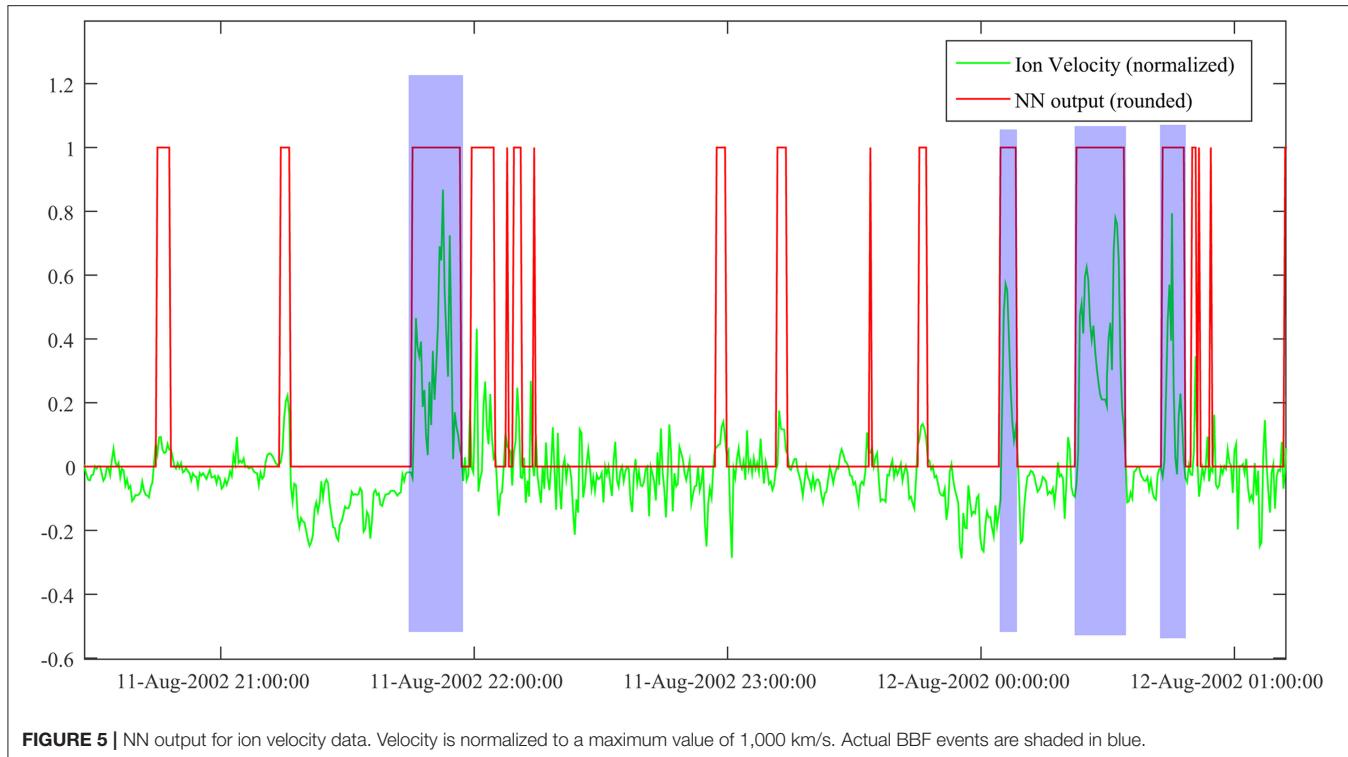
The left plot of **Figure 4** shows the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  over the selected events. As hinted at above, we picked the integrated magnitude of the events since this is less sensitive to the exact definition of an event—whether the individual spikes identified by the NN are merged together, as in the algorithmic approach, or not. We also selected the pie chart for the visual representation, even if in this case the whole pie does not have the usual meaning, namely it is not equal to the sum of the slices. Nevertheless, it provides a useful tool to quickly assess the matching of the two procedures. More exactly, dark/light blue indicates events found only by the neural network/algorithmic approach, while light green/yellow shows the cumulative sum for the common events, computed over the results of the neural network/algorithmic approach. For the first, the small light blue and large dark blue slices show that the NN events is essentially a super-set of the events derived by the algorithmic approach, which looks promising. For the latter, the positive difference between yellow and light green is consistent with the visual impression of **Figure 3**, that the “elementary” events identified by the NN do not fully fill the merged events obtained by the algorithmic approach after post-processing.

A complementary view is provided by the scatter plot on the right side of **Figure 4**, which helps to substantiate further the comparison of the results by an event-oriented perspective. Each event is indicated by an empty bullet, with the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  as derived by the algorithmic/NN approach on the x/y axis. Most bullets indicate values of  $< \sim 1,000 \text{ pW}/\text{m}^3$ , consistent with the point made before that most events are observed to be short and weak (**Figure 2**). Most of the bullets identified by both

the algorithmic and the NN approach are also slightly under the first bisector, consistent with the larger yellow slice compared to the light green one, and the respective cumulative sums of the algorithmic approach are larger than  $50 \text{ pW}/\text{m}^3$ , as required by the threshold of  $200 \text{ pJ}/\text{m}^3$ , indicated in Section Neural Network Identification of ECRs: Key Questions (which, when expressed in terms of cumulative sum, should be divided by the 4 s duration of each sample). Several bullets are aligned along the x and y axes, consistent with the light blue and dark blue slices of the pie chart, respectively, in the left plot of **Figure 4**. For these latter bullets,  $10 \text{ pW}/\text{m}^3$  were artificially added, to make them visible in logarithmic representation—whose origin here is the point  $(10, 10) \text{ pW}/\text{m}^3$ .

Since typical magnitudes of  $\mathbf{E}$  and  $\mathbf{J}$  are in the range of a few up to several  $\text{mV}/\text{m}$  and  $\text{nA}/\text{m}^2$ , respectively, the power density unit for  $\mathbf{E} \cdot \mathbf{J}$ , and for the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  in **Figure 4**, is  $\text{pW}/\text{m}^3$ . Note that by scaling  $\mathbf{E} \cdot \mathbf{J}$  with 4 s (the duration of the sampling interval), one can derive the energy density profile along Cluster crossing of the CLR, in  $\text{pJ}/\text{m}^3$ . From this perspective, by scaling the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  with 4 s and by dividing the result through the number of points (i.e., 4 s intervals), one can derive a proxy for the average energy density of each CLR event. While Cluster data provide some information on the size and lifetime of the CLR events (Hamrin et al., 2009b), the precise relevance of this proxy depends on the particular geometry of the Cluster crossing with respect to the CLR. For this reason, we prefer to show the cumulative sum of  $\mathbf{E} \cdot \mathbf{J}$  and refrain from scaling this result.

The visual information provided by the pie chart and scatter plot can aid in quickly assessing the performance of the NN and is preferred here against the standard evaluation in terms



**FIGURE 5 |** NN output for ion velocity data. Velocity is normalized to a maximum value of 1,000 km/s. Actual BBF events are shaded in blue.

of false positive and false negative events. As described above (section Neural Network Identification of ECRs: Key Questions), the definition of the “event” in the algorithmic approach—which provides our comparison benchmark—is to some extent empirical, and cannot be regarded as an “absolute” reference. The identification of the events allows for jitter in the start and end time, the NN approach can select individual smaller events corresponding to a larger, singular event of the algorithmic approach, and several new events are detected by the NN (the dark blue slice of the pie chart and the points on the y axis of the scatter plot in **Figure 4**). Therefore, comparing the results in terms of “events” is not straight forward and, in this case, performance assessment by means of the more flexible cumulative sum looks better suited than the classical manner.

## NEURAL NETWORK IDENTIFICATION OF BBFs

Using the software developed to search for ECRs, we considered testing the functionality of the selection tool provided by the neural network implementation on other time series. We decided to explore the possibility of using this setup to locate Bursty Bulk Flow (BBF; Angelopoulos et al., 1992) events in ion velocity observed by the HIA instrument (Rème et al., 2001) on Cluster. In qualitative terms, BBFs are better defined than ECRs, meaning less weak events and a better signal-to-noise ratio. Intuitively, one can expect that an automatic pattern recognition tool will work better with BBFs. Unlike for ECRs, in the case of BBFs we did not benefit from an event database at hand, to provide

the training set. On the other hand, the definition of BBF events is documented in several studies (even though there is some variability in the relevant criteria), therefore assembling a training set is significantly easier compared to ECRs. Moreover, observation of BBFs requires just single-s/c ion data, compared to multi-s/c, multi-instrument data for ECRs. This has a positive effect on the errors and on the signal-to-noise ratio.

In order to assemble the training set, we manually selected 39 events, with duration between 500 and 3,000 s and a velocity threshold of 400 km/s. The actual training set was built by extending the selection around the events to include also non-event data points and by finally concatenating the data. Similar to the ECR case, we tested several network configurations with this training set and eventually selected once again the one used with ECRs (window size 5, no hidden layer).

**Figure 5** shows a representative sample output of the trained network over a ~5 h Cluster crossing of the plasma sheet. Compared to the ECR sample in **Figure 3**, covering a shorter interval, the NN output is less abundant (fewer red events per unit time), which is a consequence of the better signal-to-noise ratio. The event selection seems reasonably accurate, but it still requires further post-processing (i.e., joining short, neighboring events; rejecting events below the velocity threshold). For this particular time interval, the actual number of BBF events is 4 (shaded in blue), while the NN output is ~20 (counting the red rectangles/spikes). This suggests a rough multiplication factor of five, between the actual number of events and the network result. Implementing the post-processing will also make possible a quantitative assessment of the NN performance, in the standard terms of false positive and false negative events.

The test data was built by using Cluster plasma sheet measurements from the tail seasons of 2001–2004, 1st of August to 10th of October. The actual tail seasons were actually somewhat longer, but we made a conservative choice, to avoid contamination with magnetosheath data on the dawn flank (before 1st of August) and on the dusk flank (after 10th of October). A preliminary count provided about 4,150 NN events, namely some 800 actual events for the rough multiplication factor of five estimated above. Compared to the training set of 39 events, the gain in the number of events is of the order of 20, which consolidates the case for using NNs in exploring time series. Obviously, in a particular case like this, the human effort to build up a large event data base, appropriate, e.g., for statistical studies, is tremendously decreased.

## CONCLUSIONS

The NN approach explored in this paper provides an efficient tool to automatically identify specific events in time series. When using supervised learning, as illustrated here, a key stage is building a representative training set, to be extended by the network later on, during its nominal use. Another essential feature is that the user must explore different network configurations, by training and testing, in order to find the right setup for the problem at hand.

The effort spent on finding the proper NN setup can be considerable, as illustrated with the case of the weaker/noisier ECRs, which are also harder to identify. In such cases, the benefits of using the NN approach may appear questionable. On the other hand, for the stronger/clearer BBF events, detection is both easier and more reliable. Opposite to ECRs, the benefits of using NNs to build up large event data bases are in such cases obvious. In the future, we plan to further explore and try to quantify the connection between the noise level and the results of neural networks approaches.

Time series data are generously supplied nowadays by a broad spectrum of high-resolution satellite (and ground) experiments.

## REFERENCES

- Angelopoulos, V., Baumjohann, W., Kennel, C. F., Coroniti, F. V., Kivelson, M. G., Pellat, R., et al. (1992). Bursty bulk flows in the inner central plasma sheet. *J. Geophys. Res. Space Phys.* 97, 4027–4039. doi: 10.1029/91JA02701
- Dunlop, M., Balogh, A., Glassmeier, K.-H., and Robert, P. (2002). Four-point cluster application of magnetic field analysis tools: the curlometer. *J. Geophys. Res.* 107:1384. doi: 10.1029/2001JA005088
- Escoubet, C. P., Fehringer, M., and Goldstein, M. (2001). The Cluster mission. *Ann. Geophys.* 19, 1197–1200. doi: 10.5194/angeo-19-1197-2001
- Hamrin, M., Norqvist, P., Marghitu, O., Buchert, S., Klecker, B., Kistler, L. M., et al. (2009a). Occurrence and location of concentrated load and generator regions observed by Cluster in the plasma sheet. *Ann. Geophys.* 27, 4131–4146. doi: 10.5194/angeo-27-4131-2009
- Hamrin, M., Norqvist, P., Marghitu, O., Buchert, S., Klecker, B., Kistler, L. M., et al. (2010). Geomagnetic activity effects on plasma sheet energy conversion. *Ann. Geophys.* 28, 1813–1825. doi: 10.5194/angeo-28-1813-2010
- Hamrin, M., Norqvist, P., Marghitu, O., Vaivads, A., Klecker, B., Kistler, L. M., et al. (2009b). Scale size and life time of energy conversion regions observed by cluster in the plasma sheet. *Ann. Geophys.* 27, 4147–4155. doi: 10.5194/angeo-27-4147-2009
- Haykin, S. S. (2009). *Neural Networks and Learning Machines*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Marghitu, O., Hamrin, M., Klecker, B., and Rönnmark, K. (2010). “Cluster observations of energy conversion regions in the plasma sheet,” in *The Cluster Active Archive: Studying the Earth’s Space Plasma Environment*, eds H. Laakso, M. Taylor, and C. P. Escoubet (Dordrecht: Springer), 453–459. doi: 10.1007/978-90-481-3499-1\_32
- Marghitu, O., Hamrin, M., Klecker, B., Vaivads, A., McFadden, J., Buchert, S., et al. (2006). Experimental investigation of auroral generator regions with conjugate Cluster and FAST data. *Ann. Geophys.* 24, 619–635. doi: 10.5194/angeo-24-619-2006
- Paschmann, G., Haaland, S. E., Phan, T. D., Sonnerup, B. U. Ö., Burch, J. L., Torbert, R. B., et al. (2018). Large-scale survey of the structure of the dayside magnetopause by MMS. *J. Geophys. Res. Space Phys.* 123, 2018–2033. doi: 10.1002/2017ja025121
- Raab, W., Branduardi-Raymont, G., Wang, C., Dai, L., Donovan, E., Enno, G., et al. (2016). “SMILE: a joint ESA/CAS mission to investigate the interaction between the solar wind and Earth’s magnetosphere,” in *Space Telescopes and*

Searching such data for specific events, as well as assembling relevant statistical sets, are notoriously time-consuming in space physics research. By taking advantage of machine learning tools, like neural networks, to automate such operations, the time share for creative (and rewarding) research will increase, to the benefit of both event-oriented and statistical studies.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://csa.esac.esa.int/csa-web/>.

## AUTHOR CONTRIBUTIONS

VC implemented the C code and batch scripts and ran tests with different network configurations. OM guided the exploration of EJ and ion velocity data, helped in manually selecting the BBF training events, and evaluated the results of different network configurations. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the PRODEX project MAGICS, ESA contract 4000127660, and by the Core Program LAPLAS VI of the Romanian Ministry of Education and Research. Earlier relevant work was performed under the PECS project ECSTRA, ESA contract 4200098048.

## ACKNOWLEDGMENTS

Helpful suggestions by Simon Wing, on optimizing the NN performance, are gratefully acknowledged. Useful discussions with Maria Hamrin and Patrik Norqvist, on the Cluster ECR database, are thankfully acknowledged too.

- Instrumentation 2016: Ultraviolet to Gamma Ray, Vol. 9905* (Edinburgh: International Society for Optics and Photonics).
- Rème, H., Aoustin, C., Bosqued, J. M., Dandouras, I., Lavraud, B., Sauvaud, J. A., et al. (2001). First multispacecraft ion measurements in and near the earth's magnetosphere with the identical Cluster ion spectrometry (CIS) experiment. *Ann. Geophys.* 19, 1303–1354. doi: 10.5194/angeo-19-1303-2001
- Wing, S., Greenwald, R. A., Meng, C.-I., Sigillito, V. G., and Hutton, L. V. (2003). Neural networks for automated classification of ionospheric irregularities in HF radar backscattered signals. *Radio Sci.* 38, 2-1-2-8. doi: 10.1029/2003RS02869

**Conflict of Interest:** 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.

Copyright © 2020 Constantinescu and Marghitu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# MMS SITL Ground Loop: Automating the Burst Data Selection Process

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

<sup>1</sup> Space Science Center, EOS, University of New Hampshire, Durham, NC, United States, <sup>2</sup> Department of Computer Science, University of New Hampshire, Durham, NC, United States, <sup>3</sup> Laboratory for Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, CO, United States, <sup>4</sup> Space Science Laboratory, University of California at Berkeley, Berkeley, CA, United States, <sup>5</sup> Goddard Space Flight Center, NASA, Greenbelt, MD, United States, <sup>6</sup> EOS-SwRI, Southwest Research Institute, Durham, NH, United States, <sup>7</sup> Southwest Research Institute, San Antonio, TX, United States

## OPEN ACCESS

### Edited by:

Enrico Camporeale,  
University of Colorado Boulder,  
United States

### Reviewed by:

Alexandros Chasapis,  
University of Delaware, United States  
Rungplophan Om Kieokaew,  
UMR5277 Institut de Recherche en  
Astrophysique et Planéologie (IRAP),  
France

### \*Correspondence:

Matthew R. Argall  
matthew.argall@unh.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 31 March 2020

**Accepted:** 16 July 2020

**Published:** 01 September 2020

### Citation:

Argall MR, Small CR, Piatt S, Breen L, Petrik M, Kokkonen K, Barnum J, Larsen K, Wilder FD, Oka M, Paterson WR, Torbert RB, Ergun RE, Phan T, Giles BL and Burch JL (2020) MMS SITL Ground Loop: Automating the Burst Data Selection Process. *Front. Astron. Space Sci.* 7:54. doi: 10.3389/fspas.2020.00054

Global-scale energy flow throughout Earth's magnetosphere is catalyzed by processes that occur at Earth's magnetopause (MP). Magnetic reconnection is one process responsible for solar wind entry into and global convection within the magnetosphere, and the MP location, orientation, and motion have an impact on the dynamics. Statistical studies that focus on these and other MP phenomena and characteristics inherently require MP identification in their event search criteria, a task that can be automated using machine learning so that more man hours can be spent on research and analysis. We introduce a Long-Short Term Memory (LSTM) Recurrent Neural Network model to detect MP crossings and assist studies of energy transfer into the magnetosphere. As its first application, the LSTM has been implemented into the operational data stream of the Magnetospheric Multiscale (MMS) mission. MMS focuses on the electron diffusion region of reconnection, where electron dynamics break magnetic field lines and plasma is energized. 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. Of all 219 SITL selections classified as MP crossings during the first five months of model operations, the model predicted 166 (76%) of them, and of all 360 model predictions, 257 (71%) were selected by the SITL. Most predictions that were not classified as MP crossings by the SITL were still MP-like, in that the intervals contained mixed magnetosheath and magnetospheric plasmas. The LSTM model and its predictions are public to ease the burden of arduous event searches involving the MP, including those for EDRs. For MMS, this 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, mission operations

## 1. INTRODUCTION

Earth is a strongly magnetized planet whose internal dynamics are largely influenced by its interaction with the solar wind and the resulting cycle of magnetic reconnection (Dungey, 1961). 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 (Nagai et al., 2011, 2013; Scudder et al., 2012; Tang et al., 2013; Oka et al., 2016) because spacecraft lacked the spacial and temporal resolution to resolve electron-scale dynamics. These limitations were overcome by the Magnetospheric Multiscale (MMS) mission. Leading up to its launch, so little was known about the EDR that it was unclear how or if EDRs could be identified in the data, which led to speculation into the best EDR indicator (Mozer, 2005; Scudder and Daughton, 2008; Scudder et al., 2008; Zenitani et al., 2011; Aunai et al., 2013; Hesse et al., 2014; Swisdak, 2016). Since launch, however, MMS has identified more than 50 EDRs (see Webster et al., 2018 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 operation's data flow in order to ensure that the selected 4% of burst data is able to address MMS science objectives (Burch et al., 2016; Torbert et al., 2016). 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 have burst mode schemes that operate only when triggered. Some burst modes are triggered on a pre-determined duty cycle, while others are triggered on an instrument-by-instrument basis when on board measurements meet certain criteria. Still others are coordinated among multiple or all instruments. 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. (2015). Triggers used on STEREO for shock detection, their evolution, and their efficacy are documented in Jian et al. (2013). 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 (Baker et al., 2016; Fuselier et al., 2016), and a human Scientist-in-the-Loop

(SITL) who examines all of the low-resolution survey data, and who manually selects and classifies burst intervals. Survey data, however, is insufficient to resolve electron dynamics. A strategy for the SITL, then, is to select all MP crossings. This has resulted numerous EDR encounters but is labor- and resource-intensive; after manual reclassification, just  $\sim 0.7\%$  of MP crossings, or  $\sim 0.0001\%$  of the mission lifetime during MMS's first two years contained an EDR. Such challenges were foreseen when designing the ABS and SITL selection processes and it was envisioned that 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 are the design, implementation, and results of the first ground loop.

Most applications of ML to magnetospheric physics problems to-date have been geared toward the prediction of catastrophic events, including geomagnetically induced currents (Wintoft et al., 2015) that threaten power grids, solar energetic particles that threaten space assets (Bouabrahimi et al., 2017), and geomagnetic indices (Lundstedt, 1997; Borovsky, 2014; Bhaskar and Vichare, 2019) that indicate when global geomagnetic activity could lead to such events. Most methods associate upstream conditions at L1 to those at geosynchronous orbit or on the ground because of the continuous data coverage linking upstream and downstream conditions. Unfortunately, because such models lack knowledge of processes internal to the MSP (e.g., Bhaskar and Vichare, 2019), they tend to suffer precisely during the extreme events they are trying to predict. Space weather prediction can be improved by creating several ML models with knowledge of 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 (Phan et al., 2015). Past attempts to identify the MP in an automated fashion used gradients in plasma parameters such as density or ion flux (Boardsen et al., 2000; Phan et al., 2015). Other methods introduce machine learning to identify the MP indirectly by classifying topologically distinct regions then locating the transition between them. The solar wind, magnetosheath, and magnetosphere were identified by applying probability functions (Jelínek et al., 2012) and support vector machines (da Silva et al., 2020) to ion density and temperature data, 3D convolutional neural network to 3D ion distribution functions (Olshevsky et al., 2019), and random forests to magnetic field and plasma data (Nguyen et al., 2019). 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.

It is clear from the number of MP classifiers above that identifying the MP is important not just for the SITL. The MP is the primary location of mass, momentum, and energy transfer into Earth's magnetosphere. Because of this, many statistical

**TABLE 1 |** Guidelines used during Phase 5A (30 September 2019 through 24 November 2019) specifying how the SITL should classify magnetopause crossings.

Event type	FOM category	Data signatures/Notes
Complete high magnetic shear magnetopause crossing	1 (1 – if very long crossing or low-shear)	Full density gradients and full magnetic field rotations, includes separatrix and exhaust boundary
Magnetopause diffusion region candidate	1+	Reversals of high speed jet and $B$ normal during magnetopause crossing
Magnetopause diffusion region candidate	1+	Magnetopause without boundary layer. At current sheet center: positive sunward pointing normal electric field. (Such events can be difficult to identify in SITL data)
Magnetopause: Kelvin-Helmholtz induced current sheet	2	Quasi-periodic magnetic field and density oscillations, field direction changes. Can select a long interval (tens of minutes) if spectacular
Magnetopause: FTE	2	Bipolar $B$ normal and strong enhancement of $ B $
Magnetopause: partial crossings	2	Incomplete $B$ rotation and density transition (i.e., not reaching magnetospheric levels)
Boundary layer traversals	3-	Excursion into the boundary layer, characterized by magnetosphere $B$ and slight increase in density and appearance of magnetosheath ion population

Events in categories 1–4 are given FOMs of 150–199, 100–149, 50–99, and 0–49, respectively. A “+” or “–” after a given category indicate a selection at the upper- or lower-range of the category. FOMs  $\geq 200$  are reserved for special events, such as calibration intervals or a definitive EDR encounter.

studies have focused on MP properties (Paschmann et al., 1993; Phan and Paschmann, 1996); MP processes like flux transfer events (Fear et al., 2012), Kelvin-Helmholtz instabilities (Kavosi and Raeder, 2015), velocity rotation events (Matsui et al., 2019), impulse events, and kinetic Alfvén waves (Wing et al., 2014); and creating MP models (Boardsen et al., 2000; Jelínek et al., 2012). Statistical studies such as these traditionally require arduous event searches. Automated algorithms and event lists can be used as a first-step data filter to make searches less burdensome. For this reason, our model and its predictions are publicly available for use in future studies (Argall et al., 2020a; Small et al., 2020).

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 outlined as follows. First, the systems for making burst selections, including the SITL, ABS, and GLS, are described in section 2. Next, an overview of the tools and processes developed to support the GLS infrastructure is provided in section 3. Then, a description of the data is given in section 4 and the model in section 5. In section 6, we present the model results and performance. Section 7 is the Discussion, and section 7.1 outlines the GLS Hierarchy, a framework needed to fully automate the SITL selection process. Finally, a summary is given in section 8. Those interested in only the model and its results are referred to sections 5 and 6.

## 2. BURST MANAGEMENT SYSTEMS

MMS burst memory management consists of 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 (GLS).

### 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 1,090 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 ground-based radio communication network and incrementally downlink data. Data is passed through to the Science Data Center (SDC) (section 3.1) where preliminary calibrations are applied and the data is made available to the SITL. The SITL then uses the EVA tool (section 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 (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. The time spent by the SITL in making such selections can be saved 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 (Fuselier et al., 2016).

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 2 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 six parameters is now employed in the search for reconnection. These involve changes in the magnetic field components, the electric field wave power, the electron pressure, and the ion density. The parameters and their corresponding weights were determined based on their ability to select the 32 intervals that contained potential dayside EDRs identified by Webster et al. (2018). The ABS as it is now configured would have selected 31 of the 32 (Webster et al., 2018) 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 resulting ABS tables currently in use would have captured two well-substantiated magnetotail EDR encounters (Torbert et al., 2018; Zhou et al., 2019) 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 empirical models that automate the event classification process using all of the data available to the SITL. Data available to the SITL is of restricted use because it is lower quality than the Level-2 science-quality data freely available to the public. ML models trained on SITL data, such as the MP model described in this manuscript, may not perform as well when applied to Level-2 data. And, vice versa, a model trained using Level-2 data may under-perform if incorporated into the GLS.

The first GLS model (section 5) uses the text description given to each burst selection by the SITL as the ground-truth manual classification for training purposes. To encourage expansion of the GLS, our model development notebooks (Small et al., 2020) can generate additional models simply by changing the text filter (e.g., replace “Magnetopause” with “Dipolarization Front”). 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 (section 7.1), thereby reducing the burden of the SITL and allowing them to spend more time looking for new science.

## 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 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 (section 3.2). The EVA tool can then 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 hr 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 hr, spacecraft memory can hold only about 48 h of burst data. The spacecraft contacts have variable schedules and cannot be optimized just for MMS. The SDC completes the GLS processing within about 2 h 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 plugin for the Space Physics Environment Data Analysis Software (SPEDAS) package (Angelopoulos et al., 2019). 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 onboard MMS. The main functions of EVA are to:

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

**TABLE 2 |** Features used for the development and application of the GLS magnetopause model.

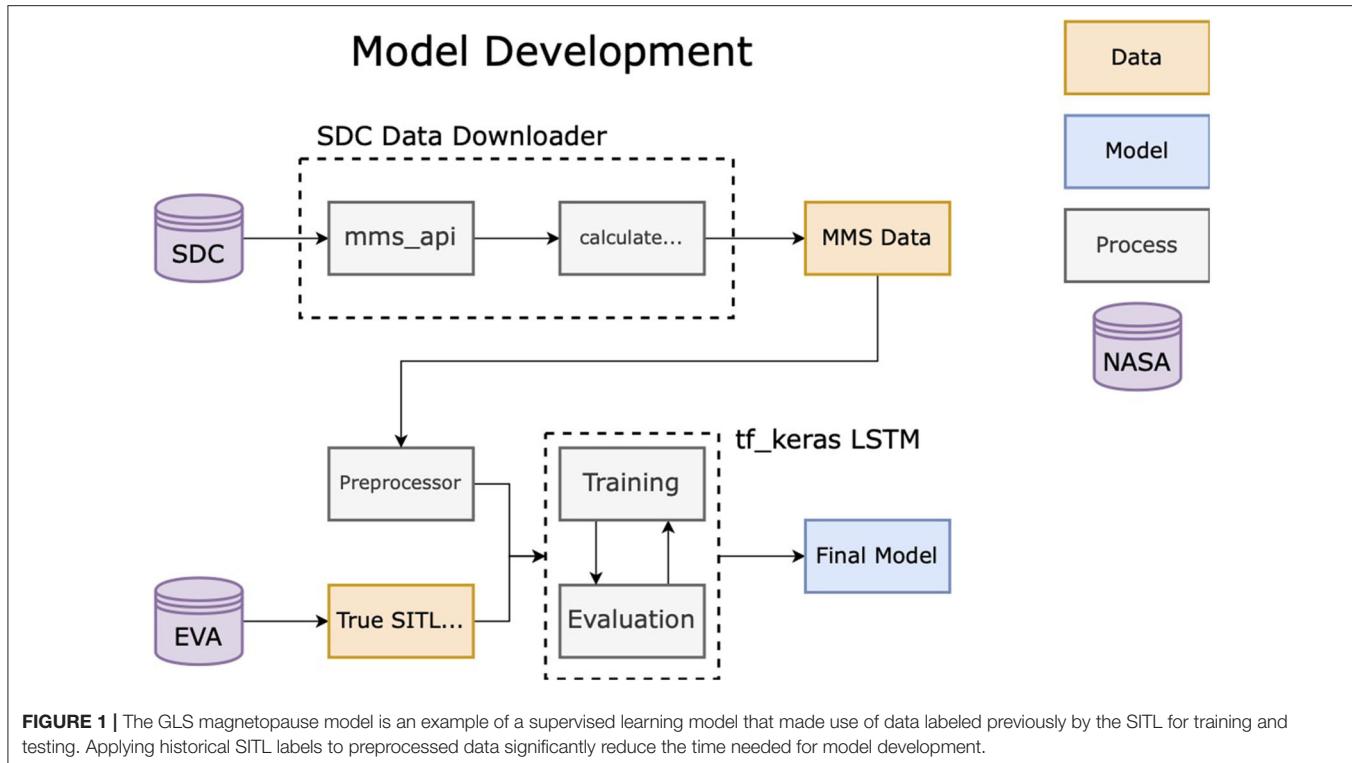
Instr.	Feature	Description
AFG		
1–3	$(B_x, B_y, B_z)$	X, Y, and Z-components of the magnetic field in DMPA coordinates
4	$ B  = \sqrt{B_x^2 + B_y^2 + B_z^2}$	Magnitude of the magnetic field
5	$Q_{\Delta B_x}$	Quality value for $B_x$
6	$P_B =  B ^2/\mu_0$	Magnetic pressure
7	$\theta_C = \arctan(B_y/B_x)$	Clock angle.
EDP		
8–10	$(E_x, E_y, E_z)$	X-, Y-, and Z-component of the DC electric field in DSL coordinates
11	$ E  = \sqrt{E_x^2 + E_y^2 + E_z^2}$	Magnitude of the electric field
DIS		
12–42	$\mathcal{E}_i$	Omni-directional energy spectrogram
43	$N_i$	Number density
44–45	$(V_{i,x}, V_{i,y})$	Bulk velocity in DBCS coordinates
46–47	$(Q_{i,xx}, Q_{i,yy})$	Heat-flux vector in DBCS coordinates
48–49	$T_{i,\parallel}, T_{i,\perp}$	Parallel and perpendicular temperatures
50–55	$\underline{P}_i$	Upper diagonal elements of pressure tensor in DBCS coordinates
55–60	$\underline{\underline{P}}_i$	Upper diagonal elements of pressure tensor in DBCS coordinates
61	$A_i = T_{i,\parallel}/T_{i,\perp} - 1$	Temperature anisotropy
62	$T_i = (T_{i,\parallel} + 2T_{i,\perp})/3$	Scalar temperature
63	$Q_{\Delta N_i}$	Quality value for $N_i$
64	$Q_{\Delta V_{i,z}}$	Quality value for the $V_{i,z}$
65	$Q_{\Delta N_i V_i }$	Quality value for ion ram pressure
DES		
66–97	$\mathcal{E}_e$	Omni-directional energy spectrogram
98	$N_e$	Number density
99–100	$(V_{e,x}, V_{e,y})$	Bulk velocity in DBCS coordinates
101–102	$(Q_{e,xx}, Q_{e,yy})$	Heat-flux vector in DBCS coordinates
103–104	$T_{e,\parallel}, T_{e,\perp}$	Parallel and perpendicular temperatures
105–110	$\underline{P}_e$	Upper diagonal components of pressure tensor in DBCS coordinates
111–116	$\underline{\underline{P}}_e$	Upper diagonal components of pressure tensor in DBCS coordinates
117	$A_e = T_{e,\parallel}/T_{e,\perp} - 1$	Temperature anisotropy
118	$T_e = (T_{e,\parallel} + 2T_{e,\perp})/3$	Scalar temperature
119	$p_e = (P_{e,xx} + P_{e,yy} + P_{e,zz})/3$	Scalar pressure
120	$Q_{\Delta N_e}$	Quality value for $N_e$
121	$Q_{\Delta V_{e,z}}$	Quality value for $V_{e,z}$ velocity
Multiple		
122	$\gamma_1 = T_i/T_e$	Custom feature
123	$\gamma_2 = 2T_i/ E $	Custom feature

Post-processing on the ground allows the raw data to be partially calibrated and expanded into a richer dataset than is available on the spacecraft. The model makes use of most data available to the SITL plus metafeatures that were used as burst triggers for previous missions.

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 (Argall et al., 2020b) 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 (section 5) was trained in this way.

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

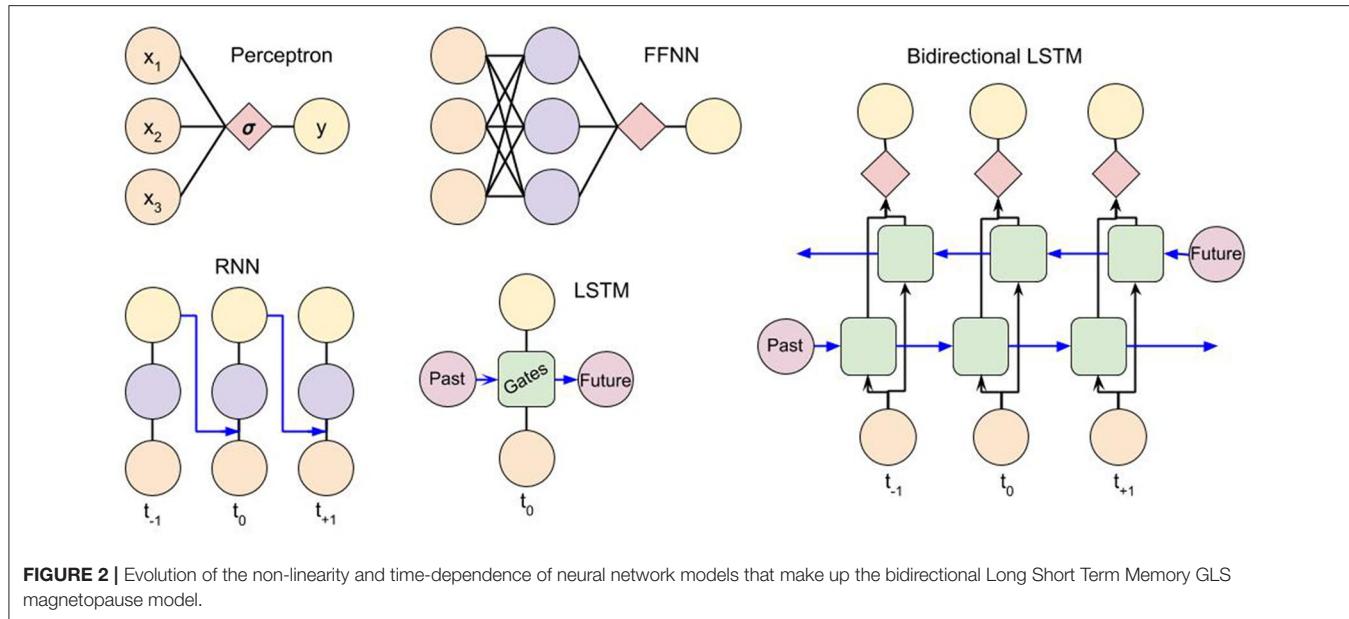
The flow of data used to train the MP model is shown in Figure 1. The SDC provides data from the Analog Fluxgate (AFG) magnetometer (Russell et al., 2014), the Electric field Double Probes (EDP) (Ergun et al., 2014; Lindqvist et al., 2014), and the Fast Plasma Investigation (FPI) Dual Ion Spectrometer (DIS) and Dual Electron Spectrometer (DES) (Pollock et al., 2016). 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 123 features listed in Table 2 were chosen to be inputs into the ML model. Most are standard products from 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  $\gamma$  values, and the  $Q_{\Delta}$  values are metafeatures, features computed from the standard features. The  $Q_{\Delta}$  features are gradient-based trigger terms used on Wind and THEMIS, 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 (Phan et al., 2016). This set of features represents a large portion of the data available to the SITL.

It is worth noting that all features listed in Table 2 are calculated from single spacecraft data; in this case, MMS1. This was done primarily for three reasons: (1) contact times for data downlink are variable so data from multiple spacecraft is not guaranteed, (2) orbit configurations can change; spatial gradients valid in a tetrahedron configuration are not valid in a strong of pearls, and (3) events may occur at different times for different spacecraft, especially in a string of pearls. If the model does not produce satisfactory results for other spacecraft, or if an instrument on MMS1 experiences a problem that invalidates any of the features in Table 2, the model can be retrained.

Data from 1 January 2017 to 30 January 2017 were used to train the model. During this time period, MMS had a single SROI keyed on apogee ( $8 < X_{GSE} < 12 R_E$ ); apogee was at  $12 R_E$  geocentric distance and was located near the subsolar point ( $-6 < Y_{GSE} < 0 R_E$ ). The amount of training data was limited by resources on the platform on which model training was



**FIGURE 2 |** Evolution of the non-linearity and time-dependence of neural network models that make up the bidirectional Long Short Term Memory GLS magnetopause model.

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 fast survey data products, then scaled and regularized. Because of an imbalance between the number of measurements that were selected as MP 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. Of all such sequences, 80% were used for training and 20% were used for testing.

## 5. GLS MAGNETOPAUSE MODEL

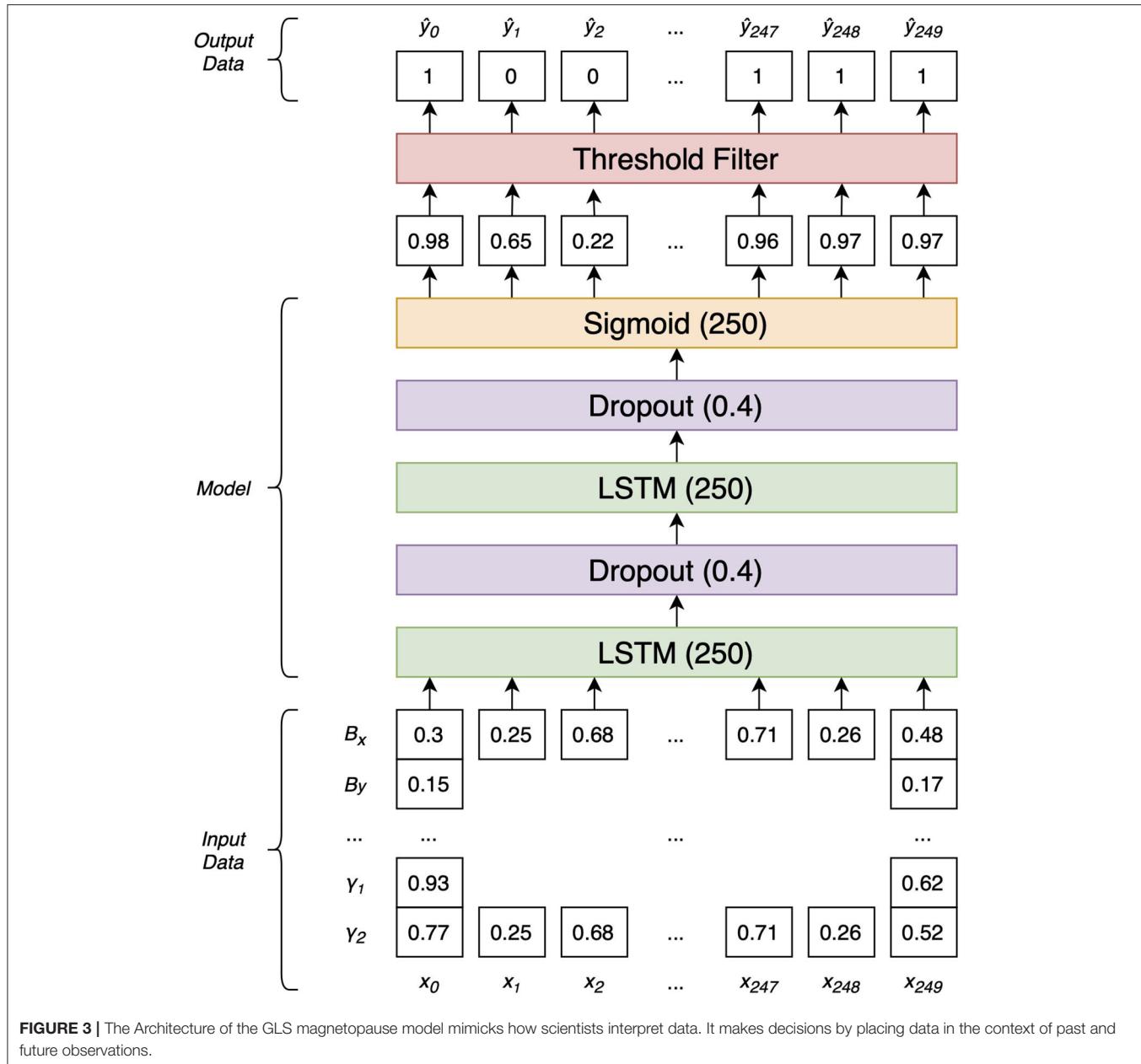
We develop an ML model that aids and automates a key task performed by the SITL. Using the same low-resolution data as the SITLs, we identify time intervals that are likely to contain MP crossings. The ML models are trained using historical data annotated by SITL selections. The input is a low-resolution (4.5 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 4.5 s interval would be selected by the SITL as an MP event.

Our machine learning model is based on neural networks (NNs) (Goodfellow et al., 2016). 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 LSTM layers (Goodfellow et al., 2016), as depicted in **Figure 3**. The output activation functions are hyperbolic tangents 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 overfitting when training the network (Srivastava et al., 2014). The output layer is a single unit with a logistic (sigmoid) activation function. The LSTM’s output is passed through a threshold filter before contiguous segments of selected points are grouped to form selections.



**FIGURE 3 |** The Architecture of the GLS magnetopause model mimicks how scientists interpret data. It makes decisions by placing data in the context of past and future observations.

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 (section 3.3) and is pre-processed (section 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.

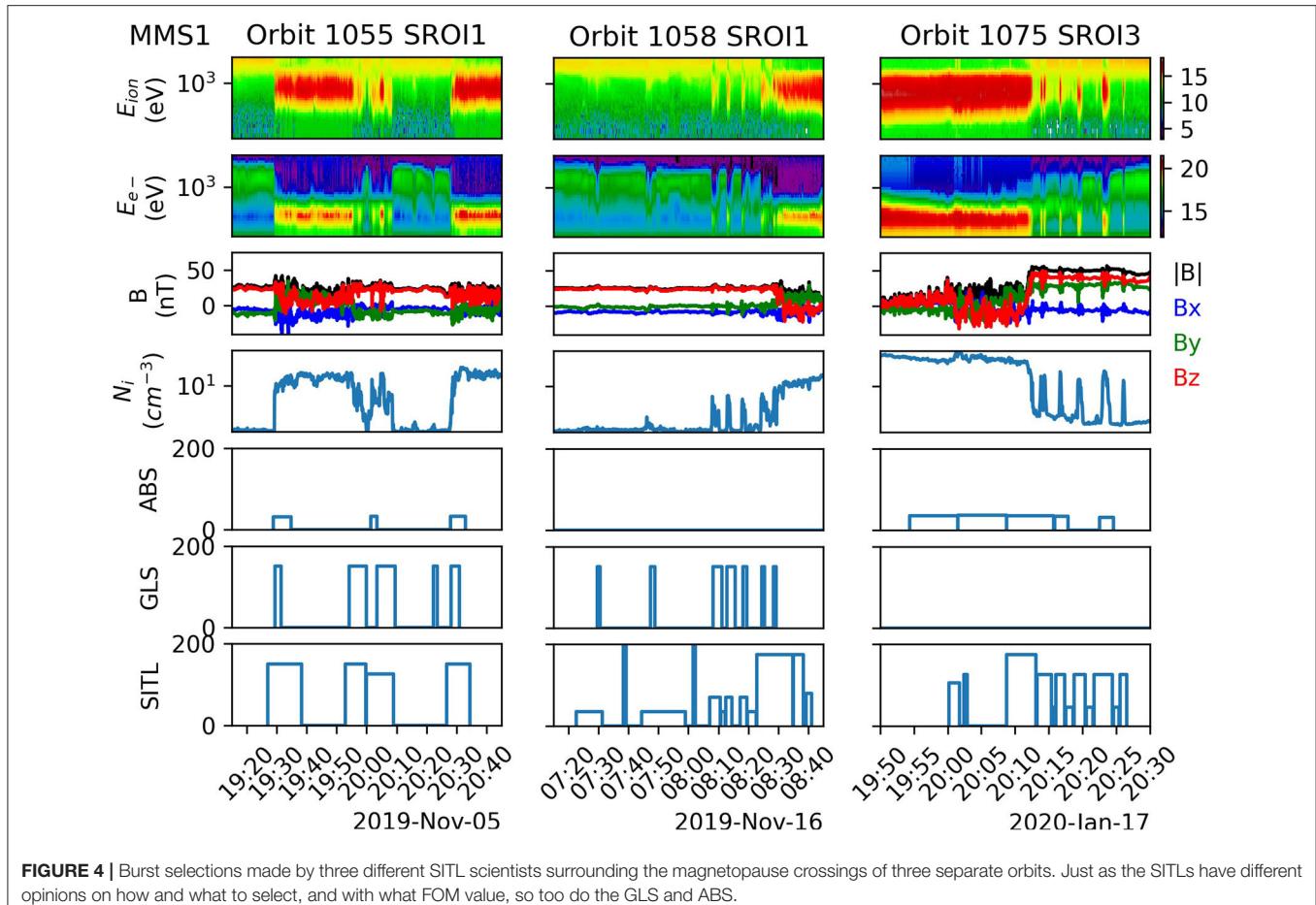
The training and validation data, notebook used to create the model, as well as the weights, scaling parameters, and notebook

to run the model are publicly available (Argall et al., 2020a; Small et al., 2020). Both notebooks have a flag to switch between SITL-quality and science-quality data, and the model creation notebook can easily be modified to generate a new model for any SITL-classified event type. Additional details about the model, threshold filter, and the hardware the model was trained on can be found in the **Supplementary Material**.

## 6. RESULTS

### 6.1. Case Studies

On 19 October 2019, the model was installed and executed at the SDC for the first time and has been providing guidance to the SITL ever since. By this time, MMS apogee was near



**FIGURE 4 |** Burst selections made by three different SITL scientists surrounding the magnetopause crossings of three separate orbits. Just as the SITLs have different opinions on how and what to select, and with what FOM value, so too do the GLS and ABS.

25  $R_E$  geocentric distance and had transitioned from one SROI to three SROIs to capture the inbound and outbound magnetopause crossings (SROI1 and SROI3), and a segment of the solar wind (SROI2). **Figure 4** shows MP crossings from SROI1 on orbits 1055 (left) and 1058 (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 (panels 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 in SROI1 and inward from the MSH to the MSP in SROI3. During the transition, the MSH and MSP plasmas are observed simultaneously,  $B_z$  (panel 3) often changes sign, and the density (panel 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 panel while the SITL-provided description of each selection is given in **Table 3**. Panels 5 and 6 show similarities and differences between the time intervals and FOM values of selections made by the SITL and those made by the ABS and GLS.

Orbits 1055 and 1058 SROI1, and orbit 1075 SROI3 were chosen because of the presence or absence of GLS and ABS selections. Orbit 1055 SROI1 consisted of three full and one partial MP crossings as the MP moved back and forth over the spacecraft during the 1.5 hour interval shown. Both the ABS

and GLS made selections similar to the SITL, but with notable differences. First, whereas the SITL selected a large portion of the MSP and MSH on either side of the MP to provide relevant contextual information, the ABS and GLS selections were more focused on the MP transition. In this case, they are under-selecting when compared to the SITL, but are correctly identifying the MP (i.e., they are not classifying the surrounding MSP and MSH as the MP like the SITL did). Second, the GLS makes a selection at  $\sim 2,022$  that the SITL does not. This selection represents a typical false-positive for the GLS despite the fact that the interval is MP-like in that the MSP and MSH plasmas are observed simultaneously and the plasma density is higher than the MSP proper. So, while differences exist between the SITL, ABS, and GLS, those differences are often subtle, and similar differences also exist between selections made by different SITL scientists. We consider the SITL to be the ground truth for the sake of model validation; however, operationally the GLS acts as a co-SITL with its own suggestions for what should be selected.

During orbit 1058 SROI1, the SITL selected intervals containing cold ions ( $\sim 0722$ – $0722$ ), the boundary layer ( $\sim 0744$ – $0820$ ), an MP crossing ( $\sim 0823$ – $0845$ ), and a flux transfer event ( $\sim 0839$ – $0841$ ; see **Table 3**). None of these intervals were selected by the ABS but the GLS captured most of them. In particular, the intervals marked “cold ions” and “boundary layer”/“BL” were

**TABLE 3 |** Selections made by the SITL during the intervals shown in **Figure 4**.

Orbit	Date	Start	End	FOM	Description
1055	2019-11-05	19:27:03	19:38:23	150	Lower shear full magnetopause crossing with Vz flow reversal
1055	2019-11-05	19:52:53	19:59:43	150	Full lower shear magnetopause crossing
1055	2019-11-05	19:59:53	20:09:03	125	Partial magnetopause crossings with deep B-minima
1055	2019-11-05	20:26:43	20:34:33	150	Full low-shear
1058	2019-11-16	07:22:33	07:31:23	35	Cold ions
1058	2019-11-16	07:38:23	07:39:13	200	FPI Burst Cal - Segment 2 (1058) - H2 - MSP
1058	2019-11-16	07:44:23	07:59:03	35	Boundary layer
1058	2019-11-16	08:01:33	08:02:23	200	FPI Burst Cal - Segment 3 (1058) - H2 - MSP
1058	2019-11-16	08:07:13	08:10:43	70	BL Traversals
1058	2019-11-16	08:10:53	08:12:23	35	Additional context between BL traversals
1058	2019-11-16	08:12:33	08:14:33	70	BL Traversals
1058	2019-11-16	08:17:13	08:19:33	70	BL Traversals
1058	2019-11-16	08:19:43	08:22:43	35	Additional context between boundary layer traversal and MP Crossing
1058	2019-11-16	08:22:53	08:34:53	175	Full MP potentially with a jet
1058	2019-11-16	08:35:03	08:38:23	175	Continuation of FULL MP
1058	2019-11-16	08:38:33	08:39:03	35	Additional context between MP and FTE
1058	2019-11-16	08:39:13	08:41:03	80	FTE
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 magnetosheath 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

"Partial", "low-shear", "high-shear", "full", and "complete" refer to classes of MP crossings 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, "FTE" = flux transfer event.

selected by the GLS because they show signs of mixing of MSP and MSH plasmas, as occurs at the MP. Such selections are similar in nature to the false positive shown in orbit 1055 SROI1.

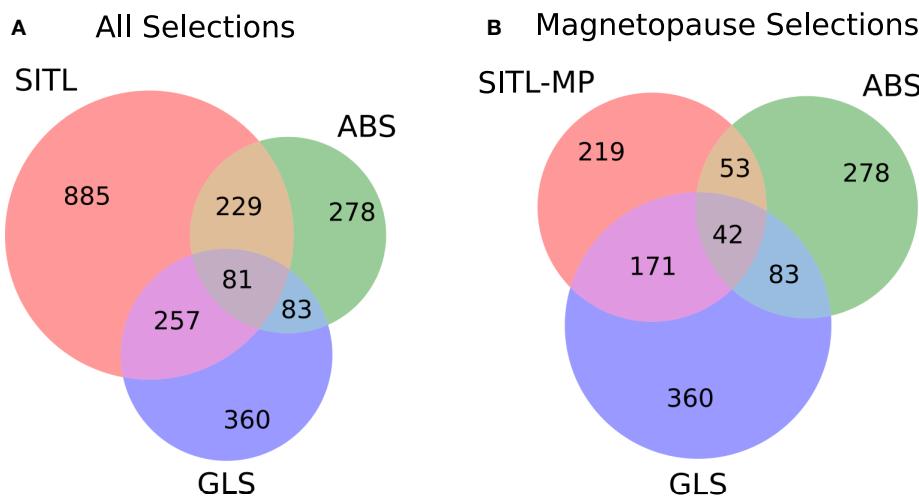
Orbit 1075 SROI3 is one in which the SITL and ABS select the MP but the GLS does not. The SITL selects a complete, high-shear MP crossing at 2010 UT followed by several partial crossings (**Table 3**), as well as reconnection-like signatures in the MSH near 2000 UT. The ABS also selects the high-shear 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 the training and validation sets need to be expanded to include data from a time period when MMS had three SROIs.

## 6.2. Statistical Study

To more broadly assess model performance, we make comparisons between the GLS, SITL, and ABS for all selections made in SROI1 between 19 October 2019 and 25 March 2020. **Figure 5** is a Venn diagram depicting (a) the number of GLS and ABS selections that have at least partial overlap with all SITL selections and (b) only those SITL selections that were identified as MP crossings. More detailed histograms showing the two-way overlap between SITL and GLS, SITL and ABS, and GLS and ABS are included in **Figures S1–S3** for SROI1, SROI3, and SROIs1 and 3, respectively. Such comparisons take into account partial- and multiple-overlaps between SITL and GLS selections, something not possible with the more traditional precision, recall, and F1-score metrics (Tatbul et al., 2018) presented in section 6.3.

### 6.2.1. SITL-GLS Comparison

Most selections made by the GLS are of interest to the SITL, but the GLS is selecting more than just MP crossings. Of the 360



**FIGURE 5 |** A comparison of SITL, GLS, and ABS segments from SROI1 that takes into account the range-based nature of event selection intrinsic to time series data. **(A)** Considers all SITL selections whereas **(B)** includes only SITL-classified magnetopause selections.

GLS selections, 257 (71%) were also selected by the SITL and 176 (49%) were classified as MP crossings by the SITL. At first glance, the latter might seem low for a model that was trained to select MP crossings. This can be explained partly because the SITL is aware of external control factors such as telemetry restrictions, and partly because some GLS segments are MP-like but are not classified as MP crossings by the SITL, as was the case for segments in orbits 1055 and 1058 SROI1 (**Figure 4**). Such selections include intervals that exhibit plasma mixing between the MSP and MSH. They also include bow shock crossings, which have field and plasma gradients similar to those present during MP crossings. While not MP crossings, these extra selections made by the GLS are still of interest to the SITL, as indicated by the larger number of overlapping segments when all of the SITL selections are included.

Again, because of the possibility of partial and multiple overlaps in time series selections, the inverse relationship relating the number SITL selection also selected by the GLS can only be qualitatively inferred (for a quantitative comparison, see **Figures S1–S3**). On one hand, the GLS selections represent only a small percentage (30%) of all SITL selections. On the other hand, the majority of SITL MP crossings (78%) are selected by the GLS. Machine learning models, therefore, are an effective means of automating subsets of SITL functions outlined in the SITL guidelines. Later (section 7.1), we discuss how multiple GLS models can be combined to assimilate more of the manual classification tasks.

### 6.2.2. SITL-ABS Comparison

Next, we compare the ABS to all SITL selections (**Figure 5A**) and to only those SITL selections that were classified as MP crossings (**Figure 5B**). The SITL selects 229 of 278 ABS segments, a larger percentage (82 vs. 71%) than GLS segments; however, only 19% of ABS segments (59 of 278) were classified as MP by the SITL. Conversely, only 28% of SITL selections and 34% of

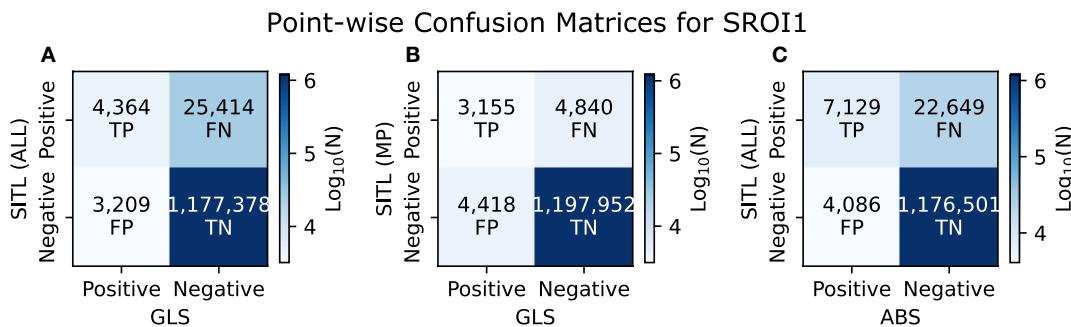
MP crossings were selected by the ABS (**Figures S1E,K**). While a majority of ABS selections are of interest to the SITL, the ABS is significantly under-selecting both in general and with respect to MP crossings.

The differences between the ABS and GLS are most likely due to how they were trained. The ABS was trained to select MP crossings that contain EDRs, which typically exhibit larger amplitude variations than MP crossings that do not. Non-EDR time intervals that exhibit large amplitude variations are still of interest to the SITL, but most MP crossings do not exhibit such activity (e.g., the MP crossings in Orbit 1058 shown in **Figure 4**), meaning they are not selected by the ABS.

### 6.2.3. ABS-GLS Comparison

We have been incorporating the ABS into the discussion so far because it is representative of how most other missions with burst memory management systems select data. Here, we compare it to the new GLS machine learning model. While the two systems were trained differently and for slightly different purposes, comparing them may provide a general impression of (a) the efficacy of a linear combination of summary data (TDNs) vs. a non-linear combination of a more robust dataset (survey data), and (b) models trained for a specific task vs. a potential catch-all model. A comparison may also be influential to future mission designs.

The Venn Diagram shows that only a small fraction of GLS segments were selected by the ABS (83 of 360, or 23%). Of those, nearly all (81) were selected by the SITL, but only about half (43) were classified as MP crossings. The GLS selects a similarly low fraction of ABS segments (30%), but of the 53 ABS segments that were also selected by the SITL as MP crossings, the GLS selected 42 of them (**Figures S1F,L**). So although both the GLS and ABS under-select compared to all SITL selections, they are not redundant; they each make useful, complementary selections that are highly relevant to the SITL.



**FIGURE 6 |** Point-by-point comparison between **(A)** all SITL and GLS selections, **(B)** SITL selections filtered for magnetopause crossings and GLS selections, and **(C)** all SITL and ABS selections. Such a point-wise metric is typical for machine learning models but does not properly weight predictions with partial or multiple overlap, as when the SITL selects additional context around a given event (**Figure 4**, Orbit 1055), or when multiple GLS selections are encompassed by one SITL selection (**Figure 4**, Orbit 1058).

### 6.3. Performance Metrics

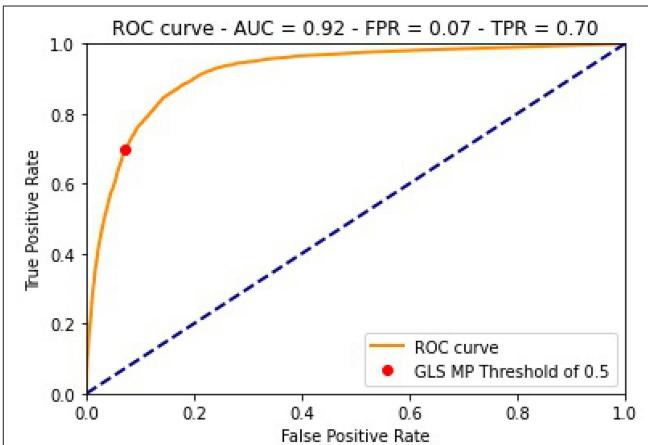
#### 6.3.1. Precision, Recall, and F1 Score

Performance of the MP model is directly tied to the threshold filter described in section 5. The filter turns the prediction into a binary classifier, where 1 indicates that the observations are from the MP and 0 indicates otherwise. If the model predicts an MP crossing (or not) that was actually classified as a MP crossing by the SITL, this is known as a true positive (false negative). Conversely, if the model classifies the observations as MP (or not) and the SITL does not, this is known as false positive (true negative). Such labels were determined for predicted GLS segments using all SITL selections and only those selections classified as MP crossings, as well as for ABS segments. They are shown in the form of a confusion matrix in **Figure 6** for SROI1 during the same date range as covered by **Figure 5**.

Results from the confusion matrix can be summarized by the precision, recall, and F1 scores. Precision is the fraction of all positive predictions that were correctly classified, while recall is the fraction of all actual positive cases that were correctly predicted and gives a sense of the number of cases missed by the model. The F1 score is a measure that captures the properties of both precision and recall.

The precision, recall, and F1 score for the GLS were (0.58, 0.15, 0.23) for all SITL selections and (0.42, 0.39, 0.41) for just the MP SITL selections. The GLS precision is higher and its recall is lower when compared to all SITL selections, alluding to the MP-like selections that were not classified as MP encounters by the SITL, as mentioned in relation to **Figure 4**. From recall, we infer that large fraction of MP points are left unselected by the GLS. This is due, in part, to the fact that the SITL is selecting contextual information than the GLS is not (again, see **Figure 4**). The F1 score is higher when only MP selections are considered, reflecting the better match between the model and the data considered.

For the ABS, the precision, recall, and F1 score were (0.64, 0.24, 0.35). High precision and low recall indicates that most ABS selections are important to the SITL, but that the SITL is selecting much more than the ABS. Similar conclusions were deduced



**FIGURE 7 |** Trade-off between over-selecting false events and under-selecting true events. For MMS, electron diffusion regions are rare and difficult to impossible to observe with the SITL data, so the SITL is willing to over-select by choosing all MP crossings, but they still has to contend with telemetry restrictions. The choice of thresholds for the GLS MP model tries to emulate this approach.

from the Venn diagram in **Figure 5**. The F1 score is better than that of the GLS when all selections are considered, but lower than when the GLS is compared to only MP points.

#### 6.3.2. ROC Curve

The ability of a model to distinguish between positive and negative cases is indicated by the Receiver Operating Characteristic (ROC) curve, which plots recall (the true positive rate) against the false positive rate, or the number of false positives out of all actual negative cases, for a variety of threshold values. If the area under the ROC curve is 1.0, the positive and negative cases are perfectly distinguishable by the model. If it is 0.5, the model has no ability to distinguish between positive and negative cases. The ROC curve from the validation data of

the GLS MP model is shown in **Figure 7**. It has an area under the curve of 0.92, indicating it can tell MP data from non-MP data accurately.

Implementing the model requires a choice in threshold values that involves a trade-off between true positive and false negative rates. The MP model uses a threshold value of 0.5, resulting in a true positive rate of 0.7 and a false positive rate of 0.07. Comparing to the confusion matrix results in **Figures 6A,B**, both the model recall and false positive rates are lower than for the validation set. The model selects fewer SITL-classified points (MP or otherwise), but also makes significantly fewer false predictions than for the validation set.

## 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 trained on all SITL selections. However, as the SITL rotates and mission priorities change, model performance would suffer. A better approach would be to create a hierarchy of classification models, as described in section 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 the model 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 orbits 1055 and 1058 SROI1 in **Figure 4**. 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.

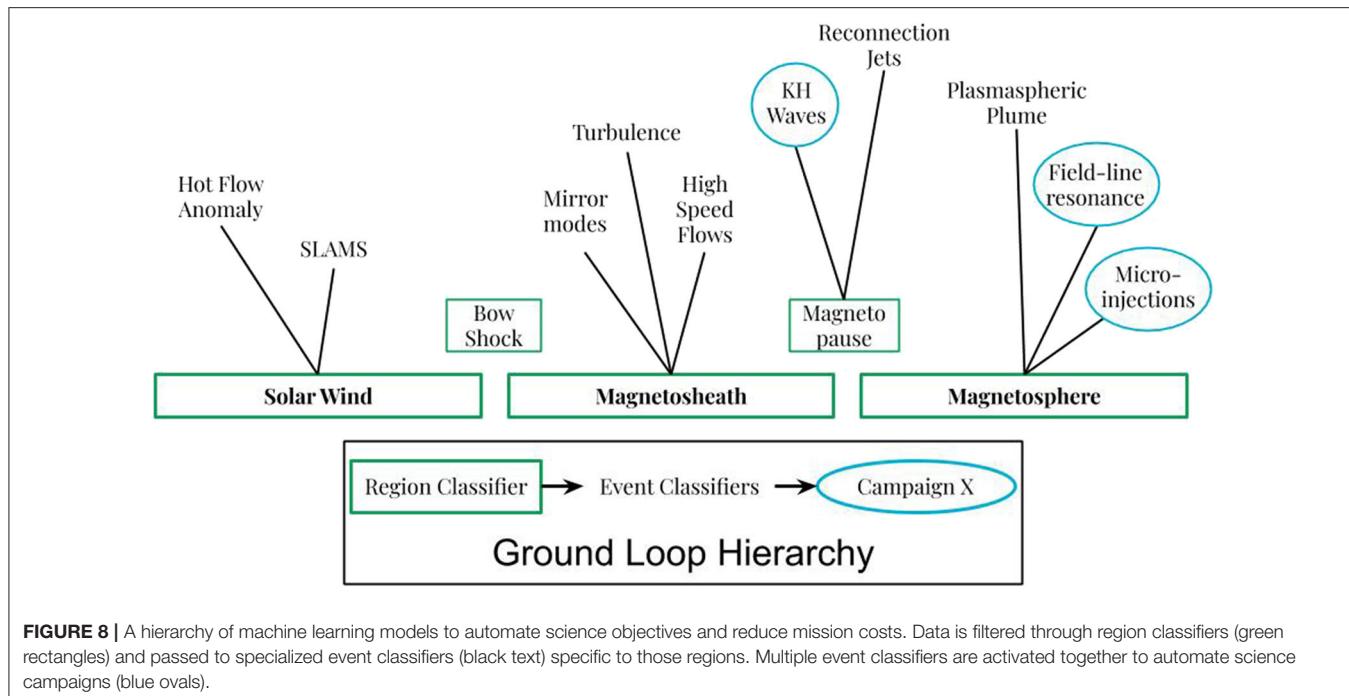
Outside of the science considerations are operational control factors (Item 4), such as the amount of on-board memory available to store selections. Such considerations often influence the number of selections that the SITL makes. Once GLS selections are made (Items 1 and 3), they can be passed through 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 h of a low-shear, slow MP crossing could potentially overwrite many other selections.

The GLS is an example of progressive autonomy (Truszkowski et al., 2005). It follows similar efforts undertaken by NASA to reduce mission costs through greater autonomy in ground control and spacecraft operations (Truszkowski et al., 2006). Autonomy can alleviate mission complexity and provide real-time decision making when communications latency exists (Truszkowski et al., 2004). 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? We propose the 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, 2015) For the GLS, this means training a hierarchy of models, as shown in **Figure 8**. At its lowest level, The Hierarchy consists of region classifiers that segregate 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 (Fennell et al., 2016). Simulations proposed that Kelvin-Helmholtz waves (KHWs) on the MP surface were the cause (Kavosi et al., 2018). To gain more insight into this multi-scale process, one could create event classifiers for micro-injections (Claudepierre et al., 2020), 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 (Nguyen et al., 2019; Olshevsky et al., 2019; Piatt, 2019; da Silva et al., 2020), and one is being adapted for that purpose (Piatt, 2019). The LSTM RNN model described above could serve as either a region or an event classifier. Other event classifiers have been developed using MMS data (e.g., Claudepierre et al., 2020), 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 (Argall et al., 2020a,b,c; Small et al., 2020) 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 78% of SITL-identified MP crossings in the outbound leg of its orbit, 44% more 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.

## DATA AVAILABILITY STATEMENT

The datasets analyzed for this study, including the Level 2 data products plotted in Figure 4 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 PyMMS (Argall et al., 2020b) or the MMS-Plugin for SPEDAS (Angelopoulos et al., 2019). Notebooks to recreate the tables and figures presented in this article (Argall et al., 2020c), to train and run the GLS MP model (Small et al., 2020), and the training and validation data, model weights, and scaling factors (Argall et al., 2020a) are also available through the open access repository Zenodo.

## AUTHOR CONTRIBUTIONS

MA, CS, SP, LB, and MP developed machine learning models and analyzed data. JB, KK, and KL developed the processes at the MMS SDC to generate, store, and distribute MMS data and burst selections. RE, FW, and MO wrote the SPEDAS and EVA software, with contributions from many of the other authors. WP determined the weights and offsets for the ABS. RT, RE, TP, BG, JLB, FW, and MO are super-SITLs responsible for the management of the MMS SITL infrastructure. MA, CS, MP, MO, FW, KK, and WP contributed to the

writing of the paper. All authors helped build the ground loop infrastructure.

## FUNDING

This work was supported by NASA grant 80NSSC19K1203 and contract NNG04EB99C. MP was funded by NSF RI 1815275.

## ACKNOWLEDGMENTS

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

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspas.2020.00054/full#supplementary-material>

## REFERENCES

- Angelopoulos, V., Cruce, P., Drozdov, A., Grimes, E. W., Hatzigeorgiu, N., King, D. A., et al. (2019). The space physics environment data analysis system (SPEDAS). *Space Sci. Rev.* 215:9. doi: 10.1007/s11214-018-0576-4
- Argall, M. R., Small, Colin, R., and Petrik, M. (2020c). MMS SITL Ground Loop: Software to Reproduce Figures and Tables.
- Argall, M. R., Small, C. R., and Petrik, M. (2020a). MMS SITL Ground Loop: Data for the GLS-MP Magnetopause Model.
- Argall, M. R., Small, C. R., and Petrik, M. (2020b). PyMMS: A Python Library for NASA's Magnetospheric Multiscale (MMS) Mission.
- Aunai, N., Hesse, M., and Kuznetsova, M. (2013). Electron nongyrotropy in the context of collisionless magnetic reconnection. *Phys. Plasmas* 20, 1–6. doi: 10.1063/1.4820953
- Baker, D. N., Riesberg, L., Pankratz, C. K., Panneton, R. S., Giles, B. L., Wilder, F. D., et al. (2016). Magnetospheric multiscale instrument suite operations and data system. *Space Sci. Rev.* 199, 545–575. doi: 10.1007/s11214-014-0128-5
- Bhaskar, A., and Vichare, G. (2019). 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:A12. doi: 10.1051/swsc/2019007
- Boardsen, S. A., Eastman, T. E., Sotirelis, T., and Green, J. L. (2000). An empirical model of the high-latitude magnetopause. *J. Geophys. Res.* 105, 23193–23219. doi: 10.1029/1998JA000143
- Borovsky, J. E. (2014). Canonical correlation analysis of the combined solar wind and geomagnetic index data sets. *J. Geophys. Res.* 119, 5364–5381. doi: 10.1002/2013JA019607
- Boubrahimi, S. F., Aydin, B., Martens, P., and Angryk, R. (2017). "On the prediction of >100 MeV solar energetic particle events using GOES satellite data," in *2017 IEEE International Conference on Big Data (Big Data)* (Boston, MA), 2533–2542. doi: 10.1109/BigData.2017.8258212
- Burch, J. L., Torbert, R. B., Phan, T. D., Chen, L.-J., Moore, T. E., Ergun, R. E., et al. (2016). Electron-scale measurements of magnetic reconnection in space. *Science* 352, 1–9. doi: 10.1126/science.aaf2939
- Claudepierre, S., Fennell, J., Sengupta, A., and Kletzing, C. (2020). Automated identification of electron microinjections in MMS/FEEPS measurements: initial results. *Earth Space Sci. Open Arch.* 10501706. doi: 10.1002/essoar.10501706.1
- da Silva, D., Barrie, A., Shuster, J., Schiff, C., Attie, R., Gershman, D. J., and Giles, B. (2020). Automatic region identification over the MMS orbit by partitioning n-T space. *arxiv [Preprint]* arXiv:2003.08822. Available online at: <https://arxiv.org/abs/2003.08822>
- Dungey, J. (1961). Interplanetary magnetic field and the auroral zones. *Phys. Rev. Lett.* 6, 47–48. doi: 10.1103/PhysRevLett.6.47
- Ergun, R. E., Tucker, S., Westfall, J., Goodrich, K. A., Malaspina, D. M., Summers, D., et al. (2014). The axial double probe and fields signal processing for the MMS mission. *Space Sci. Rev.* 199, 167–188. doi: 10.1007/s11214-014-0115-x
- Fear, R. C., Palmroth, M., and Milan, S. E. (2012). Seasonal and clock angle control of the location of flux transfer event signatures at the magnetopause. *J. Geophys. Res.* 117, 1–15. doi: 10.1029/2011JA017235
- Fennell, J. F., Turner, D. L., Lemon, C. L., Blake, J. B., Clemons, J. H., Mauk, B. H., et al. (2016). Microinjections observed by MMS FEEPS in the dusk to midnight region. *Geophys. Res. Lett.* 43, 6078–6086. doi: 10.1002/2016GL069207
- Fuselier, S. A., Lewis, W. S., Schiff, C., Ergun, R., Burch, J. L., Petrinec, S. M., et al. (2016). Magnetospheric multiscale science mission profile and operations. *Space Sci. Rev.* 199, 77–103. doi: 10.1007/s11214-014-0087-x
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Adaptive Computation and Machine Learning Series. Cambridge, MA: MIT Press.
- Hesse, M., Aunai, N., Sibeck, D., and Birn, J. (2014). On the electron diffusion region in planar, asymmetric, systems. *Geophys. Res. Lett.* 41, 8673–8680. doi: 10.1002/2014GL061586
- Jelínek, K., Néemeček, Z., and Šafránková, J. (2012). A new approach to magnetopause and bow shock modeling based on automated region identification. *J. Geophys. Res.* 117, 1–8. doi: 10.1029/2011JA017252
- Jian, L. K., Russell, C. T., Luhmann, J. G., Curtis, D., and Schroeder, P. (2013). Burst mode trigger of STEREO in situ measurements. *AIP Conf. Proc.* 1539, 195–198. doi: 10.1063/1.4811021
- Kavosi, S., and Raeder, J. (2015). Ubiquity of Kelvin?Helmholtz waves at Earth's magnetopause. *Nat. Commun.* 6:7019. doi: 10.1038/ncomms8019
- Kavosi, S., Spence, H. E., Fennell, J. F., Turner, D. L., Connor, H. K., and Raeder, J. (2018). MMS/FEEPS observations of electron microinjections due to kelvin-helmholtz waves and flux transfer events: a case study. *J. Geophys. Res.* 123, 5364–5378. doi: 10.1029/2018JA025244
- Lindqvist, P.-A., Olsson, G., Torbert, R. B., King, B., Granoff, M., Rau, D., et al. (2014). The spin-plane double probe electric field instrument for MMS. *Space Sci. Rev.* 199, 137–165. doi: 10.1007/s11214-014-0116-9
- Lundstedt, H. (1997). Solar wind magnetosphere coupling: predicted and modeled with intelligent hybrid systems. *Phys. Chem. Earth* 22, 623–628. doi: 10.1016/S0079-1946(97)00187-0
- Matsui, H., Farrugia, C. J., Goldstein, J., Torbert, R. B., Argall, M. R., Vaith, H., et al. (2019). Velocity rotation events in the outer magnetosphere near the magnetopause. *J. Geophys. Res.* 124, 4137–4156. doi: 10.1029/2019JA026548

- Mozer, F. S. (2005). Criteria for and statistics of electron diffusion regions associated with subsolar magnetic field reconnection. *J. Geophys. Res.* 110, 1–10. doi: 10.1029/2005JA011258
- Nagai, T., Shinohara, I., Fujimoto, M., Matsuoka, A., Saito, Y., and Mukai, T. (2011). Construction of magnetic reconnection in the near-Earth magnetotail with Geotail. *J. Geophys. Res.* 116, 1–18. doi: 10.1029/2010JA016283
- Nagai, T., Zenitani, S., Shinohara, I., Nakamura, R., Fujimoto, M., Saito, Y., et al. (2013). Ion and electron dynamics in the ion-electron decoupling region of magnetic reconnection with Geotail observations. *J. Geophys. Res.* 118, 7703–7713. doi: 10.1002/2013JA019135
- Nguyen, G., Aunai, N., de Welle, B., Jeandet, A., and Fontaine, D. (2019). Automatic detection of the Earth Bow Shock and Magnetopause from in-situ data with machine learning. *Ann. Geophys. Discuss.* 2019, 1–22. doi: 10.5194/angeo-2019-149
- Oka, M., Phan, T.-D., Oieroset, M., and Angelopoulos, V. (2016). *In situ* evidence of electron energization in the electron diffusion region of magnetotail reconnection. *J. Geophys. Res.* 121, 1955–1968. doi: 10.1002/2015JA022040
- Olshevsky, V., Khotyaintsev, Y.-V., Divin, A., Delzanno, G. L., Anderzen, S., Herman, P., et al. (2019). Automated classification of plasma regions using 3D particle energy distribution. *arXiv. arXiv* 1908.05715v2.
- Paschmann, G., Baumjohann, W., Sckopke, N., Phan, T.-D., and Lühr, H. (1993). Structure of the dayside magnetopause for low magnetic shear. *J. Geophys. Res.* 98, 13409–13422. doi: 10.1029/93JA00646
- Phan, T. D., Eastwood, J. P., Cassak, P. A., Oieroset, M., Gosling, J. T., Gershman, D. J., et al. (2016). MMS observations of electron-scale filamentary currents in the reconnection exhaust and near the X line. *Geophys. Res. Lett.* 43, 6060–6069. doi: 10.1002/2016GL069212
- Phan, T. D., and Paschmann, G. (1996). Low-latitude dayside magnetopause and boundary layer for high magnetic shear: 1. Structure and motion. *J. Geophys. Res.* 101, 7801–7815. doi: 10.1029/95JA03752
- Phan, T. D., Shay, M. A., Eastwood, J. P., Angelopoulos, V., Oieroset, M., Oka, M., and Fujimoto, M. (2015). Establishing the context for reconnection diffusion region encounters and strategies for the capture and transmission of diffusion region burst data by MMS. *Space Sci. Rev.* 199, 631–650. doi: 10.1007/s11214-015-0150-2
- Piatt, S. (2019). Large-scale statistical survey of magnetopause reconnection. [*arXiv e-prints*] *arXiv:1905.11359*. Available online at: <https://arxiv.org/abs/1905.11359>
- Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., et al. (2016). Fast plasma investigation for magnetospheric multiscale. *Space Sci. Rev.* 199, 331–406. doi: 10.1007/s11214-016-0245-4
- Rascoff, S., and Humphries, S. (2015). *Zillow Talk: The New Rules of Real Estate*. New York, NY: Grand Central Publishing.
- Russell, C. T., Anderson, B. J., Baumjohann, W., Bromund, K. R., Dearborn, D., Fischer, D., et al. (2014). The magnetospheric multiscale magnetometers. *Space Sci. Rev.* 199, 189–256. doi: 10.1007/s11214-014-0057-3
- Scudder, J., and Daughton, W. (2008). “Illuminating” electron diffusion regions of collisionless magnetic reconnection using electron agyrotropy. *J. Geophys. Res.* 113, 1–16. doi: 10.1029/2008JA013035
- Scudder, J. D., Holdaway, R. D., Daughton, W. S., Karimabadi, H., Roytershteyn, V., Russell, C. T., et al. (2012). First resolved observations of the demagnetized electron-diffusion region of an astrophysical magnetic-reconnection site. *Phys. Rev. Lett.* 108:225005. doi: 10.1103/PhysRevLett.108.225005
- Scudder, J. D., Holdaway, R. D., Glassberg, R., and Rodriguez, S. L. (2008). Electron diffusion region and thermal demagnetization. *J. Geophys. Res.* 113, 1–14. doi: 10.1029/2008JA013361
- Small, C. R., Argall, M. R., and Petrik, M. (2020). MMS SITL Ground Loop: Notebooks to Train and Run the GLS-MP Model.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958. doi: 10.5555/2627435.2670313
- Swisdak, M. (2016). Quantifying gyrotropy in magnetic reconnection. *Geophys. Res. Lett.* 43, 43–49. doi: 10.1002/2015GL066980
- Tang, X., Cattell, C., Dombeck, J., Dai, L., Wilson, L. B., Breneman, A., et al. (2013). THEMIS observations of the magnetopause electron diffusion region: large amplitude waves and heated electrons. *Geophys. Res. Lett.* 40, 2884–2890. doi: 10.1002/grl.50565
- Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., and Gottschlich, J. (2018). “Precision and recall for time series,” in *Advances in Neural Information Processing Systems 31*, eds S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Red Hook, NY: Curran Associates, Inc.), 1920–1930.
- Torbert, R. B., Burch, J. L., Phan, T. D., Hesse, M., Argall, M. R., Shuster, J., et al. (2018). Electron-scale dynamics of the diffusion region during symmetric magnetic reconnection in space. *Science* 362, 1391–1395. doi: 10.1126/science.aat2998
- Torbert, R. B., Russell, C. T., Magnes, W., Ergun, R. E., Lindqvist, P.-A., Le Contel, O., et al. (2016). The FIELDS instrument suite on MMS: scientific objectives, measurements, and data products. *Space Sci. Rev.* 199, 105–135. doi: 10.1007/s11214-014-0109-8
- Truszkowski, W., Hinckley, M., Rash, J., and Rouff, C. (2004). NASA’s swarm missions: the challenge of building autonomous software. *IT Prof.* 6, 47–52. doi: 10.1109/MITP.2004.66
- Truszkowski, W., Rouff, C., Bailin, S., and Riley, M. (2005). Progressive autonomy: a method for gradually introducing autonomy into space missions. *Innovat. Syst. Softw. Eng.* 1, 89–99. doi: 10.1007/s11334-005-0018-9
- Truszkowski, W. F., Hinckley, M. G., Rash, J. L., and Rouff, C. A. (2006). Autonomous and autonomic systems: a paradigm for future space exploration missions. *IEEE Trans. Syst. Man Cybern. Part C* 36, 279–291. doi: 10.1109/TSMCC.2006.871600
- Webster, J. M., Burch, J. L., Reiff, P. H., Daou, A. G., Genestreti, K. J., Graham, D. B., et al. (2018). Magnetospheric multiscale dayside reconnection electron diffusion region events. *J. Geophys. Res.* 123, 4858–4878. doi: 10.1029/2018JA025245
- Wing, S., Johnson, J. R., Chaston, C. C., Echim, M., Escoubet, C. P., Lavraud, B., et al. (2014). Review of solar wind entry into and transport within the plasma sheet. *Space Sci. Rev.* 184, 33–86. doi: 10.1007/s11214-014-0108-9
- Wintoft, P., Wik, M., and Viljanen, A. (2015). Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *J. Space Weather Space Clim.* 5:A7. doi: 10.1051/swsc/2015008
- Zenitani, S., Hesse, M., Klimas, A., Black, C., and Kuznetsova, M. (2011). The inner structure of collisionless magnetic reconnection: the electron-frame dissipation measure and Hall fields. *Phys. Plasmas* 18:122108. doi: 10.1063/1.3662430
- Zhou, M., Deng, X. H., Zhong, Z. H., Pang, Y., Tang, R. X., El-Alaoui, M., et al. (2019). Observations of an electron diffusion region in symmetric reconnection with weak guide field. *Astrophys. J.* 870:34. doi: 10.3847/1538-4357/aaf16f

**Conflict of Interest:** 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.

Copyright © 2020 Argall, Small, Piatt, Breen, Petrik, Kokkonen, Barnum, Larsen, Wilder, Oka, Paterson, Torbert, Ergun, Phan, Giles and Burch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Automatic Classification of Plasma Regions in Near-Earth Space With Supervised Machine Learning: Application to Magnetospheric Multi Scale 2016–2019 Observations

Hugo Breuillard<sup>1\*</sup>, Romain Dupuis<sup>2</sup>, Alessandro Retino<sup>1</sup>, Olivier Le Contel<sup>1</sup>, Jorge Amaya<sup>2</sup> and Giovanni Lapenta<sup>2</sup>

## OPEN ACCESS

### Edited by:

Bala Poduval,  
University of New Hampshire,  
United States

### Reviewed by:

Joseph Eric Borovsky,  
Space Science Institute, United States  
Xiangning Chu,  
University of Colorado Boulder,  
United States

### \*Correspondence:

Hugo Breuillard  
hbreull@gmail.com

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

Received: 10 April 2020

Accepted: 23 July 2020

Published: 03 September 2020

### Citation:

Breuillard H, Dupuis R, Retino A, Le Contel O, Amaya J and Lapenta G (2020) Automatic Classification of Plasma Regions in Near-Earth Space With Supervised Machine Learning: Application to Magnetospheric Multi Scale 2016–2019 Observations. *Front. Astron. Space Sci.* 7:55. doi: 10.3389/fspas.2020.00055

The proper classification of plasma regions in near-Earth space is crucial to perform unambiguous statistical studies of fundamental plasma processes such as shocks, magnetic reconnection, waves and turbulence, jets and their combinations. The majority of available studies have been performed by using human-driven methods, such as visual data selection or the application of predefined thresholds to different observable plasma quantities. While human-driven methods have allowed performing many statistical studies, these methods are often time-consuming and can introduce important biases. On the other hand, the recent availability of large, high-quality spacecraft databases, together with major advances in machine-learning algorithms, can now allow meaningful applications of machine learning to *in-situ* plasma data. In this study, we apply the fully convolutional neural network (FCN) deep machine-learning algorithm to the recent Magnetospheric Multi Scale (MMS) mission data in order to classify 10 key plasma regions in near-Earth space for the period 2016–2019. For this purpose, we use available intervals of time series for each such plasma region, which were labeled by using human-driven selective downlink applied to MMS burst data. We discuss several quantitative parameters to assess the accuracy of both methods. Our results indicate that the FCN method is reliable to accurately classify labeled time series data since it takes into account the dynamical features of the plasma data in each region. We also present good accuracy of the FCN method when applied to unlabeled MMS data. Finally, we show how this method used on MMS data can be extended to data from the Cluster mission, indicating that such method can be successfully applied to any *in situ* spacecraft plasma database.

**Keywords:** heliophysics, classification, near-Earth regions, magnetospheric multiscale mission, time series, machine learning

## 1. INTRODUCTION

Near-Earth space regions classification is a very challenging topic in space physics because of the strong plasma dynamics resulting from the interaction between the propagating solar wind and the standing Earth's magnetosphere. This obstacle forms a collisionless bow shock (Kivelson et al., 1995) which can reflect and backpropagate energetic ions into the supersonic solar wind, forming the ion foreshock region upstream. However, most of the plasma is decelerated and heated downstream to a subsonic flow which forms a sheath around the Earth's magnetosphere called the magnetosheath. The boundary on which magnetosheath and magnetospheric pressures balance is called the magnetopause. The plasma flows along the magnetopause toward the magnetotail, driving the frozen-in magnetic field lines through the lobes, plasma sheet boundary layer and ultimately the plasma sheet, in which they eventually reconnect (Dungey, 1963). This large-scale solar-terrestrial interaction process is dominated by the highly-variable solar wind conditions, but also by localized and intermittent small-scale processes in each region. As a result, the plasma throughout the near-Earth space is highly dynamic which make it impossible to classify the different regions using only spacecraft location along its orbit.

Yet, this classification is needed for statistical studies of the plasma properties of the different regions. These large-scale statistics are allowed by the ever-growing quantity of available *in-situ* data from space missions that cover most of the near-Earth space. A large amount of available data has been used to identify near-Earth space regions using human-driven (i.e., threshold-based) methods (e.g., Jelínek et al., 2012). However, finding the optimal set of thresholds which characterizes every region of the magnetosphere is time-consuming and presumably not flexible enough with regard to the large variability expected from the strong plasma dynamics. This may introduce important biases depending on the chosen thresholds, and may also result in many “unclassified” regions (i.e., a significant loss of information).

One way to bypass this issue and benefit from this large amount of data is to make use of supervised learning, the most common form of machine learning. The role of the supervised algorithms is to learn the relationship between data instances and an associated label for each data instance. For example, the text from an email may represent the data instance and the label is a binary encoding the presence or not of spam in order to detect and filter spams from emails. Supervised methods aim at expanding human knowledge automatically by identifying the intrinsic differences between labeled points in a dataset. This knowledge is then generalized to other unlabeled datasets. Such machine learning techniques show extremely promising (often state-of-the-art) results in various tasks, such as image and speech recognition, analyzing particle accelerator data, or natural language understanding (LeCun et al., 2015). They have already been used specifically in heliophysics: to classify time series into several categories defined by the solar origin of the wind (Camporeale et al., 2017), to detect magnetospheric ultralow frequency waves (Balasis et al., 2019), to forecast solar flares (Nishizuka et al., 2017), or in particular to identify different

regions of the near-Earth space (Nguyen et al., 2019; Olshevsky et al., 2019). Nevertheless, these models remain usually region- or mission-specific because the process of human-labeling is a time-consuming and tedious task and there is still a blatant lack of human labels to cover the whole near-Earth space.

However, the Magnetospheric Multi Scale mission (MMS) (Burch et al., 2016), launched in 2015 and whose orbit covers most of near-Earth space, has enabled such a dataset to be built. This mission uses the so-called Scientist In The Loop (SITL) system (see the following section for more details) to human-label phenomena and regions of interest. The SITL process involves that an expert scientist is designated to select data of interest that will be transmitted to the ground as high-time resolution (burst) data. The selected intervals are tagged with timestamps and a comment, which are reported in the SITL reports. This system has been active since July 2015 and thus provides more than 4 years of commented data as of now. This makes it the biggest dataset of commented events and regions to our knowledge. We developed an automatic parser to convert all the comments, written by different scientists, in a unified list of labels that can be used by a machine learning algorithm.

All SITL comments are associated with a time interval, therefore the temporal variations of the measurements can be used to predict the label of the signals. Thus, each data instance can be interpreted as a time series and the near-Earth classification using SITL data can be seen as a time series classification (TSC), a very challenging problem in data mining and machine learning (Esling and Agon, 2012; Fawaz et al., 2019). The main difference with classical classification problems comes from the ordering of the data. The important features helping to discriminate the labels are mainly found in the ordering of the values while classical methods only consider the values at a given time. We employ two different techniques to solve this multivariate and multiclassification task: the multilayer perceptron is applied naively on instantaneous data and is considered as a baseline for comparison and the convolutional neural network is applied on time series.

In this article, we use the magnetic field and particle measurements of the MMS mission to establish a vast and reproducible automatic detection of 8 plasma regions that cover almost the entire near-Earth space. First, we introduce the MMS and Cluster missions, and how we used the SITL reports to build a large labeled dataset of magnetic field, plasma moments and plasma distributions from the 2016–2019 period. Then, we present the two machine learning algorithms we used and their characteristics, as well as the metrics to measure their performances. Later, we present their respective results for the MMS mission and the adaptation to the Cluster mission. Finally, we discuss these results and show our conclusions.

## 2. DATA

A fundamental and time-consuming step in any machine learning application is collecting, cleaning, and labeling data. This is a very important task as the quality of the data defines accuracy and the generalization capability of the model.

## 2.1. The MMS and Cluster Missions

MMS is a NASA space mission, launched in 2015, designed to study the electron-scale physics in Earth's magnetosphere and in particular where magnetic reconnection occurs (Burch et al., 2016). Its equatorial orbit is optimized to spend extended periods in locations where reconnection is known to occur and thus covers the majority of the near-Earth space key regions. The mission is composed of four identical spacecraft flying in an adjustable tetrahedral formation and its highly-elliptic orbit also covers almost all regions of near-Earth space. The direct-current (DC) magnetic field data are provided by the Fluxgate Magnetometer (FGM) from the FIELDS instrument suite (Russell et al., 2016; Torbert et al., 2016) with a temporal resolution of 0.1 s in "survey" mode and the plasma parameters by the Fast Plasma Investigation (FPI) instrument Pollock et al. (2016) with a temporal resolution of 4.5 s in "fast" mode.

Cluster is an ESA space mission, launched in 2000, whose aim is to study the ion-scale physics of Earth's magnetic environment and its interaction with the solar wind. The mission is also composed of four identical spacecraft flying in a tetrahedral formation and its highly-elliptic polar orbit also covers almost all regions of near-Earth space. The magnetic field data are provided by the Fluxgate Magnetometer with a temporal resolution of 4 s (Balogh et al., 2001) and the plasma parameters by the Hot Ion Analyzer instrument (Rème et al., 2001) when the instrument was working under the magnetosphere or the magnetosheath mode.

## 2.2. The Labels Datasets

One of the most important innovation brought by the MMS mission is its burst data management and selection system. The MMS spacecraft collect a combined volume of  $\sim 100$  gigabits per day of particle and field data. On average, only 4 gigabits of that volume can be transmitted to the ground. With nested automation and "Scientist-in-the-Loop" (SITL) processes, this system is designed to maximize the value of the burst data by prioritizing the data segments selected for transmission to the ground.

Concretely, the SITL system consists of a manual selection process by a scientific expert designated to eyeball daily survey (low-time resolution) data and pick time intervals of interest that will be transmitted to the ground as burst (high-time resolution) data. The selected intervals are tagged with timestamps and a comment, which usually includes the type of event selected and eventually the near-Earth region in which it occurred. For each orbit, these selections are written up in a "SITL report," in the form of a text file notably. An example of these reports can be found in Argall et al. (2020).

We developed a Python code which parses these text files to extract the timestamps and comments by identifying keywords associated with the near-Earth region where the event occurred. Each time interval was then associated with a label indicating where the spacecraft was located as follows.

If a region cannot be identified from the SITL comment, then the time interval is rejected. If FGM or FPI data is not available during a labeled time interval, then the time interval is also rejected. Using this technique, from the date at which the SITL reports are available as text files, i.e., April 2016,

to the end of 2019, we collected 7,832 labeled time intervals relevant for our study. We note here that the number of occurrences for each regions were somewhat unbalanced, so we added a total of 605 time intervals labeled by hand to undersampled regions (examples of typical plasma parameters for the different magnetospheric regions can be found in textbooks, e.g., Baumjohann and Treumann, 1996), bringing the total to 8,437 labeled time intervals of various lengths. We resampled the label data to the same cadence as FPI data, i.e., 4.5 s, representing a total of 1,331,133 labeled data points. This constitutes the biggest dataset of labeled time intervals for near-Earth space to our knowledge. The labels of the different regions are sorted roughly according to their distance to the Sun, and summarized in Table 1.

Regarding the Cluster mission labels, we use the dataset presented in Nguyen et al. (2019). This dataset covers three near-Earth space regions, namely the magnetosphere, the magnetosheath and the solar wind, over a 2-year time period (2005–2006). The data is resampled to a 1 min resolution, yielding a total of 148,762 labeled points.

## 2.3. Final Data Set

Once the labels are defined, we build the final dataset that will feed the ML models. In Earth's magnetosphere, the plasma is dominated by its core magnetic field, thus the first feature we took into account is the magnetic field vector  $\vec{B}$ . Additionally, in each magnetospheric region the plasma has a rather typical distribution function. However, the full particle distribution function constitutes an enormous amount of data when dealing with several years of data (as done here), in particular with the MMS mission (several TB of data). Furthermore, important differences exist between different heliospheric missions in the specificities of the distribution functions, e.g., energy and angular ranges as well as energy and angular resolutions. Therefore, we chose not to consider those products and focused on moments only (density, bulk velocity and temperature), to keep our model lightweight and as general as possible (i.e., applicable to different heliospheric missions such as Cluster). For this latter reason, we also decided not to use the spacecraft location as a feature, because different missions cover different locations in near-Earth

**TABLE 1 |** Labels for the 10 different near-Earth regions in our model.

Regions	Labels
Solar wind	SW
Ion foreshock	FS
Bow shock	BS
Magnetosheath	MSH
Magnetopause	MP
Boundary layer	BL
Magnetosphere	MSP
Plasma sheet	PS
Plasma sheet boundary layer	PSBL
Lobe	LOBE

space (e.g., Cluster has a polar orbit while MMS has an equatorial orbit; apogees/perigees are different, etc.).

We start with loading the raw MMS magnetic field and plasma moments data from FGM and FPI instruments for the 2016–2019 period using the aidapy package<sup>1</sup>, and resample the data to the FPI cadence, i.e., 4.5 s. The resulting dataset contains 12 variables: the magnetic field magnitude  $B_{tot}$  and its components  $B_x$ ,  $B_y$ ,  $B_z$ , the ion density  $N_i$ , the bulk velocity magnitude  $V_{i,tot}$  and its components  $V_{i,x}$ ,  $V_{i,y}$ ,  $V_{i,z}$ , the parallel and perpendicular temperatures  $T_{\parallel}$ ,  $T_{\perp}$  with regards to the ambient magnetic field and the total value  $T_{tot}$ . The dataset is matched with the labels points defined in the previous section. If data gaps are present in the specified time interval, then the whole interval is rejected. This method yields a total dataset of size  $1,331,133 \times 12$ .

The TSC model requires as input arrays of equal sizes with one label per array. Thus the dataset is grouped as time series corresponding to the labeled time intervals, resulting in 6,928 data blocks of various sizes. We decided to split these time series into equal chunks of 3 min, because this time length correspond roughly to the mean value of the data blocks (i.e., it minimizes the padding) and yields enough points (40) in each block given the time sampling resolution. To do so, we apply the following scheme: if the data block is shorter than 3 min, it is padded with the wrap of the vector along the axis (i.e., the first values are used to pad the end and the end values are used to pad the beginning); and if the data block is longer than 3 min, then it is split into several time series of 3 min each (the last one being padded as described above if also shorter than 3 min). **Table 2** gives the distribution of the occurrences for each class.

This operation yields a dataset as an multidimensional array of size (34159, 40, 12). Finally, to input this array to the TSC model (and to avoid temporal bias due to the spacecraft orbit), we shuffle and randomly split the time series into training (56.25%), test (25%), and validation (18.75%) datasets. These three classical categories are defined such as:

- The training set is used to determined the weights of the model;
- The validation set is an out-of-sample set and allows to evaluate the error for data which has not been observed. In particular, the validation set is used for hyperparameter search and early stopping to avoid over-fitting ensuring a good generalization of the model;
- The test set is also an out-of-sample set which has not been used during training. The data are only used after the training to assess the performance of the final model, after training and hyperparameter optimization.

### 3. CLASSIFICATION METHODS WITH A FULLY CONVOLUTIONAL NEURAL NETWORK

In this section, we overview the machine-learning algorithm we use: the fully convolutional neural network (FCN). This

**TABLE 2 |** Number of occurrences for each class.

	Time series
SW	2,130
FS	4,714
BS	2,328
MSH	4,274
MP	3,764
BL	4,544
MSP	4,658
PS	3,576
PSBL	2,703
LOBE	1,468
Total	34,159

algorithm is assessed with several evaluation metrics. FCN works with time-dependant inputs (time series) in order to learn dynamical features. The main purpose of classifying time series and not only the instantaneous values is to learn automatically different features across the time dimension. The events can be characterized by dynamical insights at different scales learned by the model. We present a second model, called Multilayer perceptron (MLP), in the **Supplementary Materials**. It considers the input of the 12 variables as time independent. The main advantages of this method are its simplicity to prepare the machine learning pipeline and the possibility to work with any temporal resolution. However, the classification performance is much lower.

The FCN belongs to the category of artificial neural networks. They are usually characterized by their architecture: number of layers, type of connection (feedforward, feedback, convolution, etc.), and the activation functions. A layer is defined as a set of neurons which are not connected to each other. We selected neural network as they are able to learn non-linear models and can handle large datasets compared to kernel methods.

#### 3.1. Model Presentation

Time series classification (TSC) is an active field of research and hundreds of different algorithms have been developed in recent years (Bagnall et al., 2017). Classical TSC usually uses new features spaces generated from the time series. The best performing models can combine dozen of different classifiers and feature transformations, leading to significant complexity and an important computational cost (Lucas et al., 2019).

Even if deep learning has seen very successful applications in the last decades, only a few examples of algorithms for TSC exist. This lack of overview has been filled recently (Fawaz et al., 2019) and several models show very promising results. In particular, we are convinced that convolution architectures can build very accurate classifiers. The convolutional layers use convolution in place of general matrix multiplication.

Local groups of values are often highly correlated in time series and they form distinctive local motifs that are easily detected. Convolutional layers are designed to identify these patterns invariant to location: if a motif specific to a label

<sup>1</sup><https://gitlab.com/aidaspaces/aidapy>

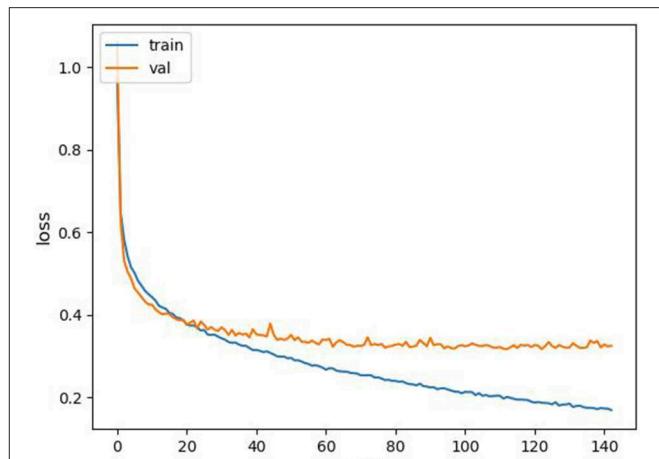
appears in one part of the time series, it could also appear anywhere. Hence a convolutional layer sharing the same weights at different locations can detect temporal patterns in different parts of the time series (LeCun et al., 2015). Moreover, several deep learning frameworks, such as Tensorflow (Abadi et al., 2016) or Pytorch (Paszke et al., 2019), are nowadays freely available. They provide differentiable programming to build very efficient neural networks. Thus, we consider deep learning for TSC for these two reasons: the availability of high-quality frameworks and the relevance of convolutional architectures.

The Fully Convolutional Network (FCN) (Wang et al., 2017) is a competitive deep-learning architecture for TSC yielding good results on large benchmarks (Fawaz et al., 2019). The network architecture is relatively simple and is comprised of a sequence of three convolution blocks followed by a global average pooling block (Lin et al., 2013). Each time convolution block is divided into a convolution layer, a batch normalization layer, and a Rectified Linear Unit (RELU) Nair and Hinton (2010) activation function. The convolutional layer is used to extract temporal features from the inputs by performing a convolution between the input signal and several filters. Each convolutional block has different filter lengths in order to analyze several time scales. The extracted features of the last convolutional block are used as inputs for the global average pooling (GAP) module to output the classification result. This last block is composed of the global average pooling layer and the Softmax layer. A more complete description of the FCN architecture can be found in Wang et al. (2017) and Fawaz et al. (2019).

As regards the numerical parameter, we performed a bayesian optimization to select the hyperparameters using the validation set for four parameters:

- The optimizer: adam or rmsprop
- The learning rate: between  $10^{-6}$  and  $10^{-3}$ ;
- The batch size: 32, 64, or 92;
- The number of filters: the baseline of the FCN paper, half the baseline and twice the baseline. The baseline given by the FCN paper (Wang et al., 2017) is the following: the first convolutional block has 128 filters with a filter length equal to 8, the second convolution extracts features with 256 filters of size 5 and the final convolutional layer is defined by 128 filters, each one with a length equal to 3.

The final FCN is trained for 1,000 epochs with the categorical cross entropy as loss function. The result of the bayesian optimization provides the different hyperparameters. The Adam optimizer is used with an initial value of the learning rate set to 0.00002. An adaptive learning rate decrease is selected (with a minimum learning rate of  $10^{-6}$ ), reducing the learning rate when the validation accuracy has stopped improving for 50 epochs. A batch size of 92 is used with twice the baseline configuration for the number of filters. The final FCN has 1,079,000 free parameters. The method of early termination is adopted to avoid over-fitting and stops the learning process when the validation loss did not decrease for 20 epochs. All the time series are standardized before the training process. The learning curve is presented in **Figure 1**. The learning process stopped after 140 epochs thanks to the early termination strategy.



**FIGURE 1 |** Learning curve history comparing training and validation loss.

### 3.2. Performance Evaluation

Assessing accurately the quality of a classification model raises significant challenges. A single and unified performance metric cannot effectively evaluate a multi-class classifier from all perspectives, such as class-balance or the number of different outcomes. Therefore, several metrics generalized from binary classification are considered.

The confusion matrix is a basic but useful tool to visualize the distribution of the predicted classes. Each column of the matrix gives the instances of a true class while each row represents the instance of a predicted class. Several metrics from binary classification can be generalized to multiclass by considering the prediction of each class as a binary classification problem: the given class has been predicted (positive) or not (negative). Then, the total number of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) can be computed and several metrics are defined:

- **Accuracy** is the most intuitive metric and gives the percentage of correct predictions among all the predictions  $N_p$ :

$$\text{accuracy} = \frac{TP + TN}{N_p}. \quad (1)$$

- **Precision**, also called positive predictive value, measures the proportion of positive values that are correctly identified as such among all predicted positives. In other words, it corresponds to the percentage of detections that are right. High precision is associated to a small amount of FP:

$$\text{precision} = \frac{TP}{TP + FP}. \quad (2)$$

- **Recall**, also called the true positive rate, measures the proportion of positives detected. A significant number of missed true values (FN) leads to low recall:

$$\text{recall} = \frac{TP}{TP + FN}. \quad (3)$$

- **F1-score** is the harmonic mean of precision and recall. Usually, high recall is detrimental to precision and vice versa. Therefore, the F1-score gives a trade-off between the two quantities and can help to compare two classifiers with different precision and recall values:

$$\text{accuracy} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \quad (4)$$

All these metrics are computed for each class. Therefore, averaged quantities can be also defined to provide a more general view of the classifier. A macro-average gives the unweighted mean of the metric without taking imbalance into account and a micro-average computes the average weighted by the number of true instances for each class. The micro-average will account for label imbalance.

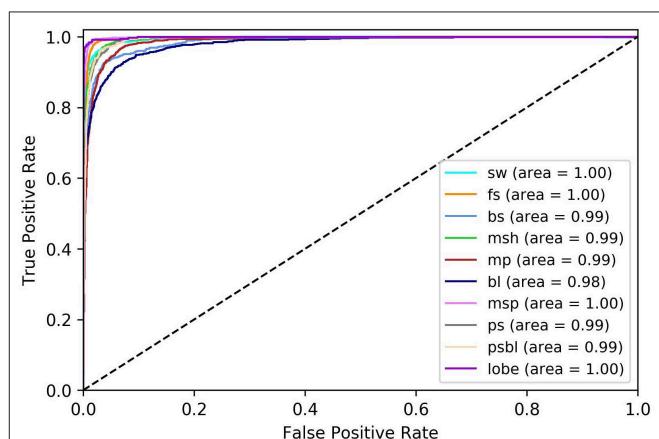
Finally, a receiver operating characteristic (ROC), also called ROC curve, is a graphical tool illustrating the performance of a binary classifier system when the discrimination threshold is varied. Indeed, each class prediction is associated to a probability. Usually, the class predicted by the classifier is chosen as the class with the highest probability. Instead, a threshold value can be used to decide if the class is effectively predicted by the classifier. The ROC curve plots the fraction of true positives out of all the predicted positives (True Positive Rate) vs. the fraction of false positives out of all the predicted negatives (False Positive rate) at various threshold values. For instance, all classes are predicted with a zero threshold (False Positive Rate and True Positive Rate of 1.0) while there are no correct predicted class and no false positive with a unit threshold (False Positive Rate and True positive rate of 0.0). A perfect classifier is located on the left corner of the curve with a true positive rate of 1.0 and a false positive rate of 0.0. A ROC curve can be drawn for each class. In case of a model with no discrimination capacity to distinguish between positive class and negative class, the ROC curve is a straight line of slope 1. The area under the ROC curve (AUC) summarized in one number if the model is capable of distinguishing between classes.

Two different groups of regions are identified in **Table 3**. The first group consisting of the FS, the MSP, and the LOBE is very well-predicted with a respective f1-score above 0.94 and a AUC of 1.00. Such regions show usually very specific patterns, explaining why the classification metrics are very high. The second group is formed by the BS, the MP, the BL, the PS, and the PSBL, with smaller metric values, such as a f-1 score between 0.82 and 0.88 and a AUC between 0.98 and 0.99. **Figure 2** shows the different ROC curves for the FCN. BL and BS are the lowest curves. They have a much higher false positive rate when they reached a true positive rate of 1.0 compared to the other classes. It means that, for some very specific predictions, the classifier associates a very low probability to the correct class. The BL and the BS are thin regions marking the boundary between larger regions. Therefore, several time intervals labeled by SITLs may overlap two or more nearby regions. It could explain the lower metric values for these two regions.

From the **Supplementary Material**, the second model (MLP) has important also problems to classify regions with strong

**TABLE 3 |** Classification report for FCN computed on the test set.

	Precision	Recall	f1-score	AUC
SW	0.88	0.92	0.90	1.00
FS	0.93	0.95	0.94	1.00
BS	0.85	0.80	0.83	0.99
MSH	0.91	0.91	0.91	0.99
MP	0.84	0.84	0.84	0.99
BL	0.84	0.84	0.84	0.98
MSP	0.96	0.95	0.96	1.00
PS	0.89	0.89	0.89	0.99
PSBL	0.86	0.87	0.87	0.99
LOBE	0.95	0.95	0.95	1.00
Macro average	0.89	0.89	0.89	0.98
Micro average	0.89	0.89	0.89	0.99
Accuracy			0.89	



**FIGURE 2 |** Receiver operating characteristic curve computed on the test set.

**TABLE 4 |** Confusion Matrix for FCN computed on the test set.

Prediction	True									
	SW	FS	BS	MSH	MP	BL	MSP	PS	PSBL	LOBE
SW	493	53	8	0	2	1	0	0	0	0
FS	42	1,140	40	0	0	0	0	0	0	0
BS	2	12	450	54	8	1	0	0	0	0
MSH	0	1	52	982	41	5	0	0	0	0
MP	0	0	10	37	801	92	5	3	2	1
BL	0	0	1	4	78	933	49	39	10	0
MSP	0	0	0	0	10	32	1,101	0	0	0
PS	0	0	0	0	5	38	5	803	54	1
PSBL	0	0	0	0	7	13	0	59	572	15
LOBE	0	0	0	0	0	0	0	1	16	354

The gray background corresponds to the correct classification in the confusion matrix.

variations, such as the BS with a f-1 score below 0.5. Thus, it seems that convolutional methods help to improve the accuracy of the classifier for the most challenging regions and it validates the choice of such algorithms designed to extract important temporal features and variations.

**Table 4** gives details on the FCN predictions with a confusion matrix. Very typical errors can be observed for nearby regions, such as: SW and FS; PS, BL, and PSBL; PSBL and LOBE; MP, MSH and BL. These errors can be explained by physical arguments. The plasma properties can be similar for several nearby regions. For instance, we may expect that MP shares common properties with MSH and MSP as the MP acts as a boundary between the two other regions. The SW region can also be very similar to the FS as the latter comes from the reflection of the SW on the BS, as stated in the introduction. Therefore, it may be concluded that the FCN model really learned the typical patterns associated to each region as the main errors are identified for regions with similar physics.

## 4.2. Examples of MMS Classification

In this section, we show some examples of time series observed by the MMS spacecraft as it spanned the different near-Earth regions, along with the labels of the classifications made by the FCN model.

**Figure 3** shows the MMS1 probe observations for the whole day of 2019-11-09, during which it flew through the dayside region. From top to bottom, the figure panels show the energy spectrum of the omnidirectional ion flux, the three components of the magnetic field ( $B_x, B_y, B_z$ ) in GSE coordinates and its magnitude  $B_{tot}$ , the ion number density, the three components of the ion bulk velocity ( $V_x, V_y, V_z$ ) in GSE coordinates, the ion temperatures parallel and perpendicular ( $T_{\parallel}, T_{\perp}$ ) to the ambient magnetic field, and finally the labels given by the FCN model (FCN labels). The time in UT format and the spacecraft position ( $X_{pos}, Y_{pos}, Z_{pos}$  in  $R_E$  and GSE coordinates) are displayed at the bottom of the X-axis.

In this example, the FCN model classifies well the regions, as its labels follow to a great extend the MMS1 observations: first the spacecraft is located in the magnetosphere, then it crosses

the magnetopause to enter the magnetosheath, afterwards it flies through a series of bow shocks and goes back and forth to the solar wind, until it finally enters the solar wind. The model is even able to precisely detect the small-scale dynamics of the boundary regions such as partial magnetopause, boundary layer and bow shock crossings, and the come and go between the solar wind and the ion foreshock. Only three points seem to be clearly misclassified (as MSP) in the ion foreshock. However, these points can be easily eyeballed as outliers or discriminated by their low quality flag.

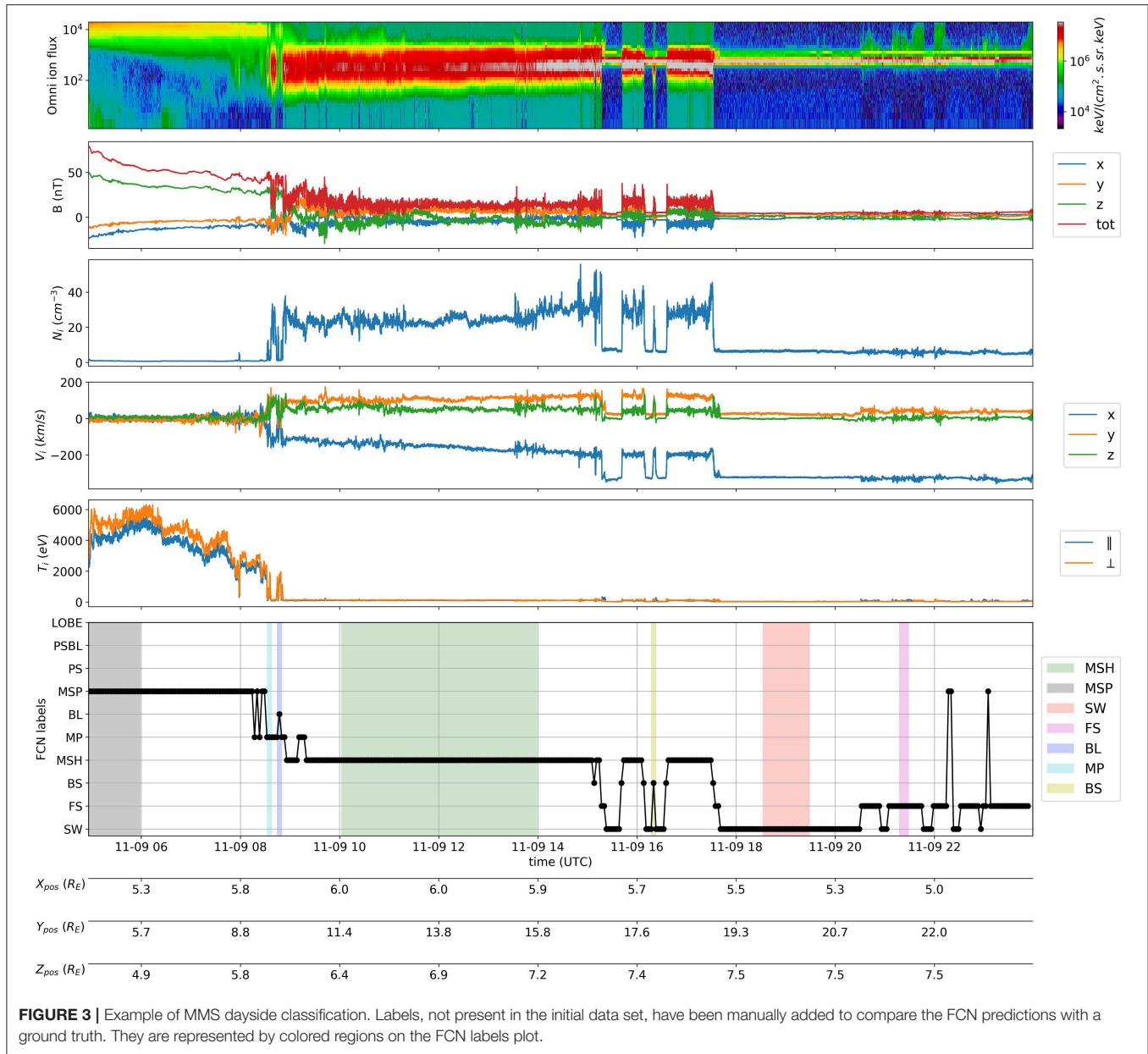
**Figure 4** shows another example of the FCN classification method described above, but this time when MMS spacecraft flew through the nightside region. The MMS1 observations take place during the whole day of 2019-07-13, and the format is the same as in **Figure 3**. In this example again, the FCN model is able to label properly the regions spanned by MMS1: at the very beginning MMS1 exits the magnetosphere to the plasma sheet boundary layer, then goes in and out the lobe regions before reaching the plasma sheet. Again, the FCN model is also able to pick up the small-scale dynamics such as short plasma sheet crossings.

The two examples shown in this section therefore support the good metrics obtained in the previous section, showing concretely that the FCN model is suited to properly label most of the MMS data with the regions spanned.

## 4.3. Locating Important Feature in Time Series With Class Activation Map

Class Activation Map (CAM) is a technique developed by Zhou et al. (2016) to get the discriminative regions used by a CNN to identify a specific class in the input data. I. CAM has been applied to TSC for one-dimensional case in Wang et al. (2017) for the first time. The objective in our case is to highlight the subsequences (i.e., sections) in each time series which are relevant to its class. Thus, this approach allows to explain the decision taken by the classifier. The mathematical description of the method can be found in Zhou et al. (2016), Wang et al. (2017), and Fawaz et al. (2019). This method works only for models with the global average pooling as last layer (Lin et al., 2013). CAM is represented by a univariate time series (with the same size than the FCN input) where the value gives the importance of the signal to classification. Low values mean the subsequence does not contribute to the decision of the classifier while high values mean the section contributes significantly.

**Figure 5** shows CAM examples for two different classes: the solar wind and the foreshock. For each class, a well-predicted time series is drawn and colored by the value of the CAM, illustrating the results from MMS data given in section 4.1. As regards the SW example, almost all the input signal contributes to the classification, as stated by the uniform CAM values above 0.8 between 20 and 160 s. It means that the values of input features are much more important than their variations or than specific patterns. On the other hand, the classifier detects the FS mainly thanks to a specific pattern around 65 s associated to extrema for  $B_x$ ,  $B_y$ ,  $N_i$ , and  $V_y$  and significant slopes for the other features. Only a few specific time



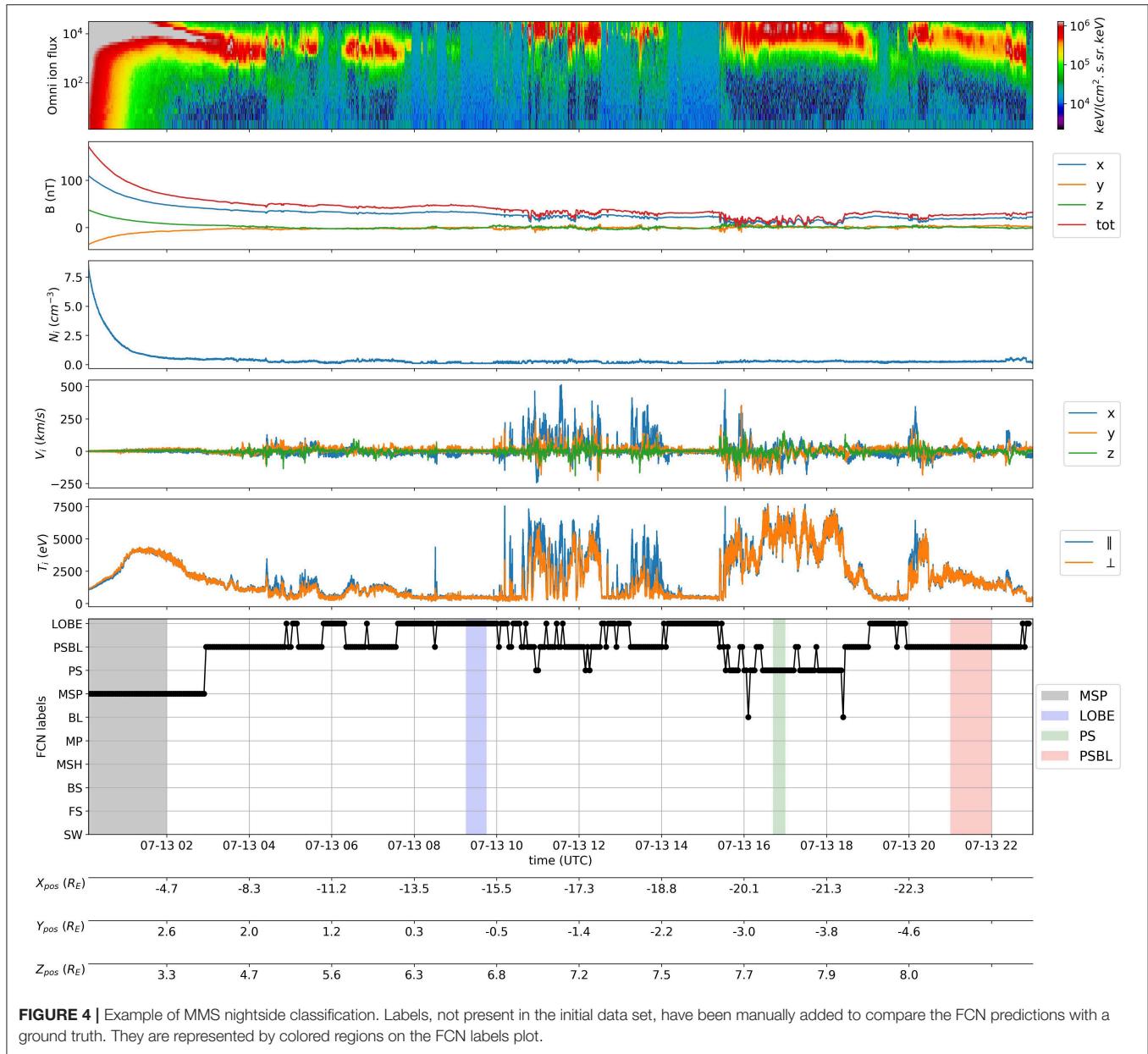
intervals are involved in the decision of the classifier for FS. The CAM analysis strengthens the conclusion that temporal and dynamical analysis are fundamental to classify nearby regions with fluctuating behaviors. For instance, the MLP results given in the **Supplementary Materials** shows that a significant number of SW region are misclassified as FS as the model works only on instantaneous quantities.

#### 4.4. Extension to Cluster Mission

In this section, we put to test the adaptability of the FCN model to different magnetospheric data, namely data from the Cluster mission. This mission has been chosen as the data is pretty close to MMS data (notably the data sampling), and a 2-year period of labeled data is available (see Nguyen et al., 2019) to audit the classification resulting from the model.

First, we qualitatively assess the quality of the different classification methods, by comparing them with C1 probe observations. We randomly select a day during which the C1 probe flew through the dayside region. This example, from the day of 2005-02-13, is shown in **Figure 6** in the same format as **Figures 3, 4**, except that we include the manual labels from Nguyen et al. (2019) (named “Man labels”) in the penultimate bottom panel.

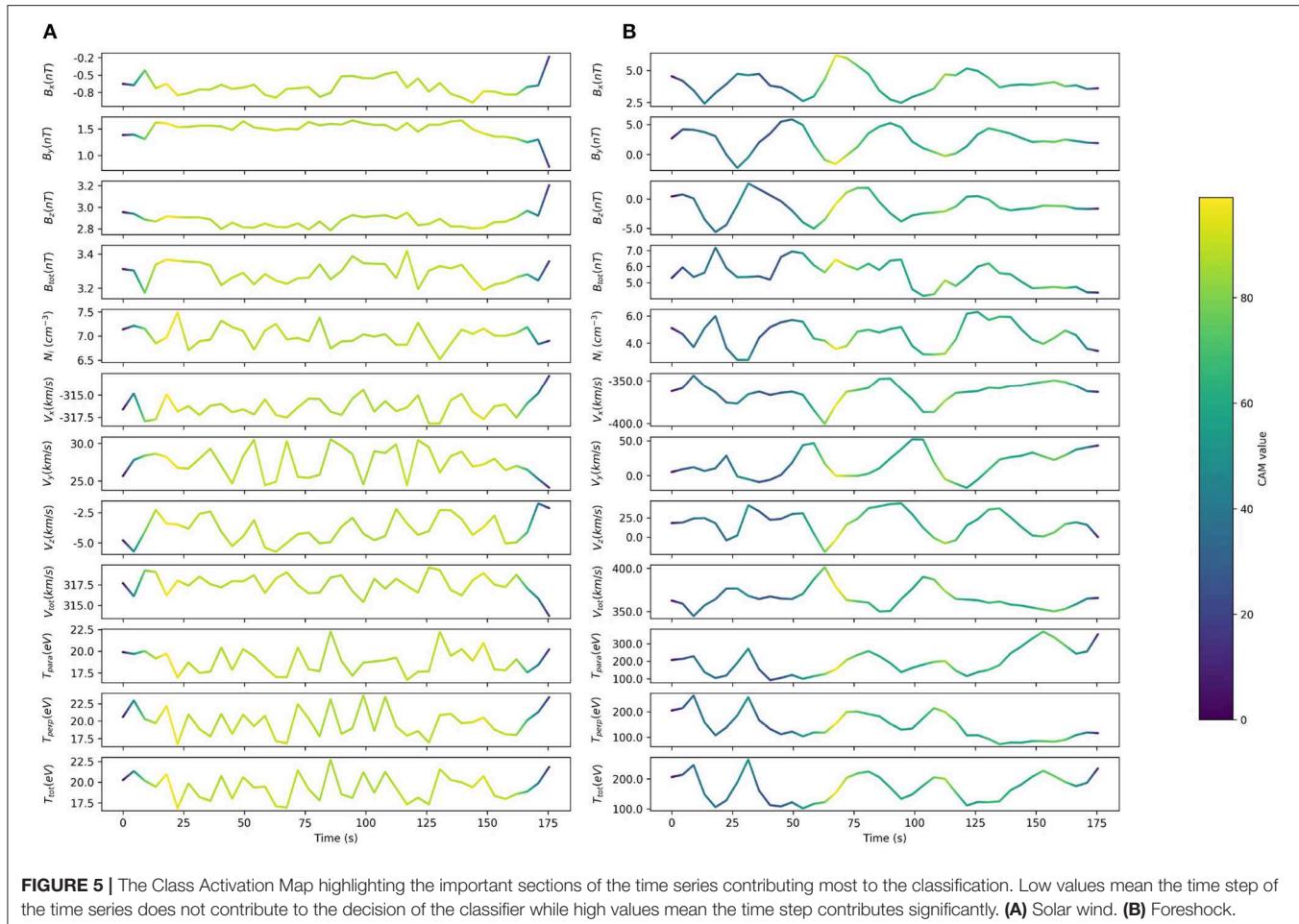
The “Man labels” are in good agreement with the observations of C1, as the probe is first located in the magnetosphere, then flies through the magnetosheath and finally reaches the solar wind. The “FCN labels” show also good agreement with C1 observations: the probe is first located in the magnetosphere and then crosses the magnetopause a few times before entering the magnetosheath, then it crosses the bow shock to enter the



solar wind before crossing several times again the bow shock and finally spanning the ion foreshock. As in the examples shown for the MMS mission, the FCN model is able to pick up the small-scale dynamics of the near-Earth frontiers: the magnetopause, the bow shock and the foreshock. As a result, the model is able to outperform the manual labels by identifying in more detail the magnetospheric regions, in particular the small-scale dynamics of the transition regions.

Then, we take advantage of the 2-year period of Cluster labeled data from Nguyen et al. (2019) to quantify the quality of the FCN model classification, by investigating the performance metrics. We select the points labeled as “MSP,” “MSH,” and “SW” by the FCN model using Cluster data from the 2005–2006 period, and compare them with the labels for these 3

regions obtained from Nguyen. Using this method, we obtain an accuracy classification score of 0.97 over a total of 10,929 common labeled points. the classification report can be found in Table 5. We note here that this score is not an exact quantization of the model’s accuracy on Cluster data, and the classification score is probably overestimated because only 3 classes (i.e., magnetospheric regions) are considered here instead of the 10 included in the FCN model. Therefore, a dataset of Cluster data labeled with these 10 near-Earth regions is required to get an absolute quantification of its accuracy regarding the Cluster mission. However, it shows globally that our FCN model classification is a reliable method for the labeling of the Cluster mission data and potentially other heliophysics missions.



**FIGURE 5 |** The Class Activation Map highlighting the important sections of the time series contributing most to the classification. Low values mean the time step of the time series does not contribute to the decision of the classifier while high values mean the time step contributes significantly. **(A)** Solar wind. **(B)** Foreshock.

## 4.5. Resources

The most time-consuming task in the present work has been the data preparation and the parsing of the SITL reports. Training a FCN model needs about 200 CPU hours on a workstation featured with an Intel Xeon E5-2670 v3 (12 cores at 2.30 GHz). These values must be multiplied by the number of different trials needed to optimize the hyperparameters (about 40 in that case).

## 5. DISCUSSION AND CONCLUSIONS

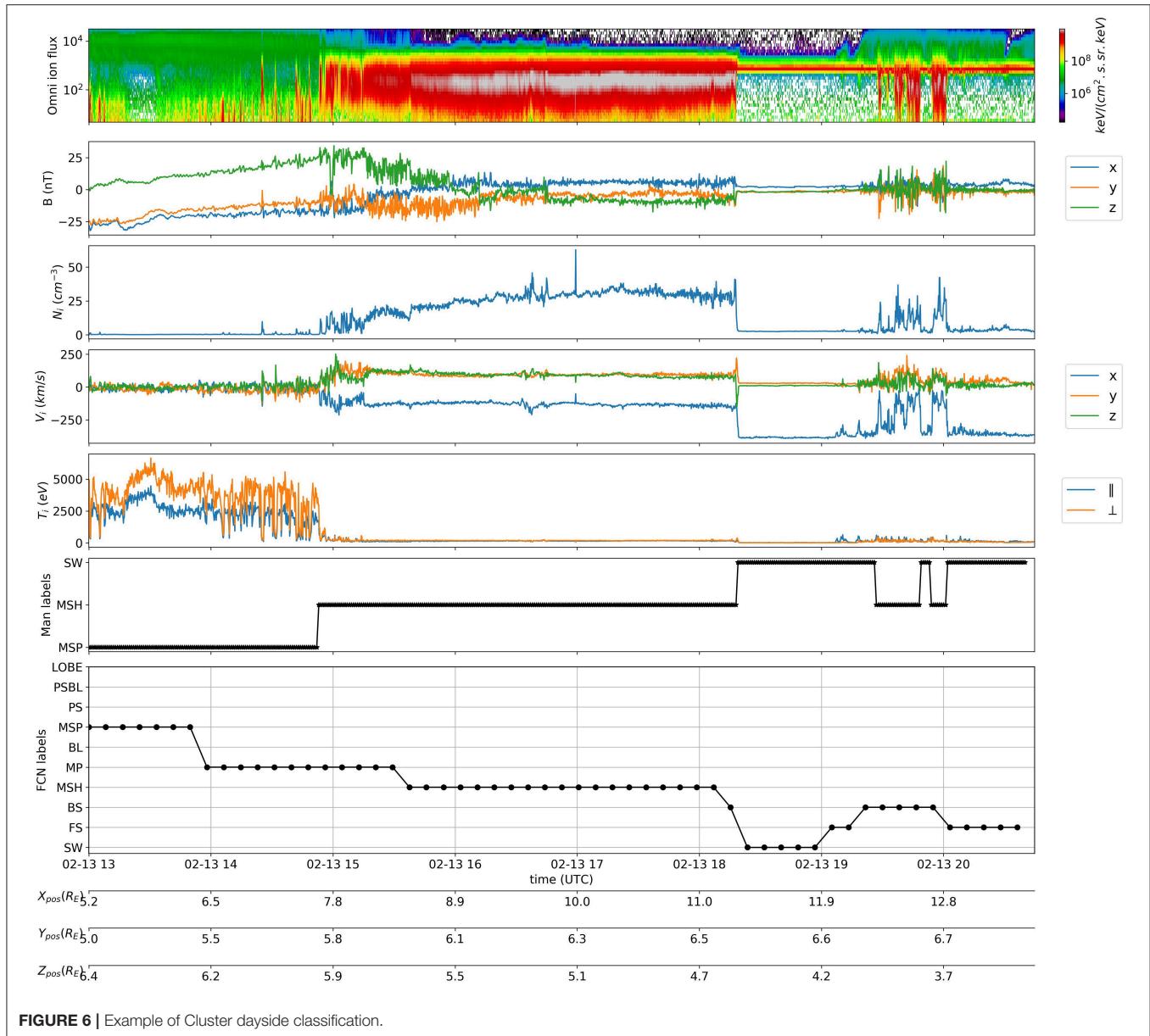
Using deep learning algorithms, namely a fully convolutional neural network (FCN), we built an automatic detection of 10 near-Earth regions: the solar wind, the ion foreshock, the bow shock, the magnetosheath, the magnetopause, the boundary layer, the magnetosphere, the plasma sheet, the plasma sheet boundary layer and the lobes.

Using more than 3 years of labeled (SITL and additional human-labeled) MMS mission data, we showed that this method are reliable to classify near-Earth regions. The FCN method (1,079,000 free parameters) has been very effective in taking into account the dynamical features of the most challenging plasma regions (important data variability due to the fluctuations of the plasma), in particular the bow shock and the boundary layer. The high accuracy of the FCN model

also highlights the quality of the labeled data set generated from SITL reports. We demonstrate the good accuracy of the classification predictions on the test and validation datasets, but also on unlabeled data from the MMS mission. We also show the adaptability of the trained FCN model by applying the classification to the Cluster ESA mission data. The predictions showed good accuracy and enhanced dynamics in comparison with previous human-labeled dataset.

These results show on one hand that the model can be applied to the whole MMS dataset, which would be of great interest to map the different magnetospheric boundaries and build empirical models of the properties and dynamics of the plasma in the near-Earth space. On the other hand, these results show that the model could be applied on different space missions orbiting around Earth (such as Cluster or THEMIS) or other planets to automatically label available data. However, we recommend to build a small labeled dataset specific to each mission with limited retraining of the last layer. This process called transfer learning could help improving the generalization of our labeled dataset to other missions such as the JUICE mission, which will be launched in 2022 and for which selective downlink strategies are being discussed.

As a matter of fact, the model could be used as a support to the scientific experts in charge of the data selection process (such

**TABLE 5 |** Classification report for FCN computed on the Cluster dataset.

	Precision	Recall	f1-score
MSP	0.84	0.97	0.90
MSH	1.00	0.97	0.98
SW	0.95	0.99	0.97
Macro average	0.93	0.97	0.95
Micro average	0.98	0.97	0.97
Accuracy		0.97	

as the SITL system), providing classification of the regions that would contribute to ease and speed up this time-consuming and tedious task. Such lightweight and easily-adaptable algorithms

could also be important for the so-called SOC (Science Operation Center) of current and future spacecraft missions, both on ground and onboard. For the latter case, these algorithms could be implemented within onboard spacecraft digital boards to automatically select regions and events of scientific interest, much reducing the complexity and the cost of science operations.

The intended integration of this model into the aidapy package<sup>2</sup>, which allows to automatically load selected data from open-access databases, will be of particular interest by providing plasma data accompanied with region labels and quality flags on the fly for use case and statistical studies. The integration of such lightweight algorithms can be also generalized to other programs.

<sup>2</sup><https://gitlab.com/aidaspaces/aidapy>

In terms of modeling, we expect to improve the FCN results by generalizing the use of hyperparameter optimization for a higher number of hyperparameters and on larger ranges. We also plan to investigate the use of convolution methods directly on the omnidirectional ion energy fluxes.

Finally, the classification of near-Earth regions represents only a first step for the time series classification in heliophysics. It paves the way to more ambitious models, such as the identification and the classification of space plasma processes (e.g., magnetic reconnection and structures, waves and turbulence or plasma jets), as well as their combinations. However, this topic is beyond the scope of the present study and is left for a forthcoming study.

## DATA AVAILABILITY STATEMENT

The models used for this study, the post-processing tools, and the link to the data can be found in the Jupyter notebook classification results.ipynb in the AIDA repository<sup>3</sup>.

## AUTHOR CONTRIBUTIONS

HB and RD built the data set, the machine learning models and wrote the manuscript. HB, RD, AR, and OL gathered information from SITL reports and provided information for their usage for machine learning. HB, AR, OL, and GL provided physical interpretation of the results. RD and JA

provided insights into the use of the different machine learning techniques. AR and GL supervised the work. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This paper has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. 776262 (AIDA, <http://www.aida-space.eu>).

## ACKNOWLEDGMENTS

Data were retrieved using HELIOPY v0.5.3 (Stansby et al., 2020) and processed using aidapy<sup>4</sup>, pandas McKinney (2010), and scikit-learn (Pedregosa et al., 2011). Figures were produced using matplotlib (Hunter, 2007). The model has been built with Keras from Tensorflow 2.2 (Abadi et al., 2016) and we modified scripts provided in the dl-4-ts repository<sup>5</sup>. The hyperparameter optimization is performed with the Python library optuna (Akiba et al., 2019). The authors would like to acknowledge Gautier Nguyen (LPP) for providing classifications on Cluster data, Nicolas Aunai (LPP) and Alexis Jeandet (LPP) for the helpful discussion and support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspas.2020.00055/full#supplementary-material>

<sup>3</sup>[https://gitlab.com/aidaspace/notebooks\\_aida/-/tree/master/04\\_sitl\\_classification\\_region](https://gitlab.com/aidaspace/notebooks_aida/-/tree/master/04_sitl_classification_region)

<sup>4</sup><https://gitlab.com/aidaspace/aidapy>

<sup>5</sup><https://github.com/hfawaz/dl-4-tsc>

## REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). “Tensorflow: a system for large-scale machine learning,” in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (Savannah, GA), 265–283.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). “Optuna: a next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK), 2623–2631. doi: 10.1145/3292500.330701
- Argall, M. R., Small, C., Piatt, S., Breen, L., Petrik, M., Barnum, J., et al. (2020). MMS SITL ground loop: automating the burst data selection process. *arXiv [Preprint]*. arXiv:2004.07199. doi: 10.3389/fspas.2020.00054
- Bagnall, A., Lines, J., Bostrom, A., Large, J., and Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining Knowl. Discov.* 31, 606–660. doi: 10.1007/s10618-016-0483-9
- Balasis, G., Aminalragia-Giamini, S., Papadimitriou, C., Daglis, I. A., Anastasiadis, A., and Haagmans, R. (2019). A machine learning approach for automated ULF wave recognition. *J. Space Weather Space Clim.* 9:A13. doi: 10.1051/swsc/2019010
- Balogh, A., Carr, C. M., Acuña, M. H., Dunlop, M. W., Beek, T. J., Brown, P., et al. (2001). The Cluster Magnetic Field Investigation: overview of in-flight performance and initial results. *Ann. Geophys.* 19, 1207–1217. doi: 10.5194/angeo-19-1207-2001
- Baumjohann, W., and Treumann, R. A. (1996). *Basic Space Plasma Physics*. London: Imperial College Press. doi: 10.1142/p015
- Burch, J., Moore, T., Torbert, R., and Giles, B. (2016). Magnetospheric multiscale overview and science objectives. *Space Science Reviews*, 199(1-4):5-21. doi: 10.1007/s11214-015-0164-9
- Camporeale, E., Caré, A., and Borovsky, J. E. (2017). Classification of solar wind with machine learning. *J. Geophys. Res. Space Phys.* 122, 10–910. doi: 10.1002/2017JA024383
- Dungey, J. W. (1963). Interactions of solar plasma with the geomagnetic field. *Planet. Space Sci.* 10, 233–237. doi: 10.1016/0032-0633(63)90020-5
- Esling, P., and Agon, C. (2012). Time-series data mining. *ACM Comput. Surv.* 45, 1–34. doi: 10.1145/2379776.2379788
- Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining Knowl. Discov.* 33, 917–963. doi: 10.1007/s10618-019-00619-1
- Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Jelínek, K., Němeček, Z., and Šáfránková, J. (2012). A new approach to magnetopause and bow shock modeling based on automated region identification. *J. Geophys. Res. Space Phys.* 117:A05208. doi: 10.1029/2011JA017252
- Kivelson, A., Kivelson, M., and Russell, C. (1995). *Introduction to Space Physics*. Cambridge University Press. doi: 10.1017/9781139878296
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

- Lin, M., Chen, Q., and Yan, S. (2013). Network in network. *arXiv [Preprint]. arXiv:1312.4400*.
- Lucas, B., Shifaz, A., Pelletier, C., O'Neill, L., Zaidi, N., Goethals, B., et al. (2019). Proximity forest: an effective and scalable distance-based classifier for time series. *Data Mining Knowl. Discov.* 33, 607–635. doi: 10.1007/s10618-019-00617-3
- McKinney, W. (2010). “Data structures for statistical computing in Python,” in *Proceedings of the 9th Python in Science Conference* (Austin, TX), 56–61. doi: 10.25080/Majora-92bf1922-00a
- Nair, V., and Hinton, G. E. (2010). “Rectified linear units improve restricted Boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa), 807–814.
- Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., and Fontaine, D. (2019). Automatic detection of the earth bow shock and magnetopause from *in-situ* data with machine learning. *Ann. Geophys. Discuss.* 2019, 1–22. doi: 10.5194/angeo-2019-149
- Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., Watari, S., and Ishii, M. (2017). Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms. *Astrophys. J.* 835:156. doi: 10.3847/1538-4357/835/2/156
- Olshevsky, V., Khotyaintsev, Y. V., Divin, A., Delzanno, G. L., Anderzen, S., Herman, P., et al. (2019). Automated classification of plasma regions using 3d particle energy distribution. *arXiv [Preprint]. arXiv:1908.05715*.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). “PyTorch: an imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems* 32, eds H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc.), 8026–8037. Available online at: <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. Available online at: <https://dl.acm.org/doi/10.5555/1953048.2078195>
- Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., et al. (2016). Fast plasma investigation for magnetospheric multiscale. *Space Sci. Rev.* 199, 331–406. doi: 10.1007/s11214-016-0245-4
- Rème, H., Aoustin, C., Bosqued, J. M., Dandouras, I., Lavraud, B., Sauvaud, J. A., et al. (2001). First multispacecraft ion measurements in and near the Earth's magnetosphere with the identical Cluster ion spectrometry (CIS) experiment. *Ann. Geophys.* 19, 1303–1354. doi: 10.5194/angeo-19-1303-2001
- Russell, C., Anderson, B., Baumjohann, W., Bromund, K., Dearborn, D., Fischer, D., et al. (2016). The magnetospheric multiscale magnetometers. *Space Sci. Rev.* 199, 189–256. doi: 10.1007/s11214-014-0057-3
- Stansby, D., Rai, Y., Argall, M., JeffreyBroll, Erwin, N., Shaw, S., et al. (2020). *heliopython/heliopy: Heliopy 0.10.0*. Available online at: <https://zenodo.org/record/3676651#.Xz38LSgzaUk>
- Torbert, R., Russell, C., Magnes, W., Ergun, R., Lindqvist, P.-A., LeContel, O., et al. (2016). The fields instrument suite on MMS: scientific objectives, measurements, and data products. *Space Sci. Rev.* 199, 105–135. doi: 10.1007/s11214-014-0109-8
- Wang, Z., Yan, W., and Oates, T. (2017). “Time series classification from scratch with deep neural networks: a strong baseline,” in *2017 International Joint Conference on Neural Networks (IJCNN)* (Anchorage, AK: IEEE), 1578–1585. doi: 10.1109/IJCNN.2017.7966039
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). “Learning deep features for discriminative localization,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 2921–2929. doi: 10.1109/CVPR.2016.319

**Conflict of Interest:** 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.

The reviewer JB declared a past co-authorship with one of the authors AR to the handling editor.

Copyright © 2020 Breuillard, Dupuis, Retino, Le Contel, Amaya and Lapenta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Convolutional Neural Networks for Predicting the Strength of the Near-Earth Magnetic Field Caused by Interplanetary Coronal Mass Ejections

Anna Malanushenko<sup>1\*</sup>, Natasha Flyer<sup>2</sup> and Sarah Gibson<sup>1</sup>

<sup>1</sup> High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO, United States, <sup>2</sup> Department of Applied Mathematics, University of Colorado, Boulder, CO, United States

## OPEN ACCESS

### Edited by:

Bala Poduval,  
University of New Hampshire,  
United States

### Reviewed by:

Emilia Kilpua,  
University of Helsinki, Finland  
Teresa Nieves-Chinchilla,  
National Aeronautics and Space  
Administration, United States

### \*Correspondence:

Anna Malanushenko  
annny@ucar.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 15 April 2020

**Accepted:** 05 August 2020

**Published:** 18 September 2020

### Citation:

Malanushenko A, Flyer N and Gibson S (2020) Convolutional Neural Networks for Predicting the Strength of the Near-Earth Magnetic Field Caused by Interplanetary Coronal Mass Ejections. *Front. Astron. Space Sci.* 7:62.  
doi: 10.3389/fspas.2020.00062

In this paper, we explore the potential of neural networks for making space weather predictions based on near-Sun observations. Our second goal is to determine the extent to which coronal polarimetric observations of erupting structures near the Sun encode sufficient information to predict the impact these structures will have on Earth. In particular, we focus on predicting the maximal southward component of the magnetic field (“ $-B_z$ ”) inside an interplanetary coronal mass ejection (ICME) as it impacts the Earth. We use Gibson & Low (G&L) self-similarly expanding flux rope model (Gibson and Low, 1998) as a first test for the project, which allows to consider CMEs with varying location, orientation, size, and morphology. We vary 5 parameters of the model to alter these CME properties, and generate a large database of synthetic CMEs (over 36k synthetic events). For each model CME, we synthesize near-Sun observations, as seen from an observer in quadrature (assuming the CME is directed earthwards), of either three components of the vector magnetic field,  $(B_x, B_y, B_z)$  (“Experiment 1”), or of synthetic Stokes images,  $(L/I, Az, V/I)$  (“Experiment 2”). We then allow the flux rope to expand and record  $\max\{-B_z\}$  as the ICME passes 1AU. We further conduct two separate machine learning experiments and develop two different regression-based deep convolutional neural networks (CNNs) to predict  $\max\{-B_z\}$  based on these two kinds of the near-Sun input data. Experiment 1 is a test which we do as a proof of concept, to see if a 3-channel CNN (hereafter CNN<sub>1</sub>), similar to those used in RGB image recognition, can reproduce the results of the self-similar (i.e., scale-invariant) expansion of the G&L model. Experiment 2 is less trivial, as Stokes vector is not linearly related to  $B$ , and the line-of-sight integration in the optically thin corona presents additional difficulties for interpreting the signal. This second CNN (hereafter CNN<sub>2</sub>), although resembling CNN<sub>1</sub> in Experiment 1, will have a different number of layers and set of hyperparameters due to a much more complicated mapping between the input and output data. We find that, given three components of  $B$ , CNN<sub>1</sub> can predict  $\max\{-B_z\}$  with 97% accuracy, and for three components of the Stokes vector as input, CNN<sub>2</sub> can predict  $\max\{-B_z\}$  with 95%, both measured in the relative root square error.

**Keywords:** coronal mass ejection, initiation and propagation, convolutional neural network (CNN), Gibson and Low model, interplanetary CMEs, spectropolarimetric data classification

## 1. INTRODUCTION

Geomagnetic storms are powerful disturbances in Earth's magnetosphere, which become increasingly more important as our technologies develop. The technological progress increases the quality of our lives and our understanding of world around us, yet it renders us increasingly more dependent on electricity, telecommunications, satellites, and air transportation. These are some of the areas most impacted by the geomagnetic storms (National Research Council, 2008; Eastwood et al., 2017). While the well-known Carrington event has resulted, in 1859, in some damage to telegraph lines and spectacular auroras seen as far as Hawaii (Cliver and Dietrich, 2013), an event of a similar magnitude happening today would be devastating for the modern civilization (National Research Council, 2008; Baker, 2013).

The mitigation efforts depend on our ability to predict the strength and the duration of a storm before it happens, in order to take costly, yet necessary protective measures. These include redirecting air traffic, taking measures to protect the power grid, or preparing for inevitable disturbances in telecommunications (e.g., Knipp and Gannon, 2019).

Geomagnetic storms are often caused by the collision of Earth's magnetic field with clouds of magnetized plasma, which are typically of solar origin. The strength of a storm crucially depends on the strength of the magnetic field in the cloud, and also on its orientation with respect to the Earth's own magnetic field. A large southward component of the magnetic field (hereafter  $B_z < 0$ ) will result in stronger storms than if the field had a comparable northward component ( $B_z > 0$ ). These clouds are associated with coronal mass ejections (CMEs), the eruptive phenomena observed on the Sun (Webb and Howard, 2012).

The ability to predict the strength of geomagnetic storms, consequently, relies on our ability to detect a CME as it departs the Sun, to predict whether Earth will lie on its way, and to predict its properties at the moment of its encounter with Earth (Kilpua et al., 2019). Typically, after a CME is detected, simulations of its propagation through interplanetary space are needed to predict its further evolution as an interplanetary CME, or ICME (e.g., Arge et al., 2004; Manchester et al., 2017).

CMEs are magnetic in nature (e.g., Chen, 2011; Bak-Stęlicka et al., 2013; Forland et al., 2013); hence, near-Sun observations sensitive to the strength of a magnetic field are needed to determine the properties of a particular erupting structure. To obtain magnetic information, measurements of spectral line profiles of polarized light must be obtained. Unfortunately, many existing observations of magnetic field regard only the solar photosphere and chromosphere, while the CMEs are observed in the solar corona. As the corona is significantly dimmer than the solar surface in visible/infrared light, observations of CMEs in these wavelengths require a coronagraph to occult the disk of the Sun.

The spectropolarimetric measurements generally include four Stokes parameters ( $I, Q, U, V$ ) (unpolarized intensity, intensity in two directions of a linear, and intensity of a circular polarization, respectively) at several locations along a spectral line. These carry information about the magnetic field in the emitting plasma. In solar photospheric observations (e.g., Schou

et al., 2012), the Stokes vector can be inverted (e.g., Ruiz Cobo and del Toro Iniesta, 1992) to obtain components of the magnetic field, ( $B_x, B_y, B_z$ ), at the solar photosphere. But in coronal observations, the inversion is greatly complicated by the fact that corona is optically thin; the observed signal is integrated not over a relatively small range of heights, like in photospheric observations, but over hundreds of megameters. Nevertheless, although the direct inversion of Stokes data in the corona is complicated, we often observe clear signatures of magnetic structures, consistent with existing models of CMEs, in coronal spectropolarimetric measurements (Gibson, 2015). This introduces the possibility of using such observations to diagnose the magnetic field at the core of the CME at its origins at the Sun. We note, however, that because the observations are at the limb of the Sun, the CME being diagnosed would be aimed at a right angle to the observer. Ideally, one would want to make the limb observations from an instrument placed in quadrature with respect to the Sun-Earth line. Such an instrument does not yet exist but the usefulness of it can be explored with the use of synthesized observations for example using forward modeling (Gibson et al., 2016).

At least two factors will affect the ability of coronal spectropolarimetric measurements to provide a good predictor of geomagnetic storms. First, since the corona is optically thin, spectropolarimetric measurements of linearly and circularly polarized light diagnose magnetic field strength and geometry in a weighted line-of-sight integral that must be inverted to obtain magnetic field. Second, evolution between the Sun and Earth will change the erupting structure in ways that will not be captured in the measurement obtained in its early stages in the solar atmosphere. The purpose of this work is to investigate how good of a predictor such coronal signatures are for the strength of the associated geomagnetic storm at the Earth if machine learning algorithms are used, bypassing the need for inversions of line-of-sight integrated spectropolarimetric signals, and also bypassing computationally expensive simulations of how ICMEs propagate through the interplanetary space.

In this paper, the first factor is examined by the following two machine learning experiments. A particular model of a CME called G&L (Gibson and Low, 1998), described in section 2, is used for generating input and output data in both experiments. The total of 36,288 different configurations of magnetic flux ropes are generated over the 5D space of parameters that control morphology, shape, and position. We then generate two kinds of synthetic *near-Sun input* data, at the time prior to eruption: either three components of the magnetic field ( $B_x, B_y, B_z$ ) on a slice at the central meridian ("Experiment 1"), or the Stokes linear and circular polarization normalized by intensity ( $L/I, Az, V/I$ ), integrated along the line of sight ("Experiment 2"). (Note that  $L/I = \sqrt{(Q/I)^2 + (U/I)^2}$  and  $Az = -\frac{1}{2}\text{atan}(U, Q)$  contain the same information as  $Q$  and  $U$  normalized by  $I$ .) We also generate synthetic *IAU output* data, common for both experiments: as the flux rope expands, we record the maximal southward component of the magnetic field (hereafter  $\max\{-B_z\}$ ) within the ICME flux rope as it impacts the Earth and drives the storm.

In Experiment 1, a binary classifier using fully connected feedforward neural network (FNN), is constructed to determine whether the ICME is launched in such a direction as to impact Earth (a “storm” or a “no storm” prediction), based on an input image of one of the three magnetic field components. We further refer to this network as FNN<sub>1</sub>. For those ICMEs classified as a “storm,” we further construct a 3-channel convolutional neural network (CNN), which inputs all three components of the magnetic field, and has the objective to predict  $\max\{-B_z\}$  at 1AU. We further refer to this network as CNN<sub>1</sub>.

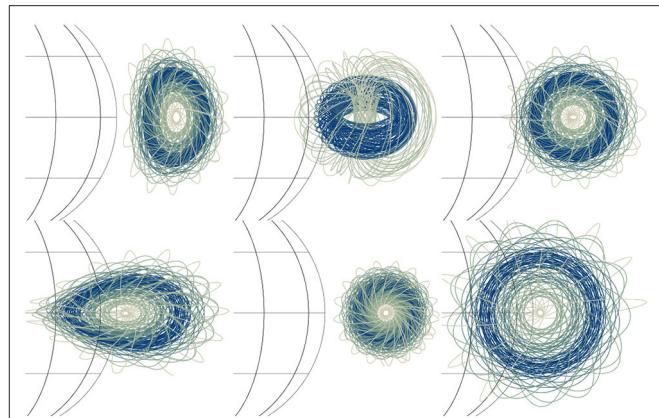
Based on the optimistic results from CNN<sub>1</sub>, we perform a separate Experiment 2 with synthetic spectropolarimetric data. As these data are of a different nature, we create a new set of neural networks, independent from Experiment 1, but with similar type architectures. For this experiment, the input data are pre-eruption near-Sun coronal spectropolarimetric measurements. Another binary classifier, hereafter FNN<sub>2</sub>, uses spectropolarimetric data in different combinations (only circularly polarized light, only linearly, and all three components), to make, as in Experiment 1, a “storm” / “no storm” prediction. Next, a new CNN, hereafter CNN<sub>2</sub>, is constructed in the similar fashion as in Experiment 1. It is again run on those events classified as a “storm,” and uses all three components of the Stokes data, to predict  $\max\{-B_z\}$ . In this way, we take into account the line of sight weighting of the observable quantity, and also consider the different diagnostic potential of circularly vs. linearly polarized light.

The paper is structured as follows. In section 2, we describe the G&L model used for synthesizing the data and describe in detail the parameters which were varied. In section 3, we describe the synthetic input and output data, and the architecture of neural networks which we constructed for each experiment. In section 4, we report the results of each experiment and demonstrate that in both cases of input data, both networks could make successful predictions. Finally, in section 5 we discuss the implication of the results and outline the path for future development of this machine learning application.

## 2. ANALYTIC CME MODEL USED FOR TRAINING AND TESTING CNN

We use a CME model called G&L (Gibson and Low, 1998) to build a database of erupting flux ropes with varying characteristics. G&L is an analytical 3D magnetohydrodynamic (MHD) model with a spheromak-like magnetic flux rope<sup>1</sup> embedded in a bipolar background magnetic field and radially symmetric background hydrostatic atmosphere. The bubble which contains the flux rope expands self-similarly with time, modeling the propagation of a CME through the interplanetary space. While the self-similar expansion is a rather idealized model of CME propagation, the current work is meant to prepare the framework for the next step of the project. In the next step,

<sup>1</sup>(A specific analytical solution for a toroidal flux rope embedded in a spherical shell, which is a common object of study, e.g., in laboratory plasma research and in solar physics, e.g., Hagenon and Krakowski , 1987; Gibson and Low, 1998; Borovikov et al., 2017, etc.)



**FIGURE 1 |** Examples of various initial G&L configurations. Blue and green lines are magnetic field lines sampling the magnetic structure of the G&L spheromak, shown here with different combinations of parameters for angular size, topology, orientation. The solar surface is shown in thin black lines for reference.

**TABLE 1 |** Ranges of the G&L parameters.

Parameter	Notation	Min	Max	<i>N<sub>steps</sub></i> <sup>a</sup>
Front height	<i>f</i>	$1.3R_{\odot}$	$2.5R_{\odot}$	6
Angular size	$\Omega$	$10^{\circ}$	$45^{\circ}$	8
Topology	$\tau_1$	0.5	$\min(4.1, \tau_1, \text{max})^b$	$6^c$
Orientation	$\sigma$	$0^{\circ}$	$330^{\circ}$	12
Latitude	$\theta$	$-0.8\Omega/2$	$0.8\Omega/2$	9

<sup>a</sup>Uniform steps for all parameters.

<sup>b</sup>See section 2.3 for description.

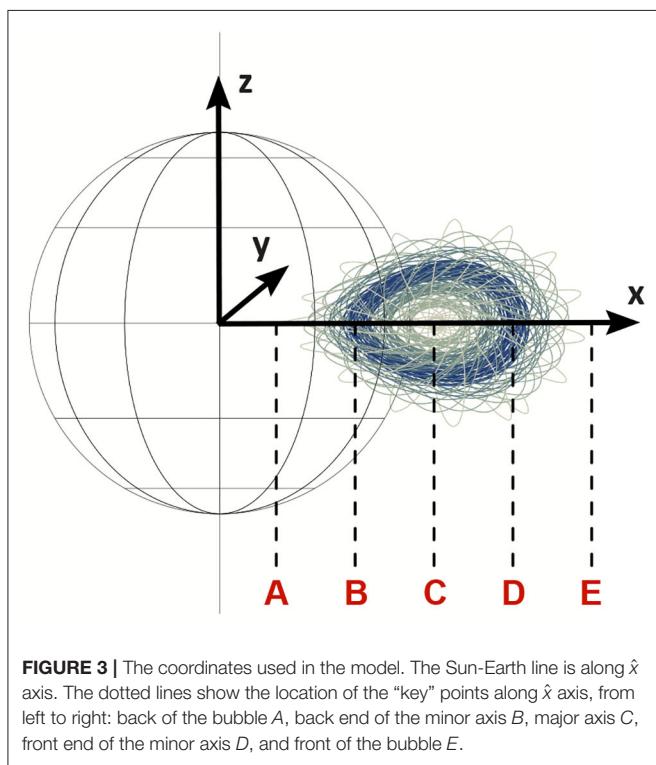
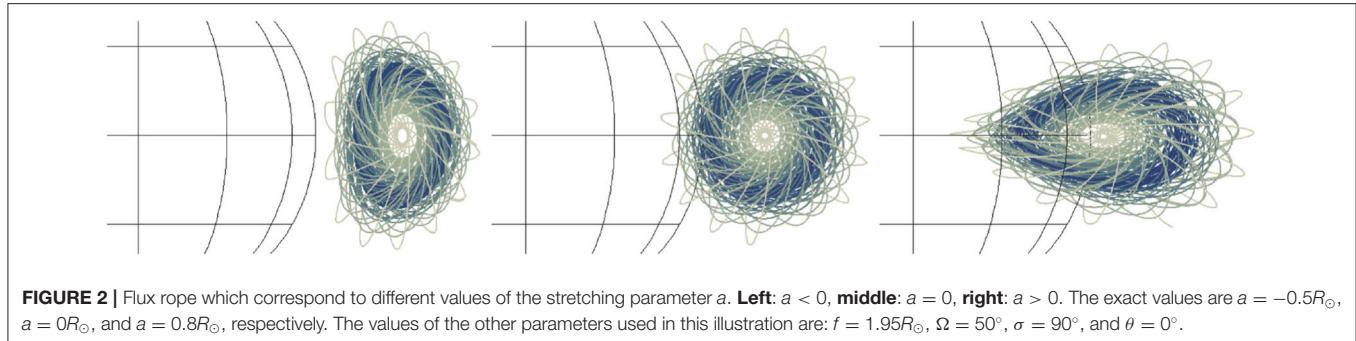
<sup>c</sup>See note on coupling of the parameters in section 2.2.1.

we will keep the structure of the project, but instead of a self-similar expansion, will use results from more realistic MHD simulations (Merkin et al., 2016) of ICME propagation, which will nonetheless use the same G&L flux ropes as initial condition. This is further explained in section 5.

A given G&L solution depends on a large number of analytical and empirical parameters (Gibson et al., 2016). For the purpose of this study, we choose to vary 5 parameters most relevant to the shape, topology, and position of the flux rope These parameters are: height of the front of the CME, its angular width, the “topology” parameter, the rotation about the Sun-to-1-AU line that passes through the center of the CME, and latitude from which the CME is launched. The further sections describe the parameters in detail. Several of the solutions from the database are shown in Figure 1. The ranges of each parameter are further listed in Table 1.

### 2.1. Size and Initial Height Parameters

The original Gibson and Low (1998) defined the geometry of the embedded spheromak through the following three parameters. The bubble of radius  $r_0$  was located at a distance  $x_0$  from the origin (Sun’s center), and the subsequent stretching transformation  $r \rightarrow r + a$  was applied in spherical coordinates,



where  $a = \text{const}$  is a stretching parameter, constant for the entire domain. Further, in the FORWARD suite which we use for the calculations (Gibson et al., 2016),  $r_0$  and  $x_0$  were replaced by two parameters which are more directly related to observations: the angular size of the bubble  $\Omega$ , defined as  $\tan(\Omega/2) = r_0/x_0$ , and the height of the front of the bubble  $f = x_0 + r_0 - a$ . By definition in the model, all points for which  $r \leq a$  are fixed at the origin ( $r = 0$ ) and stay in the origin during the self-similar expansion. The resulting solution has a spherical bubble for  $a = 0$ , teardrop-shaped bubble for  $a > 0$ , and umbrella-shaped bubble for  $a < 0$ , as shown in **Figure 2**.

## 2.2. Topology/Stretching Parameter

In this study, we introduce a new parameter to replace  $a$ , which is more directly related to the topology of the over-the-limb portion of the flux rope:  $\tau_1$ . Consider the coordinates shown in **Figure 3**,

with  $\hat{z}$  normal to the plane of the solar equator, and the flux rope propagating radially along  $\hat{x}$  axis along the Sun-Earth line.

We notice that there are several geometrically special points along  $\hat{x}$  axis, shown in **Figure 3**. E.g., consider point A which is the back end of the bubble (or closest to the origin). If the value of  $a$  is such that  $x_A > R_{\odot}$  (or  $x_0 - r_0 - a > R_{\odot}$ )<sup>2</sup>, then the entire spheromak is above solar surface (called the photosphere), and field lines are generally infinite, wrapping infinitely about the core of the spheromak. If  $a$  is such that  $x_A < R_{\odot}$ , but  $x_B > R_{\odot}$ , then some portion of the bubble will be under the solar surface prior to “eruption”; for the purpose of the study we keep track of this portion during the self-similar expansion and do not include this portion in the calculations of the strength of geomagnetic storm. The coronal portion of the bubble will contain spheromak-like field lines, as well as field lines which begin and end at the solar surface—the overall configuration will appear as a spheromak suspended in sheared-arcade of field lines. Similarly,  $x_B < R_{\odot} < x_C$  would create an apparent classical flux-rope configuration (e.g., Gold and Hoyle , 1960; Fan and Gibson, 2003), in the sense that both footpoints of field lines are anchored at the photosphere, but the field lines wrap around a common arch-like axis and have dips which potentially could support cool prominence material (Fan and Liu, 2019). Further,  $x_C < R_{\odot} < x_E$  would mean that all field lines above the  $r = R_{\odot}$  surface are arches and the overall coronal portion would appear to be a sheared arcade.

For the purpose of this study, we ignore the portion of the flux rope which is underneath the solar surface prior to eruption (see section 3 for detail on how is this implemented in the output data.) The parameter  $\tau_1 = \frac{2(-1)(k+1)}{k(f+a)}$ , where  $k = \tan(\Omega/2) = r_0/x_0$ , is an empirical dimensionless parameter related to how many special points are above the photosphere. For example,  $\tau_1 < 0$  means that all special points are below  $r = R_{\odot}$ ,  $0 < \tau_1 < 1$  means one of these points is above  $r = R_{\odot}$ , etc, and  $\tau_1 > 4$  means the entire bubble is above the surface<sup>3</sup>. Note that  $\tau_1$  is also related to the shape of the bubble: decreasing  $\tau_1$  while keeping constant  $f$  and  $\Omega$  will make the bubble more teardrop-stretched. For example, **Figure 2** shows, from left to right, solutions for  $\tau_1 \approx 4.1, 3.2, 2.2$ , with 5, 4, and 3 special points above the surface

<sup>2</sup>We hereafter assume a system of units in which  $R_{\odot} = 1$ .

<sup>3</sup>To see this, one could express both  $\tau_1$  and the coordinates of the special points via  $x_0$ ,  $r_0$ , and  $a$ :  $\tau_1 = \frac{x_0 - a - 1}{r_0/2} + 2$ ;  $x_A = x_0 - r_0 - a$ , etc.

respectively. Varying  $\tau_1$  allows us to sample qualitatively different magnetic configurations, including, for example, the classical flux rope ( $2 \leq \tau_1 \leq 3$ ) and the toroidal spheromak ( $\tau_1 \geq 4$ ). Observational evidence for different topologies including simple flux ropes (Bak-Stęślicka et al., 2013) and spheromaks (Dove et al., 2011) have been found in polarimetric observations of coronal cavities, known to be precursors of CMEs (Gibson, 2015).

Lastly, we notice that the value of  $\tau_1$  for which  $a = 0$  and the bubble is spherical is  $\tau_{1,\max} = 2(f-1)(k+1)/(kf)$ . We do not use solutions for  $a < 0$ , or equivalently, of  $\tau_1 > \tau_{1,\max}$ , as they are deemed unphysical<sup>4</sup>. In addition, choices of  $\tau_1 > 4$  result in a flux rope bubble that hovers above the photosphere at  $t = 0$ . Coronal cavity CME precursors tend to maintain some measure of contact with the photosphere—although, for highly stretched, teardrop-shaped bubbles some hovering is observed (e.g., Forland et al., 2013). Therefore, we cap the range of  $\tau_1$  by the *smallest* of two values:  $\tau_1 < \min(4.1, \tau_{1,\max})$  ensuring that none of the bubbles are in either the unphysical  $a < 0$  regime, nor too far from the photosphere before the eruption. Note that this implies that the parameter space is not uniform in  $\tau_1$ : for each pair of  $(f, \Omega)$ , the range of  $\tau_1$  is calculated independently, and uniform steps in  $\tau_1$  are taken in that range.

### 2.2.1. Coupling of the Parameters

The three parameters described above,  $f$ ,  $\Omega$ , and  $\tau_1$ , are *secondary* in the sense that they have been introduced and can be expressed through the *primary* parameters which define properties of the spheromak in the G&L model,  $x_0$ ,  $r_0$ , and  $a$ :

$$\begin{aligned} f &= x_0 + r_0 - a, \\ \tan(\Omega/2) &= k = r_0/x_0, \\ \tau_1 &= \frac{x_0-a-1}{r_0/2} + 2. \end{aligned} \quad (1)$$

However, the secondary parameters can be more easily related to the observed properties of CMEs (e.g.,  $x_0$  is the location of the spheromak's center *before* the stretching has been applied, and the resulting position of the bubble would in general be different.) To make our study more useful for future work in CME predictions, we invert the system (1) and express primary parameters through the secondary ones:

$$\begin{aligned} x_0 &= (f + a)/(k + 1), \\ r_0 &= k(f + a)/(k + 1), \\ a &= (f - k - 1 - (\tau_1 - 2)(k + 1))/k. \end{aligned} \quad (2)$$

We then vary the secondary parameters, in the ranges shown in **Table 1**, calculate the primary parameters through (2), and use these as the input to the G&L model.

### 2.3. Orientation and Latitude Parameters

The orientation of the spheromak's core, with respect to the Earth's North, is an important parameter for determining the strength of the geomagnetic storm. We vary parameter  $\sigma$  that defines rotation of the bubble about the radial direction. **Figure 4**

<sup>4</sup>In Gibson and Low (1998), the stretching parameter  $a$  is introduced in MHD equations as the additional gravity term. Therefore,  $a < 0$  will physically mean a gravity force directed away from the Sun's center.

shows several solutions with varying  $\sigma$ .

The last parameter that we vary is the starting latitude of the CME,  $\theta$ . Since the subsequent expansion is self-similar, the direction of propagation is radial and a CME remains within the same solid angle at all times. The Sun-Earth line is along  $\hat{x}$ , so CMEs launched at  $\theta > \Omega/2$  with respect to the equator will miss Earth completely, and  $\theta \approx \Omega/2$  will have a vanishingly small effect on Earth. As we show later, the CNN is taught to determine which CMEs will result in a “storm,” and which will result in “no storm” at the Earth. We adjust  $\sigma$  range so that about 82% of all CMEs result in a “storm” at the Earth, in the sense of  $\max\{-B_z\} > 10^{-7.7}\text{G} \approx 2 \times 10^{-3}\text{nT}$ .

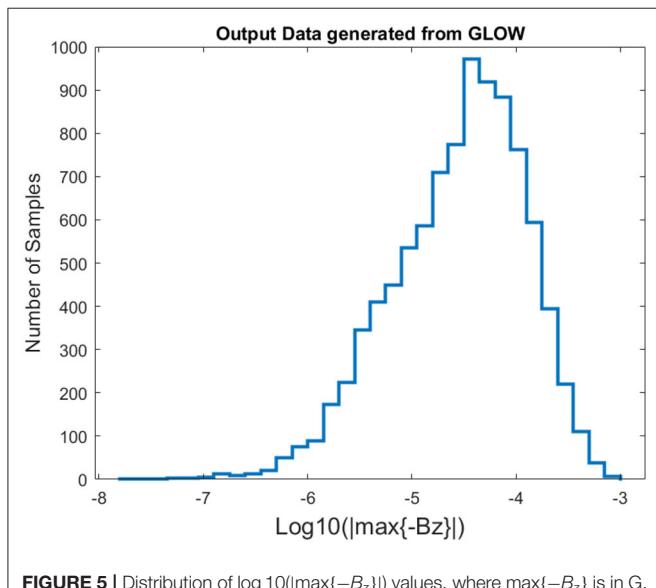
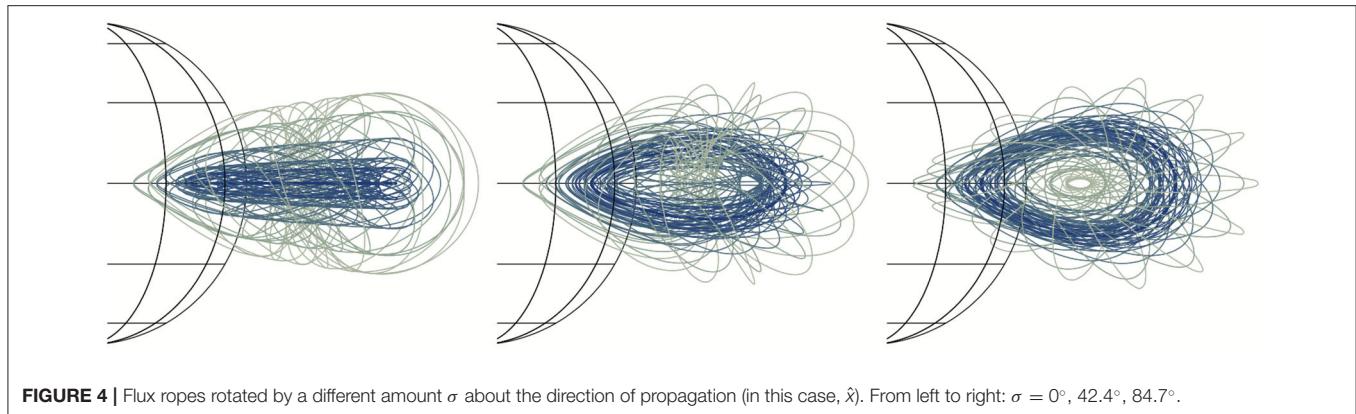
## 3. NEURAL NETWORKS: DATA AND ARCHITECTURES

Two experiments are conducted, each with two neural networks (NNs)—a feedforward fully connected neural network (FNN) and a 3 channel convolutional network (CNN) that can also accept 1 or 2 channels of data. For each experiment, these NN are different. Hence, there are a total of 4 separate NNs. The FNN acts as a binary classifier, determining whether there is a “storm”<sup>5</sup> or a “no storm.” Both  $\text{CNN}_1$  in Experiment 1 and  $\text{CNN}_2$  in Experiment 2 perform a non-linear regression to estimate  $\max\{-B_z\}$  inside the flux rope as it impacts Earth and drives the storm, for those events classified as “storms.” The big difference between the experiments is 1) the input data and 2) the NN architectures, due to a change in the characteristics of the input data. Both will be discussed in the following subsections. We refer the readers who are not familiar with FNN and CNN to textbooks and overview articles on the topic (e.g., Svozil et al., 1997; Goodfellow et al., 2016; Yamashita et al., 2018). All code was programmed in MATLAB package “Deep Learning Toolbox” (The MathWorks , 2019).

For both experiments, the output data are the same—the maximal negative (southward) amplitude of the  $B_z$  component in the flux rope at 1AU:  $\max\{-B_z(t)\}$ , which serves as a proxy for the strength of the geomagnetic storm. It is calculated as follows. The flux rope was allowed to expand self-similarly, and time profile of  $B_z(t, r = 1\text{AU})$  was stored, starting from the time the expanding bubble first encounters the  $r = 1\text{AU}$  sphere, and ending at the time when the plasma elements which were at  $r = 1R_\odot$  at  $t = 0$  first encounter the  $r = 1\text{AU}$  sphere. Note that the plasma elements which were under the solar surface prior to eruption are excluded from the time series; as the rate of expansion is known, this is a trivial task. We do this to emulate the eruption of only a portion of the spheromak geometry.

Since the non-zero values of  $\max\{-B_z\}$  range from  $10^{-7.7}\text{G}$  to  $10^{-3}\text{G}$ , we work in a  $\log_{10}$  scale. The distribution of the output,  $\log_{10}\{\max\{-B_z\}\}$ , is shown in **Figure 5**. Not shown in the figure (due to dwarfing the distribution) is a spike at 0 of 6517 samples (18% of the total number of samples) corresponding to CMEs that miss the Earth. For values less than  $10^{-6.3}\text{G}$ , there are less

<sup>5</sup>In the sense of  $\max\{-B_z\} > 10^{-7.7}\text{G} \approx 2 \times 10^{-3}\text{nT}$



than 40 samples in each histogram bin, resulting in a long tail of the distribution. Due to the low sample size, the CNN will not be able to make adequate predictions in this regime.

Regardless of the experiment:

1. The data were separated into 2 vectors, grouped by “storm” ( $\max\{-B_z\} > 10^{-7.7}\text{G}$ ; 82% of the data) and “no storm” ( $\max\{-B_z\} < 10^{-7.7}\text{G}$ ; 18% of the data).
2. The pointers to those two groups were randomly permuted, thus shuffling the data.
3. Two-thirds of the pointers were used for training, and one-third for testing, maintaining the 82 to 18% ratio of “storm” to “no storm” in both the training and testing data.
4. The input observations (be it magnetic field components or Stokes parameters) are in the solar equator plane in quadrature with Earth, along the  $\hat{y}$  direction. The field of views (FOVs) are identical in both experiments ( $x \in [1.0, 2.6]R_\odot$ ,  $z \in [-0.8, 0.8]R_\odot$ ). The observer is located at infinite distance

from the origin, so that the lines-of-sight are parallel to each other for all image pixels.

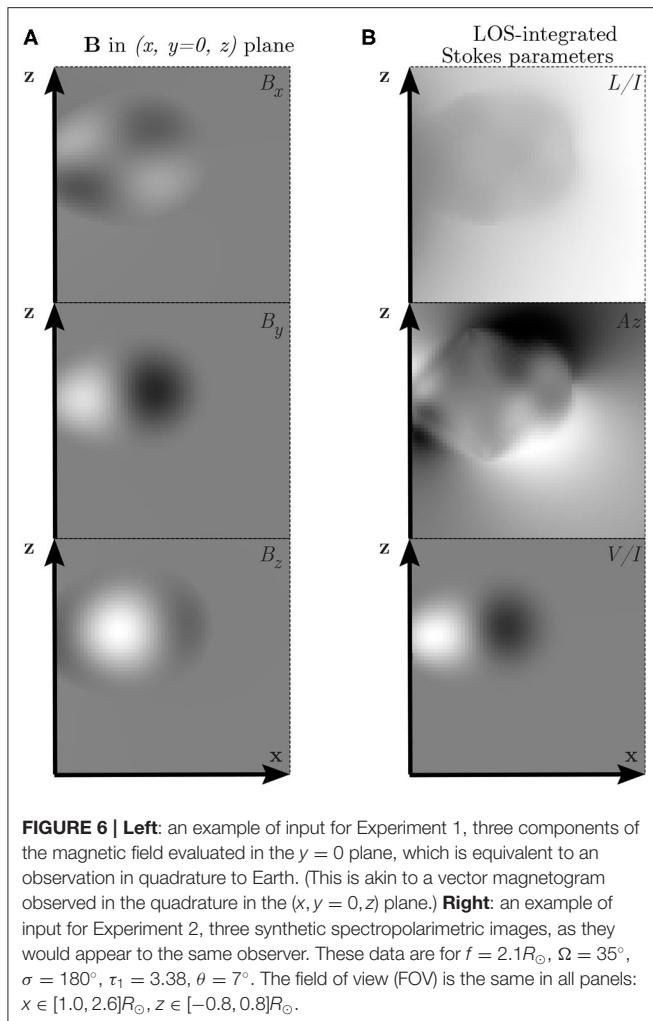
5. The architecture of the FNN in each experiment remained the same regardless of the input data. However, for each different type of input the FNN was retrained, as the numerical values of the weights and bias is dependent on the type of input data (e.g.,  $(L/I, Az)$ , or  $V/I$ ).
6. The FNN were run on an Intel CORE i7 9th generation CPU with 16 GB of RAM. The CNN were run on a NVIDIA Quadro P6000 GPU with 24 GB GDDR5X.

### 3.1. Experiment 1: Input Data

The G&L model is run to generate a total of 36288 magnetic 3D magnetic flux ropes that span a 5D parameter space (see Table 1). Two thirds (24192 flux ropes, 19847 “storm” + 4345 “no storm”) were used for training the machine learning algorithms and one third (12096 = 9923 “storm” + 2173 “no storm”) was used for testing the quality of the predictions. For Experiment 1, a single input sample showing a slice of all three magnetic field components  $B_x$ ,  $B_y$ , and  $B_z$ , is given in the left panel of Figure 6, represented by a slice of the 3D datacube at the  $y = 0$  plane. At present, no instrument can provide such input off the solar limb, but this Experiment is by design but a sanity-check test for the overall machine-learning pipeline.

Although Figure 6 shows gray-scale images, the input for Experiment 1 are  $64 \times 64$  matrices of magnetic-field values in Gauss. While We are using the concept of image pattern recognition architectures (but with regression) to see if we can, instead of the byte-scaled images, use matrices that have values of a coronal magnetic field or its corresponding Stokes parameters that, for example, can range continuously from 0.7235 to  $3.52513 \times 10^{-5}$  to -0.764515. Further, for both FNNs we convert the input from a  $64 \times 64$  matrix into a  $4096 \times 1$  vector.

For the training of the FNN for Experiment 1 (hereafter FNN<sub>1</sub>), only one of these magnetic field components should be needed as an input (given as a  $4069 \times 24192$  matrix); this is because, in this experiment, FNN<sub>1</sub> is a binary classifier for a self-similar mapping—which means that due to the underlying symmetry of the solution, the number of variables is reduced. For training of the CNN for Experiment 1 (hereafter CNN<sub>1</sub>), we use all 3 channels as input (given as a  $64 \times 64 \times 3$  matrix),



since although we still have a self-similar mapping, we are trying to estimate a number,  $\max\{-B_z\}$ , to at least one decimal place of accuracy that varies over 5 orders of magnitude, not just (1, “storm”) or (0, “no storm”) as in FNN<sub>2</sub>. Note that CNN<sub>1</sub> predicts the output that is simply a scaled-down version of the input, evaluated at a point. CNN<sub>1</sub> is simply finding this scaling or mapping, which is what NNs are very good at.

### 3.2. Experiment 1: FNN and CNN Architecture

The first network, FNN<sub>1</sub>, is an FNN with one hidden layer. Its architecture is given in Figure 7 and its properties are listed below. The figure is generated by the simple command “view [(name of the NN)]” in MATLAB’s Deep Learning Toolbox.

1. FNN<sub>1</sub> has an input of a vector 4096 by 1, representing either the B<sub>x</sub>, B<sub>y</sub>, or B<sub>z</sub> component of a GLOW flux rope.
2. It has 1 hidden layer with 10 neurons. In the Figure 7, the box with “w” and the box with “b” represent the weights and the biases for that layer, respectively.

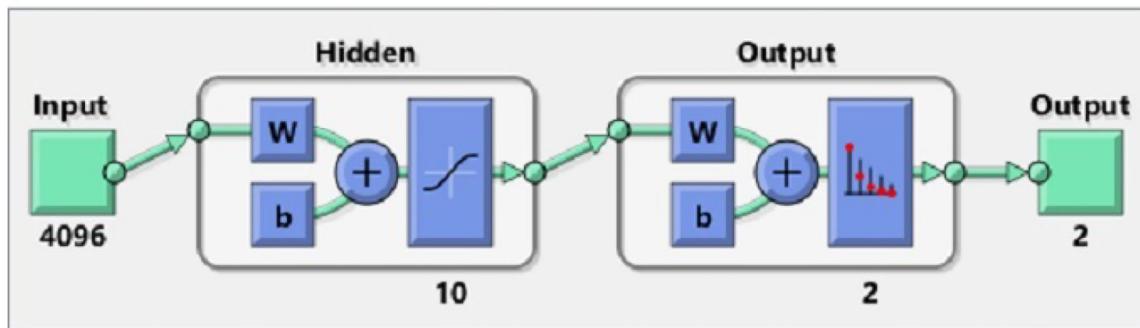
3. The hyperbolic tangent curve in the hidden layer box indicates that we are using a tanh activation function. Approximating a binary classifier, where the probabilities fall between certain ranges, is easier with combinations of tanh( $x$ ) than the more popular piecewise linear functions (ReLU, e.g., Glorot et al., 2011) that is used in more complex NN.
4. The hidden layer is then connected to the output layer that uses a soft-max activation function, to calculate the probabilities of the ICME being a “storm” or “no storm,” represented by the graph with red dots.
5. Lastly, the cross-entropy loss function is used to measure the performance of the network against the true labels.
6. The stochastic gradient descent as the optimization algorithm.
7. The network was trained for 500 epochs.
8. MATLAB’s feedforwardnet in the Deep Learning Toolbox was used.

The final output is a probability  $p$ ,  $0 \leq p \leq 1$ . For this paper, we assume that if  $p \geq 0.8$  then the output is classified as a “storm”; if  $p \leq 0.2$  it is a “no storm”; if  $0.2 < p < 0.8$  we assume the network was unable to classify the output (in other words, the results are *inconclusive*).

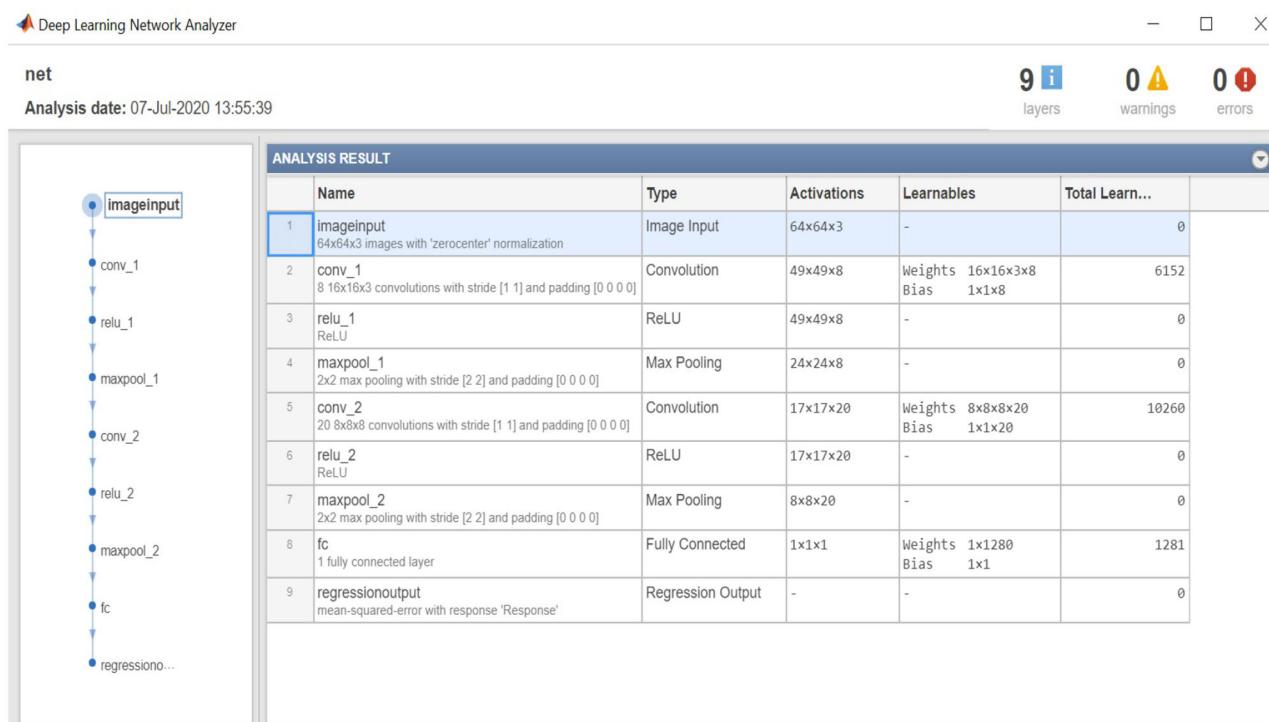
All the “storms” correctly identified by the feedforward network are then passed into CNN<sub>1</sub> for measuring the strength of  $\log_{10}\{\max\{-B_z\}\}$  of the passing CME. The architecture of CNN<sub>1</sub> for Experiment 1 is given in Figure 8 with its properties listed below. The figure is generated by the simple command “analyzeNetwork([(name of the NN)])” in MATLAB’s Deep Learning Toolbox.

1. The first layer is an image input layer of 3 channels (that is, the values of B<sub>x</sub>, B<sub>y</sub>, B<sub>z</sub> near the Sun). Consequently, it has dimension of  $64 \times 64 \times 3$ .
2. There are two 2D convolutional layers, the first with 8  $16 \times 16$  filters, the second with 20  $8 \times 8$  filters. CNN<sub>1</sub> is a shallow network due to the fact that we are simply mapping the magnetic field to a self-similar counterpart.
3. Each convolutional layer is followed by a ReLU activation function (Glorot et al., 2011).
4. Each ReLU layer is followed by a maxpooling layer, taking the maximum value from each  $2 \times 2$  pixel region, resulting in a reduction of the matrix by a factor of 4.
5. The last layer is a fully connected layer which gives a prediction for  $\log_{10}\{\max\{-B_z\}\}$ .
6. The error, compared to ground truth, is calculated by the root mean square error function and optimized through backpropagation with the Adam optimizer (Kingma and Ba, 2014).
7. The output values were normalized using a linear translation to between 0 and 1.
8. It was trained on 100 epochs.
9. A batch size of 567 was used.
10. The initial learn rate was set at 0.001.
11. The learning rate schedule was a piecewise drop with a learn rate drop of 0.75 and a learn rate drop period of 6.
12. An L2 Regularization of  $10^{-4}$  was added.

## Experiment 1: Feedforward Neural Network



**FIGURE 7 |** The feedforward fully connected network (FNN) used in Experiment 1 ( $\text{FNN}_1$ ). The letters “w” and “b” denote weights and biases, respectively; the full description of the scheme is given in the itemized list in section 3.2.



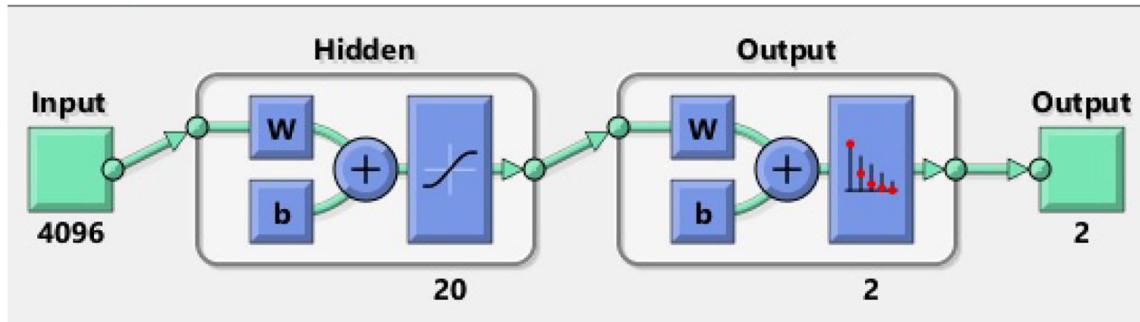
**FIGURE 8 |** The 9 layer architecture of the convolutional neural network (CNN) constructed for Experiment 1 ( $\text{CNN}_1$ ) with the number of activations per layer and total number of learnable weights and bias in the convolutional layers and fully connected layer.

### 3.3. Experiment 2: Input Data

For Experiment 2, the input data are synthetic spectropolarimetric data corresponding to the magnetic fields generated for Experiment 1. A single input sample of the 36,288 synthetic Stokes observables is shown in the right panel of **Figure 6**. Although **Figure 6** shows gray-scale images, the input for Experiment 2 are  $64 \times 64$  matrices of the Stokes vector values in the same FOV and with the same resolution as Experiment 1. We use FORCOMP (CLE) package (Judge and Casini , 2001)

of FORWARD suite (Gibson et al., 2016) in SolarSoft IDL (Freeland and Handy, 1998), to synthesize the components of the Stokes vector,  $(I, Q, U, V)$ . In this work we use an alternative vector, which is derived from the Stokes vector:  $(I, L, Az, V)$ . We further focus on the last three components of it, normalized by the intensity  $I$ ,  $(L/I, Az, V/I)$ .  $L/I = \sqrt{(Q/I)^2 + (U/I)^2}$  and  $Az = -\frac{1}{2}\text{atan}(U/Q)$  contain the same information as  $Q$  and  $U$ . We find this representation more useful because it describes the linear polarization in terms of magnitude  $(L/I)$  and polarization

## Experiment 2: Feedforward Neural Network



**FIGURE 9 |** The feedforward fully connected network used in Experiment 2 (FNN<sub>2</sub>) The full description of the scheme is given in the itemized list in section 3.4.

angle ( $Az$ ). In particular, it allows us to plot the polarization angle with respect to the local vertical (solar radial coordinate) and show how linear polarization is rotated in the presence of magnetic field. It also lets us immediately identify regions of highly reduced linear polarization associated with van Vleck angles and magnetic nulls (Gibson, 2018).

In this experiment, we are mapping observables measured in one space (Stokes data) to those measured in another space (magnetic field strength). The input and output data is no longer related by a mere scaling factor as in Experiment 1. Therefore, we will use either linearly polarized light ( $L/I, Az$ ), or circularly polarized light ( $V/I$ ), or the combination of the two, as various input data into the FNN constructed for Experiment 2 (hereafter FNN<sub>2</sub>) to understand the contribution of each to the results. Those that are classified correctly as “storms” are then inputted into the CNN architecture for Experiment 2 (hereafter CNN<sub>2</sub>).

### 3.4. Experiment 2: FNN and CNN Architecture

The fact that the functional mapping with the NNs is no longer self-similar and consequently, we do not have a reduction in the variables, will alter how we set up the architecture, requiring more degrees of freedom. In addition, we are passing the original G&L data through a highly non-linear model to synthesize Stokes images from the full-MHD variables, and yet we still expect CNN<sub>2</sub> to map back to the original  $\max\{-B_z\}$  output of the G&L model.

The architecture of FNN<sub>2</sub> for Experiment 2 is given in Figure 9. All the properties are the same as for FNN<sub>1</sub> from Experiment 1, except for those enumerated below.

1. FNN<sub>2</sub> has an input of a vector  $4096 \times 1, 8192 \times 1$ , or  $12,288 \times 1$ , depending on whether we are inputting  $V/I$ ,  $(L/I, AZ)$ , or all three channels.
2. It has 1 hidden layer but now with 20 neurons.
3. There is a L2 regularization added with a parameter of  $5e-4$ .
4. It was trained for 1,500 epochs.

As in Experiment 1, those events classified correctly as “storms” (i.e., the above mentioned  $p \geq 0.8$ ) are then inputted into CNN<sub>2</sub>.

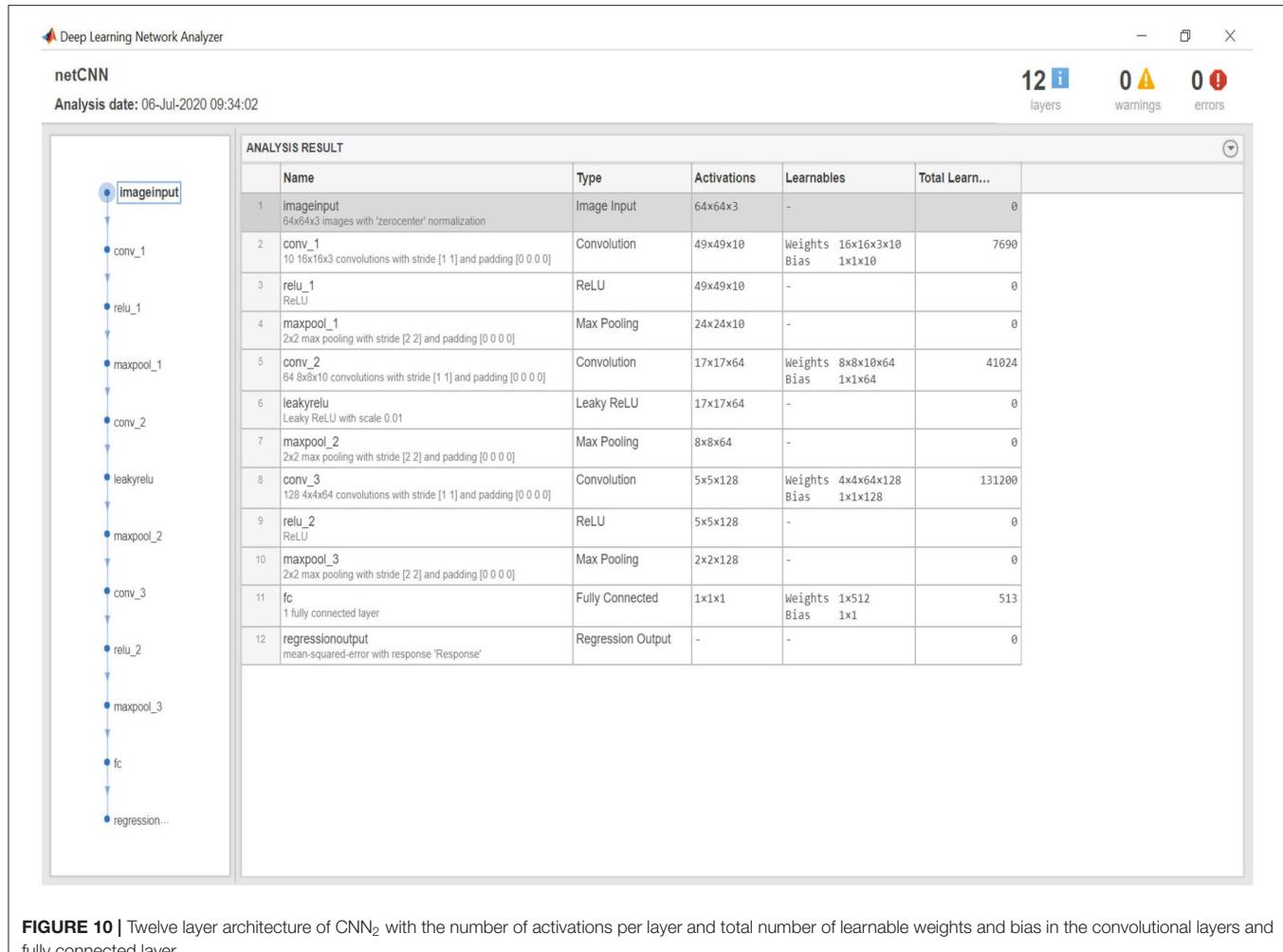
The architecture of CNN<sub>2</sub> is given in Figure 10 and its properties are listed below.

1. In the case for the Stokes parameters, the input layer can either be of size  $64 \times 64 \times 1, 64 \times 64 \times 2$ , or  $64 \times 64 \times 3$ , depending on how many channels we are using.
2. There are now three 2D convolutional layers, the first with  $10 16 \times 16$  filters, the second with  $64 8 \times 8$  filters, and the third with  $128 4 \times 4$  filters.
3. The initial random weights for each convolutional layer are defined using He initialization (He et al., 2015).
4. Each convolutional layer is followed by a ReLU activation function (Glorot et al., 2011).
5. Each ReLU layer is followed by a maxpooling layer, taking the maximum value from each  $2 \times 2$  pixel region, resulting in a reduction of the matrix by a factor of 4.
6. The last layer is a fully connected layer which gives a prediction for  $\log_{10}\{\max\{-B_z\}\}$ .
7. Again, the error was calculated by the root mean square error function and optimized with the Adam optimizer (Kingma and Ba, 2014).
8. Again, the output values for training were normalized using a linear translation to between 0 and 1.
9.  $V/I$  was multiplied by  $10^3$  so that it would be of the same order of magnitude as  $L/I$  and  $Az$  for training.
10. It was trained on 160 epochs.
11. A batch size of 735 was used.
12. The initial learn rate was set at 0.0015.
13. The learning rate schedule was a piecewise drop with a learn rate drop of 0.75 and a learn rate drop period of 10.
14. An L2 Regularization of  $3 \times 10^{-4}$  was added.

## 4. RESULTS

### 4.1. Experiment 1

This first experiment is a proof of concept. Essentially, using  $B_{xyz}$  near Sun as inputs and predicting  $B_z$  at 1AU means we are simply trying to teach NNs to predict a simple scaling of a number. If we can not succeed in this, then we cannot even hope



**FIGURE 10 |** Twelve layer architecture of CNN<sub>2</sub> with the number of activations per layer and total number of learnable weights and bias in the convolutional layers and fully connected layer.

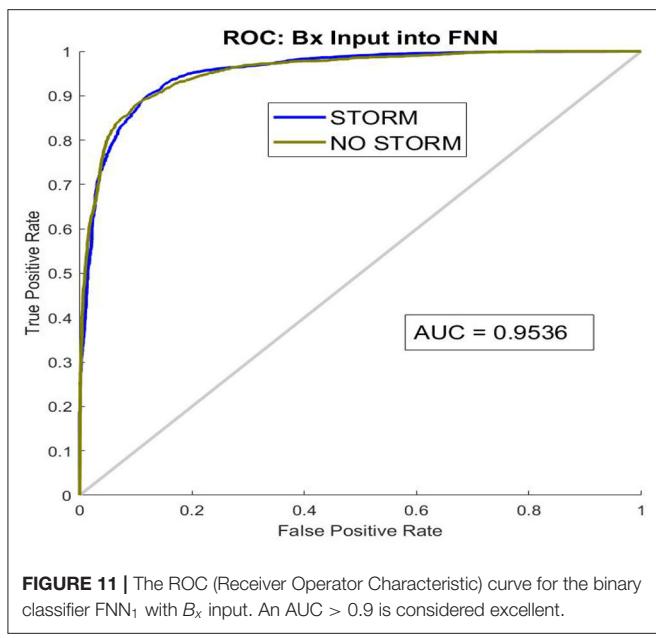
to replace the G&L model output at the Earth with model output from MHD CME solar wind models, nor can we hope to predict the strength of magnetic field in passing ICMEs from coronal spectropolarimetric observations, as is done in Experiment 2 (albeit using synthetic observations).

We compare the results of the first (predictor, “storm” / “no storm”) NN against random guesses. **Table 2** shows the outcome of random guessing: a guess was phrased (with probability of the outcomes weighted by their known ratio in the data), then an outcome was selected. The numbers in **Table 2** can be easily calculated analytically as follows. Consider  $N$  events, out of which  $qN$  are storms, and  $(1 - q)N$  are not storms ( $0 \leq q \leq 1$ ). Suppose the random guesser knows the value of  $q$  *a priori*, but makes predictions at random, weighted by  $q$ . It will classify  $qN$  events as storms. However, if predictions are truly random, amongst these  $qN$  events the fraction of actual storms will still be  $q$ ; therefore, a random guesser will correctly predict a storm for  $q^2N$  events, and will make false positive prediction for  $(1 - q)qN$  events; likewise, the amount of events correctly predicted as not storms and the amount of false negatives would be  $(1 - q)^2N$  and  $q(1 - q)N$  respectively. In our case,  $q = 9923/12096 \approx 0.82$ . A

successful FNN prediction must result in a table more diagonal than **Table 2**.

**Tables 3–5** show the results when only one of the components was taken as input. Using either of the three components, the predictor-based FNN<sub>1</sub> can predict the outcome (“storm” / “no storm”) significantly better than a random guess. The  $B_z$  input produces the best predictions for the output, which is hardly surprising. The  $B_x$  input results in the worst prediction of the three (but still significantly better than a random guess). This could be caused by the fact that the  $\hat{y}$  and  $\hat{z}$  components of the flux rope are coupled by the orientation parameter  $\sigma$ , while the  $\hat{x}$  component is independent of both. However, even in the case of using  $B_x$  as input to the binary classifier, the ROC curve (a statistical tool to give the diagnostic ability of a binary classifier) gives excellent results with an AUC (area under curve) of 0.95 (a perfect predictor would have AUC=1, that is a zero false positive rate) as shown in **Figure 11**.

We further examine the efficiency of CNN<sub>1</sub>, the regression, for predicting the strength of the storm. We use the outcome of the FNN<sub>1</sub> classifier based on  $B_x$  component as an input since, as is evident from the previous paragraph, it proves the most

**TABLE 2 |** No FNN, random guess (weighted by storm:no storm ratio).

Prediction Reality \	Storm	No storm	Inconclusive
Storm	67.3% (8,140)	14.7% (1,783)	—
No storm	14.7% (1,783)	3.2% (390)	—

Hereafter, the values are reported as: “A% (B),” where A is percentage of the total 12096 events, and B is the actual number of events.

**TABLE 3 |** Experiment 1: FNN<sub>1</sub> using  $B_x$  only.

Prediction Reality \	Storm	No storm	Inconclusive
Storm	76.3% (9,228)	1.6% (191)	7.7% (933)
No storm	2.3% (274)	12.2% (1470)	—

**TABLE 4 |** Experiment 1: FNN<sub>1</sub> using  $B_y$  only.

Prediction Reality \	Storm	No storm	Inconclusive
Storm	79.1% (9,572)	2.1% (259)	1.5% (184)
No storm	2.0% (242)	15.2% (1,839)	—

challenging scenario for the first NN, and, consequently, the results based on  $B_y$  and  $B_z$  are expected to supersede it. The results, shown in **Figure 12**, demonstrate that the predictions are successful, which ultimately means that the NN pipeline is overall working well.

## 4.2. Experiment 2

The first part of the experiment is to understand how well the binary classifier performs if the input is (a) only linearly polarized light, (b) only circularly polarized light, and (c) using both,

**TABLE 5 |** Experiment 1: FNN<sub>1</sub> using  $B_z$  only.

Prediction Reality \	Storm	No storm	Inconclusive
Storm	80.7% (9,758)	1.0% (121)	0.6% (68)
No storm	0.9% (111)	16.8% (2038)	—

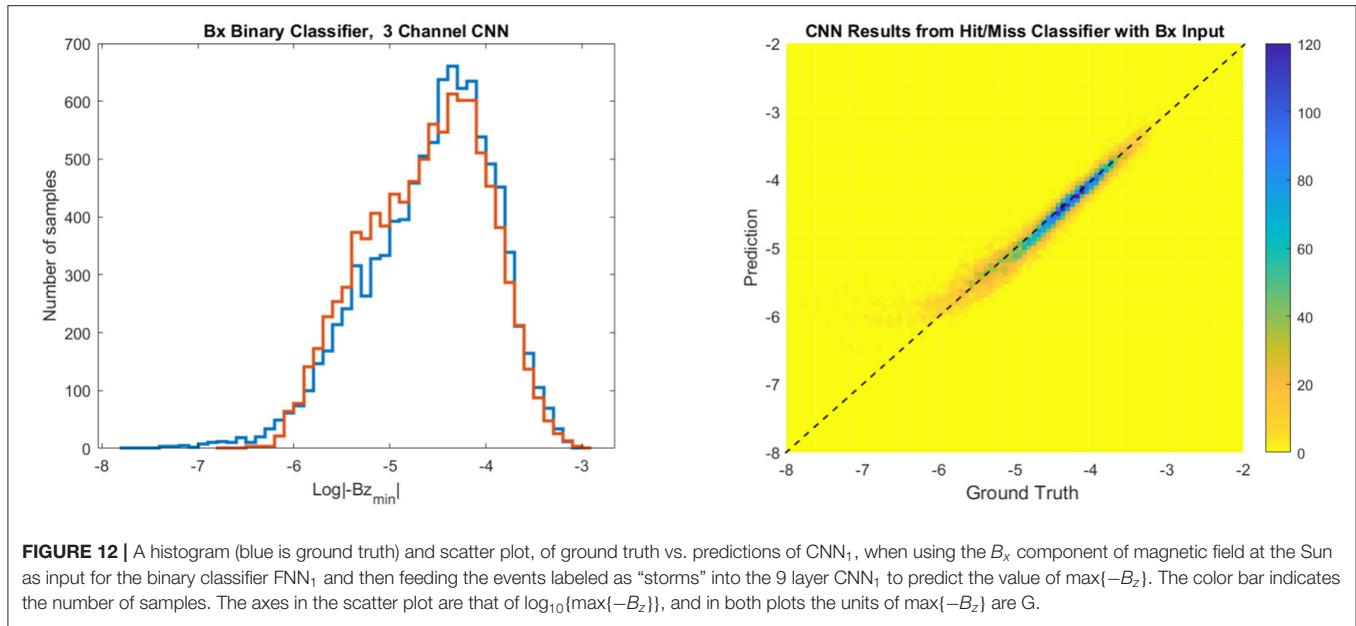
comparing to random choice shown in **Table 2**. Recall that the output of the classifier is a number  $p$  (where  $0 \leq p \leq 1$ ) indicating a probability, and we interpret  $p \geq 0.8$  as a storm,  $p \leq 0.2$  as no storm, and  $0.2 < p < 0.8$  as inconclusive, the results are given in **Tables 6–8**. First, notice that the input of just ( $L/I, Az$ ) gives results that are worse than if the guess was random. Also note that approximately 30% of the results (3570 of 12096) are inconclusive. Secondly, when using only  $V/I$  as an input, the true positives could be determined to within 93.5%, the true negatives to within 73%, and only 7% of the cases are inconclusive. By using all 3 Stokes components, we get a slight improvement of  $\leq 1\%$  which could easily be considered within the noise of the classifier. However, what is impressive is that if the ROC is plotted for the case where we use all the 3 Stokes components, as seen in **Figure 13**, the AUC is 0.99.

The second part of this experiment takes those samples that correctly classified as storms and run them through the 3 channel CNN<sub>2</sub> to predict the strength of the storm. **Figure 14** shows the results both in histogram form and as scatter plots of ground truth vs. prediction. First, note that just considering the  $V/I$  inputs from the binary classification results in a histogram that seems to overall fit the true data (blue) rather well (except for maybe accruing larger sample sizes for peak values around  $10^{-4.25}$  G). However, the scatter plot shows that the slope is not quite unity and there is a slight bias to overpredict values of  $\max\{-B_z\} < 10^{-4}$  G. In this case, the Pearson correlation coefficient is 0.95. If instead we use CNN<sub>2</sub> trained on the data from the binary classifier that includes all 3 channels ( $L/I, Az, V/I$ ), we see that the bias is corrected and the predictions line up perfectly with the ground truth. The Pearson correlation coefficient in this case is 0.98.

Finally, we calculate the relative root square error for the accuracy of predictions for both NNs. We use the following definition:

$$E = 1 - \frac{\sqrt{\sum_{i=1}^n (\Xi_{GT} - \Xi_{NN})^2}}{\sqrt{\sum_{i=1}^n \Xi_{GT}^2}}, \quad (3)$$

where for brevity we denoted  $\Xi = \max\{-B_z\}$ ,  $GT$  stands for “ground truth,” and “NN” stands for “NN-predicted value.” We find that, given three components of  $\mathbf{B}$  near Sun, CNN<sub>1</sub> can predict  $\max\{-B_z\}$  at 1AU at 97% accuracy ( $E = 0.97$ ), and for three components of the Stokes vector near Sun as input, CNN<sub>2</sub> can predict  $\max\{-B_z\}$  at 1AU with 95% accuracy ( $E = 0.95$ ).

**TABLE 6 |** FNN<sub>2</sub> using  $L/I$  and  $Az$  only.

Prediction Reality \	Storm	No storm	Inconclusive
Storm	64.6% (7816)	0.6% (73)	29.5% (3570)
No Storm	2.4% (296)	2.8% (341)	

**TABLE 7 |** FNN<sub>2</sub> using  $V/I$  only.

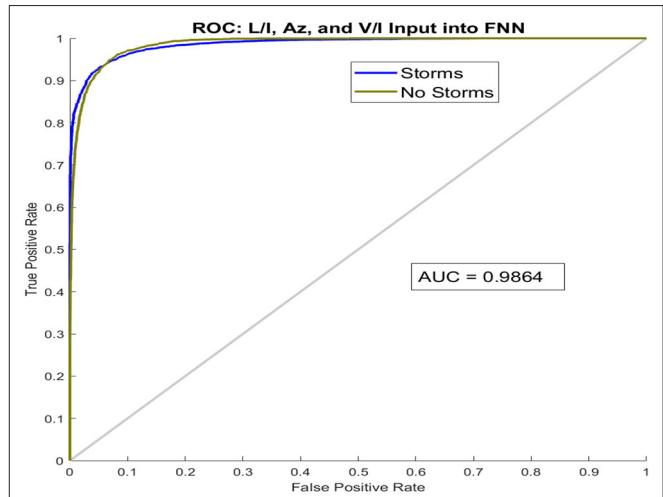
Prediction Reality \	Storm	No storm	Inconclusive
Storm	76.7% (9283)	1.4% (168)	7.0% (850)
No Storm	1.7% (211)	13.1% (1584)	

**TABLE 8 |** FNN<sub>2</sub> using  $L/I$ ,  $Az$ , and  $V/I$ .

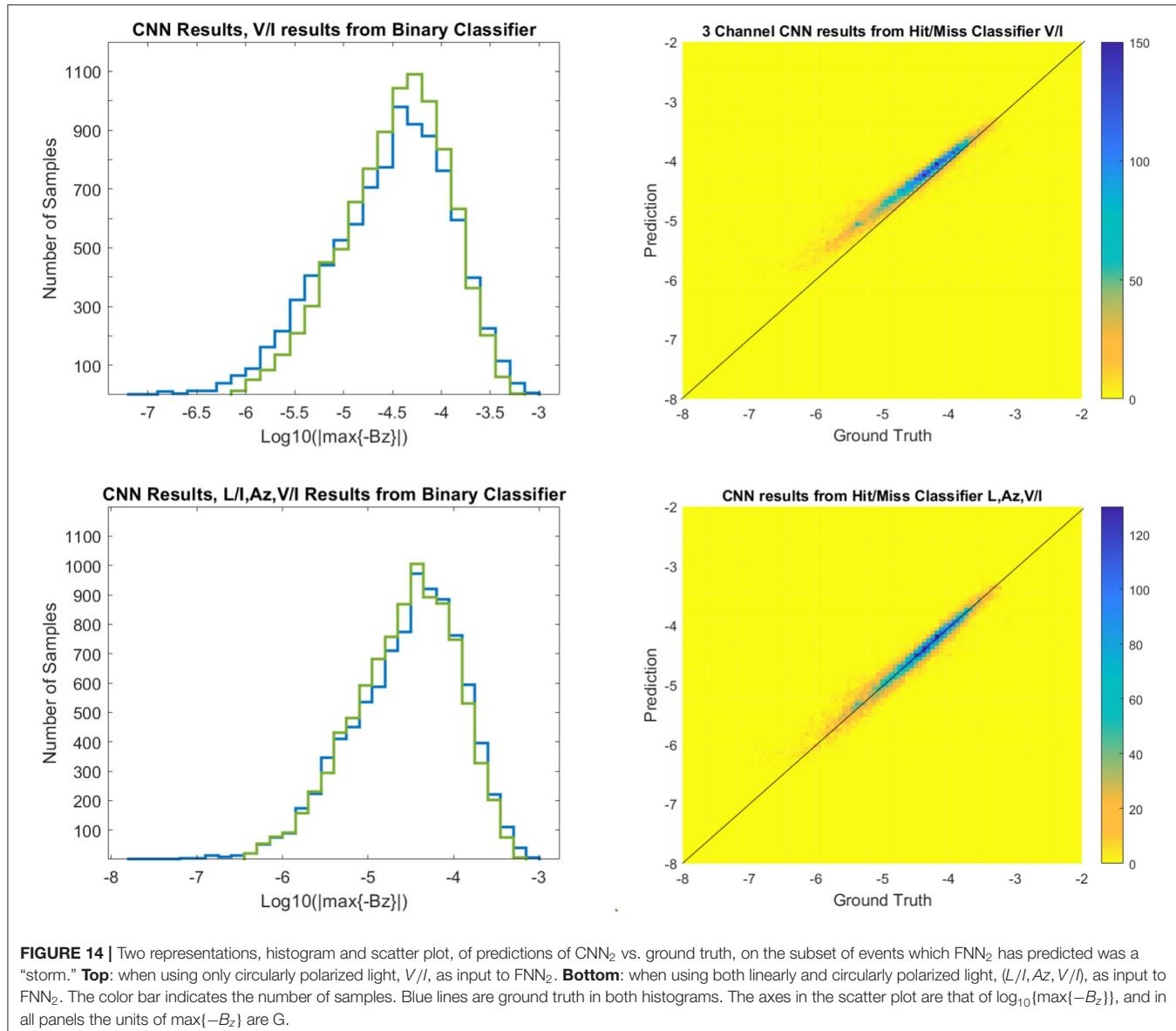
Prediction Reality \	Storm	No storm	Inconclusive
Storm	77.6% (9390)	0.9% (103)	6.8% (819)
No Storm	1.3% (155)	13.5% (1629)	

## 5. CONCLUSIONS

In this paper we have developed a machine learning algorithm to set a baseline for testing the efficacy of coronal spectropolarimetric measurements for predicting  $\max\{-B_z\}$  at the Earth. We have found that the circularly polarized light maintains the crucial magnetic field information for making a good prediction, at least for the simple model we examined. This is not unexpected, as circular polarization is directly related to the line-of-sight (in our case,  $B_y$ ) field strength (see Rachmeler et al., 2012, for further discussion). However, the



polarization signal is a result of the line-of-sight integration of the data, and disambiguation is required to derive the 3D structure of the field. The ability of neural networks to perform this disambiguation, which only yields information about  $B_y$  signal, and then to form meaningful  $B_z$  predictions at a later time, demonstrates that machine learning is a valuable asset for space weather predictions. Linearly polarized light on its own does not do so well, as it is not sensitive to the magnetic field strength but only to its geometry. It nonetheless proves to be important when considered in combination with circularly polarized light. Indeed, the most accurate prediction arises when all the components of the Stokes vector are included.



**FIGURE 14 |** Two representations, histogram and scatter plot, of predictions of CNN<sub>2</sub> vs. ground truth, on the subset of events which FNN<sub>2</sub> has predicted was a “storm.” **Top:** when using only circularly polarized light, V/I, as input to FNN<sub>2</sub>. **Bottom:** when using both linearly and circularly polarized light, (L/I, Az, V/I), as input to FNN<sub>2</sub>. The color bar indicates the number of samples. Blue lines are ground truth in both histograms. The axes in the scatter plot are that of log<sub>10</sub>{max{−B<sub>z</sub>}}, and in all panels the units of max{−B<sub>z</sub>} are G.

Full-MHD simulations of interplanetary CME evolution from Sun to Earth are typically computationally expensive. CNN allows us to explore a large space of CME models with different characteristics to study which initial states result in the strongest geomagnetic storm. The role of CNN in this case is, given either magnetic field or the synthetic spectropolarimetric observables of a CME near Sun in quadrature view, paired with max{−B<sub>z</sub>} values at Earth, to predict magnetic field inside a CME at later time and therefore to facilitate the magnetic storm predictions.

This project could be considered as preparatory work for future projects. Some of these could include, for example, exploring near-Sun signatures of CMEs in other channels (such as extreme ultraviolet), and in the addition to predicting max{−B<sub>z</sub>}, to also predict other parameters of the storm (such as its duration).

In the introduction, we mention two major factors influencing our ability to use near-Sun spectropolarimetric signatures of CMEs for space weather predictions. Our work explicitly addresses the first factor, i.e., the capability of the neural networks to use near-Sun spectropolarimetric signatures for predicting magnetic field strength in the erupting flux rope. The evolution of the CME from Sun to Earth, i.e., the second factor mentioned in the introduction, requires a model that can take into account the changes of the structure as it interacts with the solar wind. Provornikova et al. (2020, in preparation) will show an example of such a simulation, as part of a project which is currently in development (NASA award 80NSSC17K0685). The project will yield a database of tens of thousands of MHD simulations, in which various configurations of a G&L flux rope will be used as inputs, along with realistic solar wind models (Arge et al.,

2004), and their interaction will be modeled in MHD simulations (Merkin et al., 2016). We tailored our input and output for maximal compatibility with this project. Our plan is to follow up the current work with an equivalent analysis using a large number of MHD runs of CME propagation through the solar wind. By doing this, we will be able to determine how much information is retained even when non-ideal evolution of ICME in the solar wind between Sun and Earth is taken into consideration.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation.

## AUTHOR CONTRIBUTIONS

AM have designed and produced the database of GLOW flux rope models used in the project, including magnetic, spectropolarimetric, and near-Earth data. She assisted in their interpretation and analysis. NF created and trained

CNNs and FNNs, as well as analyzed their performance in predictions. SG was the author of the model and the codes used in the project, she provided consultations on GLOW model and assisted with producing and interpreting synthetic spectropolarimetric data. AM, NF, and SG participated in writing the manuscript.

## FUNDING

This work was supported by the NASA award 80NSSC17K0685 and the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

## ACKNOWLEDGMENTS

SG and AM acknowledge further support from NASA grant number 80NSSC17K0685. We thank Ricky Egeland, who was the NCAR internal referee, as well as the two external referees, for providing many valuable comments which have improved the paper.

## REFERENCES

- Arge, C. N., Luhmann, J. G., Odstrcil, D., Schrijver, C. J., and Li, Y. (2004). Stream structure and coronal sources of the solar wind during the May 12th, 1997 CME. *J. Atmos. Solar Terres. Phys.* 66, 1295–1309. doi: 10.1016/j.jastp.2004.03.018
- Baker, D. N. (2013). “The major solar eruptive event in July 2012: defining extreme space weather scenarios (invited),” in *AGU Fall Meeting Abstracts*, Vol.2013, SM13C–04.
- Bak-Steslicka, U., Gibson, S. E., Fan, Y., Bethge, C., Forland, B., and Rachmeler, L. A. (2013). The magnetic structure of solar prominence cavities: new observational signature Revealed by Coronal Magnetometry. *Astrophys. J. Lett.* 770:L28. doi: 10.1088/2041-8205/770/2/L28
- Borovikov, D., Sokolov, I. V., Manchester, W. B., Jin, M., and Gombosi, T. I. (2017). Eruptive event generator based on the Gibson-Low magnetic configuration. *J. Geophys. Res.* 122, 7979–7984. doi: 10.1002/2017JA024304
- Chen, P. F. (2011). Coronal mass ejections: models and their observational basis. *Liv. Rev. Solar Phys.* 8:1. doi: 10.12942/lrsp-2011-1
- Cliver, E. W., and Dietrich, W. F. (2013). The 1859 space weather event revisited: limits of extreme activity. *J. Space Weather Space Clim.* 3:A31. doi: 10.1051/swsc/2013053
- Dove, J. B., Gibson, S. E., Rachmeler, L. A., Tomczyk, S., and Judge, P. (2011). A Ring of Polarized Light: Evidence for Twisted Coronal Magnetism in Cavities. *Astrophys. J. Lett.* 731:L1. doi: 10.1088/2041-8205/731/1/L1
- Eastwood, J. P., Biffis, E., Hapgood, M. A., Green, L., Bisi, M. M., Bentley, R. D., et al. (2017). The economic impact of space weather: where do we stand? *Risk Anal.* 37, 206–218. doi: 10.1111/risa.12765
- Fan, Y., and Gibson, S. E. (2003). The emergence of a twisted magnetic flux tube into a preexisting coronal arcade. *Astrophys. J. Lett.* 589, L105–L108. doi: 10.1086/375834
- Fan, Y., and Liu, T. (2019). MHD simulation of prominence-cavity system. *Front. Astron. Space Sci.* 6:27. doi: 10.3389/fspas.2019.00027
- Forland, B. C., Gibson, S. E., Dove, J. B., Rachmeler, L. A., and Fan, Y. (2013). Coronal cavity survey: morphological clues to eruptive magnetic topologies. *Solar Phys.* 288, 603–615. doi: 10.1007/s11207-013-0361-1
- Freeland, S. L., and Handy, B. N. (1998). Data analysis with the solarSoft system. *Solar Phys.* 182, 497–500. doi: 10.1023/A:1005038224881
- Gibson, S. (2015). Coronal cavities: observations and implications for the magnetic environment of prominences. *Astrophys. Space Sci. Libr.* 415:323. doi: 10.1007/978-3-319-10416-4-13
- Gibson, S., Kucera, T., White, S., Dove, J., Fan, Y., Forland, B., et al. (2016b). FORWARD: a toolset for multiwavelength coronal magnetometry. *Front. Astron. Space Sci.* 3:8. doi: 10.3389/fspas.2016.00008
- Gibson, S. E. (2018). Solar prominences: theory and models. Fleshing out the magnetic skeleton. *Liv. Rev. Solar Phys.* 15:7. doi: 10.1007/s41116-018-0016-2
- Gibson, S. E., and Low, B. C. (1998). A time-dependent three-dimensional magnetohydrodynamic model of the coronal mass ejection. *Astrophys. J.* 493, 460–473. doi: 10.1086/305107
- Glorot, X., Bordes, A., and Bengio, Y. (2011). “Deep sparse rectifier neural networks,” in *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL: JMLR Workshop and Conference Proceedings), 315–323. Available online at: <http://proceedings.mlr.press/v15/glorot11a.html>
- Gold, T., and Hoyle, F. (1960). On the origin of solar flares. *MNRAS*. 120, 89–105, 1960.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press. Available online at: <http://www.deeplearningbook.org>
- Hagenson, R. L., and Krakowski, R. (1987). *The Spheromak as a Compact Fusion Reactor*. Tech. rep., Los Alamos National Lab., NM (USA).
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV ’15, p. 1026–1034, USA, IEEE Computer Society. doi: 10.1109/ICCV.2015.123
- Judge, P. G., and Casini, R. (2001). *A Synthesis Code for Forbidden Coronal Lines*, volume 236 of *Astronomical Society of the Pacific Conference Series*, p. 503.
- Kilpua, E. K. J., Lugaz, N., Mays, M. L., and Temmer, M. (2019). Forecasting the Structure and Orientation of Earthbound Coronal Mass Ejections. *Space Weather*, 17:498–526. doi: 10.1029/2018SW001944
- Kingma, D., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980.
- Knipp, D. J., and Gannon, J. L. (2019). *The 2019 National Space Weather Strategy and Action Plan and Beyond*. Washington, DC: The National Academies Press. doi: 10.1029/2019SW002254. Available online at: <https://www.whitehouse.gov/wp-content/uploads/2019/03/National-Space-Weather-Strategy-and-Action-Plan-2019.pdf>

- Manchester, W., Kilpua, E. K. J., Liu, Y. D., Lugaz, N., Riley, P., Torok, T. et al. (2017). The Physical Processes of CME/ICME Evolution. *Space Sci. Rev.* 212:1159–1219. doi: 10.1007/s11214-017-0394-0
- Merkin, V. G., Lyon, J. G., Lario, D., Arge, C. N., and Henney, C. J. (2016). Time-dependent magnetohydrodynamic simulations of the inner heliosphere. *J. Geophys. Res.* 121:2866–2890. doi: 10.1002/2015JA022200
- National Research Council. (2008). *Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report*. Washington, DC: The National Academies Press. doi: 10.17226/12507
- Rachmeler, L. A., Casini, R., and Gibson, S. E. (2012). Interpreting Coronal Polarization Observations. In T. R. Rimmele, A. Tritschler, F. Wöger, M. Collados Vera, H. Socas-Navarro, R. Schlichenmaier, M. Carlsson, T. Berger, A. Cadavid, P. R. Gilbert, P. R. Goode, and M. Knölker, editors, *Second ATST-EAST Meeting: Magnetic Fields from the Photosphere to the Corona*, volume 463 of *Astronomical Society of the Pacific Conference Series*, 227, Orem, UT, December 2012.
- Ruiz Cobo, B., and del Toro Iniesta, J. C. (1992). Inversion of Stokes Profiles. *Astrophys. J.* 398:375. doi: 10.1086/171862
- Schou, J., Scherrer, P. H., Bush, R. I., Wachter, R., Couvidat, S., Rabello-Soares, M. C. et al. (2012). Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). *Solar Phys.* 275:229–259. doi: 10.1007/s11207-011-9842-2
- Svozil, D., Kvasnicka, V., and Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems* 39:43–62. doi: 10.1016/S0169-7439(97)00061-0
- The MathWorks (2019). *Deep Learning Toolbox*. Available online at: <https://www.mathworks.com/help/deeplearning/>
- Webb, D. F., and Howard, T. A. (2012). Coronal Mass Ejections: Observations. *Living Rev. Solar Phys.* 9:3. doi: 10.12942/lrsp-2012-3
- Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629. doi: 10.1007/s13244-018-0639-9

**Conflict of Interest:** 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.

Copyright © 2020 Malanushenko, Flyer and Gibson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Visualizing and Interpreting Unsupervised Solar Wind Classifications

Jorge Amaya\*, Romain Dupuis, Maria Elena Innocenti and Giovanni Lapenta

Mathematics Department, Centre for Mathematical Plasma-Astrophysics, KU Leuven, Leuven, Belgium

## OPEN ACCESS

**Edited by:**

Veronique A. Delouille,  
Royal Observatory of Belgium,  
Belgium

**Reviewed by:**

Hui Li,  
National Space Science Center (CAS),  
China

Xochitl Blanco-Cano,  
National Autonomous University of  
Mexico, Mexico  
Verena Heidrich-Meisner,  
University of Kiel, Germany

**\*Correspondence:**

Jorge Amaya  
[jorge.amaya@kuleuven.be](mailto:jorge.amaya@kuleuven.be);  
[jorgeluis.amaya@gmail.com](mailto:jorgeluis.amaya@gmail.com)

**Specialty section:**

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 17 April 2020

**Accepted:** 17 August 2020

**Published:** 25 September 2020

**Citation:**

Amaya J, Dupuis R, Innocenti ME and  
Lapenta G (2020) Visualizing and  
Interpreting Unsupervised Solar Wind  
Classifications.  
*Front. Astron. Space Sci.* 7:553207.  
doi: 10.3389/fspas.2020.553207

One of the goals of machine learning is to eliminate tedious and arduous repetitive work. The manual and semi-automatic classification of millions of hours of solar wind data from multiple missions can be replaced by automatic algorithms that can discover, in mountains of multi-dimensional data, the real differences in the solar wind properties. In this paper we present how unsupervised clustering techniques can be used to segregate different types of solar wind. We propose the use of advanced data reduction methods to pre-process the data, and we introduce the use of Self-Organizing Maps to visualize and interpret 14 years of ACE data. Finally, we show how these techniques can potentially be used to uncover hidden information, and how they compare with previous empirical categorizations.

**Keywords:** solar wind, ACE, Self-Organizing Maps, clustering, autoencoder, PCA, unsupervised, machine learning

## 1. INTRODUCTION

The effects of solar activity on the magnetic environment of the Earth have been observed since the publication of Edward Sabine's work in 1852 (Sabine, 1852). During almost 200 years we have learned about the intimate connection between our star and the plasma environment of the Earth. Three main physical processes connect us to the Sun: the transfer of electromagnetic radiation, the transport of energetic particles, and the flow of solar wind. The later is a continuous stream of charged particles that carries the solar magnetic field out of the corona and into the interplanetary space.

The name *solar wind* was coined by Parker in 1958 because "the gross dynamical properties of the outward streaming gas [from the Sun] are hydrodynamic in character" (Parker, 1958). Over time we have learned that the wind also has many more complex properties. Initially, it was natural to classify the solar wind by defining a boundary between *fast* and *slow* winds (Neugebauer and Snyder, 1966; Schwenn, 1983; Schwenn and Marsch, 1990; Habbal et al., 1997). The former has been associated with mean speed values of 750 km/s (or in some publications with values larger than 600 km/s), while the latter shows a limit at 500 km/s, where the compositional ratio (Fe/O) shows a break (Feldman et al., 2005; Stakhiv et al., 2015). The solar wind also carries information about its origins on the Sun. At certain solar distances the ion composition of the solar wind is expected to be frozen-in, reflecting the electron temperature in the corona and its region of origin (Feldman et al., 2005; Zhao et al., 2009; Stakhiv et al., 2015). These particles have multiple energies and show a variety of kinetic properties, including non-Maxwellian velocity distributions (Pierrard and Lazar, 2010; Matteini et al., 2012).

The solar wind is also connected to the Sun by Interplanetary Magnetic Field (IMF) lines directed toward the Sun, away from the Sun, or in the case of flux ropes, connected at both ends (Gosling et al., 2010; Owens, 2016). The region separating IMF lines of opposite polarity (directed away or toward the Sun) is called the Heliospheric Current Sheet (HCS) (Smith, 2001). When a spacecraft crosses the HCS instruments onboard measure the change in polarity of the magnetic field. In quiet wind conditions the plasma around the HCS presents discontinuities in density, temperature velocity and magnetic field (Eselevich and Filippov, 1988). This perturbed region surrounding the HCS is called the Heliospheric Plasma Sheet (HPS). The passage of the spacecraft from one side of the HPS to the other is known as a Sector Boundary Crossing (SBC) (Winterhalter et al., 1994). In spacecraft observations these are sometimes confused with Corotating Interaction Regions (CIR), which are zones of the solar wind where fast flows have caught up with slow downstream solar wind, compressing the plasma (Fisk and Lee, 1980; Richardson, 2004).

From the point of view of a spacecraft SBCs and CIRs can show similar sudden changes in the plasma properties. These two in turn are often grouped and mixed with other transient events, like Coronal Mass Ejections (CME) and Magnetic Clouds (MC). Since 1981 when Burlaga et al. (1981) described the propagation of MC behind an interplanetary shock, it was suspected that CMEs and MC were coupled. However, more recent studies show that CMEs observed near the Sun do not necessarily become MC, but instead “pressure pulses” (Gopalswamy et al., 1998; Wu et al., 2006).

Much more recently it has been revealed, by observations from Parker Solar Probe, that the properties of the solar wind can be drastically different closer to the Sun, where the plasma flow is more pristine and has not yet mixed with the interplanetary environment. Patches of large intermittent magnetic field reversals, associated with jets of plasma and enhanced Poynting flux, have been observed and named “switchbacks” (Bale et al., 2019; Bandyopadhyay et al., 2020).

The solar wind is thus not only an hydrodynamic flow, but a compressible mix of different populations of charged particles and electromagnetic fields that carry information of their solar origin (helmet streamer, coronal holes, filaments, solar active regions, etc.) and is the dominion of complex plasma interactions (ICMEs, MC, CIRs, SBCs, switchbacks).

To identify and study each one of these phenomena we have relied in the past on a manual search, identification and classification of spacecraft data. Multiple authors have created empirical methods of wind type identification based on *in-situ* satellite observations and remote imaging of the solar corona. Over the years the number and types of solar wind classes has changed, following our understanding of the complexity of heliospheric physics.

Solar wind classification serves four main roles:

1. it is used for the characterization of its origins in the corona,
2. to identify the conditions where the solar wind is geoeffective,
3. to isolate different plasma populations in order to perform statistical analysis,

4. to study the basic transport effects of space plasmas of different nature.

Among the existing classifications we can include the original review work by Withbroe (1986), the impressive continuous inventory by Richardson et al. (2000) and Richardson and Cane (2010, 2012), and the detailed studies by Zhao et al. (2009) and (Xu and Borovsky, 2015). These publications classify the solar wind based on their ion composition, and on the transient events detected. Each system includes two, three, or four classes, generally involving coronal-hole origins, CMEs, streamer belt origins, and sector reversal regions.

The precise point of origin of the solar wind can be traced back from spacecraft positions to the solar corona and the photosphere: multiple authors (Neugebauer et al., 2002; Zhao et al., 2009, 2017; Fu et al., 2015) have used a ballistic approximation coupled to a Potential Field Source Surface (PFSS) model to trace back solar wind observations to their original sources on the Sun. This procedure relies on multiple assumptions, including a constant solar wind speed and a force free magnetic field configuration of the solar corona. The uncertainty on the source position is estimated around  $\pm 10^\circ$  by Neugebauer et al. (2002). This is currently the best method to acquire the *ground truth* about the origin of the solar wind. Unfortunately, to our knowledge, there is no central repository of solar wind origins for any space mission that we can use to train or verify our novel machine learning techniques.

We are moving now toward a new era of data analysis, where manual human intervention can be replaced by *intelligent* software. The trend has already started, with the work by Camporeale et al. (2017) who used (Xu and Borovsky, 2015) classes to train a Gaussian Process algorithm that autonomously assigns the solar wind to the proper class, and by Roberts et al. (2020) who used unsupervised classification to perform a 4 and 8 class solar wind classification. A recent publication by Bloch et al. (2020) uses unsupervised techniques to classify ACE and Ulysses observations, and Li et al. (2020) have successfully tested 10 different supervised techniques to reproduce the categories introduced by Xu and Borovsky (2015).

The most basic ML techniques learn using two approaches: (A) in supervised learning the algorithms are shown a group of inputs,  $\mathbf{X} \in \mathbb{R}^n$ , and outputs,  $\mathbf{Y} \in \mathbb{R}^o$ , with the goal of finding a non-linear relationship between them,  $\xi_s : \mathbf{X} \rightarrow \mathbf{Y}$ , (B) in unsupervised learning the machine is presented with a cloud of multi-dimensional points,  $\mathbf{X} \in \mathbb{R}^n$ , that have to be autonomously categorized in different classes, either performing associations with representative points in the same data space,  $\xi_u : \mathbf{X} \rightarrow \mathbf{W} \in \mathbb{R}^n$ , or by grouping neighboring data points together into an assigned set,  $\xi_u : \mathbf{X} \rightarrow g \in \mathbb{R}$ . This means that we can program the computer to learn about the different types of solar wind using the existing empirical classifications using method (a), or allowing the computer to independently detect patterns in the solar wind properties with method (b).

In the present work we show how the second method, unsupervised classification, can be used to segregate different types of solar wind. In addition, we show how to visualize and interpret such results. The goal of this paper is to introduce the

use of unsupervised techniques to our community, including the best use practices and the opportunities that such methods can bring. We promote the use of one specific type of classification, called Self-Organizing Maps, and we compare it to simpler classification techniques.

In the next sections we present in detail the techniques of data processing (section 2.1), data dimension reduction (sections 2.2.1, 2.2.2 and 2.2.3) and data clustering (section 2.2.4) that we have used. We then present in detail the Self-Organizing Map technique and all its properties in section 2.2.5. We show how to connect all of these parts together in section 2.2.6, and finally we show how the full system can be used to study 14 years of solar wind data from the ACE spacecraft in section 3.

## 2. MATERIALS AND METHODS

### 2.1. Data and Processing

#### 2.1.1. Data Set Used

The solar wind data used in this work was obtained by the Advanced Composition Explorer (ACE) spacecraft, during a period of 14 years, between 1998 and 2011. The data can be downloaded from the FTP servers of The ACE Science Center (ASC) (Garrard et al., 1998). The files in this repository correspond to a compilation of hourly average data from four instruments: MAG (Magnetometer) (Smith et al., 1998), SWEPAM (Solar Wind Electron, Proton, and Alpha Monitor) (McComas et al., 1998), EPAM (Electron, Proton, and Alpha Monitor) (Gold et al., 1998), and SWICS (Solar Wind Ion Composition Spectrometer) (Gloeckler et al., 1998). A detailed description of the entries in this data set can be found in the ASC website listed in the Data Availability Statement.

A total of 122,712 data points are available. However, routine maintenance operations, low statistics, instrument saturation and degradation produce gaps and errors in the data. The SWICS data includes a flag assessing the quality of the calculated plasma moments. We retain only *Good quality* entries. Our pre-processed data set contains a total of 72,454 points.

#### 2.1.2. Additional Derived Features

We created additional features for each entry, based on previous knowledge of the physical properties of the solar wind. Some are derived from the existing properties in the data set, others computed from statistical analysis of their evolution. We introduce here the additional *engineered* features included in our data set.

Multiple techniques have been proposed in the literature to identify ejecta, Interplanetary Coronal Mass Ejections (ICME), and solar wind origins in the ACE data. Zhao et al. (2009) suggest that, during solar cycle 23, three classes of solar wind can be identified using its speed,  $V_{sw}$ , and the oxygen ion charge state ratio,  $O^{7+}/O^{6+}$ . It has been shown that slow winds originating in coronal streamers correlate with high values of the charge state ratio and fast winds coming from coronal holes present low values (Schwenn, 1983; Withbroe, 1986; Schwenn and Marsch, 1990). Plasma formed in coronal loops associated with CMEs also show high values of the charge state ratio (Xu and Borovsky, 2015;

Zhao et al., 2017). The classification boundaries of the Z09 model, proposed by Zhao et al. (2009), are presented in **Table 1**.

Xu and Borovsky (2015) suggested an alternative four classes system based on the proton-specific entropy,  $S_p = T_p/n_p^{2/3}$  [K cm<sup>2</sup>], the Alfvén speed,  $V_A = B/(\mu_0 m_p n_p)^{1/2}$  [Km s<sup>-1</sup>], and the ratio between the expected and the measured proton temperature,  $T_{exp}/T_p = (V_{sw}/258)^{3.113}/T_p$  [-], where  $n_p$  is the proton number density,  $m_p$  is the proton mass, and  $\mu_0$  is the permeability of free space. The classification boundaries used for the X15 model, proposed by Xu and Borovsky (2015), are also presented in **Table 1**. For each entry in the data set we have included the values of  $S_p$ ,  $V_A$ ,  $T_{exp}$ ,  $T_{ratio} = T_{exp}/T_p$ , and the solar wind type.

Two additional empirical threshold methods will be included in this work for comparison. These two methods were derived from the compositional observations of the solar wind at higher heliospheric latitudes, using data from the *Ulysses* mission (Wenzel et al., 1992). The first model, that we call vS15, comes from the work by von Steiger and Zurbuchen (2015), where the first figure shows a clear division between Coronal Hole (CH) sources and non-Coronal Hole (NCH) wind. The boundary between the two classes is presented in **Table 1**. The second threshold model was presented as an example by Bloch et al. (2020). This boundary, named here B20, is an empirical approximation that divides CH and NCH origin winds. The threshold values are shown in **Table 1**.

In addition to the instantaneous properties of the solar wind used in all previous classifications, we can perform statistical operations over a window of time of 6 h, including values of the maximum, minimum, mean, standard deviation, variance, auto-correlation, and range. We expect to capture with some of these quantities turbulent signals or sudden jumps associated with different transient events. These additional rolling operations are a complement to the stationary solar wind parameters mentioned above and add information about the temporal evolution of the plasma. The selection of the statistical parameters and the window of time is arbitrary and will require a closer examination in the future.

An additional term, which has been successfully used in the study of solar wind turbulence (D'Amicis and Bruno, 2015; Zhao et al., 2018; Magyar et al., 2019; Adhikari et al., 2020), is included here to account for additional time correlations. The normalized cross-helicity,  $\sigma_c$ , is defined in Equation (1), where  $\mathbf{b} = (\mathbf{B} - \langle \mathbf{B} \rangle)/(\mu_0 m_p n_p)^{1/2}$  is the fluctuating magnetic field in Alfvén units,  $\mathbf{v} = \mathbf{V}_{sw} - \langle \mathbf{V}_{sw} \rangle$  is the fluctuating solar wind velocity, and  $\langle \cdot \rangle$  denotes the averaging of quantities over a time window of 3 h (Roberts et al., 2020).

$$\sigma_c = 2 \langle \mathbf{b} \cdot \mathbf{v} \rangle / \langle \mathbf{b}^2 + \mathbf{v}^2 \rangle \quad (1)$$

Due to gaps in the data, some of the above quantities can not be obtained. We eliminate from the data set all entries for which the derived features presented in this section could not be calculated. This leaves a total of 51,374 entries in the data set used in the present work.

**TABLE 1 |** Solar wind types and boundaries as defined by the empirical models: Z09, X15, vS15, and B20.

#	SW type	Condition	References	
0	CH	$\log_{10} O^{7+}/O^{6+} \leq 0.145$	Zhao et al., 2009	
2	NCH	$0.145 < \log_{10} O^{7+}/O^{6+} < 6.008e^{(-0.00578V_{sw})}$		
4	TR	$\log_{10} O^{7+}/O^{6+} > 6.008e^{(-0.00578V_{sw})}$		
0	CH	Not type TR, and $\log_{10}(S_p) > -0.525 \log_{10}(T_{exp}/T_p) - 0.676 \log_{10}(V_A) + 1.74$	Xu and Borovsky, 2015	
2	SB	Not type CH, TR, or SR		
3	SR	Not type TR, and $\log_{10}(S_p) < -0.125 \log_{10}(T_{exp}/T_p) - 0.658 \log_{10}(V_A) + 1.04$		
4	TR	$\log_{10}(V_A) > 0.055 \log_{10}(T_{exp}/T_p) + 0.277 \log_{10}(S_p) + 1.83$		
0	CH	$\log_{10} O^{7+}/O^{6+} \times \log_{10} C^{6+}/C^{5+} \leq$	0.01	von Steiger and Zurbuchen, 2015
2	NCH	Not type CH		
0	CH	$1.25 \times \log_{10} O^{7+}/O^{6+} + 6.75 < \log_{10}(S_p)$	Bloch et al., 2020	
2	NCH	Not type CH		

The four types are: fast solar wind of coronal hole origin (CH), slow wind of non-coronal hole origin (NCH), transients, including ejecta, ICMEs, CIRs, MCs or other sudden jumps in solar wind parameters (TR), solar wind originated in the streamer belt (SB), and solar wind of sector reversal origins (SR). The ID value in the first column is arbitrary and has been chosen to simplify the visualization of our results.

To account for the differences in units and scale, each feature column  $F$  in the data set is normalized to values between 0 and 1, using:  $f = (F - \min F) / (\max F - \min F)$ .

Not all the features might be useful and some of them can be strongly correlated. We do not perform here a detailed evaluation of the inter dependencies of the different features, and we leave that task for a future work. The present manuscript focuses on the description of the methodology and on the visualization and interpretation capabilities of unsupervised machine learning classification. We limit our work here to test and compare a single model that incorporates a total of 15 features. These are listed in Table 2.

### 2.1.3. Complementary Data Catalogs

We support the interpretation of our results using data from three solar wind event catalogs. The first is the well-known Cane and Richardson catalog that contains information about ICMEs detected in the solar wind in front of the Earth (Cane and Richardson, 2003; Richardson and Cane, 2010)<sup>1</sup>. We used the August 16, 2019 revision. As the authors state in their website, there is no spreadsheet or text version of this catalog and offline editing was necessary. We downloaded and re-formatted the catalog to use it in our application. The CSV file created has been made available in our repository. We call this, the Richardson and Cane catalog.

The second catalog corresponds to the ACE List of Disturbances and Transients<sup>2</sup> produced by the University of

New Hampshire. As in the previous case, the catalog is only available as an html webpage, so we have manually edited the file and extracted the catalog data into a file also available in our repository. This is hereafter referred to as the UNH catalog.

Finally, we also included data from the Shock Database<sup>3</sup> maintained by Dr. Michael L. Stevens and Professor Justin C. Kasper at the Harvard-Smithsonian Center for Astrophysics. Once again we have gathered and edited multiple web-pages in a single file available in our repository. In this work this database will be known as the CfA catalog.

## 2.2. Dimension Reduction and Clustering

### 2.2.1. Dimension Reduction Using PCA

Principal Component Analysis (PCA) is a mathematical tool used in data analysis to simplify and extract the most relevant features in a complex data set. This technique is used to create entries composed of linearly independent *principal components*. These are the eigenvectors,  $v$ , of the covariance matrix  $\Sigma = (\Sigma_{ij})$  applied to the centered data (Equation 2), ordered from the largest to the smallest eigenvalue,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , where  $\bar{X}$  is the mean value of each one of the  $n$  original features (Equation 3), and  $m$  is the total number of entries in the data set. The projection of the data onto the principal component space ensures a maximal variance on the direction of the first component. Each subsequent principal component is orthogonal to the previous ones and points in the direction of maximal

<sup>1</sup>Near-Earth Interplanetary Coronal Mass Ejections Since January 1996: <http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable.2.html>

<sup>2</sup>ACE Lists of Disturbances and Transients: [http://www.ssg.sr.unh.edu/mag/ace/ACElists/obs\\_list.html](http://www.ssg.sr.unh.edu/mag/ace/ACElists/obs_list.html)

<sup>3</sup>Harvard-Smithsonian, Center for Astrophysics, Interplanetary Shock Database—ACE: [https://www.cfa.harvard.edu/shocks/ac\\_master\\_data/](https://www.cfa.harvard.edu/shocks/ac_master_data/)

**TABLE 2 |** List of features used for the AE+DSOM model.

ID	Name in the database	AE + DSOM
0	proton_speed	✓
1	proton_density	✓
2	O7to6	✓
3	C6to5	✓
4	FetoO	✓
5	avqFe	✓
6	proton_temp	✓
7	sigmac <sup>(*)</sup>	✓
8	Sp	✓
9	Va	✓
10	Tratio	✓
11	proton_speed_range	✓
12	Bn_range	✓
13	FetoO_range	✓
14	O7to6_range	✓
Initial year		
Final year		
Neurons / encoding layer		
Optimizer		
Learning rate		
Lattice nodes		
$\epsilon$		
$\eta$		

The logarithm of all quantities was used, except for the features marked with an asterisk (\*). Bottom: data range and hyper-parameters of the AE and the SOM.

variance in the residual sub-space (Shlens et al., 2014).

$$\Sigma_{ij} = \frac{1}{m} \sum_{k=1}^m (X_i^k - \bar{X}_i)(X_j^k - \bar{X}_j) \quad (2)$$

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i \quad (3)$$

$$\Sigma v = \lambda v \quad (4)$$

The PCA transformation creates the same number of components in the transformed space,  $\tilde{X}$ , as features in the original data space  $X$ . However, components with small eigenvalues belong to a dimension where the variance is so small that it is impossible to separate points in the data. It is a general practice in data reduction to keep only the first  $k$  components that explain at least a significant portion of the total variance of the data,  $\sum_{i=1}^k \lambda_i / \text{Tr}(\Sigma) > \epsilon$ . This allows for a selection of information that will effectively differentiate data points, and for a reduction of the amount of data to process during analysis. Many techniques have been suggested for the selection of the values of  $k$  and the cut-off  $\epsilon$  (Rea and Rea, 2016). We use the value of  $\epsilon = 0.95$ .

## 2.2.2. Dimension Reduction Using Kernel PCA

PCA has a limitation: the principal components are a linear combination of the original properties of the solar wind. The Kernel PCA (KPCA) is an extension of the PCA that allows to perform non-linear transformations of the original data. The goal in KPCA is to perform the original PCA operations in a high dimensional space.

For a list of  $m$  data points composed of  $n$  features, it is sometimes difficult (or impossible) to build a linear hyper-plane that dissects regions of different density. However, it is possible to conceive a function,  $\xi : X \in \mathbb{R}^n \rightarrow \tilde{X} \in \mathbb{R}^m$ , that will transform all the data into a space where each cluster of points can be linearly separable. The goal is then to avoid explicitly calculating the high-dimensional function  $\xi$  by building a Kernel,  $K$ , which is the inner product of the high-dimensional space:

$$K = k(X_i, X_j) = \xi(X_i)^T \xi(X_j) \quad (5)$$

In this space the projected points are linearly separable using the same principles of the PCA. In this case the covariance matrix would be expressed as:

$$\Sigma_{ij} = \frac{1}{m} \sum_{k=1}^m \xi(X_i) \xi(X_i)^T \quad (6)$$

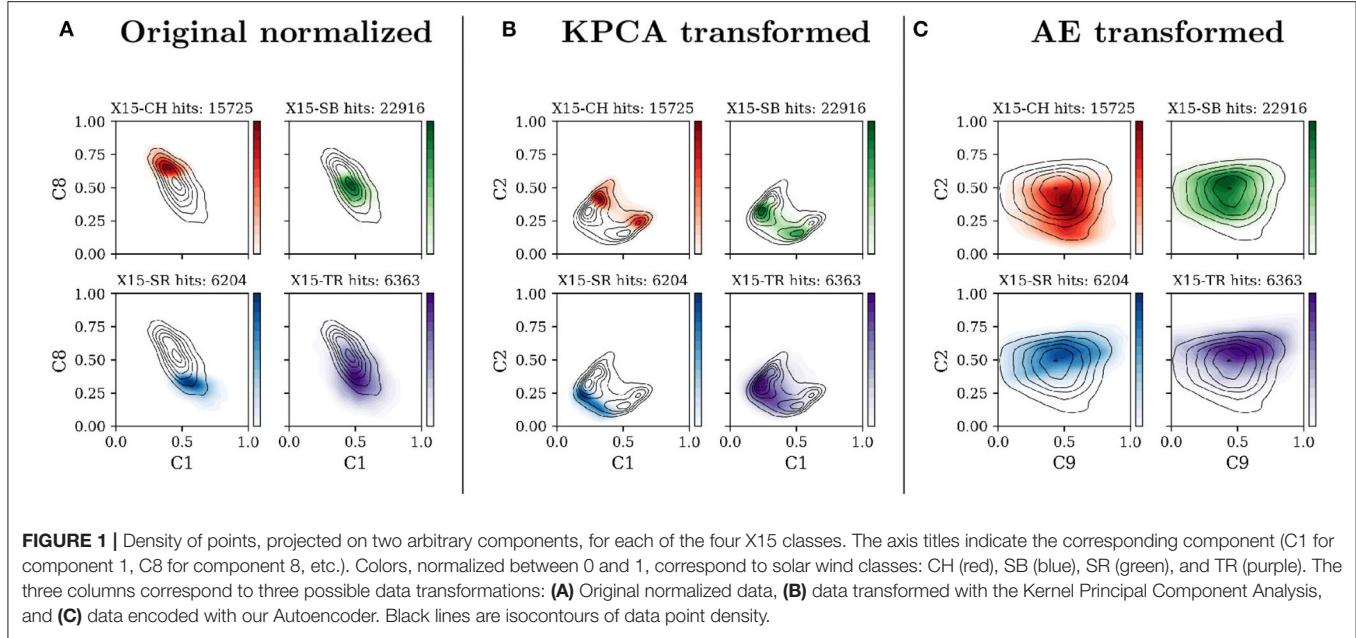
$$v = \sum_{i=1}^n a_i \xi(x_i) \quad (7)$$

Popular kernel functions include Gaussian, polynomial and hyperbolic tangent. The transformation is reduced to solving the eigenvalue problem:  $Ka = \lambda a$ , where  $a$  are the coefficients of the linear combination of the eigenvectors (Equation 7). Although a powerful tool, KPCA requires the creation of an  $m \times m$  matrix that can consume large amounts of time and memory resources.

In this work we use KPCA with a polynomial kernel of order eight (8). We also apply the procedure described before to select the total number of retained components: we impose  $\epsilon = 0.95$ . Cutting off the number of components implies a loss of data. To verify that only minimal information is lost, we perform a transformation of all our data set followed by an inverse transformation. The relative error between the two is normally distributed around zero with  $<1\%$  of variance.

## 2.2.3. Dimension Reduction Using Autoencoders

An alternative to data reduction is the use of Autoencoders (AE). These are machine learning techniques that can create non-linear combinations of the original features projected on a latent space with less dimensions (Hinton and Salakhutdinov, 2006). This is accomplished by creating a system where an encoding function,  $\phi$ , maps the original data  $X$  to a latent space,  $\mathcal{F} \in \mathbb{R}^d$  (Equation 8). A decoder function,  $\psi$ , then maps the latent space back to the original input space (Equation 9). The objective of the autoencoder is to minimize the error between the original data and the data produced by the compression-decompression



**FIGURE 1 |** Density of points, projected on two arbitrary components, for each of the four X15 classes. The axis titles indicate the corresponding component (C1 for component 1, C8 for component 8, etc.). Colors, normalized between 0 and 1, correspond to solar wind classes: CH (red), SB (blue), SR (green), and TR (purple). The three columns correspond to three possible data transformations: **(A)** Original normalized data, **(B)** data transformed with the Kernel Principal Component Analysis, and **(C)** data encoded with our Autoencoder. Black lines are isocontours of data point density.

procedure as shown in Equation (10).

$$\phi : \mathbf{X} \rightarrow \mathbf{Z} \in \mathcal{F} \quad (8)$$

$$\psi : \mathbf{Z} \in \mathcal{F} \rightarrow \mathbf{X} \quad (9)$$

$$\phi, \psi = \arg \min_{\phi, \psi} \| \mathbf{X} - (\phi \circ \psi) \mathbf{X} \|^2 \quad (10)$$

Autoencoders can be represented as feed-forward neural networks, where fully connected layers lead to a central bottleneck layer with few nodes and then expands to reach again the input layer size. An encoded element,  $z \in \mathcal{F}$ , can be obtained from a data entry,  $x \in \mathbf{X}$ , following the standard neural network function (Equation 11), where  $\mathbf{W}$  is the weights matrix,  $c$  is the bias, and  $f$  is the non-linear activation function.

$$z = f(\mathbf{W}x + c) \quad (11)$$

$$\hat{x} = f'(\mathbf{W}'z + c') \quad (12)$$

$$\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2 \quad (13)$$

The decoding procedure, shown in Equation (12), transforms  $z \rightarrow \hat{x}$ , where the prime quantities are associated with the decoder. The loss function,  $\mathcal{L}(x, \hat{x})$ , is the objective to be minimized by the training of the neural network using gradient descent. Once training is completed, the vector  $z$  is a projection of the input vector  $x$  onto the lower dimensional space  $\mathcal{F}$ .

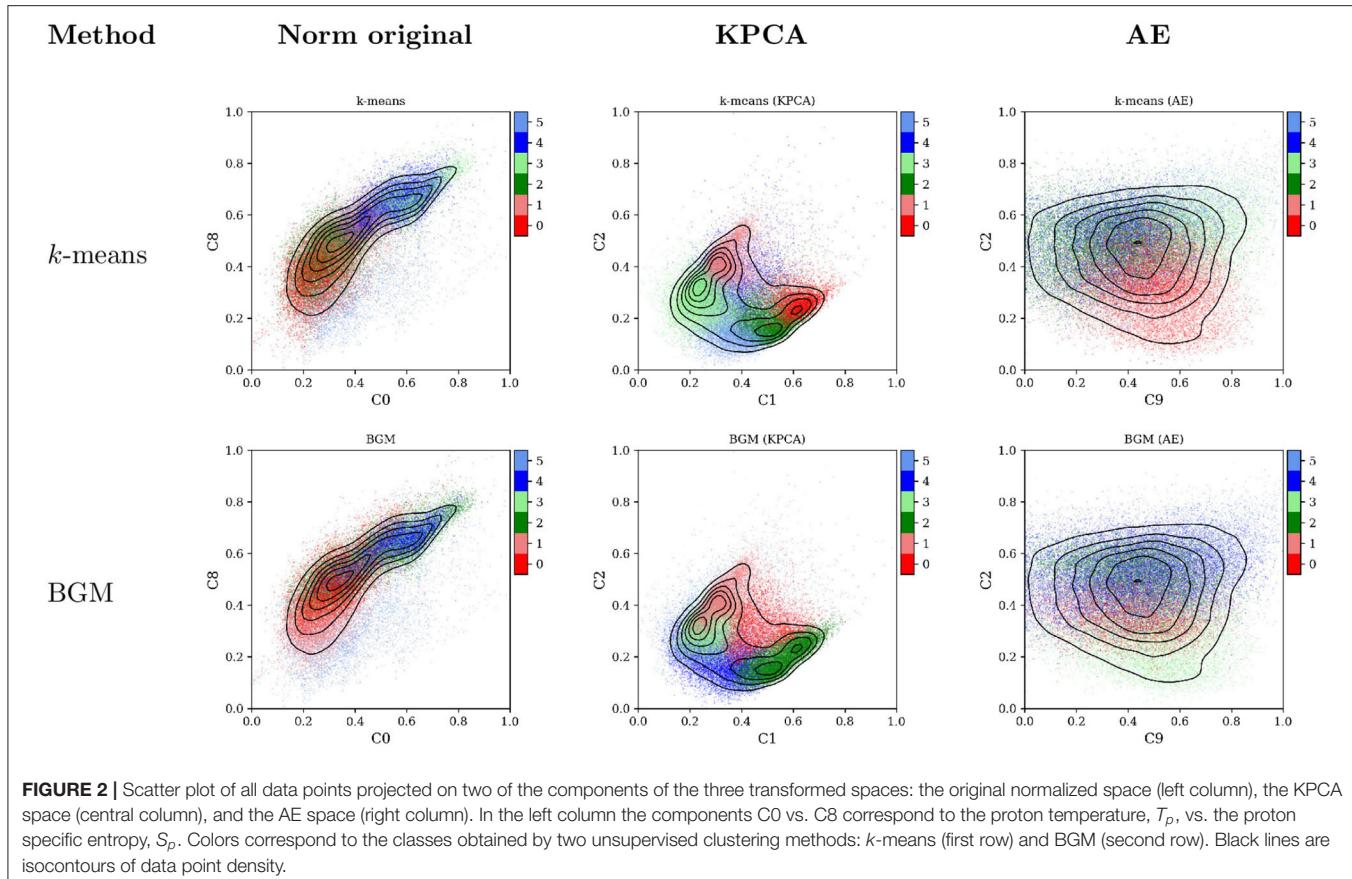
Additional enhancements and variations of this simple autoencoder setup exist in the literature, including multiple regularization techniques to minimize over-fitting (Liang and Liu, 2015), Variational Autoencoders (VAE) that produce encoded Gaussian distribution functions (Kingma and Welling, 2013), and Generative Adversarial Networks that automatically generate new data (Goodfellow et al., 2014). In this work we use the most basic form of autoencoders, presented above.

In the present work we will be showing different representations of the solar wind data, transformed with different techniques and projected on flat planes. **Figure 1** presents our data set in three different projections: (A) the original feature space, normalized between zero and one, (B) the transformed data set using the KPCA method, and (C) the AE transformed data. In each panel four histograms present the distribution of the X15 classes, on two arbitrary components identified by the axis title.

#### 2.2.4. Clustering Techniques

The goal of unsupervised machine learning is to group data points in a limited number of clusters in the N-dimensional space  $\Omega \in \mathbb{R}^n$ , where  $n$  is the number of features (components or properties) in the data set. Multiple techniques can be used to perform multi-dimensional clustering. We present in **Figure 2** the application of two basic clustering techniques to classify our data set. Following the same order as before, the first column in the figure contains all data points projected in the original normalized feature space; column two contains scatter plots of the points after KPCA transformation; column three contains the same points encoded in the AE latent space. Each row corresponds to a different clustering method. The colors in the top row were obtained using the *k*-means method (Lloyd, 1982), while the colors in bottom panels were obtained using the Bayesian Gaussian Mixture (BGM) (Bishop, 2006).

The *k*-means technique has already been used in multiple publications for the determination of solar wind states (Heidrich-Meisner and Wimmer-Schweingruber, 2018; Roberts et al., 2020). The BGM technique has also been recently used by Bloch et al. (2020) to classify solar wind observations by the ACE and Ulysses missions. Mixture models similar to the BGM have also been recently used to classify space plasma regions in magnetic reconnection zones (Dupuis et al., 2020). None of these previous



publications used data transformation to solve the classification problem in a more suitable latent space.

The colors used in **Figure 2** are assigned randomly by each clustering technique. The most glaring issue with them is that different methods can lead to different clusters of points. The BGM and the *k*-means do not agree on their classification in the PCA and the AE space. More importantly, for each technique, slight modifications of the clustering parameters, e.g., using a different seed for the random number generator, can lead to very different results. We address this last issue using an algorithm that launches the *k*-means (the BGM) algorithm 100 (30) times until the method converges to a global minimum. The final results are implementation dependent.

In the present data set, the cloud of points is convex and well-distributed in all components. This raises one additional issue, observed more clearly in the second column of **Figure 2**: when classical clustering methods are applied to relatively homogeneously dense data, it divides the feature space in Voronoï regions with linear hyper-plane boundaries. This is an issue with all clustering techniques based on discrimination of groups using their relative distances (to a centroid or to the mean of the distribution). To avoid this problem density-based techniques, such as DBSCAN (Ester et al., 1996), and agglomeration clustering methods, use a different approach.

However, we can not apply them here because in such homogeneous cloud of points these techniques lead to a trivial solution where all data points are assigned to a single class. An alternative projection was used by Bloch et al. (2020), who performed a Uniform Manifold Approximation and Projection (UMAP). We performed the same projection unsuccessfully: the Ulysses data used in that publication contains a very dense and large number of CH observations. ACE lacks such a rich variety of CH data, so applying a UMAP leads to a single class.

There is no guarantee that a single classification method, with a particular set of parameters will converge to a physically meaningful classification of the data if the points in the data do not have some level of separability, or have multiple zones of high density. This is also true for other classification methods based on *supervised learning*. The same issues will be observed when the training data include target classes derived from dense data clouds using simple hyper-plane boundaries, as done for the Z09 and X15 classes. An example of such application was published by Camporeale et al. (2017) and Li et al. (2020). The authors used the X15 classification to train supervised classifiers. No new information is gained with such methods, as the empirical boundaries are already mathematically known. A more compelling task would be to compare all classification methods against a *ground truth*, i.e., against a catalog of footprint

locations on the solar surface. But such catalog, to our knowledge, does not exist.

### 2.2.5. Self-Organizing Maps

#### 2.2.5.1. Classical SOM

Following the definitions and notations by Villmann and Claussen (2006), a class can be defined as  $C_i \stackrel{\text{def}}{=} \{x \in \Omega | \Phi(x) = \mathbf{w}_i\}$ , where  $\Phi$  is a function from  $\Omega$  to a finite subset of  $k$  points  $\{\mathbf{w}_i \in \mathbb{R}^N\}_{i=1..k}$ . A cluster  $C_i$  is then a partition of  $\Omega$ , and  $\{\mathbf{w}_i\}$  are the code words (also known as nodes, weights or centroids) associated. The mapping from the data space to the code word set,  $\Phi : \Omega \rightarrow \mathcal{W}$ , is obtained by finding the closest neighbor between the points  $x$  and the code words  $w$  (Equation 14). The code word  $\mathbf{w}_s$ , the closest node to the input  $\mathbf{x}_s$ , is called the *winning element*. The class  $C_i$  corresponds to a Voronoi region of  $\Omega$  with center in  $\mathbf{w}_i$ .

$$\Phi : x \rightarrow \arg \min_{i \in \mathcal{N}} (\|x - \mathbf{w}_i\|) \quad (14)$$

A Self-Organizing Map (SOM) is also composed of structured nodes arranged in a lattice, each one assigned to a fixed position  $\mathbf{p}_i$  in  $\mathbb{R}^q$ , where  $q$  is the dimension of the lattice (generally  $q = 2$ ). The map nodes are characterized by their associated code words. The SOM learns by adjusting the code words  $\mathbf{w}_i$  as input data  $x$  is presented.

The SOM is the ensemble of code words and nodes  $\{\mathbf{w}_i, \mathbf{p}_i\} \in (\Omega \times \mathbb{R}^q)$ . For a particular entry  $\mathbf{x}_s$ , the code word  $s \in \mathcal{N}$  is associated to the winning node  $\mathbf{p}_s$  if the closest word to  $\mathbf{x}_s$  is  $\mathbf{w}_s$ . At every iteration of the method, all code words of the SOM are shifted toward  $x$  following the rule:

$$\Delta \mathbf{w}_i = \epsilon(t) h_\sigma(t, i, s)(x - \mathbf{w}_i) \quad (15)$$

with  $h_\sigma(t, i, j)$  defined as the lattice neighbor function:

$$h_\sigma(t, i, j) = e^{-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\sigma(t)^2}} \quad (16)$$

where  $\epsilon(t)$  is the time dependent learning rate (Equation 17), and  $\sigma(t)$  is the time dependent lattice neighbor width (Equation 18). The training of the SOM is an iterative process where each data point in the data set is presented to the algorithm multiple times  $t = 0, 1, \dots, t_f$ . In these equations the subscript 0 refers to initial values at  $t = 0$  and the subscript  $f$  to values at  $t = t_f$ .

$$\epsilon(t) = \epsilon_0 \left( \frac{\epsilon_f}{\epsilon_0} \right)^{t/t_f} \quad (17)$$

$$\sigma(t) = \sigma_0 \left( \frac{\sigma_f}{\sigma_0} \right)^{t/t_f} \quad (18)$$

This procedure places the code words in the data space  $\Omega$  in such a way that neighboring nodes in the lattice are also neighbors in the data space. The lattice can be presented as a  $q$ -dimensional image, called map, where nodes sharing similar properties are organized in close proximity.

The main metric for the evaluation of the SOM performance is called the quantization error:

$$Q_E = \frac{1}{m} \sum_{i=1}^m \|x_i - \mathbf{w}_{x_i}\| \quad (19)$$

where  $m$ , is the total number of entries in the data set. It has been shown that the SOM tends to converge in the mean-square (m.s.) sense to the probabilistic density center of the multi-dimensional input subset (Yin and Allinson, 1995). This means that, if the SOM hyper-parameters are chosen correctly, the code words of the SOM will have a tendency to move toward high density regions of subsets of the input data, and will be located close to the mean of the subset points.

Once the training of the SOM is finished, the code words  $\mathbf{w}_i$  can be grouped together using any clustering technique, e.g.,  $k$ -means. The nodes of the SOM with close properties will be made part of the same class. The classes created are an ensemble of Voronoi subspaces, allowing a complex non-linear partitioning of the data space  $\Omega$ .

The final number of clusters is an input of the algorithm, but can also be calculated autonomously. The Within Cluster Sum of Squares (WCSS) can be used as a metric of the compactness of the clustered nodes. As its name implies the WCSS is the sum of the squared distances from each node to their cluster point. If only one class is selected, the large spread of the nodes would produce a high WCSS. The lowest possible value of the WCSS is obtained for a very high number of classes, when the number of classes is equal to the number of nodes. But such extreme solution is also unpractical. The optimal number of clusters can be obtained using the Kneedle class number determination (Satopaa et al., 2011). We use this automatic technique to let the machine select the optimal number of solar wind classes.

#### 2.2.5.2. Dynamic SOM

The time dependence of the SOM training allows the code words  $\mathbf{w}_i$  to reach steady coordinates by slowing down their movement over the iterations. Due to the minimization of the distance in Equation (14) code words tend to agglomerate around high density zones of the feature space. The Dynamic Self-Organizing Map (DSOM), introduced by Rougier and Boniface (2011), eliminates the time dependence and allows to cover larger zones of the space outside of the high density regions.

The DSOM is a variation of the SOM where the learning function (Equation 15) and the neighbor function (Equation 16) are replaced by Equations (20) and (21), respectively:

$$\Delta \mathbf{w}_i = \epsilon \|x - \mathbf{w}_i\|_\Omega h_\eta(i, s, x)(x - \mathbf{w}_i) \quad (20)$$

$$h_\eta(i, s, x) = e^{-\frac{1}{\eta^2} \frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{\|x - \mathbf{w}_s\|_\Omega^2}} \quad (21)$$

where  $\epsilon$  is a constant learning rate,  $h_\eta(i, s, x)$  is defined as the new lattice neighbor function, and  $\eta$  is the *elasticity* parameter. In their work (Rougier and Boniface, 2011) show that DSOM can be used to draw a larger sample of the feature space  $\Omega$ , reducing the agglomeration of code words around high density zones. The

main parameters of the DSOM,  $\eta$  and  $\epsilon$ , control the convergence of the method. A large  $\epsilon$  moves the code words,  $w$ , very fast with each new iteration; a very low value moves the points slowly in the space. A high elasticity,  $\eta$ , keeps all the nodes extremely close to each other, while a low value does not induce movement on far away code words. The best compromise is to use a very low value of the learning rate coupled with a mid-range elasticity, and a large number of training epochs. This can ensure a relative good convergence to a steady set of code words.

One special advantage of the DSOM is that it can be trained *online*, i.e., it is not necessary to re-train all the model when new data arrives: it adapts automatically to new information.

#### 2.2.5.3. Visualization of SOM and DSOM

Most clustering techniques do not guarantee to converge to a steady immutable solution. Differences in the training parameters or slight changes in the data can have an important impact on the final classification. Clustering tools can be used for statistical analysis, comparisons, data visualization and training of supervised methods. But it will be practically impossible to claim the existence of a general objective set of states discovered only by the use of these basic clustering techniques.

However, SOMs and DSOMs provide an important tool for the study of the solar wind: the maps are composed of nodes that share similar properties with its immediate neighbors. This allows for visual identification of patterns and targeted statistical analysis.

We used the python package MiniSom (Vettigli, 2013) as the starting point of our developments. Multiple methods of the MiniSom have been overloaded to implement the DSOM, and to use a lattice of hexagonal nodes. All auxiliary procedures used to calculate inter-nodal distances, node clustering, data-to-node mapping, and class boundary detection have been implemented by us. All visualization routines are original and have been developed using the python library Matplotlib (Hunter, 2007).

**Figure 3** shows the basic types of plots that can be generated using the SOM/DSOM techniques. We present in this figure the outcome of our model, combining a non-linear AE transformation of the ACE data set with the unsupervised classification of the encoded data using the DSOM method. Panel (A) shows a histogram of two components of the feature space  $\Omega$ , with dots marking the position of the code words  $w_i$ . The colors of the dots represent their DSOM classification. The red lines connect a single code word  $w_s$  with its six closest neighbors. Panel (B) shows the same information as in the previous panel, but using a scatter plot colored by the DSOM classification. This image shows the domain of influence of each one of the DSOM classes.

Panel (C) shows the *hit map* of the DSOM. It contains the lattice nodes  $p_i$  associated to the code words  $w_i$ . They are depicted as hexagons with sizes representing the number of data points connected to each node and colored by their DSOM class. The thickness of the lines between lattice nodes represent the relative distance to its neighbors in the feature space  $\Omega$ . Red lines connect the node  $p_s$ , associated to the code word  $w_s$  in panel (A), to its closest neighbors.

**Figure 3D** displays three components of the code words  $w_i$  associated to each one of the  $p_i$  nodes. The node components have been mapped to the basic colors Red, Green and Blue (RGB) and combined together to produce the composite color shown in the figure.

These four representations are only a few examples of the variety of data that can be represented using SOMs. The most important aspect of the SOMs is that data is represented in simple 2D lattices where the nodes share properties with their neighbors. Here we also decided to use hexagonal nodes, connecting 6 equidistant nodes, but other types of representations are also valid, e.g., square or triangular nodes.

#### 2.2.6. The Full Architecture

The previous sections introduced all the individual pieces that we use for the present work. Here, we give a global view of the full model. **Figure 4** shows how all the components are interconnected. At the center of the image is the processed and normalized original ACE data set. The blue dashed lines show the unsupervised techniques already presented by Heidrich-Meisner and Wimmer-Schweingruber (2018), Bloch et al. (2020), and Roberts et al. (2020). The KPCA step is added to the data pipelines used in the literature in order to project the data into a hyper-space where the class boundaries are better defined.

On the right side of the same figure we present our main approach: we perform first a data encoding using an AE, then we perform unsupervised classification of the solar wind with the  $k$ -means, BGM and DSOM methods. After training, the code words of the DSOM are clustered to group together nodes that share similar properties. This second level classification is done using the  $k$ -means++ algorithm with 100 re-initializations (it is in general recommended to use between 50 and 500 initializations, searching for a global optimum, as different random runs can lead only to a local minima). We use the Kneedle method to automatically select the number of classes that the DSOM will produce (Satopaa et al., 2011). The BGM and the  $k$ -means clustering techniques are included for comparison.

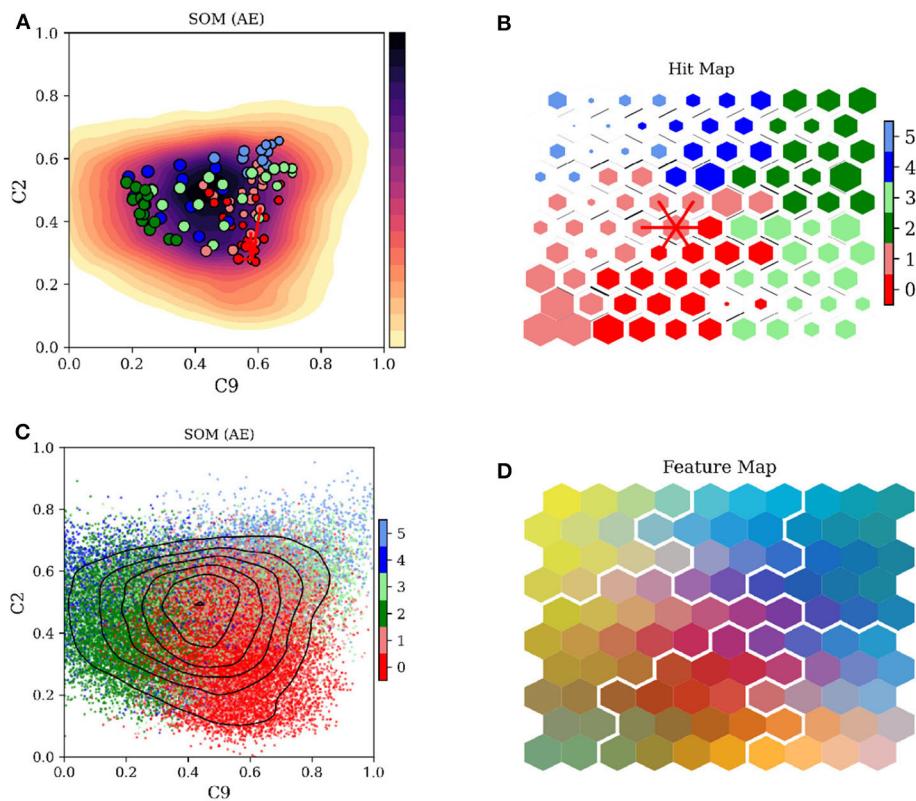
All the software was implemented in Python using as main libraries PyTorch (Paszke et al., 2019), Scikit-learn (Pedregosa et al., 2011), Matplotlib (Hunter, 2007), MiniSom (Vettigli, 2013), Pandas (McKinney, 2010), and NumPy (Oliphant, 2015).

##### 2.2.6.1. Feature selection

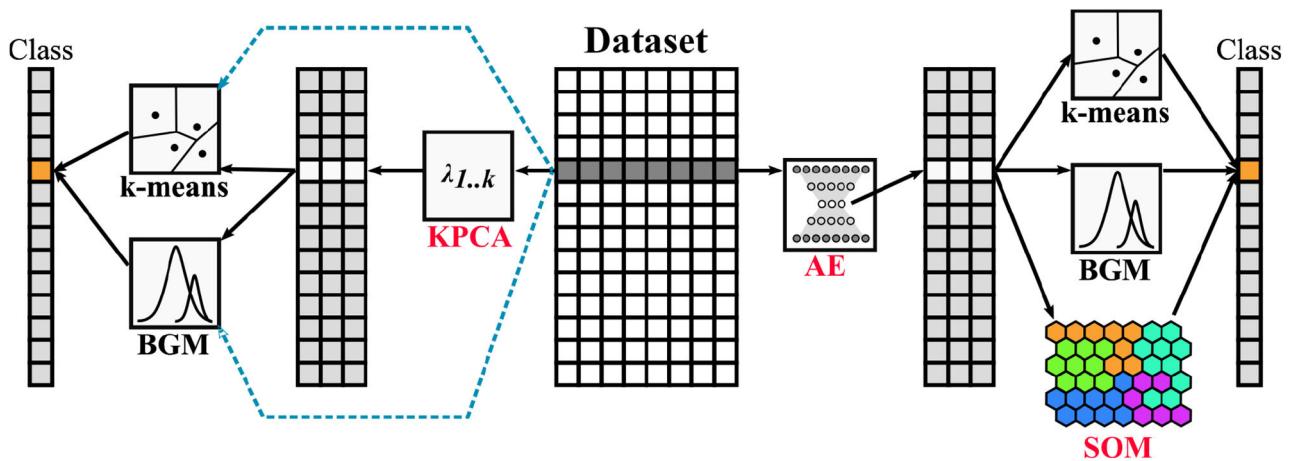
**Table 2** lists all the features used in our model. A detailed description of each feature can be found in the ACE Level 2 documentation. To spread the data over a larger range of values in each component, we have used the logarithm of all the quantities, except of those marked with an asterisk in the table.

Features 11–15 contain an additional suffix, corresponding to a statistical operation performed on the corresponding feature. In our model we only include *range* operations, but we have provided our software with the ability to calculate also the mean, the standard deviation and the auto-correlation of quantities over a window of time of 6 h. This window allows to capture temporal (spatial) fluctuations in some of the solar wind parameters.

On the lower part of **Table 2** we present the range of dates used for the model. The same table also contains the hyper-parameters



**FIGURE 3 |** Visualization of the Self-Organizing Maps. **(A)** Histogram with the normalized density of data points superposed by the code words of the DSOM, projected on two components of the latent AE space. A single node is connected to its closest neighbors by red lines. **(B)** Scatter plot of all data points, colored by the DSOM class. **(C)** Hit map: the size of the hexagon corresponds to the number of data points associated to the map node, and the color is the corresponding DSOM class. Black lines between nodes represent their relative distance. Red lines connect the nodes similarly highlighted in **(A)**. **(D)** Map of the nodes colored by three of their components, combined as a single RGB color. White lines mark the boundaries between DSOM classes.



**FIGURE 4 |** General overview of the pipelines tested in this work. Starting from the center, the ACE data set is processed and normalized. Blue dashed lines show the work done in previous publications by different authors. Black lines show how data in this work is first transformed and then classified using multiple methods. The original techniques presented in this paper are highlighted in red.

selected to run the two models. The number of neurons per layer in the encoding half of the neural network is listed in the table.

#### 2.2.6.2. Autoencoder architecture

We use a basic, fully connected feed-forward neural network for the encoding-decoding process. The neural network is symmetric in size but the weights of the encoder,  $\mathbf{W}$ , and the decoder,  $\mathbf{W}'$ , are not synchronized (see Equations 11, 12). Each layer is composed of a linear regressor, followed by a GELU activation function. The output layer of the network contains a linear regressor followed by a sigmoid activation function. The AE has been coded in python using the PyTorch framework (Paszke et al., 2019).

The final architecture of the AE and its hyper-parameters have been optimized automatically using the *Optuna* library (Akiba et al., 2019). We instructed this Hyper-Parameter Optimization (HPO) to select the optimal values for the following parameters, given the corresponding constraints:

- **Number of layers:** an integer between 2 and 6.
- **Number of neurons per layer:** it must be larger than 3 and smaller than the number of neurons in the previous encoder layer.
- **The neural network optimizer:** selected among Adam, Stochastic Gradient Descent, and RMSprop.
- **The learning rate:** a float value between  $10^{-5}$  and  $10^{-1}$ .

The automatic HPO is based on a technique called Tree-structured Parzen Estimator (TPE) (Bergstra et al., 2013), which uses Bayesian Optimization to minimize a target function,  $\mathcal{H}$ , provided by the user. We use the test loss of the AE as target function to be minimized.

The HPO performs a total of one thousand (1,000) different trials. However, to accelerate the optimization process, we built a smaller complementary data set. To avoid over-fitting on a subset of the original data we used the  $k$ -means algorithm to produce a representative sample of  $m' = 4\sqrt{m}$  data points. This allows to explore a much broader set of hyperparameters in a short period of time. This artificial data set is then discarded and the AE is trained on the real data set.

The HPO selected the Adam optimizer (Kingma and Ba, 2014) for the gradient descent with a learning rate of 0.042. The total number of layers selected is 2, and the number of nodes in the bottleneck is 10. The loss function is the Mean Squared Error (MSE). We train the network for 500 epochs, after which no additional improvement in the loss function is observed. The full data set was randomly divided 50/50% between training and testing sets. We track the evolution of both data sets during training. We did not observe any variance or bias error.

The final architecture is trained using the full data set for 500 epochs. **Figure 5** shows the distribution of data in the original feature space, panel (A), and in the AE latent space, panel (B). The data in the original space contains extreme data points far from the mean value, and most features present a normal distribution. The combination of these two properties makes it difficult for any unsupervised clustering technique to separate points and accurately categorize different kinds of solar wind.

Panel (C) shows the error in the encoding-decoding procedure of the AE. It shows a histogram of the relative error,  $\mathbf{E}_r = \hat{\mathbf{X}}/\mathbf{X} -$

1, observed between the input data,  $\mathbf{X}$ , and the decoded values,  $\hat{\mathbf{X}}$ . A normal distribution function has been fitted to the values of the histogram. It shows that the relative error is centered near zero and its variance is around 1%.

#### 2.2.6.3. Selection of parameters for the DSOM

In this manuscript we have introduced the use of the DSOMs for the classification of solar wind data. This technique requires the selection of four main Hyper-Parameters (HPs): the size of the lattice,  $(L_x \times L_y)$ , the constant learning rate,  $\epsilon$ , and the elasticity,  $\eta$ . These last two parameters were chosen manually, while the lattice size was automatically selected by Hyper-Parameter Optimization (HPO) using *Optuna* (Akiba et al., 2019).

For the selection of the number of nodes in the lattice we propose the use of the objective function,  $\mathcal{H}$ , described in Equation (22):

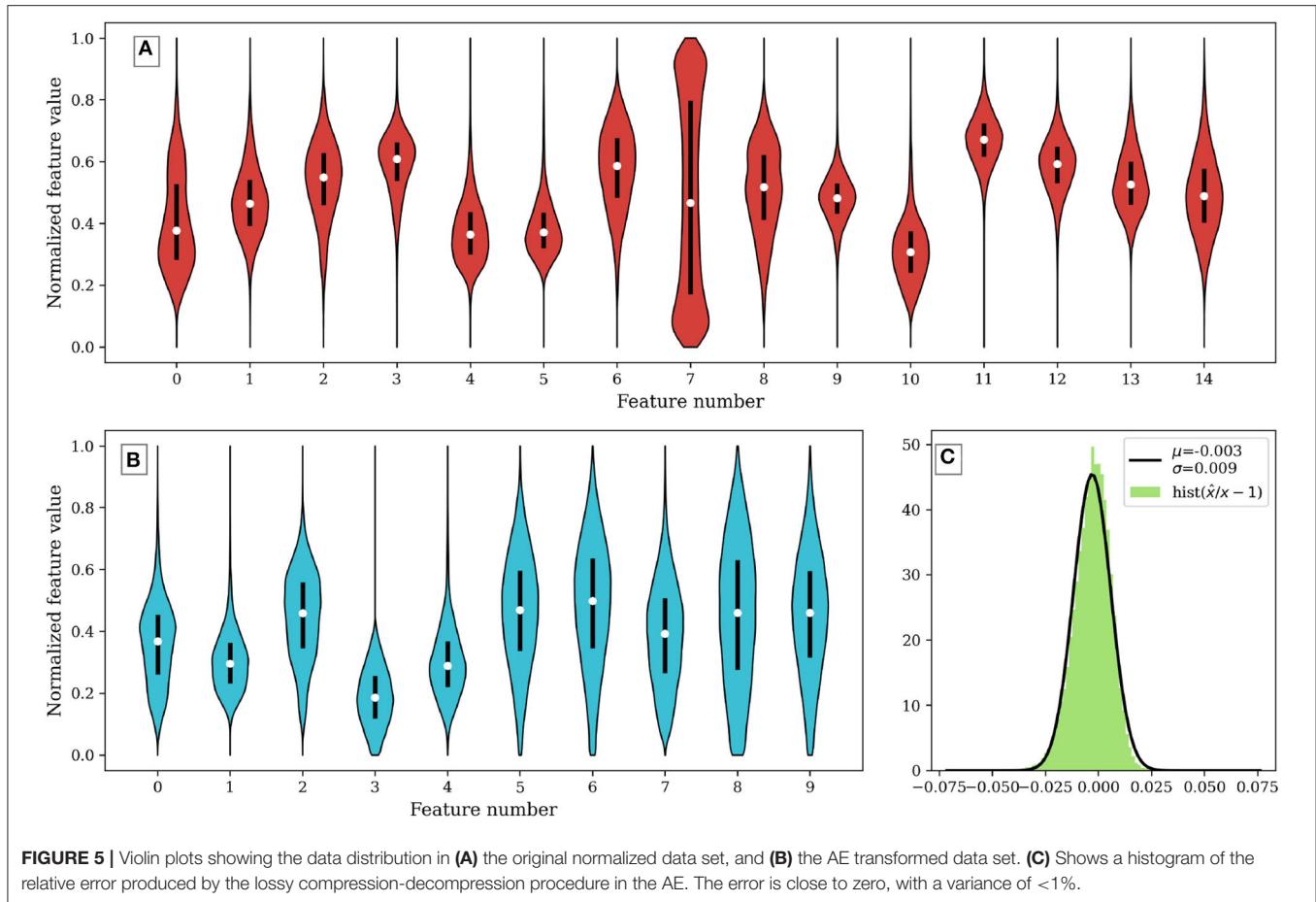
$$\begin{aligned} \mathcal{H}(\sigma, \eta, L_x, L_y) = & \frac{Q_E(\sigma, \eta, L_x, L_y)}{Q_0} + \alpha \frac{L_x}{m_{\max}} + \beta \frac{L_y}{n_{\max}} \\ & + \gamma \frac{L_x L_y}{\max(m_{\max}, n_{\max})} \end{aligned} \quad (22)$$

where  $Q_E$  is the quantization error at the end of the training,  $Q_0$  is a reference quantization error before training,  $L_x$  and  $L_y$  are the number of lattice nodes in each dimension, and  $m_{\max}$  and  $n_{\max}$  are the given maximum number of possible nodes. The weight factors  $\alpha$ ,  $\beta$ , and  $\gamma$  are used to impose restrictions on each term. We have fixed their value to  $\alpha = \beta = 0.08$  and  $\gamma = 0.6$ . When a large number of nodes is available smaller values of  $Q_E$  are automatically obtained because the mean distance from the data set entries to the code words is reduced. The second and third terms on the RHS of  $\mathcal{H}$  leads the optimizer to reduce the number of nodes in the SOM. The squaring term  $\gamma L_x L_y$  forces the map to be as squared as possible.

After a total of 500 trial runs of the model using different HPs, the optimizer selected the parameters presented in the lower section of **Table 2**. The optimization was accelerated using the same technique as in the optimization of the AE: we generated a reduced number of points using the  $k$ -means algorithm, with a total number of entries equal to one twentieth the size of the full data set,  $m' = \frac{1}{20}m$ .

The two remaining parameters of the DSOM, the elasticity  $\eta = 3.0$  and the learning rate  $\epsilon = 0.005$ , have been manually selected. These two values control the speed at which the code words move toward the data entries, and the attraction between neighboring code words. It has been shown by Rougier and Boniface (2011) that high values of the elasticity,  $\eta$ , lead to tightly packed code words, while low values lead to loose connections. The elasticity takes in general values between 1 and 10. On the other hand, the learning rate indicates to the code words how fast they should move toward new incoming data. Very small learning rates could lead to very slow convergence to a solution, while very large values might produce code words that jump from value to value without converging to a global solution. The value of the learning rate can be set somewhere between 0.001 and 0.9.

**Figure 6** shows how the elasticity and the learning rate can affect the convergence of the DSOM. In this figure we evaluate



**FIGURE 5 |** Violin plots showing the data distribution in **(A)** the original normalized data set, and **(B)** the AE transformed data set. **(C)** Shows a histogram of the relative error produced by the lossy compression-decompression procedure in the AE. The error is close to zero, with a variance of <1%.

the effect of using different values of  $\eta$  and  $\epsilon$ . Three different graphs are used to understand the evolution of the training and its convergence to a stable solution. The first row shows how the code words move away from their original position during the training: as the iterations advance the code words move until they find a stable location. It is clear that lower values of  $\epsilon$  and  $\eta$ , as presented in the left panel of the first row, lead to very long convergence times. At the other extreme, very high values of the two parameters produce strong movements with a compact group of code words, leading to a non-converging solution.

In the second row of the same figure we show the distance traveled by the code words at each iteration of the training. In the best case scenario this distance is large at the beginning of the training and converges toward zero as the iterations pass. The third panel of this row shows how large values of  $\eta$  and  $\epsilon$  produce solutions of the DSOM that do not converge.

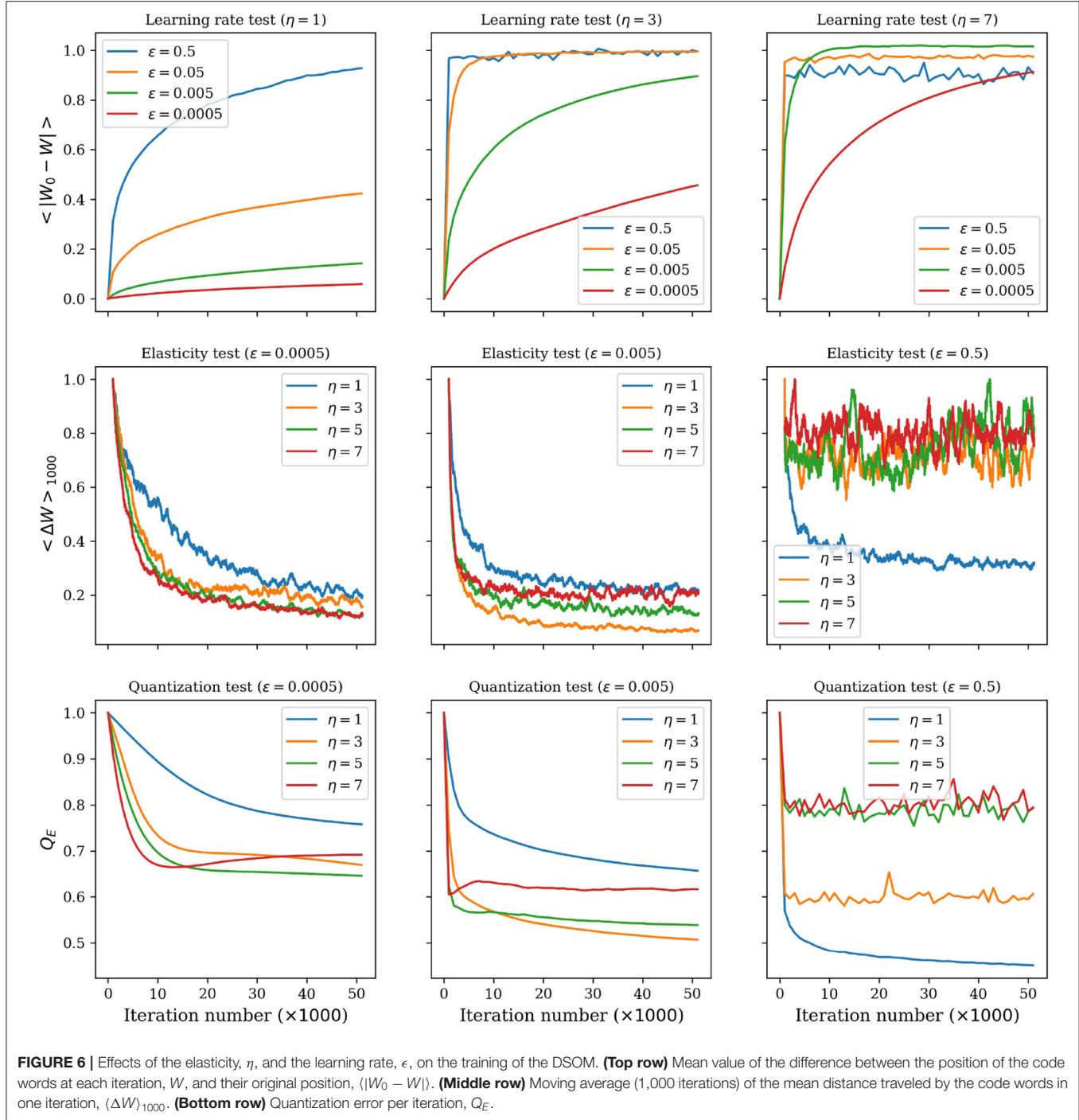
The third row of **Figure 6** shows the evolution of the quantization error (Equation 19). This value explains the compactness of the data points around the code words. Scattered points will show large  $Q_E$ , while dense clouds of points gathered around the code words will show low  $Q_E$  values. Once again in this last row we see that there is a compromise between a slow convergence with small values of  $\epsilon$ , and large values of the two parameters that can lead to unstable solutions.

This figure also shows that, even if the DSOM is a dynamic technique that does not use a decay of the learning rate with time, it is a method that converges to a steady solution, if the parameters are properly selected.

#### 2.2.6.4. Budget

Machine learning models require fine tuning of different parameters, from the selection and testing of multiple methods, to the parameterization of the final architecture. Dodge et al. (2019) suggests that every publication in machine learning should include a section on the budget used for the development and training of the method. The budget is the amount of resources used in the data processing, the selection of the model hyperparameters (HP), and its training.

The most time-consuming task in the present work has been the data preparation, the model setup and debugging and the writing of the SOM visualization routines. All the techniques described in the previous sections have been coded in python and are freely accessible in the repositories listed in the Data Availability Statement. We estimate the effort to bring this work from scratch to a total of two persons month. Of these, one person week was dedicated to the manual testing an selection of different model HPs (autoencoder architecture, feature selection, learning rates, initialization methods, number of epochs for training,



selection of data compression method, size of the time windows, etc.).

All classic clustering techniques presented in section 2.2.4 require only a few lines of code and can be trained in minutes on a mid-range workstation (e.g., Dell Precision T5600, featuring two Intel(R) Xeon(R) CPU E5-2643 0 @ 3.30GHz with four cores and eight threads each). The most time consuming tasks of our models are the training of the autoencoder (5% of the total run

time), the multiple passages of the clustering algorithms (15% of the run time), and the optimization of the hyper-parameters (80% of the run time). The training of the DSOM is performed in less than a minute.

For reference, the total run-time of our model is 30 min. The python scripts used do not contain any particular acceleration (e.g., using GPUs) or optimizations (e.g., using Numba), so there is large room for improvement of the computational efficiency.

### 3. RESULTS AND COMPARISONS

#### 3.1. Interpretation of the DSOM Plots

When the DSOM method converges to a solution, each one of the code words is a representative of their N-dimensional neighborhood. We perform then a k-means clustering of the code words and apply the Kneedle method (Satopaa et al., 2011), presented in section 2.2.5.1, to select the final number of classes. Here, the automatic procedure selects a total of six classes, numbered from 0 to 5. The *Class Map* on the first panel of **Figure 7** shows that all nodes are organized in continuous groups.

The weights of the code words can be decoded and scaled to obtain the corresponding physical properties of the associated solar wind. These physical quantities are plotted in **Figure 7** for each one of the solar wind features.

Black continuous lines in the maps mark the boundary between different DSOM classes. All of the maps show uninterrupted smooth transitions between low and high values, without sudden jumps or incoherent color changes. Inside DSOM classes solar wind properties can present variations. This is an expected consequence of projecting 15 dimensions in a 2D lattice.

The most obvious class to identify is the DSOM class 0, with clear indications of coronal hole origin. It is characterized by very low values of the O<sup>7+</sup>/O<sup>6+</sup> and C<sup>6+</sup>/O<sup>5+</sup> ratios, associated with plasma originating from open magnetic field lines (Zhao et al., 2009; Stakhiv et al., 2016), high wind speed, low proton density, high absolute values of  $\sigma_c$  (a sign of Alfvénicity), high proton entropy, high proton temperature and moderately high values of Alfvén speed [associate by Xu and Borovsky (2015) with coronal holes].

The proton density has a very broad range of values for class 1. A close examination of the map of cross-helicity,  $\sigma_c$ , shows that this class also contains Alfvénic solar wind with both polarities. Class 1 also showcases high proton temperatures, high solar wind speeds, but average oxygen and carbon ionization ratio, and average iron charge. All these observations point toward solar wind originated at the boundary of coronal holes (Zhao et al., 2017).

Class 5 can be associated to transient event, such as ICME and ejecta. It presents the very high O<sup>7+</sup>/O<sup>6+</sup> ratio values that Zhao et al. (2009), Xu and Borovsky (2015), and Stakhiv et al. (2016) associate to CME plasma, and the low proton temperature values usually found in ICMEs. It is also characterized by the high (but, quite surprisingly, lower than for class 0) solar wind velocity,  $\sigma_c \sim 0$  (Roberts et al., 2020), the high values of Alfvén speed which are usually associated to explosive transient activity (Xu and Borovsky, 2015).

Class 4 has similar properties as class 5 and can be mainly composed of transient events, but it also contains more Alfvénic plasma, and very high carbon charge state ratios, C<sup>6+</sup>/C<sup>5+</sup>. Fluctuations in this class are slightly less significant than the ones observed in class 5, except for jumps in the normal magnetic field, range  $B_n$ . These can point toward a class that contains magnetic clouds or Sector Boundary Crossing (SBC) events. Classes 4 and 5, identified as transients, remain rare, as clearly shown in **Figure 3C**.

At this point is important to remember that a different set of initial conditions or a different number of map nodes could lead to a slightly different repartition of the data, or to a different number of classes. However, points with similar properties will always remain topologically close and the interpretation of a different set of DSOM classes will lead to similar results. This is not necessarily the case with other unsupervised methods, like k-means, as the topological organization of the data is not maintained, so different runs can produce different results for which previous interpretations can not be re-cycled.

Class 2 and 3 are composed of slow, dense solar wind, the kind of wind that Zhao et al. (2017) associates to the *Quiet Sun* and that Xu and Borovsky (2015) associates to either Streamer Belt (SB) or Sector Reversal (SR) region plasma. As expected for the slow wind, the cross helicity is low, the proton temperature intermediate between the low values associated to ICMEs and the higher values observed in the fast wind, the proton entropy and the Alfvén speed are low (Xu and Borovsky, 2015). The high O<sup>7+</sup>/O<sup>6+</sup> and C<sup>6+</sup>/C<sup>5+</sup> ratios (lower only to the values associates to class mappable to transient events, class 4 and 5), point to plasma originating in closed field lines (Zhao et al., 2009; Stakhiv et al., 2016). Of the two classes, class 2 is characterized by lower wind speed, higher density, lower proton temperature, lower entropy.

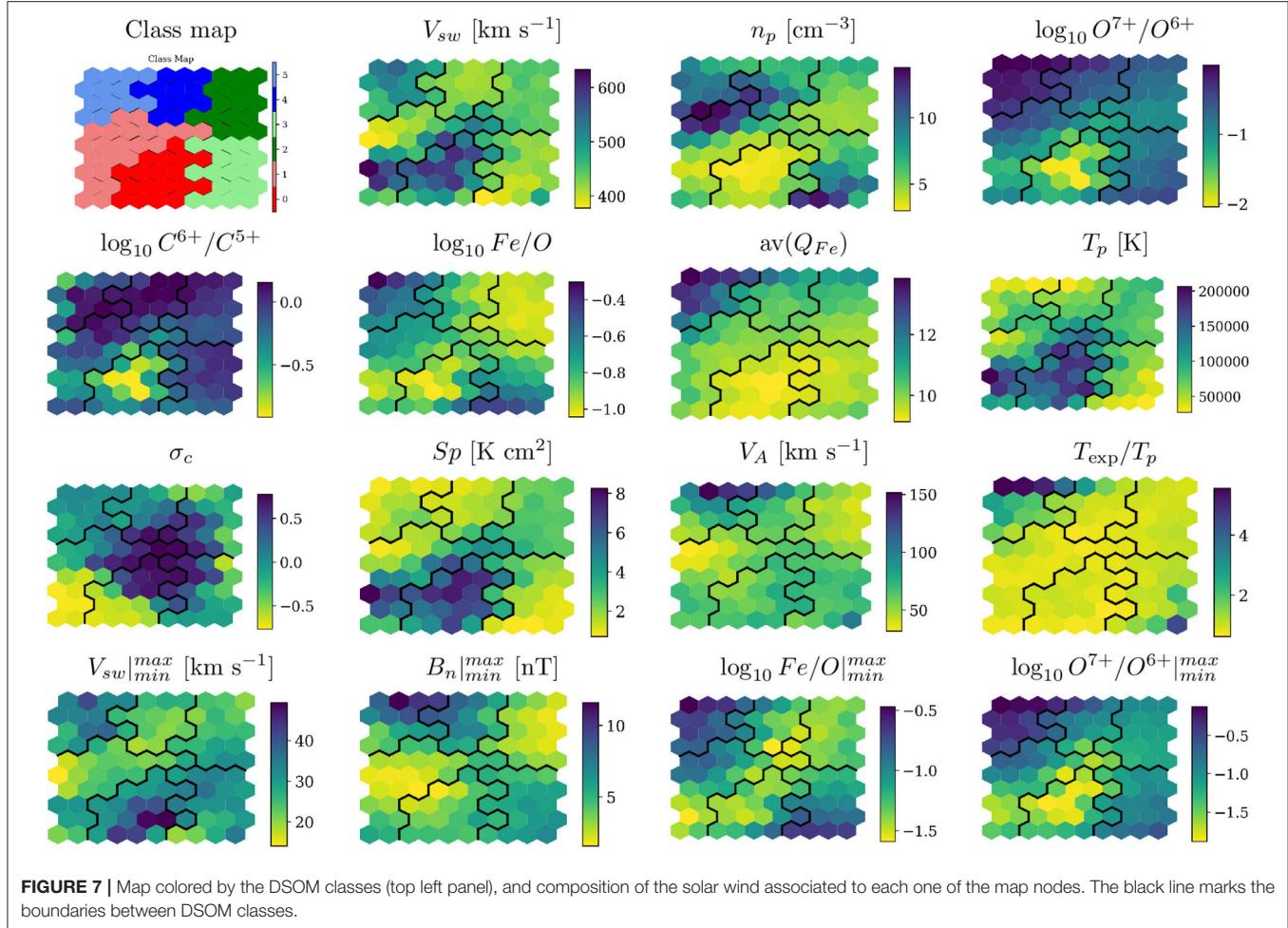
In summary we can group our classes on three major categories: CH wind (classes 0 and 1, colored in red), quiet or transitional wind (classes 2 and 3, colored in green), and transients (classes 4 and 5, colored in blue).

#### 3.2. Verification of the DSOM Classes

In addition to the interpretation of the maps presented in the previous section, we have extracted histograms of the occurrence frequency of O<sup>7+</sup>/O<sup>6+</sup> ratio (**Figure 8**) and proton speed,  $V_{sw}$  (**Figure 9**). The panels in the figures contain the histograms for six (6) different categorizations: k-means (AE), k-means (KPCA), BGM (AE), BGM (KPCA), DSOM, and the X15 classification. All the histograms have been normalized row by row (class by class), following the work done by Zhao et al. (2017). This representation of the data is inspired by Figure 5 of that paper, where the authors showed an important overlapping among different solar wind classes, and a bi-modal velocity distribution for coronal hole wind including an important population of slow wind.

The assignment of class numbers by the clustering algorithms is random. We have sorted the classes so they present an ascending value of the O<sup>7+</sup>/O<sup>6+</sup> ratio in **Figure 8**. It has been shown that solar wind originated in *Coronal Holes* present very low values of the O<sup>7+</sup>/O<sup>6+</sup>, while at the other extreme transient events present very high O<sup>7+</sup>/O<sup>6+</sup> ratios (Zhao et al., 2009, 2017; Stakhiv et al., 2016). **Figure 8** confirms the class identification we presented in the previous section.

von Steiger and Zurbuchen (2015) and Bloch et al. (2020) examine the O<sup>7+</sup>/O<sup>6+</sup> ratio in Ulysses data, which include abundant measures of wind originating from the polar CHs. Our data is composed of ACE observations from the ecliptic plane. For this reason, in all different classifications in **Figure 8**,



including the X15 empirical categorization, class 0 does not reach  $\log_{10} O^{7+}/O^{6+} \approx -2$ , where the peak of points is observed in publications using Ulysses data.

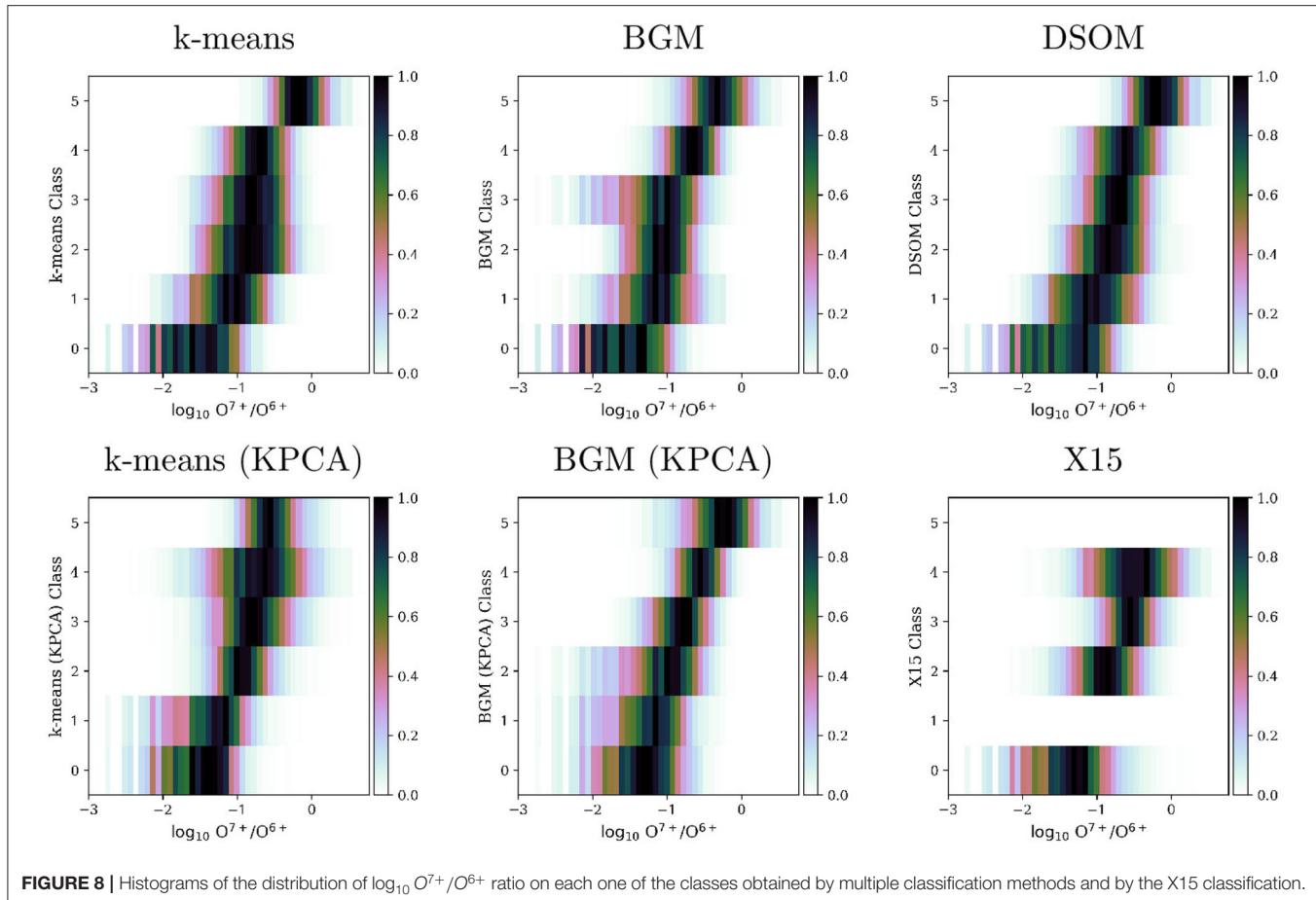
In our data set, the majority of points can be mapped to *Quiet Sun* (QS), conditions, i.e., slow solar wind. Even in these conditions, the DSOM method is able to sample and distribute enough data to each one of the classes. The BGM method applied to Kernel PCA transformed data also provide a good sample of the different classes, in particular for transient solar wind (classes 4 and 5). The X15 classification was designed with clear boundaries in  $O^{7+}/O^{6+}$ , for this reason the differences among the four classes is clear in the histograms. However, this observation contradicts the foot point back tracing performed by Zhao et al. (2017): X15 shows almost no overlap in the distribution functions between the different classes, while the back tracing shows important overlaps. We express caution in the use of this classification to train any type of supervised machine learning technique, or in the evaluation of the accuracy of unsupervised techniques.

**Figure 9** shows how velocity is distributed among the different classes for each unsupervised classification method, and for the X15 categories. Zhao et al. (2017) remarks that the different classes are more difficult to identify using the solar wind speed

histograms. We verify in these plots that three conditions are satisfied: (1) the classes we associate to the QS (class 2 and 3 in the DSOM classification) are associated to low velocity regions (Neugebauer et al., 2002), (2) high oxygen state ratios are associated with low solar wind speeds, and (3) CH wind has a highly spread velocity distribution, with two possible peaks around 400 and 600 km/s (Zhao et al., 2017).

The fact that class 0 and 1, that we associate to wind of CH origin, contains slow wind data points is particularly significant. D'Amicis and Bruno (2015) has provided proof of the presence, at 1 AU, of highly Alfvénic slow wind originating from the boundaries of coronal holes. This slow, Alfvénic wind has the same composition signature and high cross helicity that characterized the *classic* fast Alfvénic wind of CH origin, but presents lower speed and lower proton temperature. This way of visualizing our results seems to suggest that slow Alfvénic wind is classified together with fast Alfvénic wind in the classes that we associate to CH origin.

**Figure 9** shows that CH wind in the *k*-means and DSOM plots present a broad range of speeds, with a bimodal distribution. The BGM (KPCA) method separates these two populations in two different classes (0 and 1). the *k*-means (KPCA) method



differentiates the fast solar wind, in the first two classes, from the slow wind in the remaining classes.

Balancing the results from Figures 8, 9 we conclude that the BGM (KPCA) and the DSOM are the techniques that approach the most the direct observations of the solar wind origins obtained by Zhao et al. (2017). The X15 model creates a very sharp separation of solar wind types, with fast winds clearly segregated in class 0, slow winds in classes 1 and 2, and transients in class 3. The X15 model does not recognize that plasma of CH origin also contains an important population of slow winds.

### 3.3. Hit Maps of Empirical Classifications

Another advantage of the SOM/DSOM method is that it can be used to visualize additional hidden statistics. Figure 10 shows what nodes are activated by the Z09 and X15 classes. To perform this analysis, instead of using the full data set, we extract three subsets corresponding to the entries categorized as CH, ICME, and NCH wind in the Z09, and CH, SB, SR, and ICME in the X15 catalogs. Each one of these three (four) subsets is passed through the DSOM model and we observe how each one activates the nodes.

We see that CH wind, in column 1 of the figure, activates very similar nodes for both classifications, in classes 0 and 1. Most of the hits are located on nodes where the absolute value of the

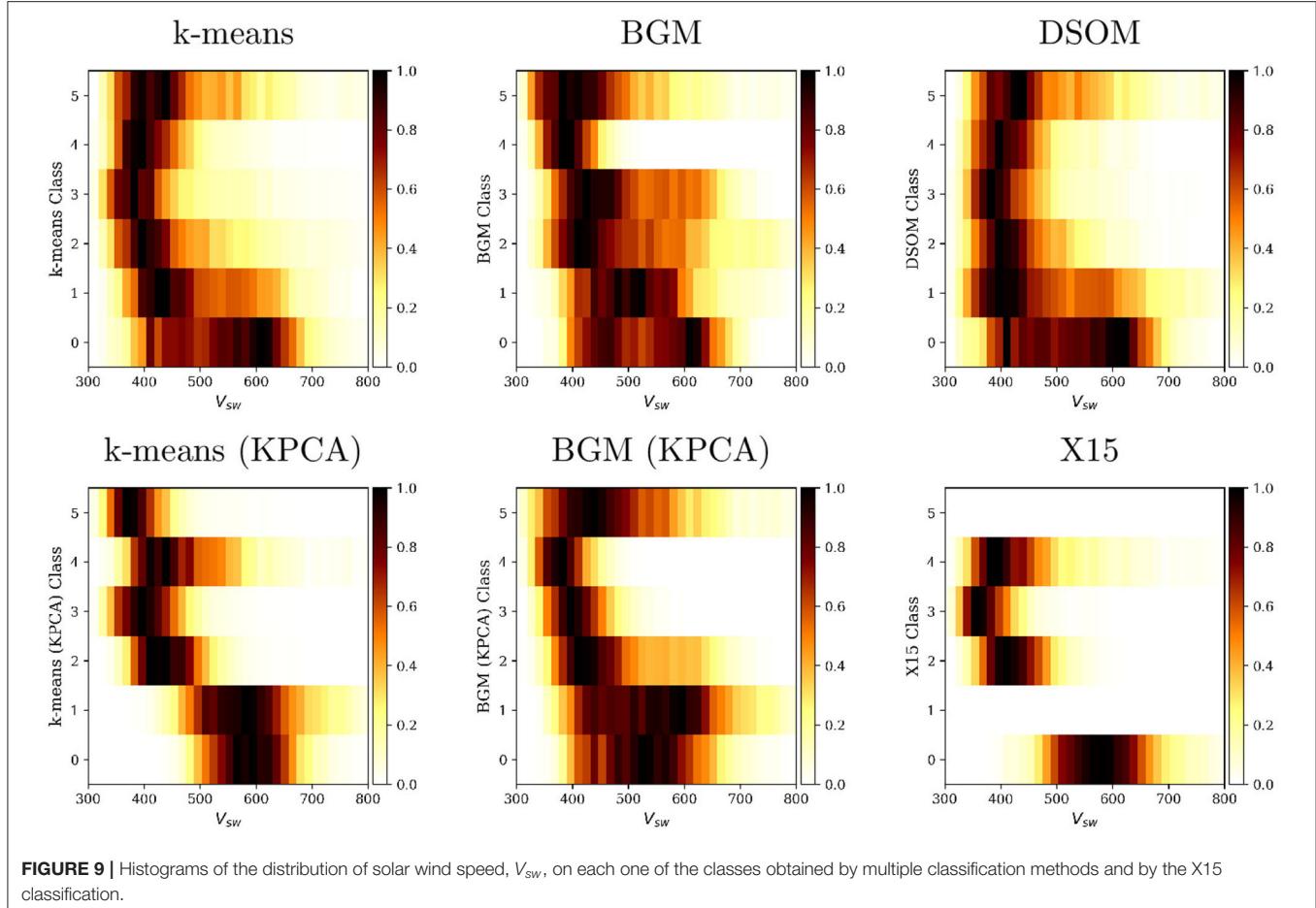
cross-helicity  $\sigma_c$  is the largest, i.e., in regions of open field lines associated with coronal holes.

NCH wind from the Z09 classification is distributed over classes 2, 3, and 4, but also includes a node from class 1 characterized by an extremely negative cross-helicity. The same zone is activated by the SB class from X15. The two affected nodes also feature a very low  $T_{\text{exp}}/T_p$ . The X15 model splits solar wind points using hyperplanes in a three-dimensional space composed by  $S_p$ ,  $O^{7+}/O^{6+}$ , and  $T_{\text{exp}}/T_p$ . None of those planes cuts the points in the  $T_{\text{exp}}/T_p$  dimension (Xu and Borovsky, 2015). However, in our maps this dimension seems to play an important role in the separation between quiet and CH winds.

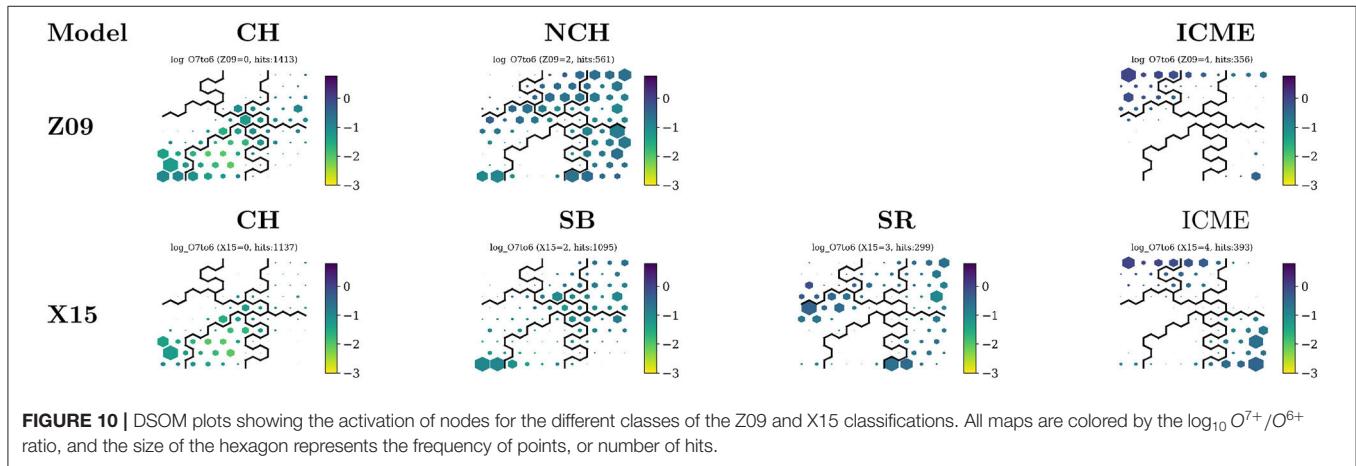
The X15 Sector Reversal (SR) class activates nodes at the boundaries of classes 1, 2, and 3. These nodes separate the quiet sun from the coronal hole wind, and coronal holes to transients. It also contains a large population of slow quiet solar wind.

Finally transients, in both the Z09 and X15 categorizations, are associated to our class 4 and 5. However, a large portion of the X15 transients is associated to class 3 of the DSOM, particularly in nodes showing low proton temperatures and specific entropy  $S_p$ , characteristics of ICMEs.

Figure 10 shows also that, on average, the values of the  $O^{7+}/O^{6+}$  ratio do not change radically among the nodes, except for small variations in the CH and the ICME classes of X15.



**FIGURE 9** | Histograms of the distribution of solar wind speed,  $V_{sw}$ , on each one of the classes obtained by multiple classification methods and by the X15 classification.



**FIGURE 10** | DSOM plots showing the activation of nodes for the different classes of the Z09 and X15 classifications. All maps are colored by the  $\log_{10} O^{7+}/O^{6+}$  ratio, and the size of the hexagon represents the frequency of points, or number of hits.

### 3.4. Quantitative Comparison With Empirical Classifications

We have included a *Matching Matrix* in **Table 3** showing the frequency of occurrences of our model with respect to the Z09 and X15 classifications. Bold numbers in the table mark the highest common frequency and regular fonts mark the second highest frequency for

each one of the columns. Matching matrices must not be confused with confusion matrices, as the later imply that there is a *ground truth*. Matching matrices are used in unsupervised learning to compare the frequency of occurrence of classes between models, so we can not perform additional metrics, like accuracy, precision, sensitivity, or specificity.

**TABLE 3 |** Matching matrix comparing the DSOM, X15, and Z09 classifications.

		DSOM Class					Z09			
		0	1	2	3	4	5	CH	NCH	TR
X15	CH	<b>7,727</b>	3,994	3,330	125	502	47	<b>14,273</b>	993	459
	SB	<u>7,423</u>	<b>6,295</b>	<b>7,194</b>	<u>916</u>	<b>950</b>	138	<u>12,904</u>	<b>9,244</b>	768
	SR	3,233	392	1,404	584	434	<u>157</u>	877	<u>4,747</u>	580
	TR	1,263	575	387	<b>2,373</b>	403	<b>1,362</b>	1,343	<u>3,011</u>	<b>2,009</b>
DSOM Class	0						<b>13,357</b>	5,966	323	
	1						<b>9,040</b>	<u>2,213</u>	3	
	2						<b>6,173</b>	<u>5,789</u>	353	
	3						637	<b>2,848</b>	513	
	4						170	987	<b>1,132</b>	
	5						20	192	<b>1,492</b>	

Values in bold (underline) font represent the highest (second highest) frequency for each column (row) in the top (bottom) half of the table.

In this matrix we see that CH and SB categories from the X15 classification are mostly associated with classes 0, 1, 2, and 4 in the DSOM model, while TR winds are associated with classes 3 and 5. No particular class is clearly associated with SR winds, but the highest frequency is observed for class 0.

CH in the Z09 classification are accurately associated with classes 0 and 1, but a big part of the NCH wind is also grouped in class 0. Transients are correctly distributed among classes 4 and 5 of the DSOM.

We highlight that the X15 and Z09 models, the two classifications most used for the verification of machine learning results (see Camporeale et al., 2017; Li et al., 2020), are not fully compatible among themselves. A large number of CH winds from the Z09 classification is associated with SB winds in the X15 classification, and a considerable number of transients are cataloged as sector boundary crossings (SB).

### 3.5. Time Series Comparisons

A complementary method to compare the different classification techniques is to visually inspect windows of time and check, with the help of a human expert, that the time series are in agreement with the previous analysis. **Figure 11** shows, in two columns, two windows of time of 4 months. The left column contains a high solar activity period, from May 2003 to September 2003, and the right column contains a period of low solar activity, between January 2008 and May 2008. Each one of the eight (8) rows contains a plot of the solar wind speed colored by a different classification method, from empirical models (Z09, vS15, and X15) to unsupervised methods (*k*-means, *k*-means, BGM, and DSOM). The colors of the empirical methods in the time series correspond to the labels assigned in **Table 1**, and the colors of the models were all assigned by manually ordering the classes following the frequency  $\log_{10}(O^{7+}/O^{6+})$ , from low values (low category number) to high values (high category number).

In the same figure vertical gray zones correspond to Richardson and Cane ICME catalog entries (Richardson and Cane, 2012), and vertical lines to entries in the UNH and CfA catalogs.

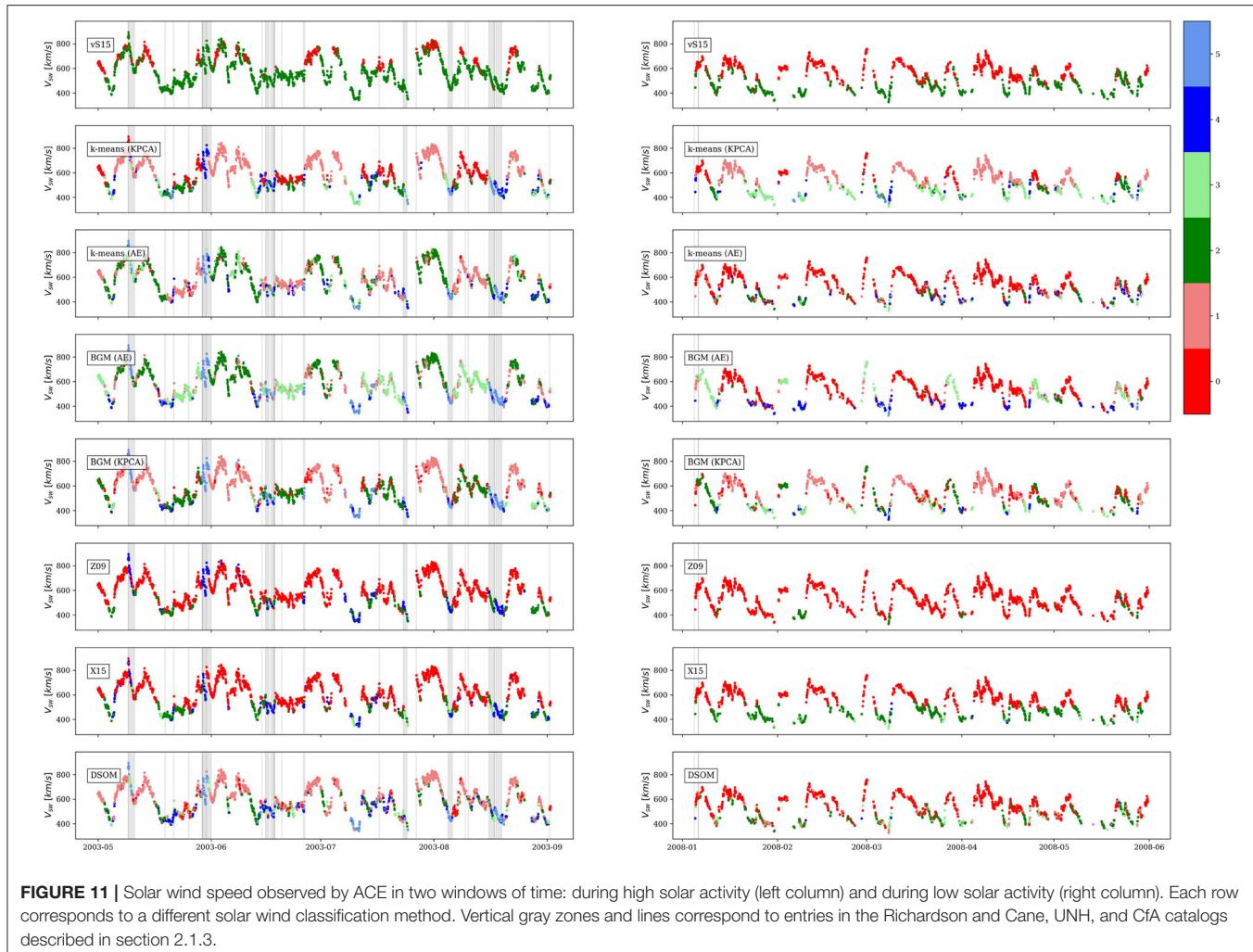
It is clear that among the empirical models, the vS15, based on observations by the Ulysses mission, is the most restrictive in the selection of CH origin winds, however during the plotted quiet time in the right column, which corresponds to the declining phase of the solar cycle, a significant part of the solar wind originates in coronal holes, and in fact High Speed Stream and Corotating Interaction regions, associated to wind of CH origin, are the main driver of geomagnetic activity during the declining phase of the cycle (Tsurutani et al., 2006; Innocenti et al., 2011). During both solar activity windows the Z09 and X15 models assign an important number of observations to coronal holes. von Steiger and Zurbuchen (2015) shows that the threshold used in the Z09 classification to identify coronal holes is not accurate and can misclassify NCH as CH. Both models accurately identify transients in the data. Quiet solar wind is more clearly visible during the low solar activity window in the X15 model.

*k*-means (KPCA) and BGM (KPCA) correctly classify CH origin winds (classes 0 and 1). A clear transition between class 1, CH wind, and class 3 can be observed on both panels. Transients are also well-captured with classes 4 and 5. On the other hand, classifications based on the *k*-means (AE) and BGM (AE), do not show high accuracy in these two windows of time, but are able to detect transients. These two methods show difficulties in discerning QS winds from CH solar winds.

The DSOM model shows good performances. The two classes associated with CH origin wind, classes 0 and 1, are more restrictive than the Z09 and X15 classes. Classes 4 and 5 distinguish between two different types of transients. ICMEs in these time windows are mainly associated with class 5, except for transients observed around 2003-05-20 and 2003-06-15. The model also detects a very slow transient around 2003-07-10. The 27 days solar period is also evident on the oscillations of the solar wind speed and the periodic nature of the solar wind types. In the low solar activity window the solar wind is more homogeneous and shows mainly CH and QS origin winds, as expected (Tsurutani et al., 2006).

Different classification methods lead to different classes with different properties. Roberts et al. (2020) performed detailed descriptions of the categorized solar wind classes based on the mean values observed in each subset of points. Zhao et al. (2017) shows that it is important to look at the frequency distribution and not only the mean. Our model shows that some features can present very large distributions inside a single class, even multiple peaks, as is the case of the solar wind speed for the CH classes.

We will perform further refinements of the model and its interpretation in a future work. These preliminary results show the great potential of the techniques introduced in this paper. DSOMs show the variability of solar wind and how it can be visually characterized. The DSOM is a helpful guide in the study of the different types of solar wind, but is not necessarily an objective, unbiased and final classification method. In our current understanding, the main factor that determines classification results is the choice of the solar wind parameters used in the DSOM training. Choosing parameters that, according to previous studies and our physical understanding of the wind, can discriminate between specific wind types can guide the classification results. On the other hand, the possibility exists that



**FIGURE 11 |** Solar wind speed observed by ACE in two windows of time: during high solar activity (left column) and during low solar activity (right column). Each row corresponds to a different solar wind classification method. Vertical gray zones and lines correspond to entries in the Richardson and Cane, UNH, and CfA catalogs described in section 2.1.3.

an unsupervised classification methods, such as the one used here will highlight the presence of solar wind types that could warrant future physical investigation. DSOMs open the possibility for a fast visual characterization of large and complex data sets.

## 4. DISCUSSION

In this paper we show how the categorization of solar wind can be informed by classic unsupervised clustering methods and Self-Organizing Maps (SOM). We demonstrate that a single technique used in isolation can be misleading in the interpretation of automatic classifications. We show that it is important to examine the SOM lattices, in conjunction with solar wind composition and velocity distributions, and time series plots. Thanks to these tools we can differentiate classes associated with known heliospheric events.

We are convinced that basic unsupervised clustering techniques will have difficulties in finding characteristic solar wind classes when they are applied to unprocessed data. A

combination of feature engineering, non-linear transformations and SOM training leads to a more appropriate segmentation of the data points.

The classification of the solar wind also depends on the objectives that want to be attained: if the goal is to classify the solar wind to study its origin on the Sun, features related to solar activity must be included in the model; however, if the goal is to identify geoeffectiveness, other parameters should be added to the list of features, including geomagnetic indices.

In this work we have presented a first test of the capabilities of the SOMs for the analysis of data from a full solar cycle. Due to the extent of the work done, in this paper we introduce all the methods and techniques developed, but we leave for a future publication a more refined selection of all the model parameters, and the corresponding interpretation of the solar wind classification.

Finally, we advocate for the creation of a catalog of foot point locations for every solar mission, that connect solar wind observations to points on the solar surface. Due to the uncertainty on the exact foot point, such catalog should be composed of

a set of probabilities for each possible solar origin. This ground truth will vastly improve the efficacy of our classification models, which in turn can be used to reduce the initial uncertainties of the catalog.

All the tools and the techniques presented here can be applied to any other data set consisting of large amounts of points with a fixed number of properties. All the software and the data used in this work are freely available for reproduction and improvement of the results presented above.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/murci3lag0/swinsom>.

## AUTHOR CONTRIBUTIONS

JA created the software used in this work, built the data sets, and wrote the manuscript. RD provided the important insights into the use of the machine learning techniques, and performed

revisions of the different drafts. MEI gathered the information from external collaborators, provided insights into the data usage, and proofread the different drafts. GL supervised the work. All authors contributed to the manuscript revision, read, and approved the submitted version.

## FUNDING

The work presented in this manuscript has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 754304 (DEEP-EST, [www.deep-projects.eu](http://www.deep-projects.eu)), and from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, [www.aida-space.eu](http://www.aida-space.eu)).

## ACKNOWLEDGMENTS

The authors would like to acknowledge the helpful advice and suggestions by Olga Panasenco (UCLA), Raffaella D'Amicis (INAF-IAPS), and Aaron Roberts (NASA-GSFC). We thank the teams of the ACE SWEPAM/SWICS/MAG/EPAM instruments and the ACE Science Center for providing the ACE data.

## REFERENCES

- Adhikari, L., Zank, G. P., Zhao, L.-L., Kasper, J. C., Korreck, K. E., Stevens, M., et al. (2020). Turbulence transport modeling and first orbit parker solar probe (PSP) observations. *Astrophys. J. Suppl. Ser.* 246:38. doi: 10.3847/1538-4365/ab5852
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). “Optuna: a next-generation hyperparameter optimization framework,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Anchorage, AK), 2623–2631. doi: 10.1145/3292500.3330701
- Bale, S. D., Badman, S. T., Bonnell, J. W., Bowen, T. A., Burgess, D., Case, A. W., et al. (2019). Highly structured slow solar wind emerging from an equatorial coronal hole. *Nature* 576, 237–242. doi: 10.1038/s41586-019-1818-7
- Bandyopadhyay, R., Goldstein, M. L., Maruca, B. A., Matthaeus, W. H., Parashar, T. N., Ruffolo, D., et al. (2020). Enhanced energy transfer rate in solar wind turbulence observed near the sun from Parker solar probe. *Astrophys. J. Suppl. Ser.* 246:48. doi: 10.3847/1538-4365/ab5daa
- Bergstra, J., Yamins, D., and Cox, D. (2013). “Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures,” in *Proceedings of the 30th International Conference on Machine Learning Volume 28 of Proceedings of Machine Learning Research*, eds S. Dasgupta and D. McAllester (Atlanta, GA: PMLR), 115–123.
- Bishop, C. M. (2006). *Machine Learning and Pattern Recognition*. Information Science and Statistics. Heidelberg: Springer.
- Bloch, T., Watt, C., Owens, M., McInnes, L., and Macneil, A. R. (2020). Data-driven classification of coronal hole and streamer belt solar wind. *Sol. Phys.* 295:41. doi: 10.1007/s11207-020-01609-z
- Burlaga, L., Sittler, E., Mariani, F., and Schwenn, R. (1981). Magnetic loop behind an interplanetary shock: Voyager, Helios, and IMP 8 observations. *J. Geophys. Res. Space Phys.* 86, 6673–6684. doi: 10.1029/JA086iA08p06673
- Camporeale, E., Carè, A., and Borovsky, J. E. (2017). Classification of solar wind with machine learning. *J. Geophys. Res. Space Phys.* 122, 10910–10920. doi: 10.1002/2017JA024383
- Cane, H. V., and Richardson, I. G. (2003). Interplanetary coronal mass ejections in the near-Earth solar wind during 1996–2002. *J. Geophys. Res. Space Phys.* doi: 10.1029/2002JA009817. [Epub ahead of print].
- D'Amicis, R., and Bruno, R. (2015). On the origin of Highly Alfvénic slow solar wind. *Astrophys. J.* 805:84. doi: 10.1088/0004-637X/805/1/84
- Dodge, J., Gururangan, S., Card, D., Schwartz, R., and Smith, N. A. (2019). Show your work: improved reporting of experimental results. *arXiv [Preprint]* arXiv:1909.03004.
- Dupuis, R., Goldman, M. V., Newman, D. L., Amaya, J., and Lapenta, G. (2020). Characterizing magnetic reconnection regions using gaussian mixture models on particle velocity distributions. *Astrophys. J.* 889:22. doi: 10.3847/1538-4357/ab5524
- Eselevich, V. G., and Filippov, M. A. (1988). An investigation of the heliospheric current sheet (HCS) structure. *Planet. Space Sci.* 36, 105–115. doi: 10.1016/0032-0633(88)90046-3
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Portland, OR), Vol. 96, 226–231.
- Feldman, U., Landi, E., and Schwadron, N. A. (2005). On the sources of fast and slow solar wind. *J. Geophys. Res. Space Phys.* 110. doi: 10.1029/2004JA010918
- Fisk, L., and Lee, M. (1980). Shock acceleration of energetic particles in corotating interaction regions in the solar wind. *Astrophys. J.* 237, 620–626. doi: 10.1086/157907
- Fu, H., Li, B., Li, X., Huang, Z., Mou, C., Jiao, F., et al. (2015). Coronal sources and *in situ* properties of the solar winds sampled by ACE during 1999–2008. *Sol. Phys.* 290, 1399–1415. doi: 10.1007/s11207-015-0689-9
- Garrard, T. L., Davis, A. J., Hammond, J. S., and Sears, S. R. (1998). *The ACE Science Center BT-The Advanced Composition Explorer Mission*. Dordrecht: Springer Netherlands.
- Gloeckler, G., Cain, J., Ipavich, F. M., Tums, E. O., Bedini, P., Fisk, L. A., et al. (1998). *Investigation of the Composition of Solar and Interstellar Matter Using Solar Wind and Pickup Ion Measurements With SWICS and SWIMS on the Ace Spacecraft BT-The Advanced Composition Explorer Mission*. Dordrecht: Springer Netherlands.
- Gold, R. E., Krimigis, S. M., Hawkins, S. E., Haggerty, D. K., Lohr, D. A., Fiore, E., et al. (1998). *Electron, Proton, and Alpha Monitor on the Advanced*

- Composition Explorer Spacecraft BT-The Advanced Composition Explorer Mission.* Dordrecht: Springer Netherlands.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Networks *arXiv [Preprint]* arXiv:1406.2661.
- Gopalswamy, N., Hanaoka, Y., Kosugi, T., Lepping, R. P., Steinberg, J. T., Plunkett, S., et al. (1998). On the relationship between coronal mass ejections and magnetic clouds. *Geophys. Res. Lett.* 25, 2485–2488. doi: 10.1029/98GL50757
- Gosling, J. T., Teh, W. L., and Eriksson, S. (2010). A torsional Alfvén wave embedded within a small magnetic flux rope in the solar wind. *Astrophys. J. Lett.* 719, 36–40. doi: 10.1088/2041-8205/719/1/L36
- Habbal, S. R., Woo, R., Fineschi, S., O’Neal, R., Kohl, J., Noci, G., et al. (1997). Origins of the slow and the ubiquitous fast solar wind. *Astrophys. J.* 489, L103–L106. doi: 10.1086/310970
- Heidrich-Meisner, V., and Wimmer-Schweingruber, R. F. (2018). “Chapter 16—solar wind classification via k-means clustering algorithm,” in *Machine Learning Techniques for Space Weather*, eds E. Camporeale, S. Wing, and J. R. Johnson (Elsevier), 397–424.
- Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647
- Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Innocenti, M. E., Lapenta, G., Vršnak, B., Crespon, F., Skandrani, C., Temmer, M., et al. (2011). Improved forecasts of solar wind parameters using the Kalman filter. *Space Weather* 9. doi: 10.1029/2011SW000659
- Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization *arXiv [Preprint]* arXiv:1412.6980.
- Kingma, D. P., and Welling, M. (2013). Auto-Encoding Variational Bayes *arXiv [Preprint]* arXiv:1312.6114.
- Li, H., Wang, C., Tu, C., and Xu, F. (2020). Machine learning approach for solar wind categorization. *Earth Space Sci.* 7:e2019EA000997. doi: 10.1029/2019EA000997
- Liang, J., and Liu, R. (2015). “Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network,” in *2015 8th International Congress on Image and Signal Processing (CISP)* (Shenyang), 697–701. doi: 10.1109/CISP.2015.7407967
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 28, 129–137. doi: 10.1109/TIT.1982.1056489
- Magyar, N., Van Doorsselaere, T., and Goossens, M. (2019). The nature of Elsässer variables in compressible MHD. *Astrophys. J.* 873:56. doi: 10.3847/1538-4357/ab04a7
- Matteini, L., Hellinger, P., Landi, S., Trávníček, P. M., and Velli, M. (2012). Ion kinetics in the solar wind: coupling global expansion to local microphysics. *Space Sci. Rev.* 172, 373–396. doi: 10.1007/s11214-011-9774-z
- McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., et al. (1998). *Solar Wind Electron Proton Alpha Monitor (SWEPAM) for the Advanced Composition Explorer BT-The Advanced Composition Explorer Mission.* Dordrecht: Springer Netherlands.
- McKinney, W. (2010). “Data structures for statistical computing in Python,” in *Proceedings of the 9th Python in Science Conference*, eds S. van der Walt and J. Millman (Austin, TX), 56–61. doi: 10.25080/Majora-92bf1922-00a
- Neugebauer, M., Liewer, P. C., Smith, E. J., Skoug, R. M., and Zurbuchen, T. H. (2002). Sources of the solar wind at solar activity maximum. *J. Geophys. Res. Space Phys.* 107, SSH 13-1–SSH 13-15. doi: 10.1029/2001JA00306
- Neugebauer, M., and Snyder, C. W. (1966). Mariner 2 observations of the solar wind: 1. Average properties. *J. Geophys. Res.* 71, 4469–4484. doi: 10.1029/JZ071i019p04469
- Oiphant, T. E. (2015). *Guide to NumPy*, 2nd Edn. North Charleston, SC: CreateSpace Independent Publishing Platform.
- Owens, M. J. (2016). Do the legs of magnetic clouds contain twisted flux-rope magnetic fields? *Astrophys. J.* 818:197. doi: 10.3847/0004-637X/818/2/197
- Parker, E. N. (1958). Interaction of the solar wind with the geomagnetic field. *Phys. Fluids* 1, 171–187. doi: 10.1063/1.1724339
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). “PyTorch: an imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems* (Vancouver, BC), 8024–8035.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195
- Pierrard, V., and Lazar, M. (2010). Kappa distributions: theory and applications in space plasmas. *Sol. Phys.* 267, 153–174. doi: 10.1007/s11207-010-9640-2
- Rea, A., and Rea, W. (2016). How Many Components Should be Retained From a Multivariate Time Series PCA? *arXiv [Preprint]* arXiv:1610.03588
- Richardson, I. G. (2004). Energetic particles and corotating interaction regions in the solar wind. *Space Sci. Rev.* 111, 267–376. doi: 10.1023/B:SPAC.0000032689.52830.3e
- Richardson, I. G., and Cane, H. V. (2010). Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): catalog and summary of properties. *Sol. Phys.* 264, 189–237. doi: 10.1007/s11207-010-9568-6
- Richardson, I. G., and Cane, H. V. (2012). Near-earth solar wind flows and related geomagnetic activity during more than four solar cycles (1963–2011). *J. Space Weather Space Clim.* 2. doi: 10.1051/swsc/2012003
- Richardson, I. G., Cliver, E. W., and Cane, H. V. (2000). Sources of geomagnetic activity over the solar cycle: relative importance of coronal mass ejections, high-speed streams, and slow solar wind. *J. Geophys. Res. Space Phys.* 105, 18203–18213. doi: 10.1029/1999JA000400
- Roberts, D. A., Karimabadi, H., Sipes, T., Ko, Y.-K., and Lepri, S. (2020). Objectively determining states of the solar wind using machine learning. *Astrophys. J.* 889:153. doi: 10.3847/1538-4357/ab5a7a
- Rougier, N., and Boniface, Y. (2011). Dynamic self-organising map. *Neurocomputing* 74, 1840–1847. doi: 10.1016/j.neucom.2010.06.034
- Sabine, E. (1852). VIII. On periodical laws discoverable in the mean effects of the larger magnetic disturbance—No. II. *Philos. Trans. R. Soc. Lond.* 142, 103–124. doi: 10.1098/rstl.1852.0009
- Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). “Finding a ‘Kneedle’ in a Haystack: detecting knee points in system behavior,” in *2011 31st International Conference on Distributed Computing Systems Workshops* (Minneapolis, MN), 166–171. doi: 10.1109/ICDCSW.2011.20
- Schwenn, R. (1983). *The Average Solar Wind in the Inner Heliosphere: Structures and Slow Variations* (Washington, DC: NASA conference publication).
- Schwenn, R., and Marsch, E. (eds.). (1990). *Physics of the Inner Heliosphere I: Large-Scale Phenomena*. Vol. 20. Springer Science & Business Media.
- Shlens, J., View, M., and Introduction, I. (2014). *A Tutorial on Principal Component Analysis*.
- Smith, C. W., L’Heureux, J., Ness, N. F., Acuña, M. H., Burlaga, L. F., and Scheifele, J. (1998). *The Ace Magnetic Fields Experiment BT-The Advanced Composition Explorer Mission.* Dordrecht: Springer Netherlands.
- Smith, E. J. (2001). The heliospheric current sheet. *J. Geophys. Res. Space Phys.* 106, 15819–15831. doi: 10.1029/2000JA000120
- Stakhiv, M., Landi, E., Lepri, S. T., Oran, R., and Zurbuchen, T. H. (2015). On the origin of mid-latitude fast wind: challenging the two-state solar wind paradigm. *Astrophys. J.* 801:100. doi: 10.1088/0004-637X/801/2/100
- Stakhiv, M., Lepri, S. T., Landi, E., Tracy, P., and Zurbuchen, T. H. (2016). On solar wind origin and acceleration: measurements from ACE. *Astrophys. J.* 829:117. doi: 10.3847/0004-637X/829/2/117
- Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L. C., Guarnieri, F. L., Gopalswamy, N., Grande, M., et al. (2006). Corotating solar wind streams and recurrent geomagnetic activity: a review. *J. Geophys. Res. Space Phys.* 111. doi: 10.1029/2005JA011273
- Vettigli, G. (2013). *MiniSom: Minimalistic and Numpy Based Implementation of the Self Organizing Maps*.
- Villmann, T., and Claussen, J. C. (2006). Magnification control in self-organizing maps and neural gas. *Neural Comput.* 18, 446–469. doi: 10.1162/089976606775093918
- von Steiger, R., and Zurbuchen, T. H. (2015). Solar metallicity derived from *in situ* solar wind composition. *Astrophys. J.* 816:13. doi: 10.3847/0004-637X/816/1/13
- Wenzel, K. P., Marsden, R. G., Page, D. E., and Smith, E. J. (1992). The ULYSSES mission. *Astron. Astrophys. Suppl. Ser.* 92:207.
- Winterhalter, D., Smith, E. J., Burton, M. E., Murphy, N., and McComas, D. J. (1994). The heliospheric plasma sheet. *J. Geophys. Res. Space Phys.* 99, 6667–6680. doi: 10.1029/93JA03481

- Withbroe, G. L. (1986). *Origins of the Solar Wind in the Corona BT-The Sun and the Heliosphere in Three Dimensions*. Dordrecht: Springer Netherlands.
- Wu, C. C., Lepping, R. P., and Gopalswamy, N. (2006). Relationships among magnetic clouds, CMES, and geomagnetic storms. *Sol. Phys.* 239, 449–460. doi: 10.1007/s11207-006-0037-1
- Xu, F., and Borovsky, J. E. (2015). A new four-plasma categorization scheme for the solar wind. *J. Geophys. Res. Space Phys.* 120, 70–100. doi: 10.1002/2014JA020412
- Yin, H., and Allinson, N. M. (1995). On the distribution and convergence of feature space in self-organizing maps. *Neural Comput.* 7, 1178–1187. doi: 10.1162/neco.1995.7.6.1178
- Zhao, L., Landi, E., Lepri, S. T., Gilbert, J. A., Zurbuchen, T. H., Fisk, L. A., et al. (2017). On the relation between the *in situ* properties and the coronal sources of the solar wind. *Astrophys. J.* 846:135. doi: 10.3847/1538-4357/aa850c
- Zhao, L., Zurbuchen, T. H., and Fisk, L. A. (2009). Global distribution of the solar wind during solar cycle 23: ACE observations. *Geophys. Res. Lett.* 36, 1–4. doi: 10.1029/2009GL039181
- Zhao, L. L., Adhikari, L., Zank, G. P., Hu, Q., and Feng, X. S. (2018). Analytical investigation of turbulence quantities and cosmic ray mean free paths from 1995–2017. *J. Phys. Conf. Ser.* 1100:012029. doi: 10.1088/1742-6596/1100/1/012029

**Conflict of Interest:** 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.

Copyright © 2020 Amaya, Dupuis, Innocenti and Lapenta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Solar-Terrestrial Data Science: Prior Experience and Future Prospects

**Daniel N. Baker\***

Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, United States

Acquisition of relatively large data sets based on measurements in the interplanetary medium, throughout Earth's magnetosphere, and from ground-based platforms has been a hallmark of the heliophysics discipline for several decades. Early methods of time series analysis with such datasets revealed key causal physical relationships and led to successful forecast models of magnetospheric substorms and geomagnetic storms. Applying neural network methods and linear prediction filtering approaches provided tremendous insights into how solar wind-magnetosphere-ionosphere coupling worked under various forcing conditions. Some applications of neural net and related methods were viewed askance in earlier times because it was not obvious how to extract or infer the underlying physics of input-output relationships. Today, there are powerful new methods being developed in the data sciences that harken back to earlier successful specification and forecasting methods. This paper reviews briefly earlier work and looks at new prospects for heliophysics prediction methods.

## OPEN ACCESS

### Edited by:

Veronique A. Delouille,  
Royal Observatory of  
Belgium, Belgium

### Reviewed by:

Nickolay Ivchenko,

Royal Institute of Technology, Sweden

Victor Sergeev,

Saint Petersburg State  
University, Russia  
Georgios Balasis,  
National Observatory of  
Athens, Greece

### \*Correspondence:

Daniel N. Baker  
[daniel.baker@lasp.colorado.edu](mailto:daniel.baker@lasp.colorado.edu)

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 03 March 2020

**Accepted:** 17 August 2020

**Published:** 02 October 2020

### Citation:

Baker DN (2020) Solar-Terrestrial Data  
Science: Prior Experience and Future  
Prospects.

*Front. Astron. Space Sci.* 7:540133.  
doi: 10.3389/fspas.2020.540133

## INTRODUCTION

Modern information and communication technologies have created an interoperable era in which access to data can be essentially universal. Open access to these data and related services enables the research and applications community to meet new challenges of understanding Earth and its space environment (Baker, 2008). Given the complex system, understanding the Earth and near-Earth space requires managing and accessing large data sets. It also requires acquisition of progressively higher spatial and temporal resolution measurements. In the modern era, there is particular need for very rapid (near real-time) response modalities. In many instances, acquired data must be assimilated into empirical or physics-based models. Such work often requires crossing of disciplinary boundaries in order to achieve the ultimate research and applications goals.

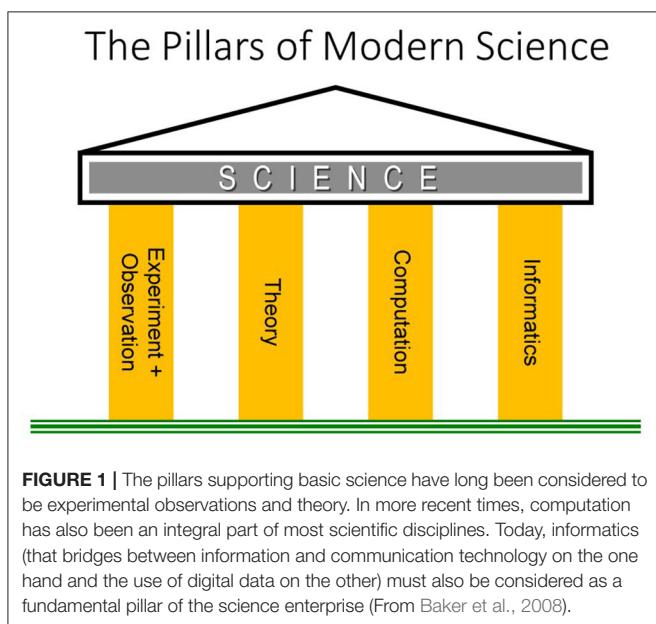
Traditionally, science has been viewed as being based on two fundamental pillars: Experimentation (i.e., observations) and theory. In the past century or so, computational methods and related modeling have been added as an essential pillar of basic science. In fact, most domains of science today could simply not be viable without the availability of computational modeling. In recent times, it has been further recognized that “informatics”—the science of processing data for storage and retrieval—is an indispensable fourth pillar of modern science (Baker et al., 2008). Thus, for the present-day purposes of understanding and predicting the behavior of the coupled Sun-Earth system, information science is as much a key foundation as are the more traditional elements of science (see **Figure 1**).

In this brief review, the goal is to use the coupled solar wind-magnetosphere-ionosphere system to illustrate the application and utilization of various data analysis methods. Many topics to be discussed were based on time series analysis and prediction filter methods. This work led

naturally to the understanding that the magnetosphere-ionosphere system often evolves toward highly non-linear states that must be examined carefully. Methods employed decades ago have now become quite relevant again in the present era of machine learning. Appreciation of the roots of solar-terrestrial data science are important looking toward the future.

## THE DRIVEN MAGNETOSPHERE-IONOSPHERE SYSTEM

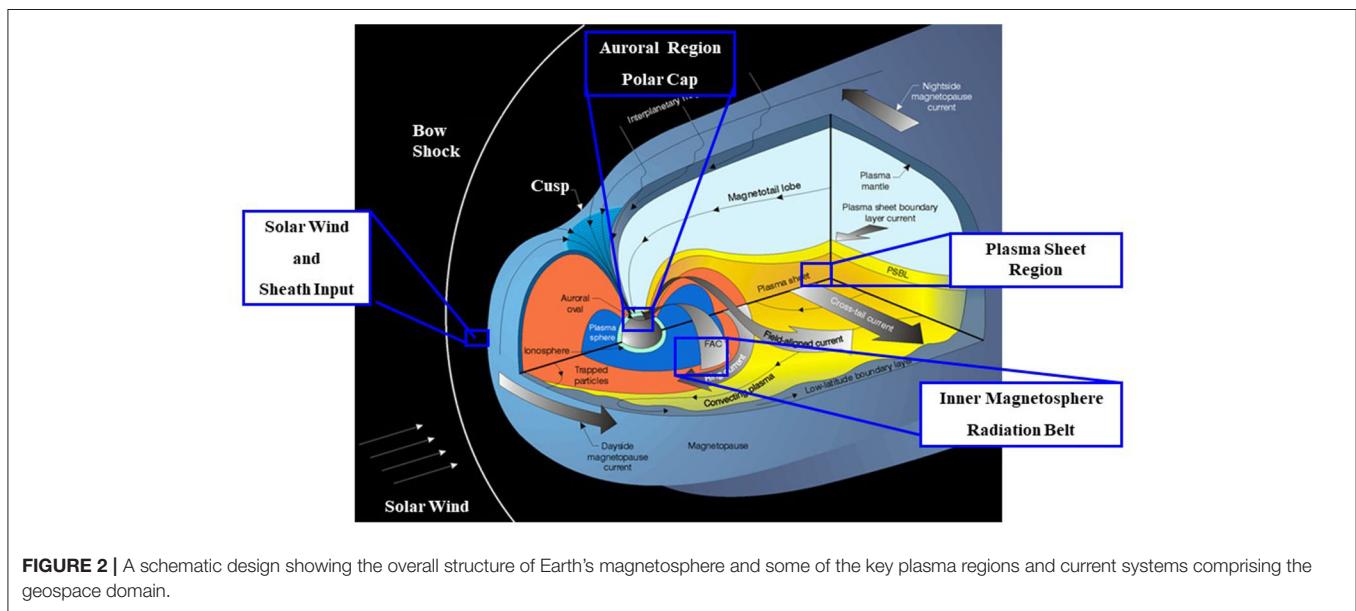
**Figure 2** is a schematic diagram of the Earth's magnetosphere and some of its key plasma physical regions. The diagram

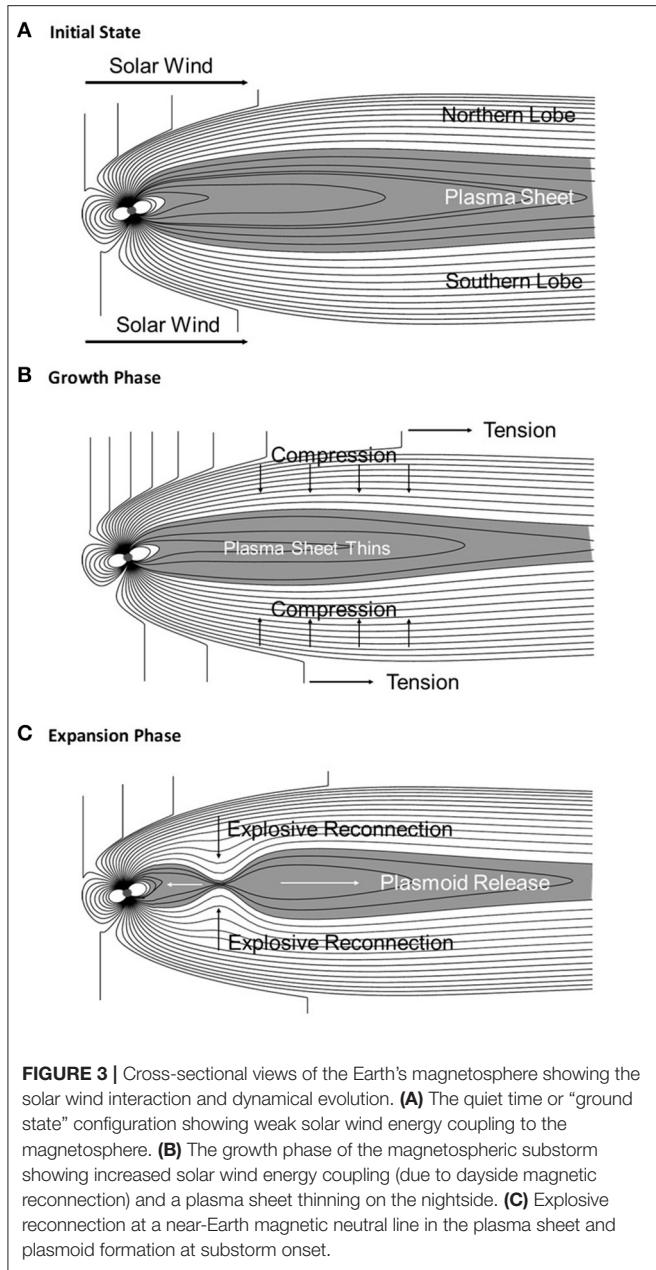


also illustrates the external solar wind flow and many of the large-scale current systems that shape the overall magnetospheric cavity. From this picture, we see that the Earth's magnetosphere-ionosphere system is a large, highly-coupled plasma domain. In its ground state, the magnetosphere is characterized by a relatively stable configuration standing off the solar wind at  $\sim 12 R_E$  (Earth radii) at its sub-solar point. It extends into an elongated magnetotail on the nightside. The basic steady-state relationship with the interplanetary magnetic field (denoted  $B_{IMF}$ ) during quiet times gives rise to a coherent global structure with persistent boundary features.

As the solar wind impinging on the magnetosphere changes velocity or density and, notably, when the IMF changes magnitude or direction, the magnetosphere undergoes substantial evolution. The size of the magnetospheric cavity changes, strong ionospheric currents are set up, and large configurational changes can be seen throughout the magnetotail. The elemental and repeatable sequence of events that occurs starts with a southward turning of the interplanetary magnetic field. This causes enhanced coupling from the solar wind into the magnetosphere.

**Figure 3** shows the commonly accepted picture of the resulting magnetospheric substorm sequence (Baker et al., 1993, 1996). **Figure 3A** shows a cross-sectional view of the magnetotail and illustrates its relaxed or quiescent state. The plasma sheet is rather thick and there is little excess “free energy” in the system. **Figure 3B** shows the magnetotail as the interplanetary field has turned southward and a large amount of magnetic energy has been added to the tail. This growth phase is characterized by considerable excess stored energy in the magnetotail lobes and a thin, stressed plasma sheet. Finally, in **Figure 3C** we show the explosive onset of near-Earth magnetic reconnection (x-line formation) which gives rise to a plasmoid that pinches off and leaves the system. This entire sequence constitutes a magnetic “loading-unloading” system. As such, it can be modeled using

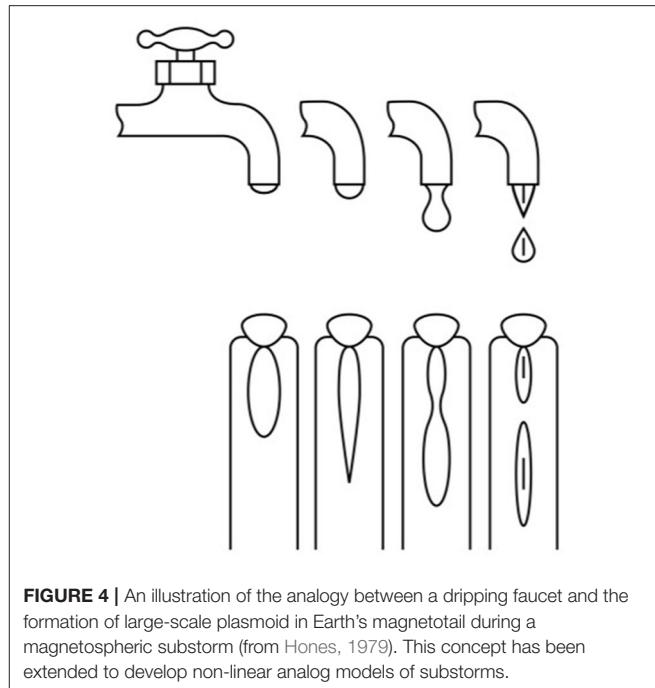




non-linear dynamic methods (Sharma et al., 2005) in analogy with concepts developed in statistical physics and in branches of applied mathematics.

The magnetospheric energy unloading process early on was considered to be analogous to a dripping faucet (Hones, 1979) (see Figure 4): The plasma sheet in the magnetotail distends as part of the substorm growth phase. Then a portion of the plasma sheet pinches off to form the separated plasmoid at substorm onset. The plasmoid moves tailward and leaves the magnetotail and the Earthward part of the plasma sheet snaps back toward the Earth, in analogy with a dripping faucet.

Figure 5 illustrates conceptually the steps of the solar wind-magnetosphere coupling process. Basically, the coupling effects are controlled by the dawn-to-dusk component of the



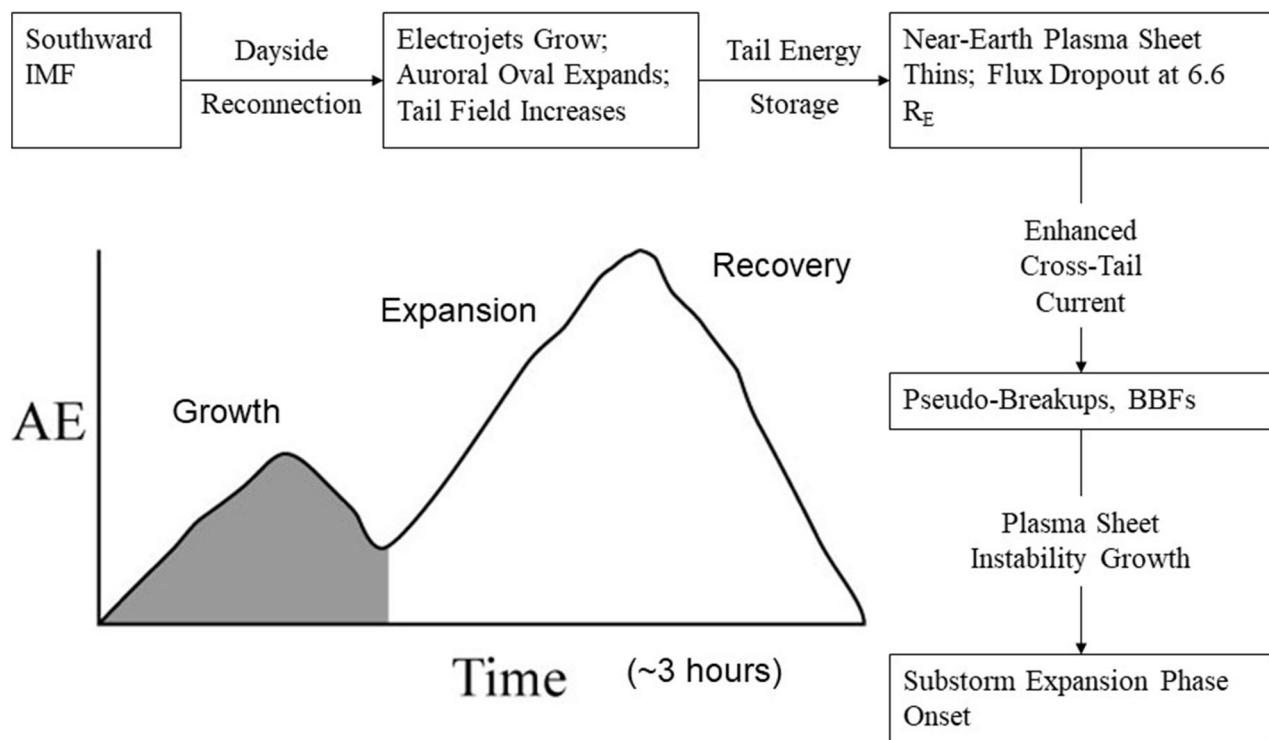
**FIGURE 4 |** An illustration of the analogy between a dripping faucet and the formation of large-scale plasmoid in Earth's magnetotail during a magnetospheric substorm (from Hones, 1979). This concept has been extended to develop non-linear analog models of substorms.

interplanetary electric field  $E_{SW}$  ( $= -V_{SW} \times B_{IMF}$ ) with  $V_{SW}$  the solar wind speed. Therefore, taking into account this “rectifier” effect, many studies have shown that the main parameter driving substorms is the solar wind speed ( $V_{SW}$ ) multiplied by the southward IMF component ( $B_s$ ) ( $B_s = -B_{IMF}$  for  $B_{IMF} < 0$ , and  $B_s = 0$  for  $B_{IMF} > 0$ ). Changing of  $V_{BS}$  ( $= V_{SW} B_s$ ) causes variations in the dayside magnetic merging rate. As shown by Figure 5, this change in dayside reconnection directly drives ionospheric currents which show up in the auroral electrojet (AE) indices. The westward electrojet (AL) index is particularly indicative of the coupling process (Baker et al., 1996).

As energy is transferred to the magnetosphere, merged flux is transported from the dayside to the nightside. This leads to enhanced magnetotail flux and enhanced convection from the distant x-line toward the Earth. The addition of tail flux is indicative of an intensification of the cross-tail electrical currents flowing in the plasma sheet. After about 1 h of tail flux loading, (i.e., a 1-h substorm growth phase) there normally is a substorm expansion phase onset. In the near-Earth neutral line (NENL) model of substorms (Baker et al., 1996), the expansion phase onset is produced by the sudden appearance of a new x-line which causes strong magnetic reconnection in the relatively near-Earth portion of the plasma sheet. The reconnection process and the accompanying cross-tail current disruption drives further flow of currents through the nightside ionosphere to form the “unloading” currents. These energy-dissipation processes show up prominently in the AE index.

It is worth noting that there can be many variants of substorm-like activity in the magnetosphere. There can be so-called steady convection events and other energy dissipation events that may not clearly exhibit all the “phases” noted above (see Shukhtina et al., 2014). Nonetheless, the basic pattern of energy

# Magnetospheric Dynamical Sequence



**FIGURE 5 |** A notational sketch of an isolated substorm as seen in a time series of auroral electrojet (AE) values vs. time. The boxes in the flow chart around this schematic describe the key, repeatable physical phenomena occurring during substorms. BBFs are bursty bulk flows of plasma. These are localized, high-speed particle transport regions.

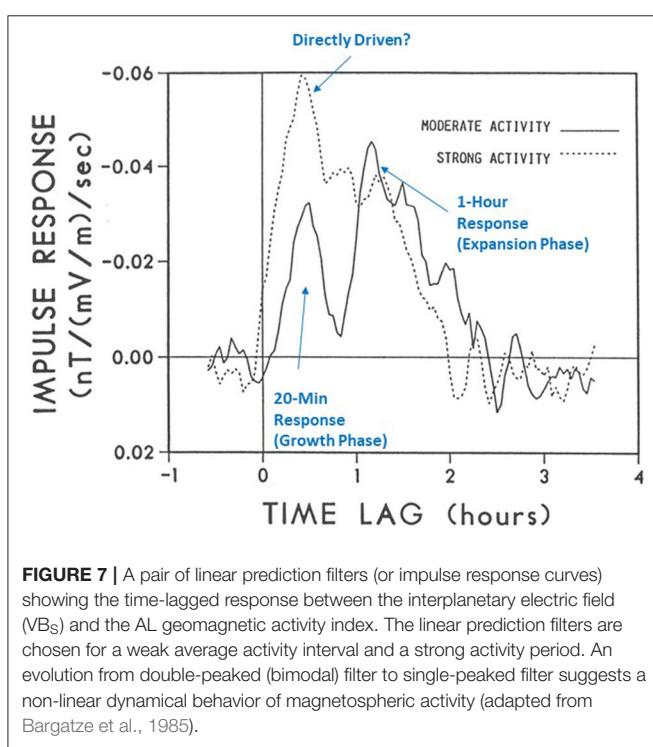
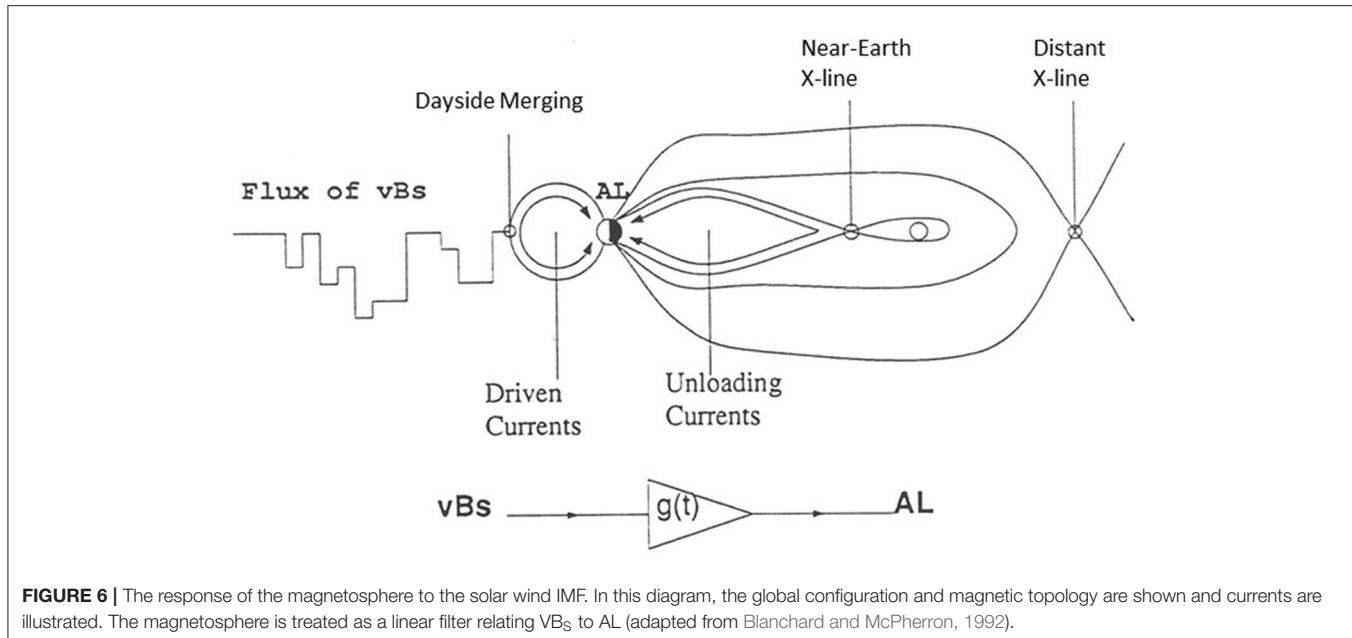
loading, rapid energy dissipation, and relaxation back toward a ground state seems clearly established in many cases. Obviously, there is great value in examining events that depart from the classic pattern.

The energy flow from the solar wind through the magnetosphere into the ionosphere has been addressed in terms of a global-scale convection process. This has been modeled in terms of linear filter relationship between VB<sub>S</sub> as an input time series and AL as a magnetospheric output time series (Blanchard and McPherron, 1992). As shown here in Figure 6, the transfer (or filter) function g(t) can be viewed as a general linear relationship between the solar wind driver and the magnetospheric response (Bargatze et al., 1985). When many different intervals of geomagnetic disturbance were considered, Bargatze et al. (1985) found that periods of weak or moderate disturbances showed linear response filters with two peaks (Figure 7). These were interpreted as the directly-driven response (at 20-min lag time) and the unloading response (at 60-min lag). During high geomagnetic activity periods, the 60-min response peak was seen to disappear or even to merge into the 20-min peak. This was interpreted by Baker et al. (1990a) as evidence of a non-linear evolution of geomagnetic activity as the system moved from weak to strong disturbance levels.

## LOW-DIMENSIONAL ANALOG MODELS

Taking into account observed evolution of linear prediction filters with increasing geomagnetic activity (Bargatze et al., 1985), Baker et al. (1990a) developed a mechanical analog model of substorm dynamics. As shown in Figure 8A, the model considered a mass on a spring. The mass increases at a fixed loading rate,  $dm_L/dt$ , until a critical distention  $D = D_C$  is reached. Then a portion of the mass is released from the spring at a rate ( $dm_u/dt$ ) that is governed by the velocity of the mass ( $dD/dt = p/m$ ) at the critical displacement point. In this formulation,  $p$  represents the momentum of the weight on the spring. The movement of this mass on the spring is described by  $dp/dt = Gm - \kappa D - \eta p/m$ . The spring constant,  $\kappa$  and the frictional coefficient,  $\eta$ , are assumed fixed. The set of equations for  $dD/dt$  and  $dp/dt$ , as closed by the conditions on the  $dm/dt$ , give rise to a set of non-linear equations. When  $dm_L/dt$  increases toward large volumes, the system moves from weak periodic unloading to highly chaotic, non-linear behavior. In essence, the addition of mass and its subsequent unloading gives rise to complex interactions as the loading rate changes.

The mechanical analog model of Baker et al. (1990a) was inspired by a similar model of drippy faucets developed by Shaw



(1984). Baker et al. took rather literally the analogy between substorm unloading processes and a dripping faucet (Hones, 1979). As noted, the plasma sheet distends considerably and in the substorm growth phase. During the near-Earth reconnection onset, a part of the plasma sheet pinches off to form the substorm plasmoid. The remaining plasma sheet snaps back sunward. The upper part of Figure 4 illustrates that the dripping faucet behaves analogously to the Earth's plasma sheet during substorms.

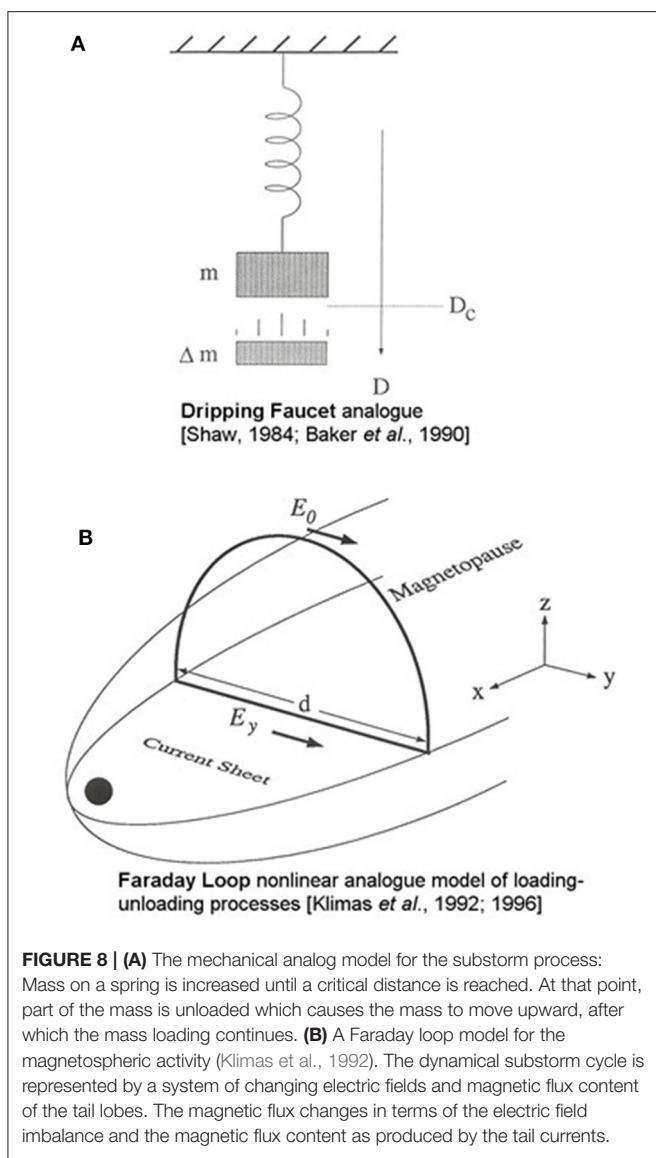
The mass-on-spring analog model of Baker et al. (1990a) was capable of reproducing aspects of magnetospheric substorms;

the sequence shown above in Figure 3 were reproduced rather well. Moreover, when a loading rate increased, the analog system exhibited non-linear behavior: period doubling and “bifurcation” resulted such that fully developed chaotic behavior ensued. Given the elementary,  $2\frac{1}{2}$  dimensional nature of the model, this was a quite successful simulation. Nonetheless, the mechanical analogy was limited. As a follow-on approach, Klimas et al. (1992) developed the Faraday loop model (Figure 8B) which was based more on a plasma analog.

The Faraday loop approach was a dynamical convection model with three degrees of freedom: (1) The average cross-tail electric field in the near-Earth current sheet,  $E_y$ ; (2) The variation of  $E_y$  due to solar wind input, noted  $E_0$  on Figure 8B, and  $E_{sw}$  earlier in the text, to the tail near the X-line position; and (3) variable magnetic flux content of magnetotail. The Faraday loop model was very much like the drippy faucet model. Flux was loaded into the tail at a rate determined by the (solar wind) input,  $E_0$ . In the model, the tail unloads by forming a plasmoid, thereby dumping much of the tail flux in the tail (as in Figure 3 above).

One of the most important findings of the early linear prediction filter analysis was the “bimodal” response behavior (see Figure 7). The 20-min driven response time scale and the 60-min unloading time scale emerged as clear features during moderate activity periods. Using the Faraday loop model, we were able to use the observed  $vB_S$  values measured upstream of the Earth to drive the Faraday loop model with realistic inputs. This allowed examination of analog model outputs as was done by Bargatze et al. (1985) using linear filter analysis. Of course, it was also possible to compute the linear filter elements in the Faraday Loop model as for the real magnetospheric response (as measured by the AL index).

Important studies in this early period went further to examine underlying dynamics as revealed within indicial time series data. Consolini et al. (1996) used the auroral electrojet indices to infer multifractal aspects of the dynamical system behavior. These authors examined the inferred turbulent character involved



**FIGURE 8 | (A)** The mechanical analog model for the substorm process: Mass on a spring is increased until a critical distance is reached. At that point, part of the mass is unloaded which causes the mass to move upward, after which the mass loading continues. **(B)** A Faraday loop model for the magnetospheric activity (Klimas *et al.*, 1992). The dynamical substorm cycle is represented by a system of changing electric fields and magnetic flux content of the tail lobes. The magnetic flux changes in terms of the electric field imbalance and the magnetic flux content as produced by the tail currents.

with the magnetosphere-ionosphere dynamics. Subsequently in important studies of the nature of the Earth's plasma sheet behavior during active times, Angelopoulos *et al.* (1999) found evidence of intermittency in the plasma flows. This work was extended by Vörös *et al.* (2003) who confirmed such intermittent multiscale behavior of the magnetic field in the near-Earth plasma sheet (see section Multi-Scale Aspects).

Follow-on work (e.g., Freeman and Morley, 2004, 2009) has sought to assemble "minimal" substorm models that contain the key aspects of substorm dynamical processes and timing properties. While substorms in the real world can have many complex variations, there is a basic, repeatable underlying pattern of substorms that is usually present. While early studies tended to focus on the auroral electrojet index time series, subsequent studies also examined other available indices. Wanliss (2005) found fractal behavior in the stormtime (SYM-H) index time series and this fractal nature of the dynamics was

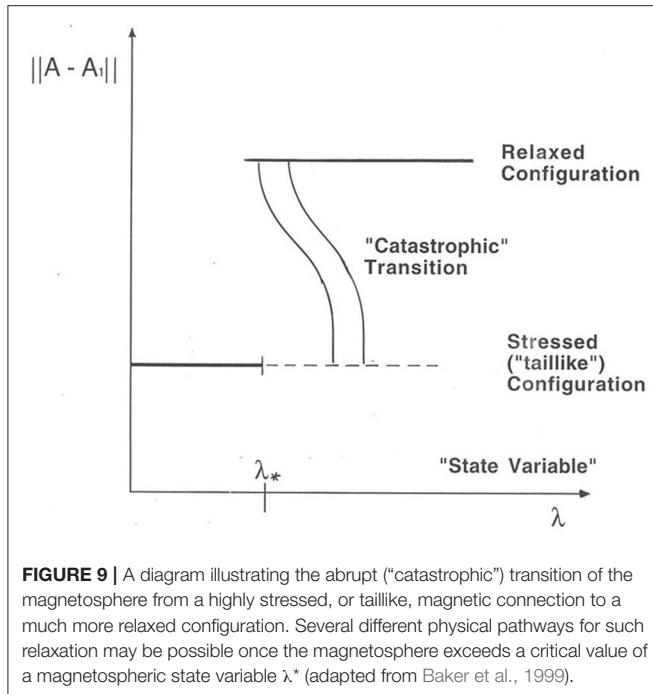
amplified upon in a broad study of magnetic storm development (Balasis *et al.*, 2006).

## MULTI-SCALE ASPECTS

The traditional methods of studying the solar wind interaction with Earth's magnetosphere have been to take a magnetohydrodynamic (MHD) approach. This framework was described in the foregoing sections of this review. The MHD method has had many successes and has helped, for example, to illuminate much about the fundamental substorm dynamical cycle. However, MHD models use different computational techniques and approaches which can lead to different predicted dynamical (substorm) behavior (see Gordeev *et al.*, 2016). Moreover, as we have also detailed in this paper, there are complex and non-linear aspects of solar wind-magnetosphere coupling that are not always adequately captured by the straightforward MHD models and simulations (Baker *et al.*, 1999 and references therein). This has led various authors to take different tacks and these alternatives have provided valuable and novel understanding of the coupling issues. For example, Chapman *et al.* (1998) and Chang (1999) employed analog models to characterize the most general dynamical properties of the terrestrial plasma sheet and its coupling in the ionosphere. These ideas are based on the slow buildup of a system until it reaches an unstable state at which point an "avalanche" occurs and the system collapses. Chapman *et al.* (1998) gained valuable insight into the energy buildup and release in the Earth's magnetotail using such a "sand pile" approach. The basic advances from this kind of work have been the realization that the growth phase of substorms is an essential element of the substorm cycle. Either a spontaneous relaxation of the system must eventually occur after sufficient loading or else even the slightest perturbation of the system can trigger a massive energy release.

Another approach considered for the magnetospheric system was predicated on a variant of "catastrophe theory." This was based on the idea of a system that has evolved to a highly stressed state. Then, after further forcing, the system undergoes a catastrophic transition to a much more relaxed state. As shown here in **Figure 9** from Baker *et al.* (1999), the magnetosphere can be thought of late in the substorm growth phase as being primed for catastrophic collapse. Then by means of one or more plasma physical processes, the magnetotail can collapse into a relaxed configuration. This kind of reasoning was also explored by Lewis (1991) in an attempt to help resolve some of the longstanding debates about the exact mechanisms and physical instabilities that lead to substorm onsets. Baker *et al.* (1999) noted that it might be a variety of different instabilities that could play important triggering roles in the state transition, depending on the circumstances.

Based upon decades of *in situ* and remote sensing observations of the global magnetospheric system and based upon the revealed complexity of the underlying magnetospheric dynamics, one can say that many different approaches have been necessary to disentangle the various facets of geomagnetic activity. But

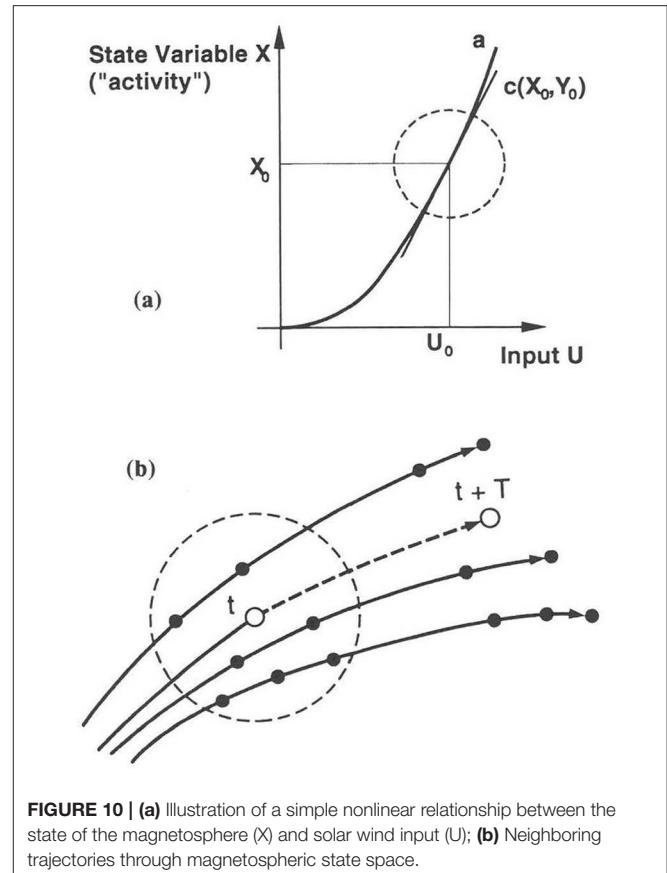


quite clearly, all the evidence suggests that the solar wind-magnetosphere-ionosphere interactions have strongly non-linear aspects to them. In light of this recognition, space plasma physics, as applied to magnetospheric dynamics, must move away from traditional plasma stability analyses and must utilize techniques that incorporate global, non-linear interactions. In trying to embrace such approaches, ideas from other branches of physics, engineering, applied mathematics, etc., can provide valid and extremely useful insights.

Adopting a more generalized point of view has certainly driven home the point that multiscale phenomena play a crucial role in magnetospheric dynamics. For example, global auroral images have been examined on a wide range of temporal and spatial scales (Lui, 2002; Uritsky et al., 2002). The results of such analysis point clearly toward essentially scale-free distributions over a wide spectrum of system conditions. This may further imply that the magnetotail can undergo strong dissipation events in part due to local plasma instabilities and in part due to control elements remote from the central plasma sheet. These elements may largely lie in the incident solar wind driver or they may reside in the distant ionosphere. Under such circumstances, it may appear that regional plasma instabilities are spontaneous or are even occurring in an essentially random fashion (Lewis, 1991). Thus, a lesson from many prior, diverse types of studies is that non-traditional analysis and modeling can reveal important aspects of our terrestrial space environment.

## STATE SPACE ANALYSIS AND PHASE SPACE RECONSTRUCTION

Phase space reconstruction and analog model development—as discussed here—give considerable insight into all aspects



of magnetospheric dynamics. The evidence has suggested that substorms are the manifestation of a fundamental dynamical cycle. But this work also has shown that the magnetosphere is never really linear in its interaction with the solar wind since the nature of the substorm response varies as the general level of the activity increases.

In light of such analysis, one can speak in terms of a certain "state" of the magnetospheric system (see Vassiliadis et al., 1995). This state can be characterized by the level of geomagnetic activity (as might be measured by AE, AL, K<sub>p</sub>, or other global indices). One can identify a state variable,  $X(t)$ , which characterizes the global magnetospheric condition, but which also takes cognizance of the recent past history of the solar wind input [ $U(t)$ ] and the resultant magnetospheric response. **Figure 10a** illustrates a simple non-linear (i.e., quadratic) relationship between  $U(t)$  and  $X(t)$ . At a particular point  $(X_0, U_0)$  one approximates the relationship by a "local linear" filter (LLF). In general, the LLF varies depending on the input level.

Pursuing the notion of a magnetospheric state, Vassiliadis et al. (1995) asserted that the system evolved according to  $dX/dt = F(X; U)$ . It assumed that an index such as AL contains sufficient information to characterize the whole magnetospheric system. The basis for this assumption has been examined in earlier studies and analogous questions have been addressed for many complex systems (Gleick, 1987). It seems clear that global geomagnetic

indices embed within them the “shadow” of all the key dynamical processes. Proceeding on this presumption, the recent history of input [ $U(t)$ ] and output [ $X(t)$ ] specifies a clear trajectory in state space. This concept is illustrated in **Figure 10b** where the open dot at time,  $t$ , is followed by another point in state space at time  $= t + T$ . Because of the repeatable cycle in substorm dynamics, there is a reasonable possibility of extrapolating the dynamical evolution of the magnetosphere into the future. By examining the “nearest neighbors” of a given point ( $X_0, U_0$ ) in state space, and using the average state space trajectories represented by the nearest neighbor points, one can predict the future evolution of the solar wind-magnetosphere system (Vassiliadis et al., 1995).

The method of state space construction proved to be a powerful way of predicting future geomagnetic activity. As described above, one can use a large data base of previous input-output (e.g.,  $VB_s$ -AL) relationships to construct a multidimensional state space. One can then consider the recent values of  $VB_s$  and AL leading up to this point. For example, consider the previous 15 min of AL behavior. Then looking back in the historical data base, find all previous examples of  $VB_s$ -AL behavior that closely parallel the one in question. By averaging together these “nearest neighbor” trajectories in state space one can predict with some assurance the future evolution of the ongoing geomagnetic activity pattern. In many ways this is similar to predicting terrestrial weather based upon previous similar patterns of season, temperature, pressure, humidity, etc.

All of this work may be termed “data-mining” and it has shown significant successes over many years. There have been crucially important models of the terrestrial magnetic field built on such data mining efforts (Tsyganenko and Sitnov, 2007; Sitnov et al., 2008). More recently, using this kind of global magnetic field information has allowed extraction of magnetospheric substorm growth phase and expansion phase patterns (Stephens et al., 2019). Going beyond simple index data to global *in situ* data exploitation may hold even greater modeling promise.

## NON-LINEAR DYNAMICS AND COMPLEXITY

As discussed above, linear prediction filters can be convolved with input time-series in order to approximate the output of a specified system process. This was the approach taken for substorm onset forecasting (Bargatze et al., 1985; Blanchard and McPherron, 1992) and the method met with considerable empirical success. However, the approach was also criticized because it did not reveal in great detail the underlying physical processes producing substorm onsets.

Many other magnetospheric properties have been modeled using linear filter methods. For example, Nagai (1988) applied this approach to the problem of specifying and forecast energetic electron fluxes at the geostationary orbit using  $K_p$  geomagnetic indices as the driving input. Baker et al. (1990b) applied linear prediction filter methods to the same problem but using solar wind speed data as the driver input. This latter approach proved to be very useful and powerful. Even today—three decades later—the Space Weather Prediction Center of the

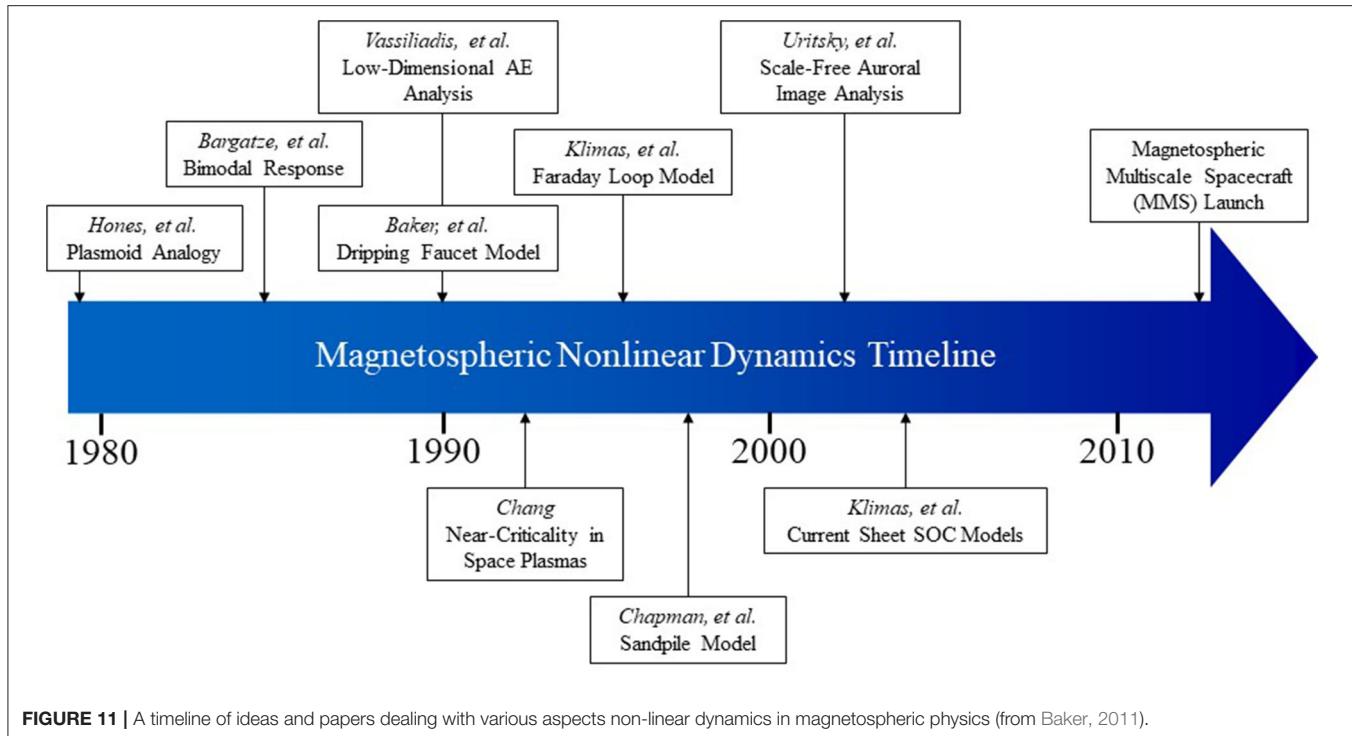
National Oceanic and Atmospheric Administration (NOAA) uses the Baker et al. (1990b) method as the basis of its Relativistic Electron Forecast Model (REFM). This model provides 1–8 day predictions of omni-directional electron fluxes at geostationary orbit using real-time solar wind data from NASA spacecraft at L1.

While linear analyses can work well in such system applications, there is ample evidence that non-linear behavior is exhibited in several ways by the Earth’s magnetosphere. This was well-documented by the Bargatze et al. (1985) work concerning magnetospheric substorms described above. Much work over the past several decades has further demonstrated that geomagnetic activity can exhibit output that is not proportional to input. There are clear examples in the ionosphere and the magnetospheric proper where there are feedbacks (that is, output influences input to some greater or lesser degree). The near-Earth system thus can exhibit what seems to be random behavior and there can be immense sensitivity to initial conditions. These are all properties of non-linear dynamics (Gleick, 1987). This has also been termed low-dimensional behavior or “deterministic chaos.” Such non-linearity can lead to self-organization in which global patterns emerge from local interactions among many subunits (see Pepper and Hoelzer, 2001). In these cases—as in the magnetosphere-ionosphere system—the interactions are often shaped by multiple feedback loops.

As described by Klimas et al. (2000), the magnetotail and its embedded plasma sheet (see **Figure 2**) exhibit many of the properties just discussed. Klimas et al. noted that the magnetotail is a spatially distributed loading-unloading system. The hypothesis—well-supported by observations—is that magnetic flux is the relevant conserved quantity. Klimas et al. also demonstrated that the magnetotail is often near a threshold instability that can produce localized magnetic reconnection. Thus, the system exhibits self-organized criticality (SOC) with a level of global coherence in a broadly distributed spatial region that is near instability (see, also, Sitnov et al., 2001). Often in the magnetotail system, localized reconnection, “pseudo-breakups,” and full-fledged magnetospheric substorms produce a rather “scale-free” cascade (avalanche) of geomagnetic disturbances (Klimas et al., 2000).

A persistent question in system analysis is how much information content may be inherently contained in a given parameter sequence or activity time series. Measures of ‘entropy’ are often used to assess such information content and many studies have suggested that the magnetosphere exhibits clear evolution from more complex to a more orderly state as the near-Earth space moves from pre-storm to full geomagnetic storm conditions (Balasis et al., 2008, 2009). More recent work building on such entropy assessments have examined the information transfer via solar wind forcing leading to radiation belt flux enhancements (Wing et al., 2016). Entropy analyses have also helped understand more deeply the substorm-storm relationship that lies at the heart of the magnetospheric dynamical pattern (Runge et al., 2018).

As shown schematically in **Figure 11** (from Baker, 2011), there has been quite a progression of thought about non-linear magnetospheric dynamics over the past several decades. From



**FIGURE 11 |** A timeline of ideas and papers dealing with various aspects non-linear dynamics in magnetospheric physics (from Baker, 2011).

the rudimentary analogy of the dripping faucet by Hones (1979) through the several models and approaches of the 1990s and early 2000s, many methods of data analysis have shown that the magnetosphere-ionosphere system is a complex, non-linear domain. With new large data sets from observing platforms such as the Magnetospheric Multiscale (MMS) mission (Burch et al., 2015), new opportunities now exist to examine questions that were not accessible in earlier times.

## NEURAL NETWORKS, MACHINE LEARNING, AND FUTURE PROSPECTS

In the 1990s and into the 2000s, several authors and research groups began to employ artificial neural network (ANN) methods to study such things as solar wind driving of radiation belt particle fluxes (e.g., Stringer and McPherron, 1993) and geomagnetic activity (e.g., Gleisner and Lundstedt, 1997; Takalo and Timonen, 1997). These multi-layer feed-forward neural nets were able to provide—in many cases—impressive specification and forecasts of important space weather indicators such as the AE/AL indices or geostationary-orbit electron fluxes. Somewhat later, related ANN methods were able to specify and predict up to 1-h ahead the ionosphere total electron content (TEC) (see, e.g., Tulinay et al., 2006). While these approaches often gave remarkably good forecasts of geophysical conditions, there were also criticisms that the methods did not reveal deep insight into the underlying physics that produced specific responses.

In today's era of machine learning (ML), there seems to be less concern about always needing to understand precisely *why* a given method works. Rather there is more concern about

how well a ML tool might be able to do. Recent studies have used vast new spacecraft data sets for a wide range of purposes such as assessing magnetospheric field models (Yu et al., 2014) and identifying radiation belt pitch angle distribution patterns (Souza et al., 2016). In these instances, and many more that could be cited, neural network tools permit data examination and classification in rather automatic ways.

Machine learning has reached even more full flower in just the last few years. Combining physical models with machine learning techniques has provided the capability to analyze and understand relationships between cold plasma properties, local wave characteristics, and relativistic electron flux distributions throughout the inner magnetosphere (Chu et al., 2017; Bortnik et al., 2018). Particularly notable successes have been achieved using non-linear autoregressive moving average with exogenous inputs (NARMAX) methods to predict energetic particle events (e.g., Boynton et al., 2011, 2013). Both acceleration and loss processes for relativistic electrons, including the amount of "memory" in the system, have been more deeply understood through the use of ML techniques.

Looking to the future, it is obvious that both space missions and ground-based systems will continue to increase in their capabilities. This means even higher volumes of solar, interplanetary, and geospace data. "Non-linear" approaches (spectral methods, filter techniques, ANNs, etc.) will also probably keep increasing in complexity and in their "data-mining" capabilities. Many useful applications of these capabilities will probably result. They likely will include:

- Data conditioning (gap filling, noise reduction, data smoothing, etc.);

- Empirical model development and coupling to other models (either first-principles or other empirical models);
- Data assimilation (related to non-linear dynamics, control theory, radiation belt behavior, thermospheric properties, GPS performance, etc.); and
- Prediction algorithms.

## SUMMARY AND CONCLUSIONS

Advances in computer technology in the 1980s allowed scientists and mathematicians to solve numerically many vexing problems posed decades earlier. This meant that fields of fractal geometry, non-linear dynamics, chaos theory, wavelet (and other) transform methods, neural networks, etc., could be formulated as extensions (or juxtapositions) to traditional linear and spectral approaches. Hence, concepts such as the “state” of a plasma, fractals, chaos (as opposed to turbulence), cellular automata, and several other ideas trace back to that time. From these ideas, especially in the 1990s, sprang a plethora of methods for data analysis. In this brief review, we have touched on some examples in geomagnetic field studies, radiation belt behavior, ionospheric and thermospheric changes, and solar wind-magnetosphere coupling, among others. These approaches often gave rise to improved methods of numerical simulations of the magnetospheric domain.

While some of the methods explored in earlier decades found only limited use and therefore did not develop further into the 2000s, others revealed great utility. This often sprang from data analysis methods developed in earlier times that now have found applications in processing large volumes of data returned by spacecraft missions and ground-based observation platforms. In this machine learning era, this is an extremely important outcome. These earlier methods now are aiding in producing new (and effective) empirical and physical models. These results hold

great promise for predicting the state of geoplasma systems going forward. Thus, space situational awareness and space weather applications seem to be on the threshold for a new and highly productive phase.

It is perhaps worth noting that much of the early work described in this brief review was built upon analysis and modeling of geomagnetic index information. With years of *in situ* measurements throughout the magnetosphere-ionosphere system, we now have plasma, energetic particle, and magnetic field data sets that could be organized and used to examine dynamical properties of the solar-terrestrial system that were not accessible readily based only on ground indices such as AE or Dst. It would be quite valuable (in the author’s opinion) to use different long-term descriptors of the magnetospheric system and repeat some of the kinds of studies that have been reviewed here.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## FUNDING

This work was supported by Director’s Discretionary funding from the University of Colorado Laboratory for Atmospheric and Space Physics.

## ACKNOWLEDGMENTS

The author thanks D. Vassiliadis for extremely useful conversations. The author acknowledges the invited talk, Heliophysics Data Science: Past Experience and Future Prospects, given at the Machine Learning in Heliophysics Conference in Amsterdam, The Netherlands. The conference was held September 16–20, 2019.

## REFERENCES

- Angelopoulos, V., Mukai, T., and Kokubun, S. (1999). Evidence for intermittency in earth’s plasma sheet and implications for self-organized criticality. *Phys. Plasmas* 6, 4161–4168. doi: 10.1063/1.873681
- Baker, D. N. (2008). A 21st century vision for geophysical data management. *Phys. Today* 61, 54–55. doi: 10.1063/1.2982123
- Baker, D. N. (2011). “Perspectives on geospace plasma coupling,” in *Modern Challenges in Nonlinear Plasma Physics: A Festschrift Honoring the Career of Dennis Papadopoulos*. AIP conference proceedings/plasma physics, eds D. Vassiliadis, S. F. Fung, X. Shao, I. A. Daglis, and J. D. Huba (College Park, MD: American Institute of Physics). doi: 10.1063/1.3544314
- Baker, D. N., Barton, C., Peterson, W., and Fox, P. (2008). Informatics and the 2007–2008 electronic geophysical year. *EOS* 89, 485–486. doi: 10.1029/2008EO480001
- Baker, D. N., Klimas, A. J., McPherron, R. L., and Buchner, J. (1990a). The evolution from weak to strong geomagnetic activity: an interpretation in terms of deterministic chaos. *Geophys. Res. Lett.* 17, 41–44. doi: 10.1029/GL017i001p00041
- Baker, D. N., Klimas, A. J., Pulkkinen, T. I., and McPherron, R. L. (1993). Re-examination of driven and unloading aspects of magnetospheric substorms. *Adv. Space Res.* 13, 75–83. doi: 10.1016/0273-1177(93)90316-4
- Baker, D. N., McPherron, R. L., Cayton, T. E., and Klebesadel, R. W. (1990b). Linear prediction filter analysis of relativistic electron properties at 6.6 RE. *J. Geophys. Res.* 95, 15133–15140. doi: 10.1029/JA095iA09p15133
- Baker, D. N., Pulkkinen, T. I., Angelopoulos, V., Baumjohann, W., and McPherron, R. L. (1996). The neutral line model of substorms: past results and present view. *J. Geophys. Res. Space Phys.* 101, 12975–13010. doi: 10.1029/95JA03753
- Baker, D. N., Pulkkinen, T. I., Büchner, J., and Klimas, A. J. (1999). Substorms: a global instability of the magnetosphere-ionosphere system. *J. Geophys. Res. Space Phys.* 104, 14601–14611. doi: 10.1029/1999JA900162
- Balasis, G., Daglis, I. A., Kapiris, P., Mandea, M., Vassiliadis, D., and Eftaxias, K. (2006). From pre-storm activity to magnetic storms: a transition described in terms of fractal dynamics. *Ann. Geophys.* 24, 3557–3567. doi: 10.5194/angeo-24-3557-2006
- Balasis, G., Daglis, I. A., Papadimitriou, C., Kalimeri, M., Anastasiadis, A., and Eftaxias, K. (2008). Dynamical complexity in  $D_{st}$  time series using non-extensive Tsallis entropy. *Geophys. Res. Lett.* 35:L14102. doi: 10.1029/2008GL034743
- Balasis, G., Daglis, I. A., Papadimitriou, C., Kalimeri, M., Anastasiadis, A., and Eftaxias, K. (2009). Investigating dynamical complexity in the magnetosphere using various entropy measures. *J. Geophys. Res.* 114:A00D06. doi: 10.1029/2008JA014035
- Bargatze, L. F., Baker, D. N., McPherron, R. L., and Hones, E. W., Jr. (1985). Magnetospheric impulse response for many levels of geomagnetic activity. *J. Geophys. Res. Space Phys.* 90, 6387–6394. doi: 10.1029/JA090iA07p06387
- Blanchard, G. T., and McPherron, R. L. (1992). “A bi-modal linear prediction filter relating the solar wind electric field to the westward electrojet,” in *Solar Terrestrial Prediction Proceedings* (Ottawa, ON).

- Bortnik, J., Claudepierre, S., Chu, X., Tobiska, K., and Camporeale, E. (2018). A Machine Learning Based Specification and Forecast Model of the Inner Magnetospheric Radiation Environment. Space Weather Week presentation. Available online at: <https://www.swpc.noaa.gov/sites/default/files/images/u59/04%20Jacob%20Bortnik%20Official.pdf>
- Boynton, R. J., Balikhin, M. A., Billings, S. A., and Amariutei, O. A. (2013). Application of nonlinear autoregressive moving average exogenous input models to geospace: advances in understanding and space weather forecasts. *Ann. Geophys.* 31, 1579–1589. doi: 10.5194/angeo-31-1579-2013
- Boynton, R. J., Balikhin, M. A., Billings, S. A., Wei, H. L., and Ganushkina, N. (2011). Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere. *J. Geophys. Res.* 116, 1–8. doi: 10.1029/2010JA015505
- Burch, J. L., Moore, T. E., Torbert, R. B., and Giles, B. L. (2015). Magnetospheric multiscale overview and science objectives. *Space Sci. Rev.* 199, 5–21. doi: 10.1007/s11214-015-0164-9
- Chang, T. (1999). Self-organized criticality, multi-fractal spectra, sporadic localized reconnections and intermittent turbulence in the magnetotail. *Phys. Plasmas* 6, 4137–4145. doi: 10.1063/1.873678
- Chapman, S. C., Watkins, N. W., Dendy, R. O., Helander, P., and Rowlands, G. (1998). A simple avalanche model as an analogue for magnetospheric activity. *Geophys. Res. Lett.* 25, 2397–2400. doi: 10.1029/98GL51700
- Chu, X. N., Bortnik, J., Li, W., Ma, Q., Angelopoulos, V., and Thorne, R. M. (2017). Erosion and refilling of the plasmasphere during a geomagnetic storm modeled by a neural network. *J. Geophys. Res. Space Phys.* 122, 7118–7129. doi: 10.1002/2017JA023948
- Consolini, G., Marcucci, M. F., and Candidi, M. (1996). Multifractal structure of auroral electrojet index data. *Phys. Rev. Lett.* 76, 4082–4085. doi: 10.1103/PhysRevLett.76.4082
- Freeman, M. P., and Morley, S. K. (2004). A minimal substorm model that explains the observed statistical distribution of times between substorms. *Geophys. Res. Lett.* 31:L12807. doi: 10.1029/2004GL019989
- Freeman, M. P., and Morley, S. K. (2009). No evidence for externally triggered substorms based on superposed epoch analysis of IMF  $B_z$ . *Geophys. Res. Lett.* 36:L21101. doi: 10.1029/2009GL040621
- Gleick, J. (1987). *Chaos, Making a New Science*. New York, NY: Viking Books.
- Gleisner, H., and Lundstedt, H. (1997). Response of the auroral electrojets to the solar wind modeled with neural networks. *J. Geophys. Res. Space Phys.* 102, 14269–14278. doi: 10.1029/96JA03068
- Gordeev, E., Sergeev, V., Tsyanenko, N., Kuznetsova, M., Rastäetter, L., Raeder, J., et al. (2016). The substorm cycle as reproduced by global MHD models. *Space Weather* 15, 131–149. doi: 10.1002/2016SW001495
- Hones, E. W. Jr. (1979). Transient phenomena in the magnetotail and their relation to substorms. *Space Sci. Rev.* 23, 393–410. doi: 10.1007/BF00172247
- Klimas, A. J., Baker, D. N., Roberts, D. A., Fairfield, D. H., and Buchner, J. (1992). A nonlinear dynamical analogue model of geomagnetic activity. *J. Geophys. Res.* 97, 12253–12266. doi: 10.1029/92JA00794
- Klimas, A. J., Valdivia, J. A., Vassiliadis, D., Baker, D. N., Hesse, M., and Takalo, J. (2000). Self-organized criticality in the substorm phenomenon and its relation to localized reconnection in the magnetospheric plasma sheet. *J. Geophys. Res.* 105, 18765–18780. doi: 10.1029/1999JA000319
- Klimas, A. J., Vassiliadis, D., Baker, D. N., Robert, D. A. (1996). The organized nonlinear dynamics of the magnetosphere. *J. Geophys. Res.* 101, 13089–13113. doi: 10.1029/96JA00563
- Lewis, Z. V. (1991). On the apparent randomness of substorm onset. *Geophys. Res. Lett.* 18, 1627–1630. doi: 10.1029/91GL01781
- Lui, A. T. Y. (2002). Multiscale phenomena in the near-earth magnetosphere. *J. Atmos. Solar-Terrestrial Phys.* 64, 125–143. doi: 10.1016/S1364-6826(01)00079-7
- Nagai, T. (1988). “Space weather forecast”: prediction of relativistic intensity at synchronous orbit. *Geophys. Res. Lett.* 15, 425–428. doi: 10.1029/GL015i005p00425
- Pepper, J. W., and Hoelzer, G. (2001). Unveiling mechanisms of collective behavior. *Science* 294, 1466–1467. doi: 10.1126/science.1066235
- Runge, J., Balasis, G., Daglis, I. A., Papadimitriou, C., and Donner, R. V. (2018). Common solar wind drivers behind magnetic storm–magnetospheric substorm dependency. *Sci. Rep.* 8:16987. doi: 10.1038/s41598-018-35250-5
- Sharma, A. S., Baker, D. N., and Borovsky, J. E. (2005). “Nonequilibrium phenomena in the magnetosphere,” in *Nonequilibrium Phenomena in Plasmas*. Astrophysics and space science library, Vol. 321, eds W. B. Burton, J. M. E. Kuijpers, E. P. J. Van Den Heuvel, H. Van Der Laan, I. Appenzeller, and J. N. Bahcall (Dordrecht: Springer), 3–22. doi: 10.1007/1-4020-3109-2\_1
- Shaw, R. (1984). *The Dripping Faucet as a Model Chaotic System*. The science frontier express series. Santa Cruz, CA: Aerial Press.
- Shukhta, M. A., Dmitrieva, N. P., and Sergeev, V. A. (2014). On the conditions preceding sudden magnetotail magnetic flux unloading. *Geophys. Res. Lett.* 41, 1093–1099. doi: 10.1002/2014GL059290
- Sitnov, M. I., Sharma, A. S., Papadopoulos, K., and Vassiliadis, D. (2001). Modeling substorm dynamics of the magnetosphere: from self-organization and self-organized criticality to nonequilibrium phase transitions. *Phys. Rev. E* 65:016116. doi: 10.1103/PhysRevE.65.016116
- Sitnov, M. I., Tsyanenko, N. A., Ukhorskiy, A. Y., and Brandt, P. C. (2008). Dynamical data-based modeling of the storm-time geomagnetic field with enhanced spatial resolution. *J. Geophys. Res.* 113. doi: 10.1029/2007JA013003
- Souza, V. M., Vieira, L. E. A., Medeiros, C., Da Silva, L. A., Alves, L. R., Kog, D., et al. (2016). A neural network approach for identifying particle pitch angle distributions in Van Allen probes data. *Space Weather* 14, 275–284. doi: 10.1002/2015SW001349
- Stephens, G. K., Sitnov, M. I., Korth, H., Tsyanenko, N. A., Ohtani, S., Gkioulidou, M., et al. (2019). Global empirical picture of magnetospheric substorms inferred from multimission magnetometer data. *J. Geophys. Res.* 124, 1085–1110. doi: 10.1029/2018JA025843
- Stringer, G. A., and McPherron, R. L. (1993). “Neural networks and predictions of day-ahead relativistic electrons at geosynchronous orbit,” in *Proceedings of the International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics*, Lund, Sweden, eds J. A. Joselyn, H. Lundstedt, and J. Trolinger (Boulder CO: NOAA Space Environment Center), 139–143.
- Takalo, J., and Timonen, J. (1997). Neural network prediction of AE data. *Geophys. Res. Lett.* 24, 2403–2406. doi: 10.1029/97GL02457
- Tsyanenko, N. A., and Sitnov, M. I. (2007). Magnetospheric configurations from a high-resolution data-based magnetic field model. *J. Geophys. Res.* 112:A06225. doi: 10.1029/2007JA012260
- Tulunay, E., Senalp, E. L., Radicella, S. M., and Tulunay, Y. (2006). Forecasting total electron content maps by neural technique. *Radio Sci.* 41:RS4016. doi: 10.1029/2005RS003285
- Uritsky, V. M., Klimas, A. J., Vassiliadis, D., Chua, D., and Parks, G. (2002). Scale-free statistics of spatiotemporal auroral emissions as depicted by POLAR UVI images: the dynamic magnetosphere is an avalanching system. *J. Geophys. Res. Space Phys.* 107, SMP 7-1–SMP 7-11. doi: 10.1029/2001JA000281
- Vassiliadis, D., Klimas, A. J., Baker, D. N., and Roberts, D. A. (1995). A description of the solar wind-magnetosphere coupling based on nonlinear prediction filters. *J. Geophys. Res. Space Phys.* 100, 3495–3512. doi: 10.1029/94JA02725
- Vörös, Z., Baumjohann, W., Nakamura, R., Runov, A., Zhang, T. L., Volwerk, M., et al. (2003). Multi-scale magnetic field intermittence in the plasma sheet. *Ann. Geophys.* 21, 1955–1964. doi: 10.5194/angeo-21-1955-2003
- Wanliss, J. A. (2005). Fractal properties of SYM-H during quiet and active times. *J. Geophys. Res.* 110:A03202. doi: 10.1029/2004JA010544
- Wing, S., Johnson, J. R., Camporeale, E., and Reeves, G. D. (2016). Information theoretical approach to discovering solar wind drivers of the outer radiation belt. *J. Geophys. Res. Space Phys.* 121, 9378–9399. doi: 10.1002/2016JA022711
- Yu, Y., Koller, J., Jordanova, V. K., Zaharia, S. G., Friedel, R. W., Morley, S. K., et al. (2014). Application and testing of the  $L^*$  neural network with the self-consistent magnetic field model of RAM-SCB. *J. Geophys. Res.* 119, 1683–1692. doi: 10.1002/2013JA019350

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Baker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparison of Deep Learning Techniques to Model Connections Between Solar Wind and Ground Magnetic Perturbations

Amy M. Keesee<sup>1\*</sup>, Victor Pinto<sup>1</sup>, Michael Coughlan<sup>1</sup>, Connor Lennox<sup>2</sup>, Md Shaad Mahmud<sup>2</sup> and Hyunju K. Connor<sup>3</sup>

<sup>1</sup> Department of Physics & Astronomy and Space Science Center, University of New Hampshire, Durham, NH, United States,

<sup>2</sup> Department of Electrical and Computer Engineering, University of New Hampshire, Durham, NH, United States,

<sup>3</sup> Department of Physics and Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AL, United States

## OPEN ACCESS

### Edited by:

Enrico Camporeale,  
University of Colorado Boulder,  
United States

### Reviewed by:

Gabor Toth,  
University of Michigan, United States  
Christian L. Vásconez,  
National Polytechnic School, Ecuador

### \*Correspondence:

Amy M. Keesee  
amy.keesee@unh.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and Space  
Sciences*

**Received:** 10 April 2020

**Accepted:** 28 August 2020

**Published:** 06 October 2020

### Citation:

Keesee AM, Pinto V, Coughlan M, Lennox C, Mahmud MS and Connor HK (2020) Comparison of Deep Learning Techniques to Model Connections Between Solar Wind and Ground Magnetic Perturbations. *Front. Astron. Space Sci.* 7:550874.  
doi: 10.3389/fspas.2020.550874

Geomagnetically induced currents (GIC) can drive power outages and damage power grid components while also affecting pipelines and train systems. Developing the ability to predict local GICs is important to protecting infrastructure and limiting the impact of geomagnetic storms on public safety and the economy. While GIC data is not readily available, variations in the magnetic field,  $dB/dt$ , measured by ground magnetometers can be used as a proxy for GICs. We are developing a set of neural networks to predict the east and north components of the magnetic field,  $B_E$  and  $B_N$ , from which the horizontal component,  $B_H$ , and its variation in time,  $dB_H/dt$ , are calculated. We apply two techniques for time series analysis to study the connection of solar wind and interplanetary magnetic field properties obtained from the OMNI dataset to the ground magnetic field perturbations. The analysis techniques include a feed-forward artificial neural network (ANN) and a long-short term memory (LSTM) neural network. Here we present a comparison of both models' performance when predicting the  $B_H$  component of the Ottawa (OTT) ground magnetometer for the year 2011 and 2015 and then when attempting to reconstruct the time series of  $B_H$  for two geomagnetic storms that occurred on 5 August 2011 and 17 March 2015.

**Keywords:** space weather, GIC, geomagnetic storms, ground magnetic field, machine learning, neural network, LSTM

## 1. INTRODUCTION

Geomagnetically induced currents (GICs) are one of the most significant space weather effects due to their potential to damage the power grid and can cause widespread, long-term power outages. Thus, the ability to forecast GICs is of significant interest to the space weather community, industry partners, and national interests. The intensity of GICs is determined by the strength of the geoelectric field. However, neither measurements of GICs nor the geoelectric field are readily available. The geoelectric field is driven by temporal changes in the magnetic field and the local geology. Thus, measurements of  $dB/dt$  using ground magnetometers are used as a proxy for studying GICs. Ngwira et al. (2018) studied two storms during which intense  $dB/dt$  peaks occurred and indicated that substorms appear to be the driver of GICs, but state that it's not clear how the widespread features of substorms lead to localized peaks in  $dB/dt$ . They theorize that it could be

due to “the mapping of magnetospheric currents to local ionospheric structures,” but indicate that further study is needed. Physics based models are used to determine magnetic field fluctuations, but high resolution models are needed to obtain the spatially localized variations (Welling et al., 2019). Such models are computationally expensive and take longer time to run, posing challenges for their use as a forecasting tool. Machine learning based models have the potential for providing efficient, computationally inexpensive forecasts. Wintoft et al. (2015) developed models using Elman neural networks to predict the 30-min maximum of  $dB_H/dt$  (horizontal component of  $dB/dt$ ) from ACE solar wind and magnetic field measurements. Their models generally predict the timing of GICs caused by sudden impulses well, even when they train the model using only ACE magnetic field measurements.

While many studies of GICs focus on high magnetic latitudes ( $>60^\circ$ ) that lie under the auroral oval, it has been shown that mid- ( $50^\circ$ – $60^\circ$ ) and low- ( $<50^\circ$ ) latitude regions are also at risk (Gaunt and Coetzee, 2007; Ngwira et al., 2008; Pulkkinen et al., 2010; Oliveira et al., 2018). Lotz and Cilliers (2015) developed a neural network based model using solar wind and IMF inputs and  $dB/dt$  measurements at a Southern hemisphere mid-latitude station as outputs. They developed separate models for the north and east components of the geomagnetic field and found that fluctuations in the eastward component are more dependent on the interplanetary magnetic field (IMF)  $B_z$ . Similar to Wintoft et al. (2015), they found reasonable predictions of the timing of intense fluctuations, with less accuracy as the storm evolved.

More complex neural network architectures can be used to improve the predictions for time-series data. For example, recurrent neural networks such as long short-term memory (LSTM) techniques are used to “remember” parameters from earlier times that have a strong influence on the output features. In this study we present a comparison of models using a feed-forward artificial neural network (ANN) with a built-in time dependence and a LSTM neural network to predict the ground magnetic field north and east components (and therefore the perturbations  $dB_H/dt$ ) at the mid-latitude ground magnetometer station located in Ottawa (OTT). We then discuss the performance of the models by using two of the benchmark geomagnetic storms suggested by Welling et al. (2018) and Pulkkinen et al. (2013). Finally, we discuss several model variations that were implemented during the course of this study to determine possible improvements to the performance of the models.

## 2. DATA

For this study, we use solar wind and interplanetary magnetic field (IMF) data obtained from the OMNIWeb dataset available through NASA’s Space Physics Data Facility from 1995 through 2010 for the purpose of training and validation of the models, and from 2011 and 2015 for testing. These 2 years were selected for testing because they include storms from the Pulkkinen-Welling validation set for ground magnetic perturbations (Pulkkinen et al., 2013; Welling et al., 2018). Baseline-removed ground

magnetometer data from OTT has been obtained from SuperMag (Gjerloev, 2012). The Ottawa ground magnetometer is located at magnetic latitude  $54.98^\circ$  N and lags UT by 5 h (meaning local midnight occurs at 05:00 UT). The choice of using solar wind data from OMNI instead of the more traditional and continuously available geomagnetic indices is based on the long-term goal of being able to forecast variations in the ground magnetic field ahead of time, and real time solar wind parameters obtained at the L1 position tend to give a 30–40 min window to distribute a warning. However, the OMNI dataset has been mostly avoided in the past as it contains approximately 20% of missing data distributed roughly evenly through the years in the plasma parameters and  $\sim 8\%$  in the IMF measurements. This was noted by Wintoft et al. (2015), leading them to compare models trained on just the magnetic field vs. combined magnetic field and plasma measurements. Our intention is to use both IMF and plasma parameters. Since a relatively continuous dataset is preferred for training, some linear interpolation has been done in the training/validation dataset of up to 10 min in all missing parameters, which reduces the missing values to  $\sim 6\%$  in both IMF and plasma measurements. OTT ground magnetic field measurements have less than 1% of missing data during the period of study. Those missing data points have been removed from the training set. For the testing periods of 2011 and 2015 a full linear interpolation has been performed to the solar wind data to achieve a completely continuous dataset.

Although  $dB_H/dt$  is the best proxy measurement to GIC forecasting, it is also very noisy and therefore difficult to forecast directly with data-driven models. Tóth et al. (2014) also found this to be true for first principles-based models. We therefore aim to first predict the northward and eastward components of the baseline-removed ground magnetic field,  $B_N$  and  $B_E$ , respectively, using two independent models and then combine them to obtain the predicted horizontal component

$$B_H = \sqrt{B_N^2 + B_E^2} \quad (1)$$

For the purpose of comparison with the metrics defined by Pulkkinen et al. (2013) we also need to obtain  $(dB/dt)_H$ , which is calculated as

$$\left( \frac{d\mathbf{B}}{dt} \right)_H = \sqrt{\frac{dB_N}{dt}^2 + \frac{dB_E}{dt}^2} \quad (2)$$

where

$$\frac{dB_i}{dt} = \frac{[B_i(t+1 \text{ min}) - B_i(t)]}{1 \text{ min}}$$

$i$  is determined by the ground data resolution (1 min), and  $i$  represents the components  $N$  or  $E$ . We emphasize that large variations are most likely to result in significant GIC events.

## 3. MODELS

### 3.1. Feed-Forward Artificial Neural Network

In our first attempt to forecast the  $B_E$  and  $B_N$  components we have chosen to train a fully-connected, feed-forward, artificial

neural network (ANN) developed using the TensorFlow-Keras environment (Abadi et al., 2015). The Tensorflow-Keras environment is highly modular and allows the easy integration of data and different types of networks. It also allows integration and computation through GPUs instead of the traditional CPU computing which reduces significantly (in our case, up to a factor of 10) the training times. For the network architecture we have chosen a 4-layer deep network with hidden layers of 291-146-73-36 neurons plus a dropout layer (rate 0.1) in between the first and second layer to avoid overfitting. Such selection of neurons matches the dimensionality of the feature vectors and then each layer is halved. We have selected mean square error (MSE) as our loss function, ADAM as our optimizer, and a REctified Linear Unit (RELU) as our activation function. We track the loss in the validation dataset and stop the training after the MSE has not decreased for 25 epochs to avoid overfitting. To incorporate time dependence in the ANN, the input vector for  $t$  includes features from previous time steps, e.g.,  $t - 1$ ,  $t - 2$ . We have chosen to include a 2-h time-history for solar wind speed ( $V_T$ ,  $V_x$ ,  $V_y$ ,  $V_z$ ), IMF ( $B_T$ ,  $B_x$ ,  $B_y$ ,  $B_z$ ), proton density, dynamic pressure, temperature and solar wind electric field using a 1-min cadence for the first 12 preceding minutes (i.e., up to  $t - 12$ ) plus 10-min averages over the entire interval (yielding 12 additional values). Additionally, ground magnetometer sin(MLT) and cos(MLT) values have been included to ensure a cyclical dependence over the Earth's rotation and solar zenith angle as a proxy of both longitude and yearly seasonality. The resulting feature vector thus contains 291 features and that explains our choice of neurons in the first layer. OMNI data from 1995 to 2010 was split sequentially 70% for training and 30% for validation. The model was then trained and the result used on the test data from 2011 to 2015.

**Figure 1** (top row) presents a density plot of the  $\log_{10}$ (real) vs.  $\log_{10}$ (predicted)  $B_H$  values calculated using the validation and test results obtained for  $B_E$  and  $B_N$  and Equation (1). The ANN performs a relatively good job at predicting values for the validation set with a root mean square error (RMSE) of  $\sim 8.7$  nT and an explained variance of 39%, resulting in a correlation coefficient of 0.61. For the test cases, the respective RMSE values are  $\sim 9.2$  nT with explained variance of 36% and correlation coefficient 0.60 for the year 2011 and an RMSE of  $\sim 14.7$  nT and explained variance of 44% with a correlation of 0.66 for the year 2015. While the predictions could be improved, the consistency of the values of **Figure 1** indicate that the model is not overfitting. The correlation coefficient values are a bit lower to those obtained by Lotz et al. (2017) of 0.71 and 0.69 for models predicting separate components of the horizontal magnetic field at a mid-latitude station. Wintoft et al. (2015) obtain much higher correlation coefficients, although they are only considering the maximum value within each 30-min window. The slightly worse RMSE values obtained for the year 2015 could be due to the higher variability of a year in the solar maximum. We have also plotted the error distributions for the predictions of the validation and 2011 and 2015 test sets (See **Supplementary Material**). The similarity in the error distributions is also an indication of a model that is not overfitting, consistent with **Figure 1**.

### 3.2. Long Short Term Memory

The long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) recurrent neural network was developed using TensorFlow-Keras (Abadi et al., 2015). The input features used in the LSTM model were the same as were used for the ANN as was the training-validation split of the 1995-2010 OMNI data. The LSTM model requires an extra dimension in the input vector to build in the time history. To keep the models consistent with each other, the LSTM model also uses the preceding 12-min plus 10-min averages over the preceding 2 h time history. Memory limitations required the use of a custom data generator which fed batch sizes of 512 training samples sequentially into the model for training. The network consists of a single LSTM layer with 147 neurons and a single dense layer with one neuron. A RELU is used as the activation function, the optimizer is ADAM, and the loss function MSE. As with the ANN, validation loss was monitored and training was stopped after the MSE had not decreased in 25 epochs. Although recurrent neural networks, and in particular LSTM, are capable of utilizing the time history of the target parameter (in this case,  $B_N$  or  $B_E$ ) to improve the prediction, in this study we have chosen not to use it in order to obtain a closer comparison with the ANN model. Once the model was trained, it was tested on data from 2011 and 2015.

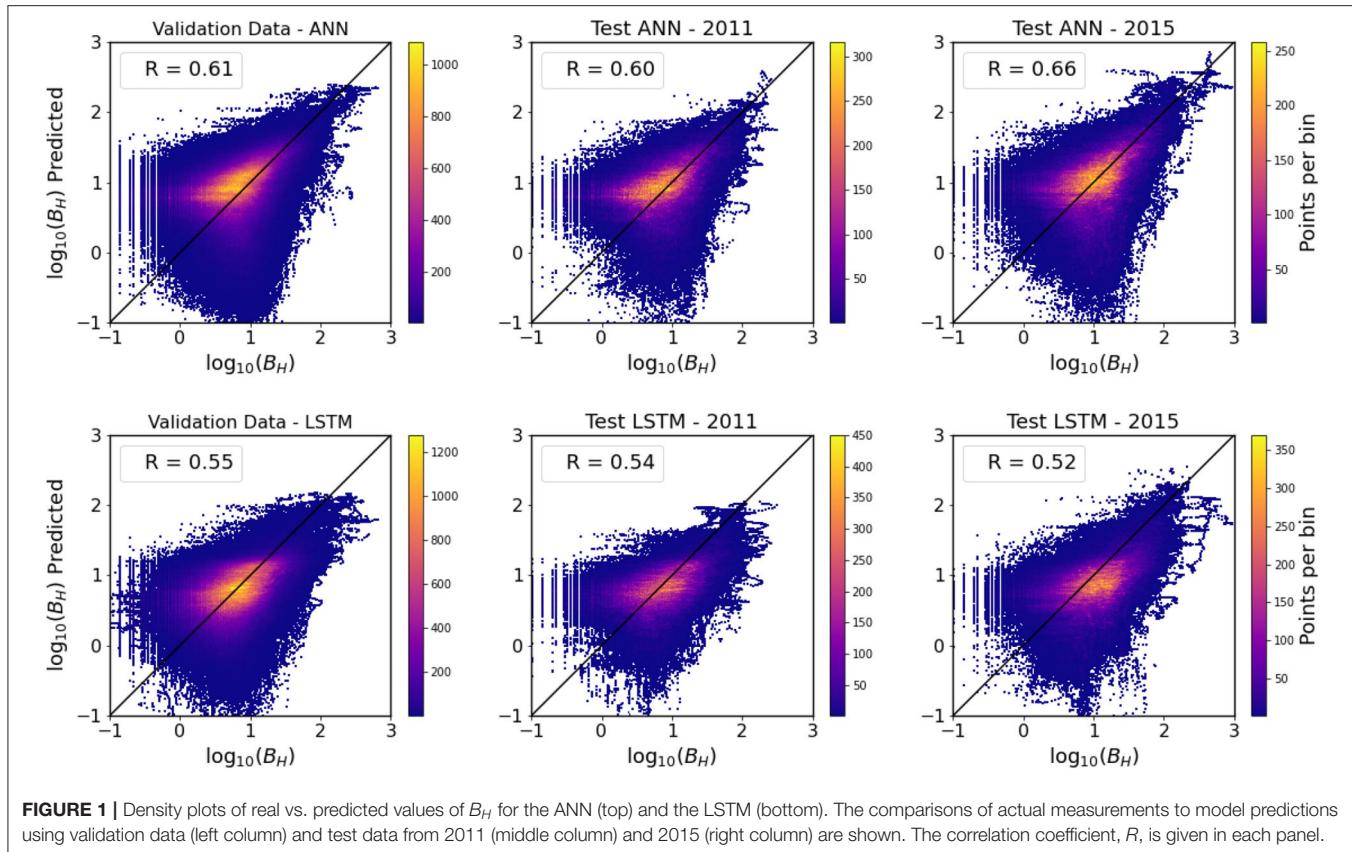
The real vs. predicted values of  $B_H$  for the validation and test datasets are shown as a density plot in **Figure 1** (bottom row). The RMSE of the validation set is 8.7 nT with a 33% explained variance, 9.7 nT for the 2011 dataset with a 35% explained variance and 16.8 nT for the 2015 dataset with a 36% explained variance. The consistency of the correlation coefficients (shown in **Figure 1**) indicates that the model did not suffer from overfitting, which could have been a concern with the amount of input data. Considering these parameters, the LSTM model seems to under-perform the ANN. Similarly to the ANN case, the error distribution for the predictions of the LSTM model further indicate that the model is not severely overfitting (see **Supplementary Material**).

## 4. RESULTS

The two types of models we trained are used to predict  $B_N$  and  $B_E$  during storms that occurred on 5 August 2011 and 17 March 2015. The selection was based on the recommendations from the Pulkkinen-Welling validation set for ground magnetic perturbations (Pulkkinen et al., 2013; Welling et al., 2018). These two storms were selected because they are outside of the time range used to train and validate the models and because they correspond to two very different years in terms of geomagnetic activity, 2011 being on the minimum-ascending part of the solar cycle, and 2015 corresponding to the solar cycle maximum. The predicted values are used to calculate  $B_H$  (Equation 1) and  $dB_H/dt$  (Equation 2), and both values are compared to the real measurements. The results for a third storm from the validation set, 17 March 2013, are shown in the **Supplementary Material**.

### 4.1. August 5 2011 Storm

**Figure 2** shows the temporal evolution of the 5 August 2011 geomagnetic storm, including SYM-H index, solar wind speed  $V_x$ ,



**FIGURE 1 |** Density plots of real vs. predicted values of  $B_H$  for the ANN (top) and the LSTM (bottom). The comparisons of actual measurements to model predictions using validation data (left column) and test data from 2011 (middle column) and 2015 (right column) are shown. The correlation coefficient,  $R$ , is given in each panel.

IMF magnetic field  $z$ -component, along with the measured  $B_H$  and  $dB_H/dt$  at Ottawa station, and the predicted  $B_H$  and  $dB_H/dt$  from the ANN and LSTM models. This is a strong storm with a minimum SYM-H value of  $-126$  nT, driven by a combination of a coronal mass ejection and high speed stream, with the shock arriving at Earth leading to a sudden storm commencement around 18 UT on August 5. In this event,  $B_H$  presents a clear response to the storm that shows as repeated  $dB_H/dt$  variations of up to  $\sim 50$  nT/min occurring for several hours during the storm main phase, following the period of quick increase in solar wind speed and mostly during the time in which the IMF  $B_z$  component is strongly and persistently southward.

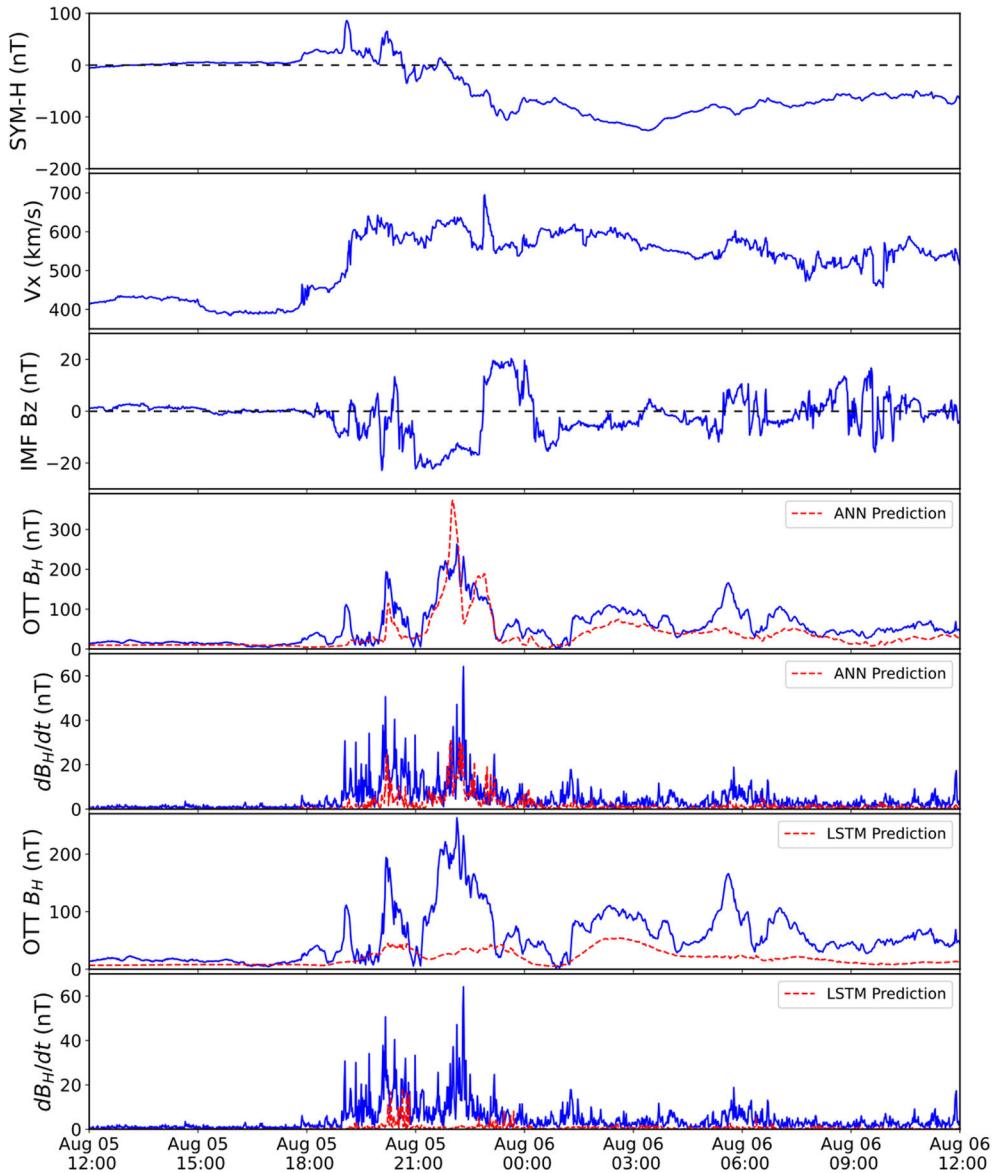
In terms of reconstructing the  $B_H$  evolution, neither model does a good job of predicting the first two enhancements. The ANN does a decent job predicting the third and fourth enhancements. After  $\sim 23:00$  UT on August 5, the ANN predicts some increases but does not match the magnitude likely due to these being later in the storm when the magnetic field perturbations are controlled more by parameters within the magnetosphere or ionosphere, although this is also a period in which the actual fluctuations in  $B_H$  are relatively small. The LSTM doesn't do well predicting the enhancements during this storm in general. Considering the  $dB_H/dt$ , the ANN comes close to predicting the timing of the biggest spike in  $dB_H/dt$  at  $\sim 20:00$  UT on August 5, at about half the magnitude, while the LSTM misses this completely.

## 4.2. March 17 2015 Storm

Figure 3 shows the evolution of the 17 March 2015 geomagnetic storm using the same format as Figure 2. This was the largest geomagnetic storm of solar cycle 24, with minimum SYM-H index of  $-234$  nT. Carter et al. (2016) analyzed the ground magnetic perturbations during this storm, showing that the mid- and low-latitude fluctuations predominantly occurred at the sudden storm commencement. This indicates that the solar wind parameters are most important for predicting GICs at mid-latitude.

Both models miss the initial spike at the sudden storm commencement just before 5:00 UT, and start to predict enhanced  $B_H$  about 2–3 h later. Both models predict some enhancement near 8:00 UT, though matching the real enhancement is unlikely due to this being an interval of linear interpolation of the input data (as seen in the straight line in the solar wind velocity and IMF  $B_z$  over 7:00–9:00 UT). This linear interpolation results in the ANN overpredicting the magnitude and the LSTM predicts significant spikes. The LSTM does a better job of predicting the enhancement just after 12:00 UT, even getting close in overall magnitude of  $B_H$ . Again, the linearly interpolated input data ( $\sim 15:00$ – $17:00$  UT) reduces the ability to accurately predict the second half of this enhancement. Only the ANN predicts the largest enhancement at  $\sim 21:00$  UT, but does not match the shape of the peak or timing.

It is important to note that in this particular storm, there are some gaps in the solar wind measurements, particularly



**FIGURE 2 |** August 5, 2011 storm measurements and predictions, including Sym-H index, solar wind velocity, interplanetary magnetic field z-component, measured (blue solid) and predicted (red dashed) horizontal magnetic field and time-dependent variation at OTT ground station for the ANN and LSTM models.

between 06–08 UT and then on 16–18 UT on March 17. In order to generate a prediction, the solar wind data has been linearly interpolated. Despite the missing data, a simple linear interpolation allows the models to recover part of the variability of the ground magnetic field fluctuations, suggesting that efforts on gap-filling, including empirical modeling of the solar wind data could still yield positive results in the risk assessment of GICs.

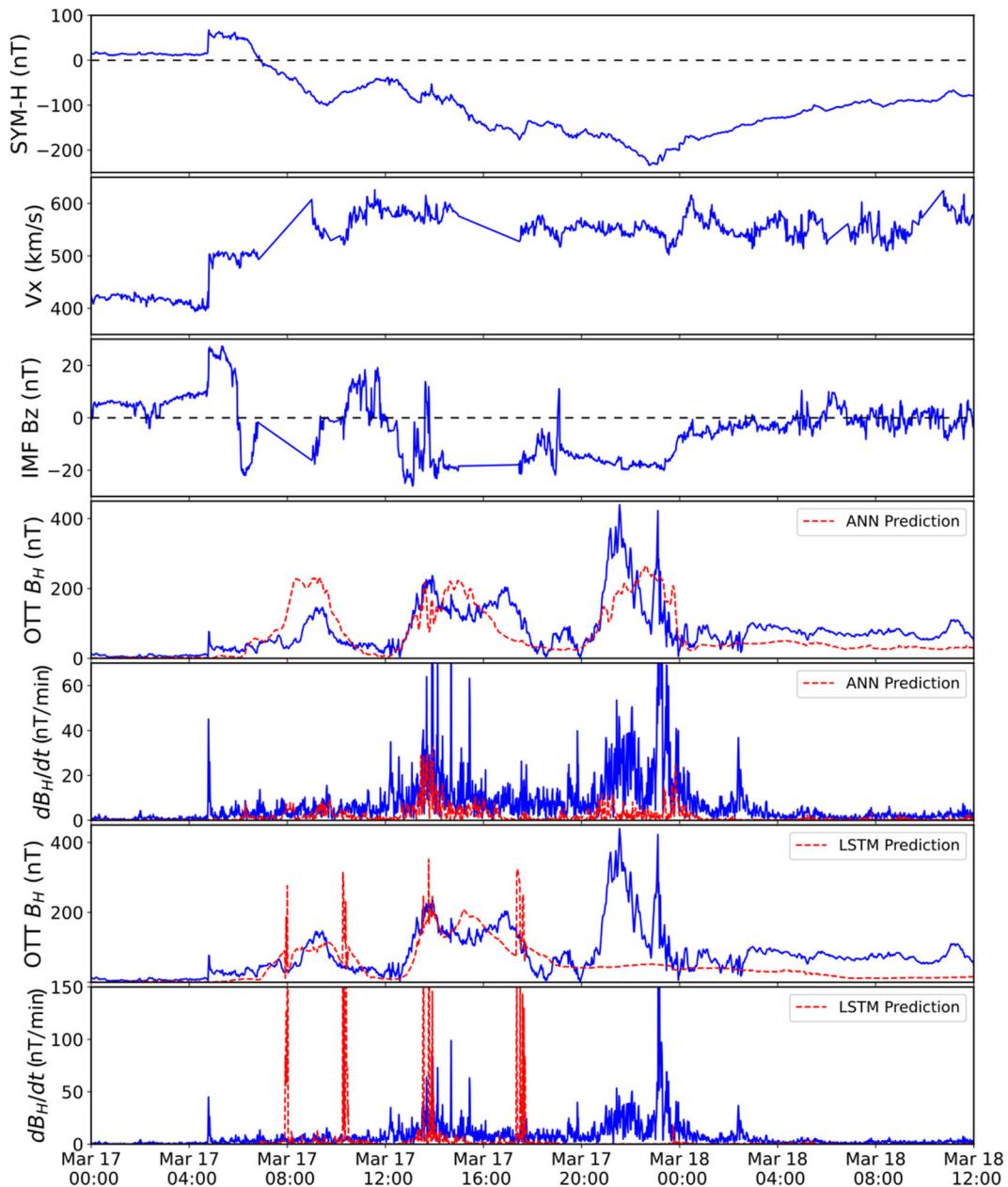
### 4.3. Validation Metrics

The Pulkkinen-Welling recommendations include four metrics for validating predictions of ground magnetic perturbations (Pulkkinen et al., 2013; Welling et al., 2018) using binary event

analysis. For a particular storm, the interval of interest is divided into non-overlapping, 20-min windows, and for each window, a  $dB_H^*/dt$  value is calculated as the maximum measured  $dB_H/dt$  within the window. In addition to the method of calculating  $(dB/dt)_H$  using Equation (2), we use the power law empirical fitting for the OTT station described by Tóth et al. (2014),

$$\frac{dB_H}{dt} = \left( \frac{B_H}{248 \text{ nT}} \right)^{1.04} \frac{\text{nT}}{\text{s}}, \quad (3)$$

and present both results for comparison. Pulkkinen-Welling propose four different thresholds of 0.3, 0.7, 1.1, and 1.5 nT/s to evaluate if the model is able to predict a variation of that



**FIGURE 3 |** March 17, 2015 storm measurements and predictions, including SYM-H index, solar wind velocity, interplanetary magnetic field z-component, measured (blue solid) and predicted (red dashed) horizontal magnetic field and time-dependent variation at OTT ground station for the ANN and LSTM models.

magnitude. Since the thresholds are in nT/s, and our calculated  $dB_H/dt$  is in nT/min, we multiply the thresholds by 60. Using the values of the predicted  $dB_H^*/dt$ , we determine whether the model accurately predicts the events by calculating true positives, false positives, true negatives, and false negatives. Accuracy is determined by calculating the following metrics: probability of detection (POD), probability of false detection (PFD), proportion correct (PC), and Heidki Skill Score (HSS). **Table 1** displays the metrics for the ANN and LSTM for the 2011 and 2015 storms. The missing values are due to no occurrences of the real and

predicted values crossing the higher thresholds. The metrics for the 2013 storm are also shown in the **Supplementary Material**, and although they are a bit better than for the 2011 and 2015 storms, it suffers from having very few thresholds crossings beyond 0.3 nT/s for the Ottawa station.

The poor prediction of the LSTM for the 2011 storm is evident in the low POD and HSS values. The low PFD (lower is better) and high PC are indicative of the low numbers of real crossings of the threshold levels. The Tóth et al. (2014) empirical fitting method results in improved metrics for these

**TABLE 1 |** Validation metrics for the 05 August 2011 and the 17 March 2015 geomagnetic storms using  $dB_H/dt$  maximum values every 20 min calculated from the Pulkkinen definition (Equation 2) and the Tóth empirical fitting (Equation 3).

		Pulkkinen				Tóth			
		18	42	66	90	18	42	66	90
2011 Storm	ANN POD	0.33	0.00	—	—	0.50	0.50	—	—
	LSTM POD	0.08	0.00	—	—	0.00	0.00	—	—
	ANN PFD	0.02	0.00	0.00	0.00	0.02	0.03	0.03	0.01
	LSTM PFD	0.08	0.00	—	—	0.00	0.00	—	—
	ANN PC	0.88	0.97	1.00	1.00	0.90	0.96	0.97	0.99
	LSTM PC	0.85	0.97	1.00	1.00	0.84	0.97	1.00	1.00
	ANN HSS	0.41	0.00	—	—	0.58	0.37	—	—
	LSTM HSS	0.13	0.00	—	—	0.00	0.00	—	—
	ANN POD	0.09	0.00	0.00	0.00	0.69	0.77	0.00	0.00
2015 Storm	LSTM POD	0.13	0.11	0.20	0.00	0.38	0.33	0.20	0.00
	ANN PFD	0.00	0.00	0.00	0.00	0.13	0.13	0.00	0.00
	LSTM PFD	0.05	0.07	0.07	0.07	0.19	0.08	0.03	0.00
	ANN PC	0.73	0.92	0.95	0.98	0.82	0.86	0.95	0.98
	LSTM PC	0.71	0.86	0.90	0.92	0.68	0.87	0.94	0.98
	ANN HSS	0.13	0.00	0.00	0.00	0.56	0.42	0.00	0.00
	LSTM HSS	0.09	0.04	0.10	0.00	0.19	0.23	0.19	0.00

Four threshold values in nT/min are used.

models, as they found for a first principles-based model. In fact, the ANN has a similar HSS for the 0.3 nT/s threshold as they report for mid-latitude stations (0.583). However, our methods have a lot of potential for improvement. We had originally trained and optimized a single model of each type that predicted  $B_H$ , rather than independent models for  $B_N$  and  $B_E$ . The single models of each type had better correlation coefficients, explained variance, and RMSE than the values discussed in section 3, but much lower validation metrics than those shown in Table 1. (We note that predicting only  $B_H$  requires calculation of  $dB_H/dt$  directly, rather than being able to use Equation 2, such that the metrics do not have a one-to-one comparison.) We implemented separate modeling of the north and east components for more direct comparison with other models since the Pulkkinen et al. (2013) method is widely accepted.) Using this single model method, we also trained models that use only 24 min of time history but all at a 1-min cadence. The use of the 10-min averages produced higher explained variance scores of  $B_H$ . However, this resulted in a smoothing of the  $B_H$  prediction. This smoothing causes less variation in the  $dB_H/dt$  predictions and decreases the HSS. Additionally, the LSTM performs poorly, likely due to the fact that we did not use the time history of the target parameter, despite that being the strength of the LSTM.

## 5. CONCLUSIONS

We have developed and compared two types of models that predict the north and east components of the ground magnetic

field,  $B_N$  and  $B_E$ , at a single mid-latitude ground station. One model is a feed-forward artificial neural network that includes time dependence as input features and the other is a long short-term memory neural network. The predictions from each model are compared to real measurements for 2 years, 2011 and 2015, including a storm during each year. There is some ability for each of the models to predict the timing of magnetic field perturbations, though this ability is not consistently better for either model between the storms and neither is able to predict the magnitude of the enhancements or predict enhancements later in the storm. Validation metrics indicate that the LSTM is barely more skilled than random or constant predictions, and that using an empirical fitting improves HSS as it does for first principles-based models. Next steps to improve the models include adjustments of the input parameters, increased time history cadence, and comparison to additional time series techniques. Another limitation of these models is the use of only one ground magnetometer station. We expect better predictions if we include more ground stations in the mid-latitude range to get more MLT coverage.

## DATA AVAILABILITY STATEMENT

The solar wind, IMF, and Sym-H index data are available from OMNIWeb at <https://omniweb.gsfc.nasa.gov> and the ground magnetometer data are available from SuperMAG at <http://supermag.jhuapl.edu>.

## AUTHOR CONTRIBUTIONS

AK was the primary author of the paper, provided guidance on model development and analysis, and participated in interpretation. VP conducted data preparation, model development, analysis, and interpretation and assisted with writing. MC conducted model development and analysis and assisted with writing. MM provided guidance on data preparation and model development. CL assisted with data preparation. HC assisted with overall project discussions and provided editorial comments. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by NSF EPSCoR Award OIA-1920965.

## ACKNOWLEDGMENTS

We thank members of the project team at UNH and UAF and Chigo Ngwira that participated in discussions for this work. We thank OMNIWeb and SuperMag for providing the data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fspas.2020.550874/full#supplementary-material>

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.
- Carter, B. A., Yizengaw, E., Pradipta, R., Weygand, J. M., Piersanti, M., Pulkkinen, A., et al. (2016). Geomagnetically induced currents around the world during the 17 March 2015 storm. *JGR Space Phys.* 121, 10496–10507. doi: 10.1002/2016JA023344
- Gaunt, C. T., and Coetze, G. (2007). “Transformer failures in regions incorrectly considered to have low GIC-risk,” in *2007 IEEE Lausanne POWERTECH, Proceedings*, 807–812.
- Gjerloev, J. W. (2012). The SuperMAG data processing technique. *J. Geophys. Res. Space Phys.* 117:A09213. doi: 10.1029/2012JA017683
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Lotz, S. I., and Cilliers, P. J. (2015). A solar wind-based model of geomagnetic field fluctuations at a mid-latitude station. *Adv. Space Res.* 55, 220–230. doi: 10.1016/j.asr.2014.09.014
- Lotz, S. I., Heyns, M. J., and Cilliers, P. J. (2017). Regression-based forecast model of induced geoelectric field. *Space Weather* 15, 180–191. doi: 10.1002/2016SW001518
- Ngwira, C. M., Pulkkinen, A., McKinnell, L. A., and Cilliers, P. J. (2008). Improved modeling of geomagnetically induced currents in the South African power network. *Space Weather* 6, 1–8. doi: 10.1029/2008SW000408
- Ngwira, C. M., Sibeck, D., Silveira, M. V. D., Georgiou, M., Weygand, J. M., Nishimura, Y., et al. (2018). A study of intense local d B /d t variations during two geomagnetic storms . *Space Weather* 16, 676–693. doi: 10.1029/2018SW001911
- Oliveira, D. M., Arel, D., Raeder, J., Zesta, E., Ngwira, C. M., Carter, B. A., et al. (2018). Geomagnetically induced currents caused by interplanetary shocks with different impact angles and speeds. *Space Weather* 16, 636–647. doi: 10.1029/2018SW001880
- Pulkkinen, A., Kataoka, R., Watari, S., and Ichiki, M. (2010). Modeling geomagnetically induced currents in Hokkaido, Japan. *Adv. Space Res.* 46, 1087–1093. doi: 10.1016/j.asr.2010.05.024
- Pulkkinen, A., Rastatter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., et al. (2013). Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space Weather* 11, 369–385. doi: 10.1002/swe.20056
- Tóth, G., Meng, X., Gombosi, T. I., and Rastatter, L. (2014). Predicting the time derivative of local magnetic perturbations. *J. Geophys. Res. Space Phys.* 119, 310–321. doi: 10.1002/2013JA019456
- Welling, D. T., Dimmock, A. P., L, R., Morley, S., and Yordanova, E. (2019). “Resolving small scale gic effects: what is our capability?,” in *Abstract SH32B-05 Presented at 2019 Fall Meeting* (San Francisco, CA:AGU).
- Welling, D. T., Ngwira, C. M., Opgenoorth, H., Haidupek, J. D., Savani, N. P., Morley, S. K., et al. (2018). Recommendations for next-generation ground magnetic perturbation validation. *Space Weather* 16, 1912–1920. doi: 10.1029/2018SW002064
- Wintoft, P., Wik, M., and Viljanen, A. (2015). Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *J. Space Weather Space Clim.* 5, A7–P9. doi: 10.1051/swsc/2015008

**Conflict of Interest:** 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.

Copyright © 2020 Keesee, Pinto, Coughlan, Lennox, Mahmud and Connor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Data Mining Reconstruction of Magnetotail Reconnection and Implications for Its First-Principle Modeling

Mikhail Sitnov<sup>1\*</sup>, Grant Stephens<sup>1</sup>, Tetsuo Motoba<sup>1</sup> and Marc Swisdak<sup>2</sup>

<sup>1</sup>Applied Physics Laboratory, The Johns Hopkins University, Laurel, MD, United States, <sup>2</sup>Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD, United States

## OPEN ACCESS

**Edited by:**

Enrico Camporeale,  
University of Colorado Boulder,  
United States

**Reviewed by:**

Thomas Berger,  
University of Colorado Boulder,  
United States  
Anton Artemyev,  
Space Research Institute (RAS),  
Russia

**\*Correspondence:**

Mikhail Sitnov  
*Mikhail.Sitnov@jhuapl.edu*

**Specialty section:**

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Physics*

**Received:** 22 December 2020

**Accepted:** 05 February 2021

**Published:** 21 April 2021

**Citation:**

Sitnov M, Stephens G, Motoba T and Swisdak M (2021) Data Mining Reconstruction of Magnetotail Reconnection and Implications for Its First-Principle Modeling. *Front. Phys.* 9:644884.  
doi: 10.3389/fphy.2021.644884

Magnetic reconnection is a fundamental process providing topological changes of the magnetic field, reconfiguration of space plasmas and release of energy in key space weather phenomena, solar flares, coronal mass ejections and magnetospheric substorms. Its multiscale nature is difficult to study in observations because of their sparsity. Here we show how the lazy learning method, known as K nearest neighbors, helps mine data in historical space magnetometer records to provide empirical reconstructions of reconnection in the Earth's magnetotail where the energy of solar wind-magnetosphere interaction is stored and released during substorms. Data mining reveals two reconnection regions (X-lines) with different properties. In the mid tail ( $\sim 30R_E$  from Earth, where  $R_E$  is the Earth's radius) reconnection is steady, whereas closer to Earth ( $\sim 20R_E$ ) it is transient. It is found that a similar combination of the steady and transient reconnection processes can be reproduced in kinetic particle-in-cell simulations of the magnetotail current sheet.

**Keywords:** data mining and knowledge discovery, nearest neighbor method, magnetosphere, magnetotail, magnetic reconnection, space weather, particle-in-cell simulations

## 1 INTRODUCTION

Charged particles, electrons and ions forming space plasmas usually drift in the ambient magnetic field making plasmas frozen in that field [1]. The frozen-in condition may be broken when oppositely directed field lines approach each other so closely that particles become unmagnetized and their orbits become different from conventional drift motions. As a result, magnetic field lines may change their connectivity near so-called X-lines in the process of magnetic reconnection. This process was introduced to explain major sources of space weather disturbances on the Sun, solar flares [2, 3] and coronal mass ejections (CMEs) [4]. It was also invoked by Dungey [5] to describe the structure of the Earth's magnetosphere, the plasma bubble surrounding our planet and protecting its life from the hazardous stream of high-energy particles emitted by our star. According to Dungey, reconnection takes place on the day side of the magnetospheric boundary, the magnetopause, to provide the solar wind plasma entry into the magnetosphere through the reconnected magnetic flux tubes. Then the flux tubes reconnect again on the night side, in the region where the Earth's dipole magnetic field lines are stretched in the antisunward direction forming the magnetotail. Finally, to explain a delayed explosive response of the polar regions of the magnetosphere to solar wind disturbances during

substorms [6], Hones [7] proposed that the substorm explosions are powered by the unsteady reconnection in the tail due to the formation of another “near-Earth” X-line.

The magnetotail reconnection is important not only as a key element of the space weather chain. It occurs in space plasma practically in the absence of particle collisions. Similar collisionless reconnection processes are expected to occur in the solar corona during flares and CMEs, where in-situ observations are impossible [1]. They are also expected in sufficiently hot laboratory plasmas that are investigated on the way to controlled nuclear fusion [8]. Thus, the magnetotail represents a natural space laboratory for collisionless reconnection due to many dedicated missions, such as Geotail [9], Cluster [10], THEMIS [11] and MMS [12].

The magnetotail is also very interesting because it reveals different regimes of reconnection. On the one hand, it must experience steady reconnection, which was conjectured by Dungey [5] in his description of the magnetospheric convection cycle and later confirmed in observations of steady magnetospheric convection (SMC) regimes [13]. On the other hand, the magnetotail experiences unsteady reconnection during substorms [7].

Both the first-principle modeling and the empirical reconstruction of magnetotail reconnection are very difficult to perform because of its multiscale nature. It links global reconfigurations of the nightside magnetosphere to kinetic processes on the scales of ion or even electron gyroradii that provide irreversibility for global reconfigurations. As a result, its kinetic particle-in-cell (PIC) simulations describing the full dynamics of electrons and ions (largely protons) and their self-consistent electromagnetic fields [14] are usually limited to the immediate X-line vicinity [15] and the moments after the X-line formation in global magnetohydrodynamic models [16], where the reconnection onset is provided due to numerical or ad hoc plasma resistivity. Moreover, it is very difficult to take into account that the magnetotail itself becomes multiscale prior to the reconnection onset. In-situ observations suggest that it may contain thin (ion-scale) current sheets (TCS) embedded into a thicker current sheet (CS) [17–21]. The latter may also be split in two current layers forming bifurcated TCSs [19, 22–24].

The major problem in the empirical reconstruction of magnetotail reconnection, common to all in-situ space observations, is the extreme sparsity of these observations with fewer than a dozen probes available at any moment. To solve this problem, it has recently been proposed to mine data in the multi-mission database covering many years of historical spaceborne magnetometer observations [25, 26]. It was found that such a data-mining (DM) method resolves the formation of embedded TCSs in the growth phase of substorms and their decay after the substorm onset. It also resolves the formation of the near-Earth X-lines during substorms [27]. Here we show that the DM approach allows one to resolve the formation of two different X-lines in the magnetotail during substorms. Moreover, it becomes possible to quantitatively assess their steadiness. We also show that PIC simulations

guided by the DM reconstruction of the magnetotail reproduce the formation of X-lines and reconnection regimes similar to those found in the DM analysis.

## 2 DATA MINING METHOD

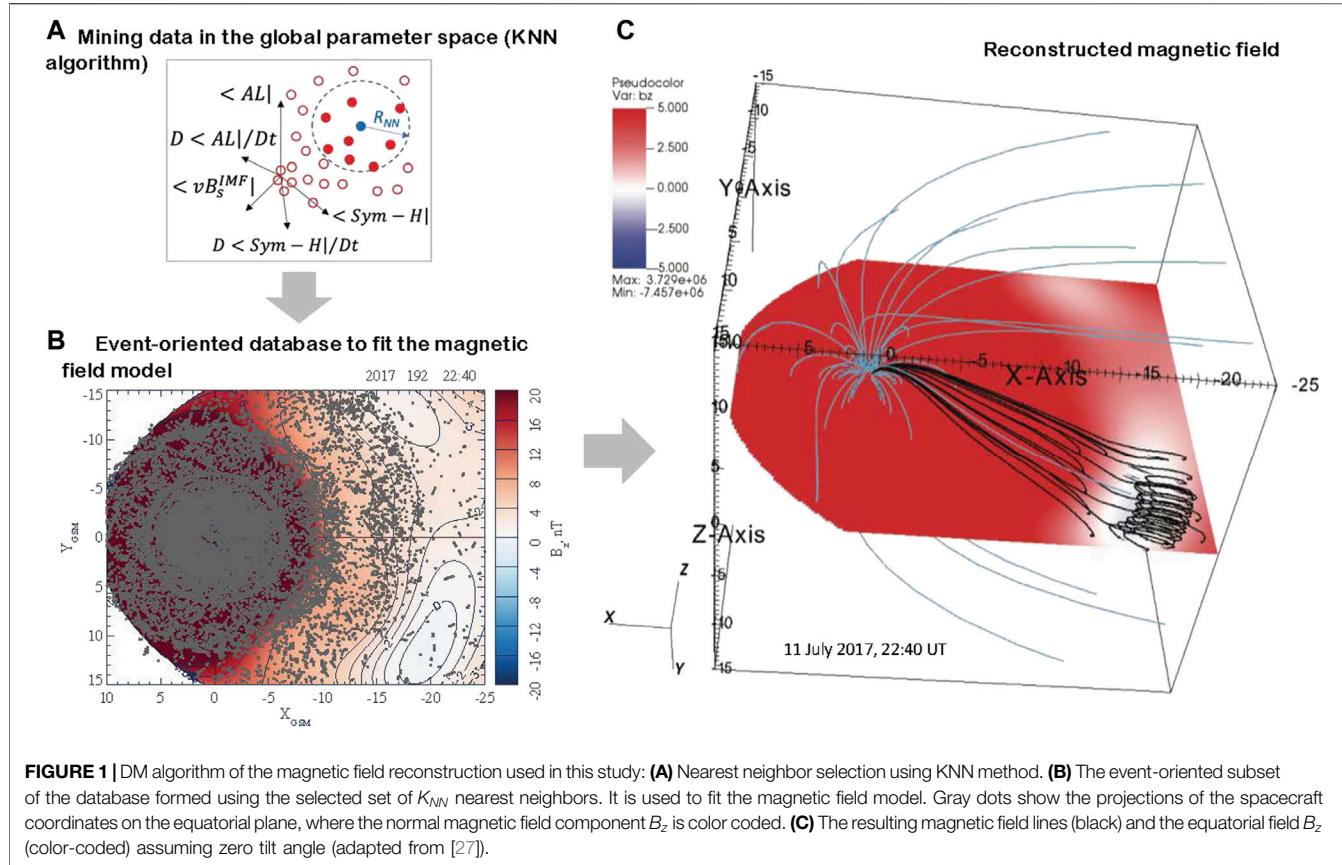
In the DM approach, the geomagnetic field is reconstructed using not only a few points of spaceborne magnetometer measurements available at the moment of interest, but also a much larger number of other measurements made at the  $K_{NN} \gg 1$  moments in the past. These moments called “the nearest neighbors” or NNs are similar to the event of interest in terms of similar values of the geomagnetic indices  $Sym-H$  and  $AL$ , their time derivatives and the solar wind input parameter  $vB_s^{IMF}$ . Here  $B_s^{IMF}$  is the southward component of the Interplanetary Magnetic Field (IMF):  $B_s^{IMF} = -B_z^{IMF}$  if  $B_z^{IMF} < 0$  and  $B_s^{IMF} = 0$  otherwise (The Geocentric Solar Magnetospheric coordinate system (GSM) coordinate system is used throughout this paper. Its origin is at the center of the Earth; the  $X$ -axis is directed toward the Sun; the  $y$ -axis is defined as the cross product of the GSM  $x$ -axis and the magnetic dipole axis, directed positive toward dusk; the  $z$ -axis is defined as the cross product of the  $x$ - and  $y$ -axes.) The large number of NNs is at the same time much smaller than the size of the database  $K_{DB} \gg K_{NN}$ . This allows one to fit with the NN subset a complex empirical magnetic field model [26], and at the same time, to make the model reconstructions sufficiently flexible to reflect the characteristic variations of the magnetosphere during storms and substorms.

This approach resembles very much the “lazy-learning” pattern recognition technique known as the K-nearest neighbor (KNN) learning [28, 29]. At the same time, our DM approach differs from conventional KNN regression methods, where both finding the NNs (“mining”) and regressions (model fitting) are made in the same space. Here, as is illustrated in **Figure 1**, we first detect NNs as a (sub)set of  $K_{NN}$  present and historical moments in similar phases of similar substorms. Their similarity (“neighborhood”) is quantified by the closeness of the corresponding global magnetospheric activity parameters and their time derivatives to their values at the moment of interest (**Figure 1A**). Then we use these  $K_{NN}$  moments to form an event-oriented subset of the original database of magnetic field observations (**Figure 1B**) and to fit our magnetic field model with this subset (**Figure 1C**).

The NN subset is formed by points  $G^{(i)} = [G_1(t_i), \dots, G_5(t_i)]$ ,  $i = 1, \dots, K_{NN}$ , in the 5-D space that are closest to the query point (moment of interest  $t_q$ )  $G^{(q)} = [G_1(t_q), \dots, G_5(t_q)]$  by the Euclidean metric

$$R_q^{(i)} = \sqrt{\sum_{k=1}^5 (G_k^{(i)} - G_k^{(q)})^2 / \sigma_{G_k}^2}. \quad (1)$$

where  $\sigma_{G_k}$  is the standard deviation of the component  $G_k$  and the coordinates  $G_1-G_5$  are defined by the formulae:



**FIGURE 1 |** DM algorithm of the magnetic field reconstruction used in this study: **(A)** Nearest neighbor selection using KNN method. **(B)** The event-oriented subset of the database formed using the selected set of  $K_{NN}$  nearest neighbors. It is used to fit the magnetic field model. Gray dots show the projections of the spacecraft coordinates on the equatorial plane, where the normal magnetic field component  $B_z$  is color coded. **(C)** The resulting magnetic field lines (black) and the equatorial field  $B_z$  (color-coded) assuming zero tilt angle (adapted from [27]).

$$G_1(t) = \langle Sym - H^* \rangle \propto \int_{-\Pi_{st}/2}^0 Sym - H^*(t + \tau) \cos(\pi\tau/\Pi_{st}) d\tau \quad (2)$$

$$G_2(t) = D\langle Sym - H^* \rangle / Dt \propto \int_{-\Pi_{st}/2}^0 Sym - H^*(t + \tau) \sin(2\pi\tau/\Pi_{st}) d\tau \quad (3)$$

$$G_3(t) = \langle AL \rangle \propto \int_{-\Pi_{sst}/2}^0 AL(t + \tau) \cos(\pi\tau/\Pi_{sst}) d\tau \quad (4)$$

$$G_4(t) = D\langle AL \rangle / Dt \propto \int_{-\Pi_{sst}/2}^0 AL(t + \tau) \sin(2\pi\tau/\Pi_{sst}) d\tau \quad (5)$$

$$G_5(t) = \langle vB_s^{IMF} \rangle \propto \int_0^{\tau_\infty} vB_s^{IMF}(t - \tau_\infty + \tau) \exp[(\tau - \tau_\infty)/\tau_0] d\tau \quad (6)$$

Here  $Sym - H^* = A \cdot Sym - H - B \cdot \sqrt{P_{dyn}}$  is the pressure-corrected  $Sym-H$  index [30],  $P_{dyn}$  is the solar wind dynamic pressure (in nPa) and the values of  $A$  and  $B$  are taken to be 0.8 and 13.0, respectively. The functions  $G_1$  and  $G_3$  in Eqs. 2, 4 describe weighted moving averages of the indices  $Sym-H$  and  $AL$  limited to their past values (see [25] for further details), while  $G_2$  and  $G_4$ , defined by Eqs. 3, 5, describe the corresponding smoothed time derivatives. Weighting in moving averages (2–5) is provided by the sine and cosine kernel functions and by the exponential function in Eq. 6. The averaging scaling parameters  $\Pi_{st} = 12$  hr and  $\Pi_{sst} = 2$  hr reflect the characteristic

storm and substorm scales. The parameter  $G_5$  defined by Eq. 6 describes the integral effect of the magnetic flux accumulation in the tail during the growth phase due to the dayside reconnection. Its scale  $\tau_0 = 0.5$  hr is selected based on observed values of a typical growth phase duration [31]. The selected upper integration limit in Eq. 6  $\tau_\infty = 6\tau_0$  corresponds to six e-folding times.

In the 5-D space of the binning parameters (2)–(6), the  $AL$  index and its time derivative ( $G_3, G_4$ ) determine the strength and phase of the substorm activity, because the  $AL$  index reflects the strength of the substorm electrojet [32]. These parameters may still be insufficient to capture the substorm growth phase, which is characterized by the accumulation of the magnetic flux in the tail lobes without any significant electrojet enhancement. To take this effect into account, we involve in the analysis the solar wind electric field parameter through the binning variable  $G_5$ . Furthermore, many substorms occur at the moments of the storm activity, which may substantially modify the substorm evolution of the magnetosphere [33]. To take these effects into account, we further extend the binning space at the expense of the parameters  $G_1$  and  $G_2$  reflecting the storm-time index  $Sym-H$  and its time derivative (to distinguish between main and recovery storm phases).

The conjecture that the substorm dynamics of the magnetosphere is coherent and hence the distribution of its magnetic field can be determined by a few control parameters had been formulated many years ago (e.g., [34], and refs. therein).

Later, the singular spectrum analysis of substorms [35] revealed that the mean-field dynamics of the magnetosphere can be described as a motion on a folded 2-D surface in a 3-D state space formed by the average  $AL$  index, average  $vB_s^{IMF}$  parameter and its average time derivative. An increase of the dimensionality through the  $Sym-H$  index and its time derivative, to take magnetic storms into account and to distinguish between their main and recovery phases, is consistent with the original DM-based storm-time model, TS07D [36]. The latter was also justified by the empirical relationship between the  $vB_s^{IMF}$  parameter and the  $Dst$  index, a 1-h time resolution analog of  $Sym-H$  [37]. Independent description of storms and substorms assuming their common solar wind driver is consistent with recent analysis of the storm-substorm relationship using a multivariate information-theoretic approach [38]. The further increase of state space dimensionality (e.g., using higher time derivatives of storm and substorm indices as well as the solar wind input parameter) is also possible (e.g., [25]). However, it was found [39] that the effect of higher dimensions often resembles the second-order phase transition fluctuations that require a probabilistic description of the magnetospheric states [40].

The database consists of  $K_{DB} = 3,668,101$  records of the magnetic field vector with 5 and 15-min cadence inside and outside  $5R_E$ , respectively, in archived data from IMP-8, Geotail, Polar, GOES-08, GOES-09, GOES-10, GOES-12, Cluster, THEMIS, Van Allen Probes and MMS missions covering more than two decades (1995–2017) of observations. The KNN subsets are selected using  $AL$ ,  $Sym-H$  and  $vB_s^{IMF}$  time series with 5-min cadence. At every moment  $t_q$ , the subset is found as  $K_{NN}$  points  $i = 1, \dots, K_{NN}$  satisfying the condition  $R_q^{(i)} < R_{NN}(K_{NN})$ , where  $R_q^{(i)}$  is defined by Eq. 1, as is illustrated in Figure 1A. Since the resulting magnetic field geometry is determined by the instantaneous KNN swarm of virtual probes, its time resolution is largely determined by the global parameter cadence. This is seen, for instance, from rapid substorm dipolarizations reproduced by the KNN method in [27] (Fig. 8i), when the  $B_z$  field increases from 3 to 10 nT in 5 min over a significant part of the magnetotail. It was found [26, 27] that the use of NN subsets with  $K_{NN} \sim 32,000$  and the magnetic field model parameters specified below provides both sufficient selectivity of the model, which allows one to distinguish different substorm phases, and the high spatial resolution to resolve the distinctive features of the magnetospheric morphology in these phases, such as TCSs (and their buildup and decay), flux accumulation regions and X-lines. Smaller  $K_{NN}$  values were found to cause overfitting.

At every moment of interest  $t_q$ , the  $K_{NN}$  subset of the database, whose elements neighbor  $t_q$  in the state and input space of the magnetosphere by the metric (1), is used to fit the geomagnetic field model SST19 [26]. Since  $K_{NN} \gg 1$ , its architecture can be made quite complex and flexible (compared, for instance, with the event-oriented models using only a few points of data available at the moment of interest [41, 42]) to capture key features of the substorm current system. In fact, we only assume that the magnetic field is formed by two major current systems inside the magnetosphere, equatorial and field-aligned currents, whose contributions are presented as sums of basis

functions with the corresponding amplitude coefficients (more general 3-D expansions of the magnetic field using radial basis functions were considered in [43]). Moreover, to describe the multi-scale structure of the equatorial currents, including the formation of embedded and bifurcated TCSs, these currents are described by two independent expansions:

$$B^{(eq)}(\rho, \phi, z) = B^{(eq)}(\rho, \phi, z; D) + B^{(eq)}(\rho, \phi, z; D_{TCS}) \quad (7)$$

where  $(\rho, \phi, z)$  are cylindrical coordinates in a system with the origin at the center of the Earth and the  $z$  axis normal to the equatorial plane. They represent the magnetic field of thick and thin current sheets with the same structure determined by the approximate solution for the magnetic field of an arbitrary distribution of equatorial currents [44] with different thickness parameters  $D$  and  $D_{TCS}$  to be derived from the fitting with the NN subset. Each expansion is a finite-sum approximation of an integral solution of the Ampère's equation for the magnetic field of an infinitely thin CS ( $D = 0$ ) above and below the equatorial plane  $z = 0$  by separation of variables. Tsyganenko and Sitnov [44] showed that the sum consists of  $N$  azimuthally symmetric radial expansions and  $2M$  angular Fourier harmonics (even and odd parity in  $\phi$ ) with the total number of  $N + 2M \cdot N$  elements. The basis functions of the solution for the vector potential with an infinitely thin CS contain factors like  $\exp(-k|z|)$ . Their regularization comes from assuming the finite CS half-thickness  $D$  and it can be provided by replacing the function  $|z|$  by the smooth function  $\zeta = \sqrt{z^2 + D^2}$ . The radial expansions include Bessel functions and they can be exemplified by the azimuthal component  $A_\phi$  of the vector potential corresponding to the azimuthally symmetric group of basis functions  $B_{0n}^{(s)}$ :  $(A_\phi)_{0n}^{(s)} = J_1(k_n \rho) \exp(-k_n \sqrt{z^2 + D^2})$ , where  $J_1$  is the Bessel function of the first order,  $k_n = n/\rho_0$ , and  $\rho_0$  is the radial scale, corresponding to the largest mode in the radial expansion.

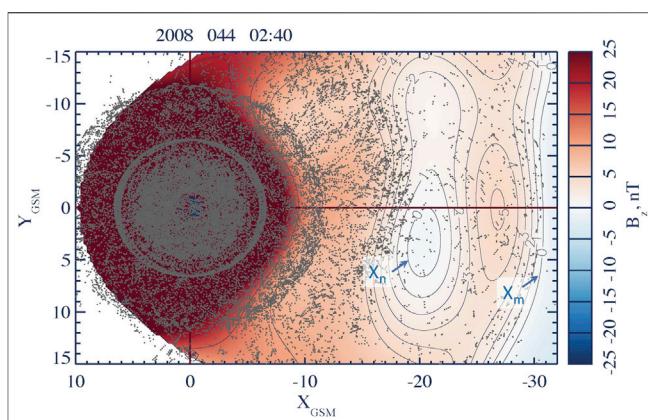
The parameters  $\rho_0$ ,  $N$  and  $M$  are fixed because they determine the adopted resolution of the expansions in Eq. 7. Other parameters, such as the weights of individual radial and azimuthal harmonics, as well as the CS thickness parameters  $D$  and  $D_{TCS}$ , are determined from fitting the model to data. In particular, to distinguish between thick current sheets and TCS, we impose the condition  $D_{TCS} < D_{TCS}^{(\max)} = 1R_E$ . The latter value is intermediate between the observed thick and thin current sheet values [19] and it does not significantly constrain the specific values of  $D$  and  $D_{TCS}$  inferred from data. Thus, the spatial resolution of such an expansion is determined by the number of terms in expansions (7) and can be increased to any desired level, commensurate with the data density. To take the global scaling of currents due to variations of the solar wind dynamic pressure  $P_{dyn}$  into account, each amplitude coefficient in expansions (7) is further expanded in two parts, one of which is constant and another is a linear function of  $\sqrt{P_{dyn}}$ . The equatorial expansion has several other nonlinear parameters to take into account global deformations of the tail CS along the dawn-dusk direction and arising from the finite dipole tilt angle, which are described in [44].

Another major group of currents are the field-aligned currents (FACs), connecting the ionosphere with the magnetopause and

the tail CS. It is described in SST19 using a similar system of finite current elements [45], sufficiently flexible to reproduce the spiral FAC structure at low latitudes [46] whose night-side part is likely associated with the Harang discontinuity [47]. Each element of the FAC system is described as the magnetic field of two deformed conical surfaces corresponding to Region 1 (R1) and Region 2 (R2) FACs [48]. The size of each system is an adjustable parameter, while their azimuthal distribution is controlled by the relative contributions of two groups of basis functions with odd and even symmetry due to factors  $\sin(l\phi)$  and  $\cos(l\phi)$ , ( $l = 1, 2, \dots$ ). The first group represents the main part of the FAC system, in which the dusk-side currents have the same magnitude but opposite direction to those at dawn, while the second group has an even distribution of currents with respect to the noon-midnight meridian plane, which allows one to model the azimuthal rotation of the FACs.

Originally two groups of such FAC elements were proposed in [44] to describe R1 and R2 systems in their DM-based storm-time model, TS07D [36]. Later, it was proposed [45] to use more elements similar to the original TS07D FAC basis functions, shifted in latitude to describe more complex FAC distributions. Eventually, Stephens and coauthors [26] showed that the FAC system can be described with many details important for substorm reconstructions, including the Harang discontinuity and the substorm current wedge [49], with the following set of elements. It consists of  $N_{FAC} = 16$  basis functions with the first two Fourier harmonics ( $l = 1, 2$ ) for R1 and R2, as well as their latitude-shifted clones. Each element in equatorial and FAC expansions is independently shielded (has its own subsystem of Chapman-Ferraro-type currents at the magnetopause).

Thus, the resulting DM algorithm, which links the SST19 model with KNN binning, represents a typical “gray box” model combining empirical algorithms with physics-based constraints [50]. As is shown in [26, 27], it reproduces the multiscale CS thinning process with the formation of an ion-scale TCS ( $D_{TCS} \ll 1R_E$ ) inside a much thicker CS ( $D > 2R_E$ ), which takes place in the substorm growth phase and causes stretching of the tail magnetic field lines in the antisunward direction. In particular, Figs. 11a–11c in [27] show the current distribution in the equatorial plane during the growth phase of the 13 February 2008 substorm discussed below. The corresponding current distributions in the meridional plane presented in their Figs. 12a–12c reveal the multiscale CS structure with an ion-scale TCS embedded into a thick CS halo. The peak TCS current density  $\sim 8 \text{ nA/m}^2$  is consistent with in-situ Cluster observations (see, for example, Figs. 2–4 and 9 in [19]). Further quantitative analysis made in [27] showed that while the TCS thickness in DM reconstructions remains approximately constant  $D_{TCS} \approx 0.2R_E$  (Fig. 8e), consistent with Cluster and THEMIS observations [19, 20, 51], their strength (measured as the TCS contribution to the total tail current) changes drastically with the substorm phase (Fig. 12d). At the same time, the contribution of the TCS

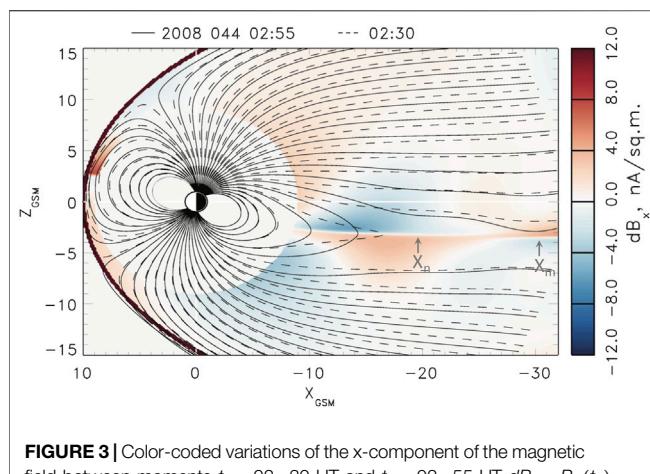


**FIGURE 2** | Two X-lines,  $X_n$  and  $X_m$  resolved in the equatorial distribution of the magnetic field  $B_z$  (using 0-degree tilt angle for the sake of visualization simplicity) as earthward parts of the contours  $B_z$  in case of the 13 February 2008 substorm. The format of this figure is similar to that of **Figure 1B**. The projections of the spacecraft coordinates on the equatorial plane (gray dots) show that the NN subset of data for the moment considered is sufficient to resolve both X-lines.

to the total tail current is relatively small ( $\approx 1/6$ ). It is worth noting that, as is seen from the comparison of Figs. 10a and 11a in [27], the extended TCS forms earthward of the flux accumulation region ( $B_z$  hump).

The DM SST19 algorithm also describes the magnetic field dipolarization in the expansion phase (see, for example, Fig. 4 in [26]) with the formation of a substorm current wedge seen as a *curl* of the difference between the expansion and growth phase magnetic field distributions ([26], Fig. 10). The disappearance of TCS after the dipolarization can be seen, for instance, from the comparison of Fig. 12d in [27] with other panels in Fig. 12. It can also be seen from their Fig. 8f, where the relative strengths of thin and thick CSs are quantified by integrating the current density over the regions  $|z| < 1R_E$  and  $|z| < 5R_E$ .

The new DM reconstruction has a characteristic property of machine learning algorithms [29, 52]: given more data in the database it may provide more details about the magnetospheric structure and evolution. In particular, as is shown in [27], with adding to the database first two years of the MMS mission data (2016–2017), it becomes possible for the same SST19 model with the parameters  $(M, N) = (6, 8)$  and  $K_{NN} = 32,000$  to resolve more details of the magnetotail structure and evolution. In particular, the 2017 MMS data help resolve the X-lines forming largely beyond  $20R_E$ , where the pre-MMS database had a substantial drop in the occurrence rate distribution ([27], Fig. 1). This is seen in particular, from the comparison of the SST19 validation using THEMIS data beyond  $20R_E$  in Fig. 2e of [26] with THEMIS validation in a similar region in Fig. S6 of [27]. The former reveals clear signatures of overfitting while the latter does not. In spite of a relatively small total number of the new MMS data, they fill the main gap in the existing database distribution ([27], Fig. 1) and thus become particularly important in solving the overfitting problem.



**FIGURE 3** | Color-coded variations of the x-component of the magnetic field between moments  $t_0 = 02:30$  UT and  $t_2 = 02:55$  UT  $dB_x = B_x(t_2) - B_x(t_0)$  in the noon-midnight meridional plane with overplotted magnetic field lines (solid lines for the moment  $t_2$  and dashed lines for the moment  $t_0$ ) for the 13 February 2008 substorm. The approximate location of the X-lines  $X_n$  and  $X_m$  is marked by gray arrows. Magnetic field lines start from the ionosphere at  $60^\circ$  with  $2^\circ$  step in latitude. White disks  $1R_E < r < 9R_E$  in panels **(B)** and **(C)** mask magnetic field variations in the inner magnetosphere.

### 3 13 FEBRUARY 2008 SUBSTORM: STEADY AND UNSTEADY X-LINES

In this section we describe the global structure and dynamics of reconnection on the example of a relatively small and short substorm (13 February 2008 02:05–02:55 UT) considered earlier in [27] with the reconstruction parameters  $(M, N) = (6, 8)$ ,  $K_{NN} = 32,000$  and  $N_{FAC} = 16$ . The latter analysis is extended here by increasing the maximum radial distance of the spacecraft data used in the reconstruction from  $31R_E$  to  $35R_E$  (largely, due to IMP8 data). **Figure 2** shows the equatorial magnetic field distribution at the moment 02:40 UT in the expansion phase of this substorm. It reveals the formation of two X-lines  $X_n$  and  $X_m$  in the near-Earth ( $x \approx -20R_E$ ) and midtail ( $x < -27R_E$ ) regions, respectively. They are seen as earthward parts of the  $B_z = 0$  contours in the distribution of the equatorial north magnetic field component  $B_z$  in **Figure 2** and they are additionally marked by blue arrows. The tailward parts of the  $B_z = 0$  contours represent O-lines.

This global X-line reconstruction is quite unique. In fact, because of the extreme sparsity of in-situ space observations, such reconstructions were not available before the machine learning era. Earlier, Nagai et al. [53, 54] described the location and the dawn-dusk extension of X-lines using single-point observations. More recently, the reconstructions of the X-line vicinity were made by processing multi-probe MMS data with Grad-Shafranov [55] and polynomial [56] techniques. However, these were still very local reconstruction, largely limited to the size of the MMS tetrahedron ( $<30$  km). Here we demonstrate for the first time how the DM approach based on the KNN algorithm resolves simultaneously two X-lines in the near-Earth and midtail regions.

The formation of transient near-Earth X-lines, which was proposed by Hones [7] as a mechanism of substorms, has

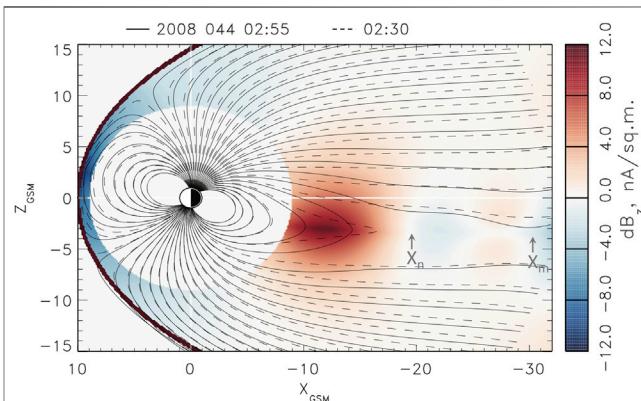
been discussed since that time in many studies, including correlated multi-probe and remote sensing analyses (see, for instance [57, 58], and references therein). At the same time, persistent reconnection in the midtail around  $30R_E$  follows from THEMIS and ARTEMIS statistics of traveling compression regions [59, 60]. However, neither the co-existence of the second, midtail X-line  $X_m$  with  $X_n$  nor its relatively steady reconnection, as suggested by Dungey's convection cycle [5], have ever been demonstrated. Here we not only resolve two X-lines in the tail but also propose a method to quantify their steadiness.

This can be done using the Faraday's law, which in the 2-D picture of reconnection takes the form

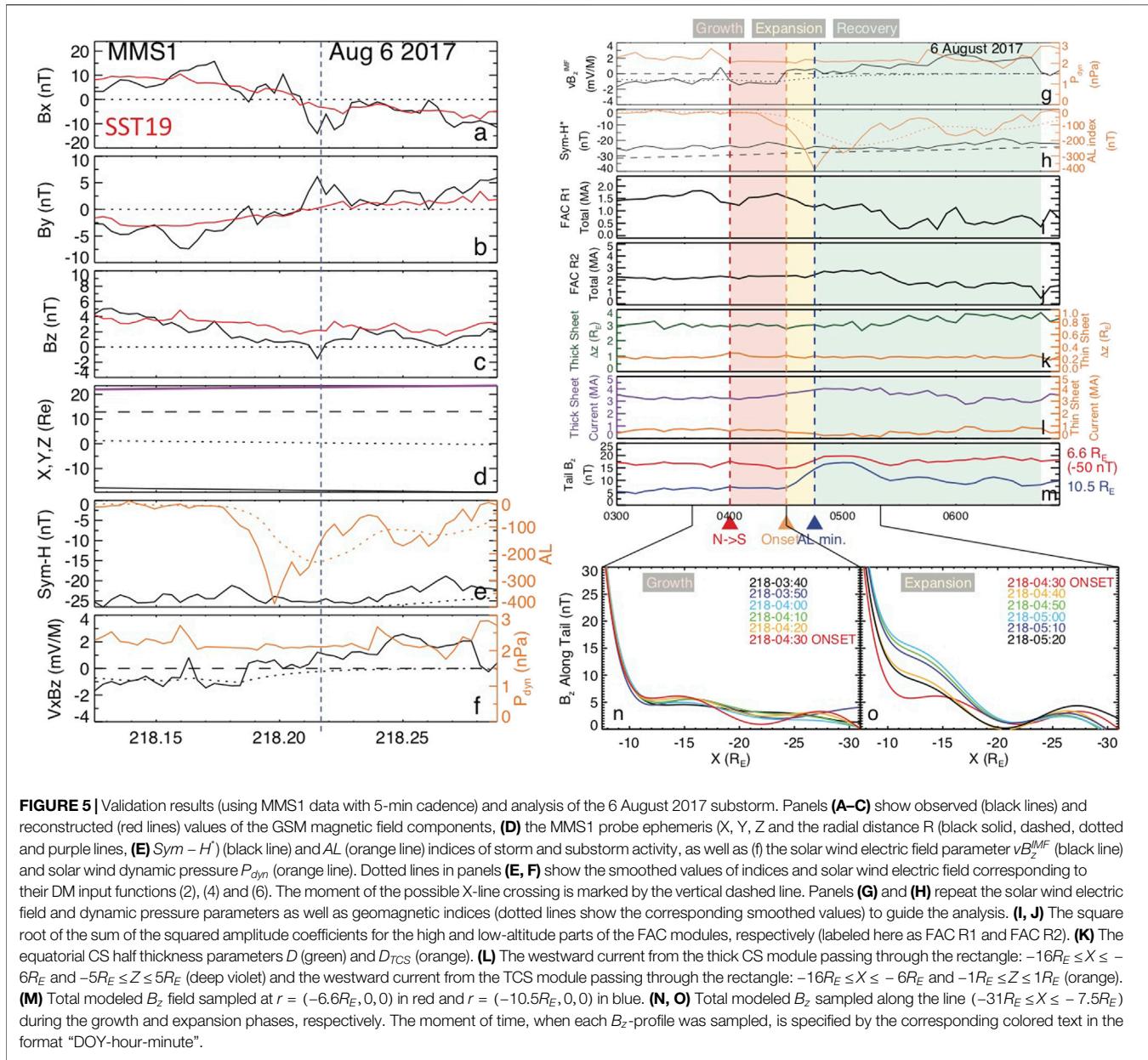
$$\partial E_y / \partial x = -\partial B_z / \partial t, \partial E_y / \partial z = \partial B_x / \partial t, \quad (8)$$

It suggests that the temporal variations of  $B_x$  and  $B_z$  magnetic field components determine the spatial gradients of the dawn-dusk (reconnection) electric field. If the magnetic field varies slowly, the corresponding reconnection electric field is broadly distributed over the whole reconnection region. This justifies the concept of the reconnection rate, one of the key global parameters characterizing steady reconnection regimes [61–63]. Reconnection can also be unsteady with the electric field being localized in space and the magnetic field changing in time consistent with (8). For example, Sitnov and Swisdak [64] showed reconnection regimes with the electric field localized near dipolarization fronts (DFs) [65–67] with their values strongly exceeding the steady reconnection values. Localization of the dawn-dusk component of the electric field near DFs was later confirmed by Cluster [68] and THEMIS [69] observations.

The noon-midnight meridional maps of magnetic field lines presented in **Figures 3, 4** reveal interesting distinctions of magnetic reconnection in the mid tail region and closer to Earth ( $X_m$  and  $X_n$  vicinities) in this substorm. As is seen from the comparison of solid and dashed field lines in these figures, reconnection near  $X_n$  is accompanied by strong changes of the



**FIGURE 4** | Color-coded variations of the z-component of the magnetic field between moments  $t_0 = 02:30$  UT and  $t_2 = 02:55$  UT  $dB_z = B_z(t_2) - B_z(t_0)$  in the noon-midnight meridional plane with overplotted magnetic field lines in the format similar to **Figure 3** in case of the 13 February 2008 substorm.



**FIGURE 5 |** Validation results (using MMS1 data with 5-min cadence) and analysis of the 6 August 2017 substorm. Panels (A–C) show observed (black lines) and reconstructed (red lines) values of the GSM magnetic field components, (D) the MMS1 probe ephemeris ( $X$ ,  $Y$ ,  $Z$ ) and the radial distance  $R$  (black solid, dashed, dotted and purple lines), (E)  $Sym - H^*$  (black line) and  $AL$  (orange line) indices of storm and substorm activity, as well as (F) the solar wind electric field parameter  $V_{Bz}^{IMF}$  (black line) and solar wind dynamic pressure  $P_{dyn}$  (orange line). Dotted lines in panels (E, F) show the smoothed values of indices and solar wind electric field corresponding to their DM input functions (2), (4) and (6). The moment of the possible  $X$ -line crossing is marked by the vertical dashed line. Panels (G) and (H) repeat the solar wind electric field and dynamic pressure parameters as well as geomagnetic indices (dotted lines show the corresponding smoothed values) to guide the analysis. (I, J) The square root of the sum of the squared amplitude coefficients for the high and low-altitude parts of the FAC modules, respectively (labeled here as FAC R1 and FAC R2). (K) The equatorial CS half thickness parameters  $D$  (green) and  $D_{TCS}$  (orange). (L) The westward current from the thick CS module passing through the rectangle:  $-16R_E \leq X \leq -6R_E$  and  $-5R_E \leq Z \leq 5R_E$  (deep violet) and the westward current from the TCS module passing through the rectangle:  $-16R_E \leq X \leq -6R_E$  and  $-1R_E \leq Z \leq 1R_E$  (orange). (M) Total modeled  $B_z$  field sampled at  $r = (-6.6R_E, 0, 0)$  in red and  $r = (-10.5R_E, 0, 0)$  in blue. (N, O) Total modeled  $B_z$  sampled along the line  $(-31R_E \leq X \leq -7.5R_E)$  during the growth and expansion phases, respectively. The moment of time, when each  $B_z$ -profile was sampled, is specified by the corresponding colored text in the format “DOY-hour-minute”.

magnetic field geometry, especially, earthward of that  $X$ -line, while near  $X_m$  the geometry barely changes, which is seen particularly well in the lobe region. The color-coded variations of the  $z$ - and  $x$ -components of the magnetic field between moments  $t_0 = 02:30$  UT and  $t_2 = 02:55$  UT  $dB_{x,z} = B_{x,z}(t_2) - B_{x,z}(t_0)$  in the same noon-midnight meridional plane in Figures 3, 4 quantify these steady and unsteady reconnection regimes.

The difference in  $dB_{x,z}$  values in regions  $x \approx -20R_E$  and  $x \approx -31R_E$  seen from Figures 3, 4 suggests that the reconnection process near  $X_m$  is more steady-state than near  $X_n$ . The unsteady nature of the near-Earth reconnection is particularly well seen from  $B_z$  variations earthward of  $X_n$  in Figure 4. Moreover, the analysis of the equatorial  $B_z(x)$  profiles

with the 5-min cadence provided in Fig. 8i of [27] shows that the main part of the  $B_z$  changes earthward of  $X_n$  shown in Figure 4 occurs in the 5-min interval between 02 : 35 UT and 02 : 40 UT. Furthermore, the analysis of the magnetic flux redistribution in the lobes made in [27] gives an estimate of the average electric field in the steady-state reconnection region  $E_y \sim 0.01v_A B_0/c$  for  $B_0 = 40$  nT and  $v_A = 1,000$  km/s, consistent with the theoretical estimates that impose the upper limit for the reconnection rate  $E_y/(v_A B_0/c) \sim 0.1$  [61, 62, and refs. therein] or  $\approx 0.2$  [63]. Therefore, one can expect the reconnection near  $X_m$  to be steady and its electric field homogeneous in space, whereas near  $X_n$  to be more transient and structured. Below we show that a similar combination of steady and unsteady reconnection regions can be reproduced in PIC simulations of weakly driven

magnetotail equilibria with some of the observed growth phase features.

## 4 6 AUGUST 2017 SUBSTORM: COMPLEX RECONNECTION PICTURE RESOLVED USING ADVANCED DM METHOD

In this section we consider another substorm event with a more complex structure of dipolarizations and X-lines. It occurred on 6 August 2017 and is also interesting because of a possible X-line crossing detected by the MMS mission. Its signatures were the  $B_z$  reversal (**Figure 5C**), the ions bulk flow reversal and the large downward electron bulk flow velocity (not shown). At the same time, at the moment of the reversal the  $|B_x|$  and  $B_y$  magnetic field components were relatively large (**Figures 5A,B**) so that the total magnetic field exceeded 10 nT. We reconstruct this event using an advanced version of the KNN algorithm where the statistical weights of NNs depend on their proximity to the event of interest (e.g., [29]). In this algorithm, the model magnetic field  $B^{(\text{mod})}$  is determined by minimizing the RMS of its deviation from observations  $B^{(\text{obs})}$

$$M_{\text{err}}^{(\text{NN})} = \sqrt{\sum_{j \in S_{\text{NN}}} \sum_{i=x,y,z} w_j w_{(0)}(r) [B_i^{(\text{mod})}(r^{(j)}) - B_i^{(\text{obs})}]^2}, \quad (9)$$

where  $S_{\text{NN}}$  is a set of  $K_{\text{NN}}^{(B)}$  magnetometer measurements of the magnetic field components  $B_i^{(\text{obs})}$  with ephemeris  $r^{(j)}$ , corresponding to the selected set of  $K_{\text{NN}}$  nearest neighbors;  $w_{(0)}$  is the original weighting factor, which is a function of the real-space distance  $r$  from Earth, introduced in [44] to mitigate the spatial inhomogeneity of observations, especially at geosynchronous orbit. A distinctive feature of the weighted KNN algorithm is that each term in the sum in **Eq. 9** has now an additional weighting factor

$$w_j = \exp \left[ - \left( R_q^{(j)} / \sigma R_{\text{NN}} \right)^2 / 2 \right], \quad (10)$$

Here  $R_q^{(j)}$  is the distance (1) of the corresponding NN from the query point  $q$  and  $R_{\text{NN}}$  is the radius of the sphere containing NNs in the binning space ( $G_1, \dots, G_5$ ). When  $\sigma \gg 1$ , all distance-modulated weights  $w_j \approx 1$  and NNs are not weighted. In contrast, for  $\sigma < 1$ , the new weights  $w_j$  are well modulated within the sphere  $R_q^{(j)} < R_{\text{NN}}$ . This increases the statistical weight of measurements that were made at the more similar state and input conditions of the magnetosphere, according to the metric (1).

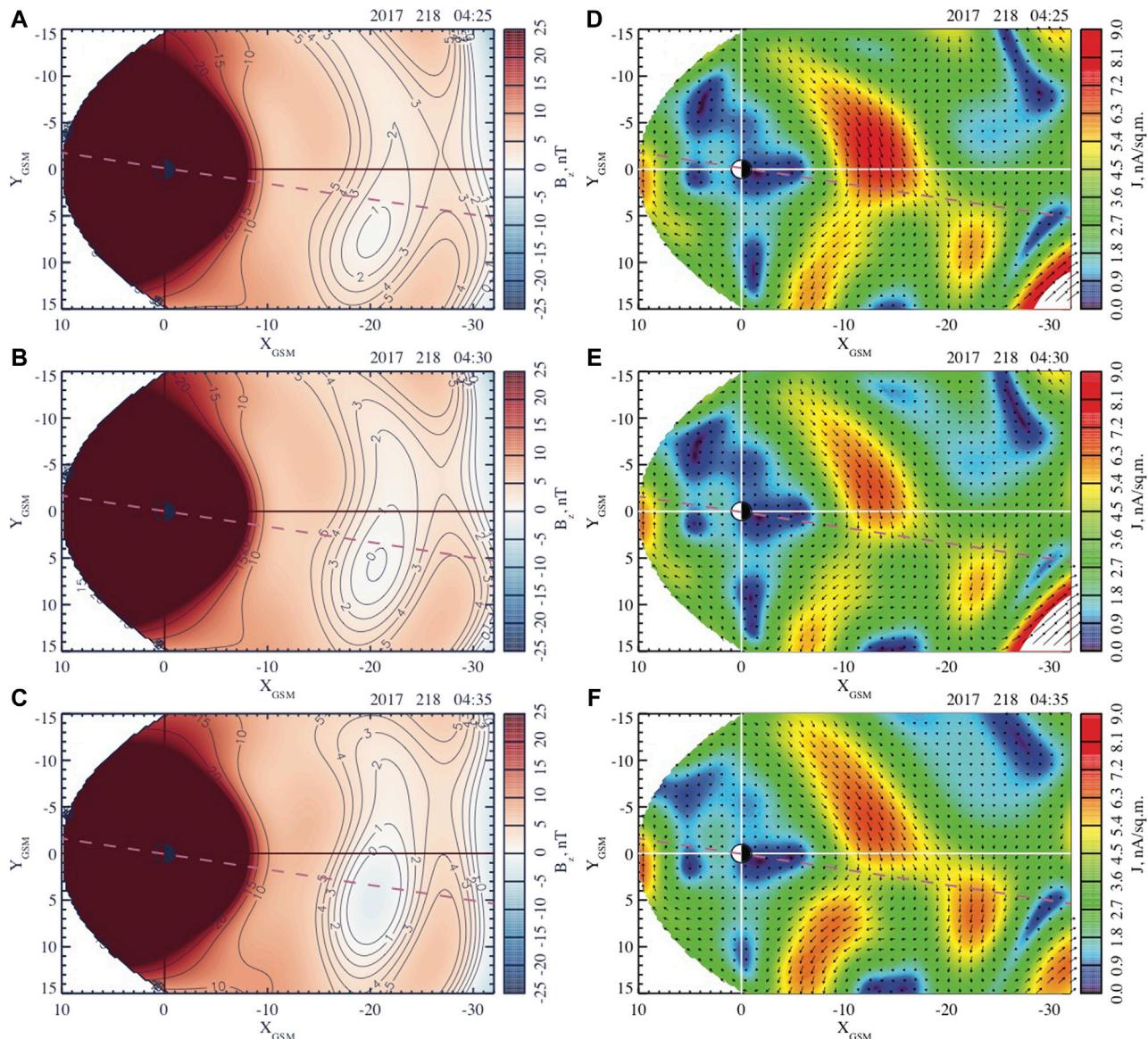
The weighted KNN approach is shown to result in better sensitivity of the model to variations of the magnetospheric state (e.g., storm or substorm phase) by using effectively much smaller numbers of NNs without overfitting [70]. Below we provide the DM reconstruction results with the parameters similar to those used in the previous section and with the weighting factor  $\sigma = 0.5$ . Validation results for this event using the MMS1 probe data are presented in the left panels of **Figure 5** and they show a reasonable agreement, especially for the  $B_z$  component, where it does not exceed  $\sim 2$  nT.

The reconstruction summary for this substorm in the format used earlier in [26, 27] is presented in the right panels of **Figure 5**. Following [26], we consider the growth phase starting from the first point with  $B_z^{\text{IMF}} < 0$  in the 5-min cadence series (vertical red dashed line corresponding to  $t = 04:00$  UT). The onset time 04:20 UT (vertical orange dashed line) is selected because of the strong change of the negative slope of the  $AL(t)$ . The start of the recovery phase (23:40 UT, vertical blue dashed line) corresponds to the minimum of the  $AL$  index. The recovery phase is postulated to end when the  $AL > -25$  nT, in accordance with [26, 27].

**Figures 5H–J** show weak storm activity: small and constant values of  $-Sym - H^*$  and low-latitude field-aligned currents FAC R2. In the expansion phase (yellow zone) the amplitude of the TCS (**Figure 5L**) decreases, consistent with the earlier DM analyses [26, 27] (with small variations of the thickness parameters  $D$  and  $D_{\text{TCS}}$ , according to **Figure 5K**), while the equatorial magnetic field in the near-Earth tail (**Figure 5M**) increases making the magnetic field more dipole-like. At the substorm onset, the evolution of the equatorial magnetic field along the midnight meridian (red lines in **Figures 5N, O**) reveals wavy perturbations similar to the tearing mode (e.g., Fig. 6.2.9 in [71]). However, their wavelength is rather macroscopic, in contrast to the electron- or ion-scale tearing modes discussed earlier in theory and kinetic simulations of the magnetotail reconnection onset ([72, 73] and refs. therein). Further it results in the formation of new X-lines (**Figure 5O**) whose structure and evolution are better seen in **Figures 6, 7**.

An interesting feature of this event is that the magnetic field dipolarization in the expansion phase has two sub-phases: 04:20–04:35 UT and 04:40–04:55 UT. Indeed, **Figures 6, 7**, which describe the evolution of the equatorial magnetic field and current, reveal two successive reconfigurations developing in the premidnight and postmidnight sectors. **Figure 6** shows that during the first dipolarization a new X-line forms at  $x \sim -20R_E$  along with the pre-existing X-line near  $x \sim -30R_E$ . As it is shown in **Figure 8**, the used NN subsets are sufficiently extended over the tail to resolve both X-lines. As is seen from **Figures 6D–F**, the equatorial current during this dipolarization becomes bifurcated.

The second dipolarization described in **Figure 7** causes stronger and more global changes of the near-Earth magnetic field (regions  $R \lesssim 15R_E$  in **Figures 7A–C**). It also causes not only the formation of new flux ropes in the postmidnight sector but the azimuthal extension of the region of the depressed or even reversed equatorial magnetic field. According to **Figures 7D–F**, this is accompanied by the reduction of the equatorial current density. To quantify these processes, we integrated the equatorial field  $B_z$  over arcs similar to dashed blue lines in **Figure 7A** from the dawn to dusk magnetopause boundaries. Each arc represents a part of the circle with the center  $(x, y) = (3R_E, 0)$  (the shift is used to avoid integration over whole circles within the magnetopause). As was argued in [27], the distribution along the tail of the corresponding integral parameter  $\text{Int}(B_z) = \int B_z ds$  (where  $ds$  is the arc length element) may be a good proxy of the magnetic flux evolution in the closed field line region of the magnetotail. The distributions of  $\text{Int}(B_z)$  along the tail shown in **Figure 9** as functions of the arc's

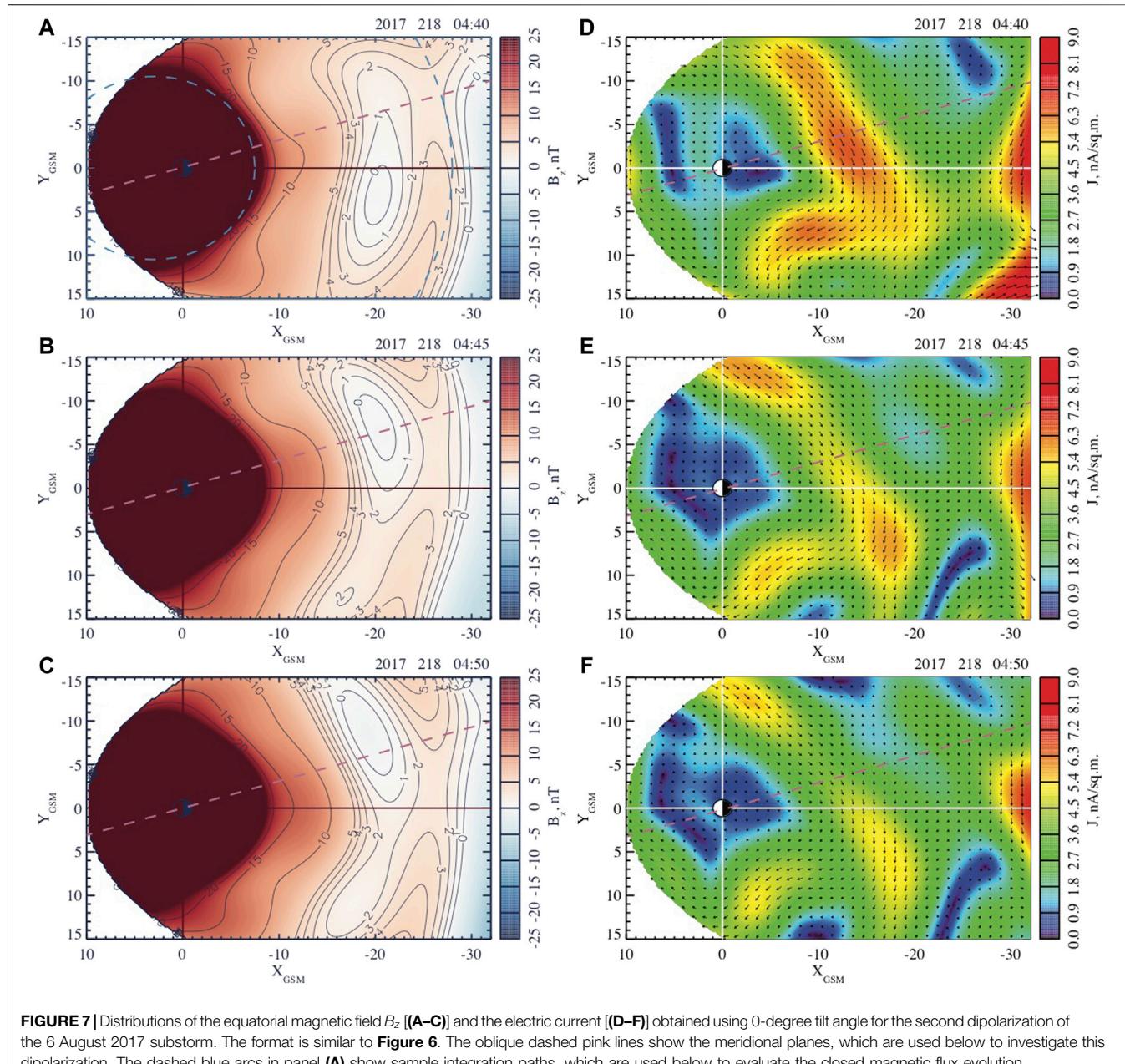


**FIGURE 6** | Distributions of the equatorial magnetic field  $B_z$  [(A–C)] and the electric current [(D–F)] obtained using 0-degree tilt angle for the first dipolarization of the 6 August 2017 substorm. The arrows are vectors of the equatorial current density whose absolute value is color coded. The oblique dashed pink lines show the meridional planes, which are used below to investigate this dipolarization.

most tailward value of  $x$  indicate that the main flux redistribution in this substorm is provided by the second dipolarization. They also suggest that the main part of the near-Earth dipolarization is provided by the redistribution of magnetic flux in the closed field line region. Indeed, the area under these curves is now magnetic flux. According to **Figure 9B**, the increase in flux during the second dipolarization in the region  $10 - 17R_E$  is roughly equivalent to the decrease in flux at  $17 - 26R_E$ . If the dipolarization were provided by an increase of the lobe field reconnection at  $\gtrsim 30R_E$  that would be seen as a net increase of flux within  $\sim 30R_E$ .

To further investigate two dipolarizations occurring during this substorm, we present in **Figures 10, 11** the meridional cuts of the cross current and in-plane magnetic field components  $B_{x',x''}$  and

$B_z$  in the planes marked by dashed lines in **Figures 6, 7** ( $x'_{GSM}$  and  $x''_{GSM}$  are the coordinates along the dashed lines in **Figures 6, 7**). These figures show processes similar to the 13 February 2008 dipolarization and shown in Figs. 12 and 13 in [27] as well as in **Figures 3, 4**. In particular, both dipolarizations reveal stronger variations of  $dB_{x,x'}$  and  $B_z$  magnetic field components in the near-Earth reconnection region (around  $X_n$ ) compared to the midtail one (around  $X_m$ ). The differences between the magnetic field variations in **Figures 10, 11** and those in **Figures 3, 4**, such as for instance, different relative phases of  $dB_x$  and  $dB_z$  variations can be explained by the larger time difference used in case of the 13 February substorm to calculate these variations. In fact, **Figures 3, 4** describe the magnetic field variations during the 25-min long



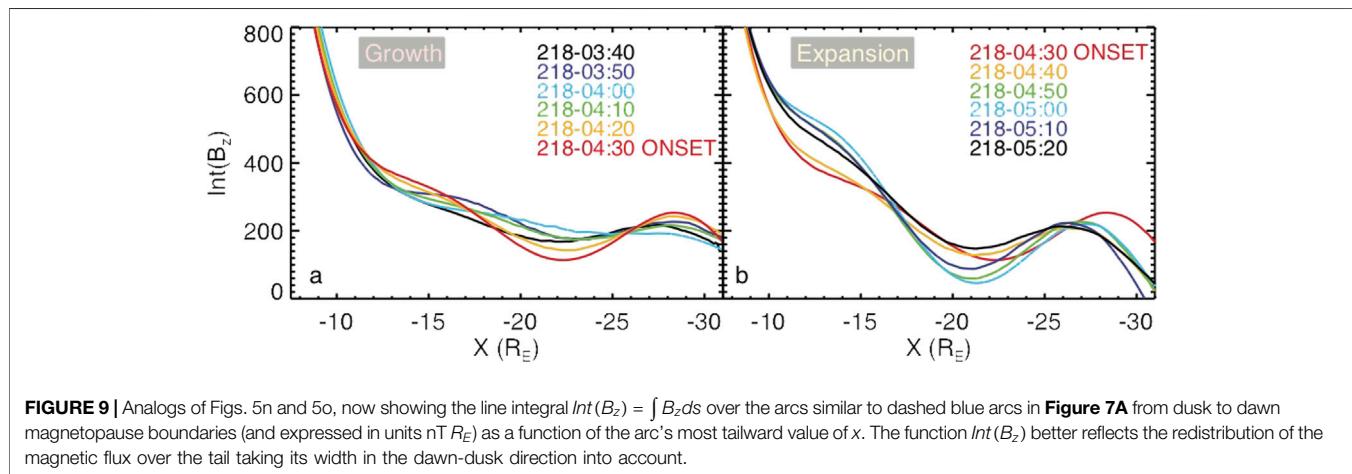
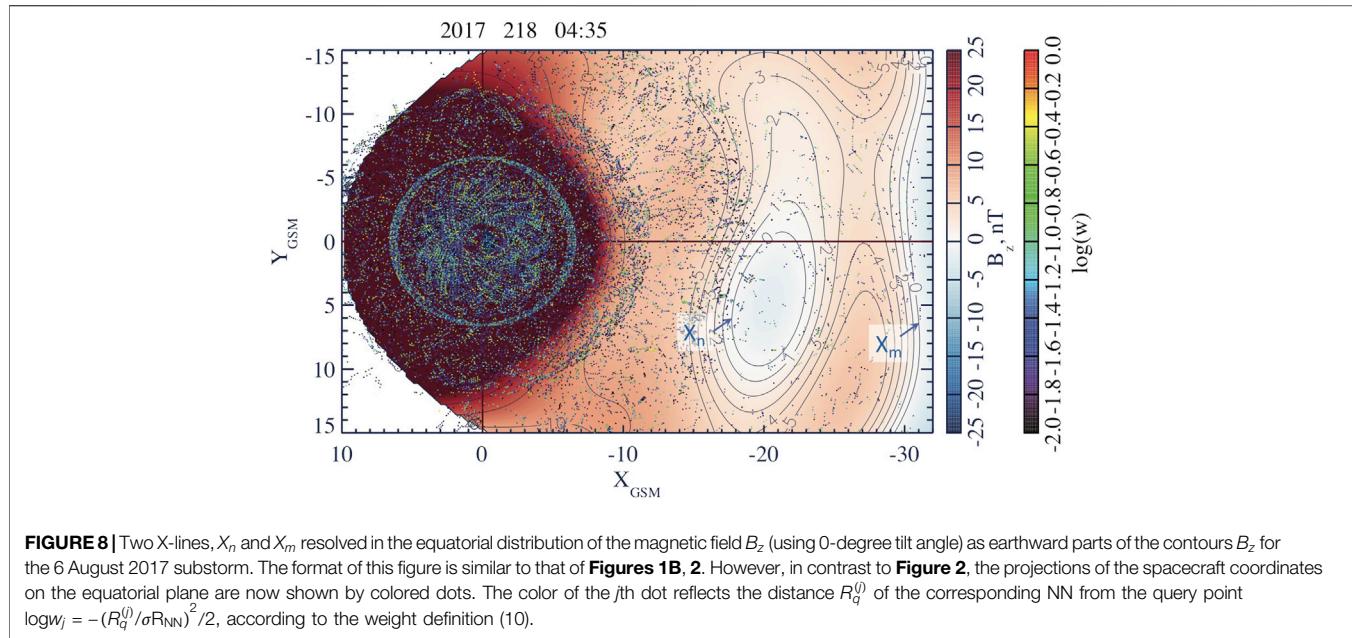
interval covering the whole expansion phase of that substorm, whereas **Figures 10, 11** describe 15-min long partial dipolarizations that constitute the more complex tail reconfiguration during the 6 August 2017 substorm.

As one can seen from the comparisons of **Figures 10A,B**, the first dipolarization is relatively weak, and it does not cause any significant flux redistribution, according to **Figure 9B**. During the first dipolarization, the magnetic field variations near  $X_n$  (**Figures 10C,D**) are confined to the region  $x > -27R_E$  and  $-2R_E \leq z \leq 4R_E$ . The near-Earth X-line during this dipolarization forms in the center of the TCS, which extends

from  $-28R_E$  to  $-9R_E$  (**Figure 10A**). It only moderately redistributes its current density (**Figure 10B**).

In contrast, during the second dipolarization, the (already shorter, less than  $\sim 10R_E$  in the radial extension) TCS disappears (**Figures 11A,B**), the near-Earth X-line forms at its tailward end and these processes are associated with a significant flux redistribution shown in **Figure 9B** (compare yellow and green curves).

It is important to note that these processes of the tail thinning and dipolarization often occur under weak variations of the lobe magnetic field. Its weak variations in the growth phase were

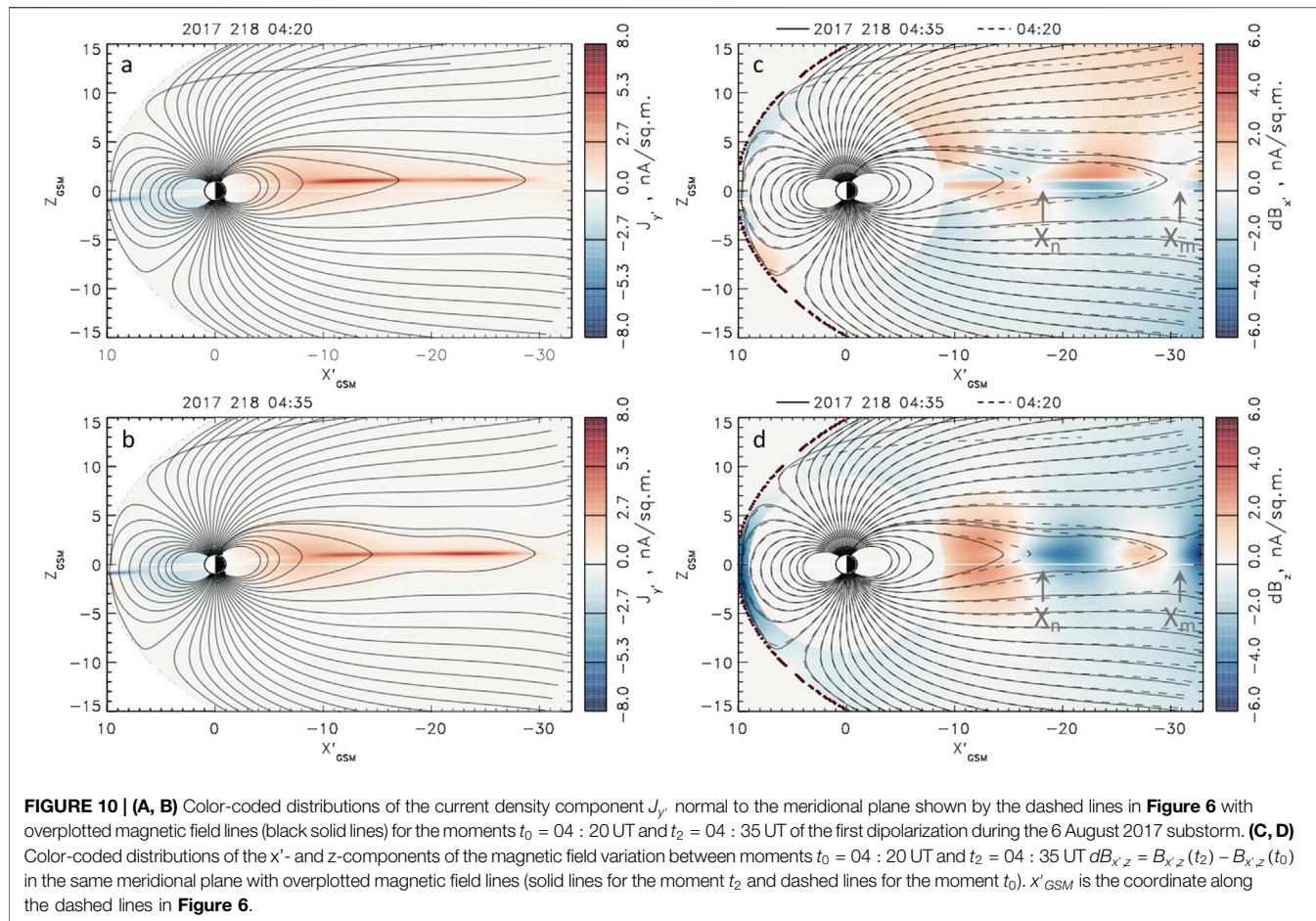


reported in [74–76] and they are seen in **Figure 10C** as well as in [27] (Figs. 12, 16, S4 and S13). Even rapid dipolarization processes shown in **Figures 4, 11C** are accompanied by more gradual lobe field variations, consistent with other data analyses [58, 74].

## 5 KINETIC SIMULATIONS OF MAGNETOTAIL RECONNECTION GUIDED BY EMPIRICAL RECONSTRUCTIONS

In order to understand the physical mechanisms of the formation of several X-lines in the magnetotail and their different reconnection regimes revealed in the DM analysis, we performed PIC simulations of the tail current sheet equilibrium sharing some of the observed pre-onset tail features. In particular, the reconstruction of the February 13

event discussed above in ([27], Fig. 8h) suggests that the near-Earth reconnection is preceded by the formation of a flux accumulation region near  $x \approx -22R_E$ . According to **Figure 6A**, similar pre-onset features in the form of a wide valley with small  $B_z$  values at  $R \sim 22R_E$  and the enhanced  $B_z$  ridge earthward of that valley took place prior to the 6 August 2017 substorm. To take these features into account, the PIC simulations start from a 2-D equilibrium with a  $B_z$  hump described by the vector potential  $A^{(0)} = [0, -\psi(x, z), 0]$ , where  $\psi = LB_0 \ln(\beta(x) \cosh[z/[L\beta(x)]])$ ,  $L$  is the characteristic current sheet thickness parameter, and the  $x$ -axis points from Earth to Sun. Its variation along the tail is determined by the function  $\beta(x) = \exp[\varepsilon_1 g(\xi)]$ , with  $\xi = x/L$ ,  $\varepsilon_1 \ll 1$  and  $-g(\xi) = \xi + (\alpha/\varepsilon_2)\{1 + \tanh[\varepsilon_2(\xi - \xi_0)]\}$ , which provides a region of accumulated magnetic flux near  $\xi = \xi_0$ . This is seen from the magnetic field profile  $B_z(x, z=0) = \varepsilon_1 B_0 \{1 + \alpha \cosh^{-2}[\varepsilon_2(\xi - \xi_0)]\}$  having a characteristic hump. The

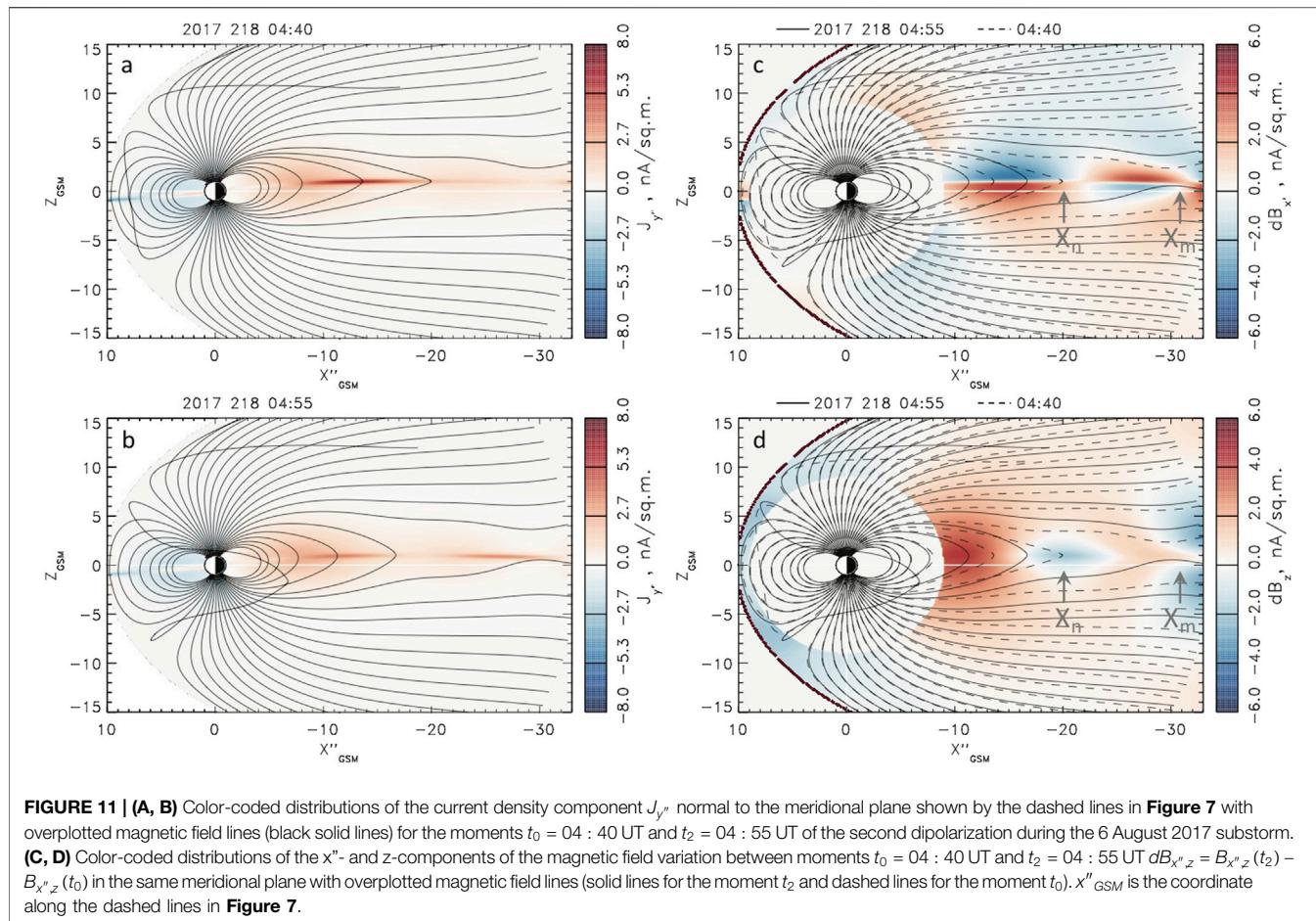


corresponding class of isotropic plasma equilibria was first proposed in [77] based on the 2-D generalization [78] of the 1-D Harris model [79] to describe spontaneous onset of the ion tearing instability.

The PIC simulations were performed using an open boundary modification [64, 80] of the explicit massively parallel code P3D [81] in a 3-D box with dimensions  $L_x^{(s)} \times L_y^{(s)} \times L_z^{(s)} = 80d_i \times 5d_i \times 20d_i$ , where  $d_i = c/\omega_{pi}$  is the ion inertial scale and  $\omega_{pi} = (4\pi e^2 n_0/m_i)^{1/2}$  is the plasma frequency;  $n_0$  is the plasma number density at the earthward side of the simulation box near the neutral plane ( $z = 0$ ). The choice of such a relatively long in  $x$  and narrow in the  $y$ -direction box was motivated by the available computer resources and the necessity to cover a large portion of the tail containing both X-lines resolved by the DM method and described in the previous section. In particular, with  $d_i \sim 500$  km  $\sim 0.1R_E$  [82], the distance between X-lines in our run is  $30d_i \sim 3R_E$ , that is only 3–4 times smaller than in the DM reconstruction. At the same time, our previous simulations of similar equilibria with shorter in  $x$  and wider in  $y$  boxes, up to  $L_y^{(s)} = 20d_i$  (see, for instance, Fig. 13 in [82]) suggest that the selected value of  $L_y = 5d_i$  with periodic boundaries in the  $y$ -direction is sufficient to reproduce major structuring in that direction, including ballooning/interchange and flapping motions.

The plasma parameters include the mass ratio  $m_i/m_e = 128$ , ion-to-electron temperature ratio  $T_i/T_e = 3$  and the effective Alfvén speed  $v_A = B_0/\sqrt{4\pi n_0 m_i} = c/15$  where  $c$  is the speed of light. The equilibrium magnetic field parameters are  $\varepsilon_1 = 0.03$ ,  $\varepsilon_2 = 0.2$ ,  $\alpha = 3$ , and  $\xi_0 = -30$  with the CS thickness parameter  $L = 1d_i$ . The magnetic and electric fields are normalized, respectively, by  $B_0$  and  $v_A B_0/c$ . The coordinates are normalized by  $d_i$  and velocity components by  $v_A$ . The simulation grid has  $2560 \times 160 \times 640$  cells with  $\approx 230$  particles per cell corresponding to  $n = n_0$ . The magnetic field configuration at the early stage of the run is shown in **Figure 12A**.

In contrast to earlier simulations ([83], and refs. therein) that described spontaneous onset regimes, here we drive the system by imposing a weak external electric field  $E_y^{(dr)}$  at top and bottom boundaries. This setup resembles earlier simulations of the externally driven electron tearing [73], and the whole setup is therefore a combination of the earlier ion and electron tearing modeling efforts. Still, in contrast to earlier setups with localized in  $x$  driving fields [73, 84, 85] and similar to [64], we do not assume any localization of the driving electric field along the tail. It remains constant throughout the main part of the box length  $L_x$ , being only attenuated near open boundaries. For example near the left boundary  $E_y^{(dr)} = E_0 [\tanh(x - \delta)/\delta]^2$  with  $\delta = 0.1L_x$ . The actual structure of the driving electric field remains



**FIGURE 11 | (A, B)** Color-coded distributions of the current density component  $J_y''$  normal to the meridional plane shown by the dashed lines in **Figure 7** with overplotted magnetic field lines (black solid lines) for the moments  $t_0 = 04:40$  UT and  $t_2 = 04:55$  UT of the second dipolarization during the 6 August 2017 substorm. **(C, D)** Color-coded distributions of the  $x''$ - and  $z$ -components of the magnetic field variation between moments  $t_0 = 04:40$  UT and  $t_2 = 04:55$  UT  $dB_{x''z} = B_{x''z}(t_2) - B_{x''z}(t_0)$  in the same meridional plane with overplotted magnetic field lines (solid lines for the moment  $t_2$  and dashed lines for the moment  $t_0$ ).  $x''_{GSM}$  is the coordinate along the dashed lines in **Figure 7**.

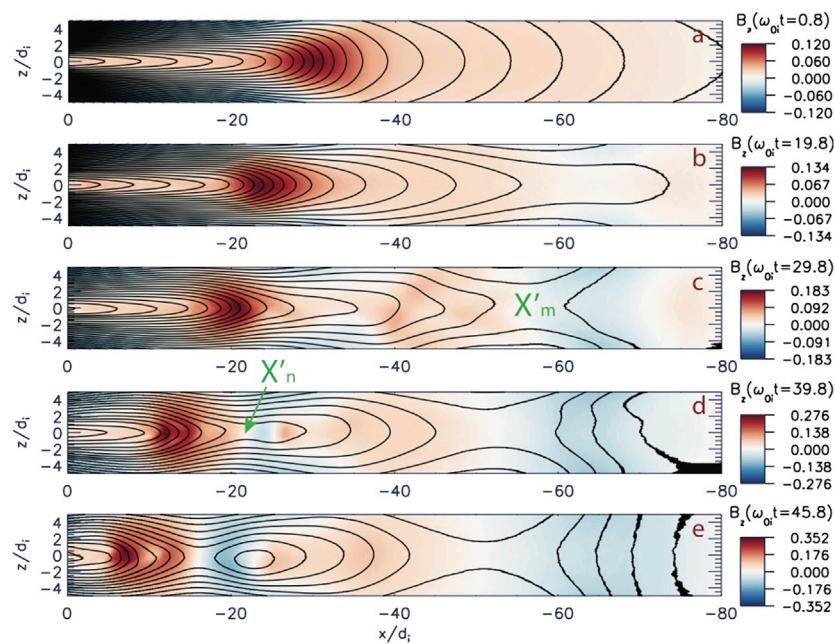
insufficiently resolved in observations and it can only be conjectured from global MHD simulations (e.g., [86]). In this situation, the assumption of the homogeneous electric field appears to be the most plausible ad hoc assumption. The driving field amplitude  $E_0$  smoothly increases in a half of the ion gyrotime  $1/\omega_{0i}$  at the beginning of the run and then remains constant with  $E_0 = 0.05$ .

The external driving first results in the CS thinning and stretching, which are seen particularly well in the tailward part of the box (**Figure 12B**). It also causes the buildup of the plasma pressure in the region  $x \lesssim -24d_i$  (not shown), consistent with previous studies of the driven reconnection regimes (e.g., Fig. 9 in [73]). This makes the CS configuration more similar to empirical reconstructions with extended TCS, such as for instance in **Figures 10A, 11A** (see also [27], Figs. 12b). At some point, the first X-line  $X'_m$  forms in the “tailward” part of the simulation box (**Figure 12C**). However, the second X-line  $X'_n$  forming later in the left (“earthward”) part of the box (**Figure 12D**) is not the secondary X-line caused by the tearing instability of the reconnection exhausts (e.g., [87]), because it also forms in the absence of any primary X-lines [88–90] or when the primary X-line shows no reconnection signatures [72, 82].  $X'_n$  rather forms because of the flux starvation effect created by the earthward-moving DF in its

trailing part. As it was shown in [72, 88, 89], the DF appears from the original  $B_z$  hump due to its spontaneous acceleration and further localization in  $x$ .

It is very interesting that the magnetic field perturbations shown in **Figure 12E** strongly resemble the DM reconstructions of substorm dipolarizations shown in **Figures 4, 10D, 11D** with much stronger bipolar  $B_z$  perturbations around the near-Earth X-line compared to the midtail region. This suggests that reconnection near  $X'_n$  is unsteady, in contrast to the steady midtail reconnection process at  $X'_m$ . This conclusion is further confirmed in our simulations by the analysis of the electric field and plasma parameters.

**Figure 13A** shows that the electric field distributions around the X-lines are indeed drastically different. Around  $X'_m$  ( $x \sim -50d_i$ ) the distribution of  $E_y(x, z)$  is homogeneous and its value  $E_y(x, z) \approx 0.1$  is consistent with the theoretical estimates [61–63]. These are strong indications of the steady reconnection process. In particular, the broad distribution of the electric field  $E_y$  over a large area in the plane  $(x, z)$  justifies the concept of the reconnection rate, measured by  $E_y$ , as a global parameter, which characterizes the reconnection process as a whole. In contrast, near  $X'_n$  ( $x \sim -15d_i$ ) the reconnection electric field strongly varies in space. However, not all these variations are related to unsteady reconnection. In particular, the sign-alternating



**FIGURE 12 | (A)** The initial magnetic field configuration in PIC simulations of magnetotail reconnection, which is shown here as a color-coded distribution of the  $B_z$  component with overplotted magnetic field lines in the plane  $y = 2.5d_i$  (“2D-equivalent” field lines calculated by treating a slice of the 3-D domain as if it were a 2-D simulation). **(B–E)** magnetic field distributions at later moments showing **(B)** the CS thinning, **(C)** the formation of the “mid-tail” X-line  $X'_m$ , **(D)** another “near-Earth” X-line  $X'_n$  near  $x = -20d_i$ , **(E)** the fully developed tail reconnection picture at the moment, which is further explored in more detail.

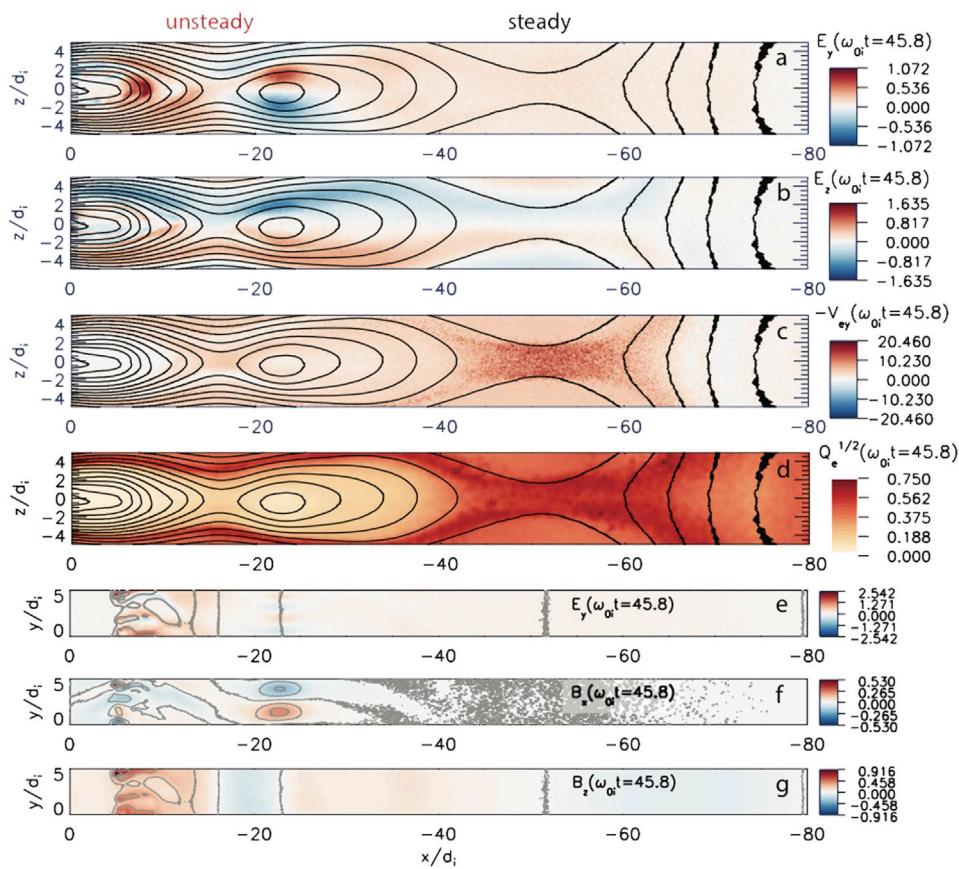
variations of  $E_y$  near the O-line ( $x \sim -23d_i$ ) describe north-south flapping motions of the CS as a whole, which are also seen in **Figure 13F** as strong variations of the magnetic field  $B_x$  ( $\sim 0.5B_0$ ) without noticeable  $B_z$  variations in the same region (**Figure 13G**). The properties of non-reconnection flapping and ballooning/interchange motions (seen in the region  $x \sim -5d_i$  in **Figure 13G**) in this run with a relatively small extension in the  $y$ -direction are similar to non-reconnection motions investigated with larger in  $y$  boxes in PIC simulations of spontaneous reconnection onset regimes [82], where they are compared with the corresponding magnetotail observations and other kinetic simulations.

At the same time, earthward of  $X_{n'}$  and near the DF, the electric field is structured in the  $y$ -direction due to ballooning/interchange perturbations that are best seen in variations of the magnetic field  $B_z$  (**Figure 13G**). All in all, the electric field associated with the earthward motion of the DF is highly localized near its  $B_z$  peak and its value strongly exceeds the steady-state reconnection limit 0.2 [63]. Note, that such strong values of the reconnection electric field were reported before in simulations of the ion tearing instability ([82], Fig. 5) and interchange-driven reconnection ([92], Fig. 11). Thus, the kinetic reconnection picture in our PIC simulation, which combines steady and unsteady reconnection regions, is quite consistent with the empirical DM-based reconstructions described in the previous section. Moreover, kinetic simulations reveal its features that cannot be captured from the empirical geomagnetic field analysis, because they represent spontaneous or small-scale plasma modes or they

are not reflected in the magnetic field data at all. The examples of the first group of such phenomena are flapping and ballooning/interchange motions seen in **Figures 13A,B**. They are indeed observed in the tail [93–97], although their relation to substorms and their reconnection modes remains a topic of ongoing discussions [98]. Another example is DFs, with their ion-scale leading edges and fast ( $v_x \sim v_A$ ) earthward propagation (e.g. [93]), forming out of relatively stationary and macroscopic  $B_z$ -humps (compare, for instance, **Figures 12A,E**).

In **Figures 13B–D** we present another group of signatures, which cannot be resolved using the DM analysis. **Figure 13B** shows the electric field directed toward the neutral plane  $z = 0$  and arising in ion and sub-ion-scale TCS due to different motions of electrons and ions on those scales [85, 99–102]. Similar effects of the electric field directed toward a negatively charged TCS were shown in PIC simulations ([101], Fig. 8) and in observations ([23], Fig. 9). **Figures 13C,D** show plasma signatures that are usually associated with the electron diffusion region (EDR) in steady reconnection regimes: The first shows super-Alfvénic downward electron flows [103] that have been found one of the key distinctive EDR features in recent MMS observations of the magnetotail reconnection [104]. The second reveals non-gyrotropic electron motions that are quantified using the agyrotropy parameter  $\sqrt{Q_e}$  proposed by [91] and shown later in MMS observations as a distinctive EDR signature [105].

Finally, in **Figure 14** we present the kinetic dissipation parameters for the unsteady part of this run and compare them with similar parameters inferred from MMS

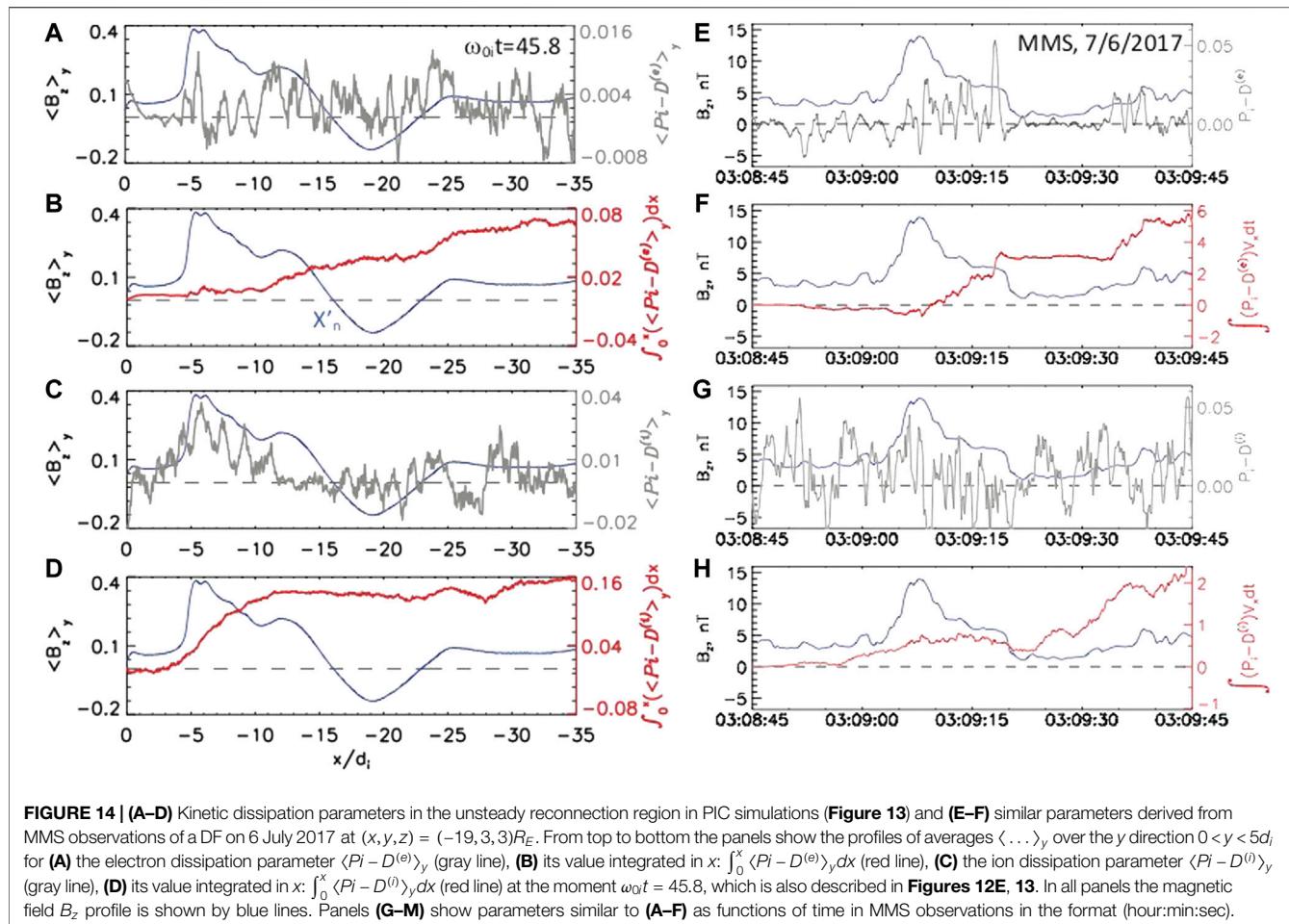


**FIGURE 13 |** Steady and unsteady reconnection regions in weakly driven magnetotail at the moment  $\omega_0 t = 45.8$  corresponding to **Figure 12E**. **(A–D)** The distributions in the plane  $y = 2.5d_i$  of the electric field components  $E_y$  and  $E_z$ , the electron bulk flow velocity  $= -V_{ey}$  and the electron agyrotropy parameter  $\sqrt{Q_e}$  [91] marking the localization of the electron diffusion region (in the latter case, to reduce noise in simulation outputs, the original numerical distributions are averaged over  $20 \times 20$  grid cells). **(E–G)** The distributions of the electric field  $E_y$  and the magnetic field components  $B_x$  and  $B_z$  in the neutral plane  $z = 0$ .

observations. In contrast to the steady-state reconnection area near  $X'_m$ , the unsteady reconnection region near  $X'_n$  does not reveal impressive EDR signatures, such as the super-Alfvénic downward electron flows or the agyrotropy enhancement (left parts of **Figures 4B,C**). This is because the main process in this region is the formation and fast earthward movement of a DF and the resulting dipolarization of the magnetic field configuration [72, 88, 89]. Moreover, many key aspects of these processes can be described by ideal MHD models [106, 107], whereas the DF formation and acceleration processes are shown to resemble the ion tearing instability [64, 72] supported by the ion Landau dissipation [108]. However, quantifying the latter in simulations and observations is a challenging problem because the conventional single-fluid measure, the Joule heating rate cannot distinguish between ion and electron Landau dissipation in collisionless magnetospheric plasmas. Indeed, the energy conversion rates in the frame moving with ions or electrons  $j \cdot E'_{e,i}$  (where  $j = j_i + j_e$ ,  $E'_{e,i} = E + v_{e,i} \times B/c$ ,  $j_{e,i}$  are the electron/ion currents in the laboratory frame of reference and  $v_e$  and  $v_i$  are the electron and ion bulk flow velocities) are same for ion and electron species  $j \cdot E'_e \approx j \cdot E'_i$  assuming plasma quasi-neutrality  $n_e \approx n_i$ .

To solve this problem, it has recently been proposed [109] to employ the new kinetic parameter  $Pi - D^{(\alpha)} = -\Pi_{ij}^{(\alpha)} D_{ij}^{(\alpha)}$  ( $\alpha = e, i$ ), the double contraction of deviatoric pressure tensor  $\Pi_{ij}^{(\alpha)} = p_{ij}^{(\alpha)} - \delta_{ij} p^{(\alpha)}$  (where  $p^{(\alpha)} = p_{ii}^{(\alpha)}/3$ ) and traceless strain-rate tensor  $D_{ij}^{(\alpha)} = (\partial_i v_j^{(\alpha)} + \partial_j v_i^{(\alpha)})/2 - \delta_{ij} \theta^{(\alpha)}/3$  (with  $\theta^{(\alpha)} = \nabla \cdot v^{(\alpha)}$ ), which was introduced earlier in [110]. It was demonstrated [109] that the  $Pi - D$  parameters represent direct analogs of the MHD Joule heating as an entropy variation measure and that they have different distributions for electrons and ions. It was shown that in the regions with ion  $Pi - D^{(i)}$  peaks, at the leading part of the DF, ion distributions show signatures of multi-flow motion, including ions reflected from the DF. Such multi-flow ion motions have indeed been detected at DFs in Cluster, THEMIS, and MMS observations [111–114].

In **Figures 14A–D** we present kinetic dissipation measures obtained in PIC simulations and averaged over the  $y$  direction  $0 < y < 5d_i$ , along with the corresponding profile of the magnetic field  $B_z$  shown here to provide the global context for this local investigation. As one can see from **Figures 14A,B**, while the linear distribution of the electron dissipation parameter  $\langle Pi - D^{(e)} \rangle_y$  remains irregular and not obviously positive, its



integration along the tail reveals its persistent accumulation upstream of the DF structure (red line in **Figure 14B**). The increase starts from the  $X'_n$  vicinity with another buildup near the corresponding O-line. The ion dissipation parameter is even more impressive: Already its average over the  $y$ -coordinate reveals a peak near the DF (**Figure 14C**), and when integrated along the tail  $\int_0^x \langle P_i - D^{(i)} \rangle_y dx$  builds up near the DF and remains elevated farther in the tail (red line in **Figure 14D**).

**Figures 14E–H** show the dissipation parameters similar to those in **Figures 14A–D** but now derived from MMS observations of a DF on 6 July 2017, a relatively rare case of a slow moving DF with the ion bulk flow speed smaller than 200 km/s. The four-probe sub-ion-scale MMS observations of the electromagnetic field and plasma parameters provide the unique opportunity to measure the kinetic dissipation parameters  $P_i - D$  for both electrons and ions. At the same time, even with the MMS capability of calculating higher moments of the plasma distribution, the assessment of the kinetic dissipation parameters remains a challenging problem. In particular, even in the MMS burst mode with the sampling time  $\delta t = 0.15$  s [115] and probe spacing  $\delta r \lesssim 20$  km, any velocity gradient estimates necessary for calculation of the tensor  $D_{ij}^{(i)}$  may give trivial results for structures moving much faster than  $V_{max} = \delta r / \delta t \approx 133$  km/s. Thus, MMS data is only appropriate

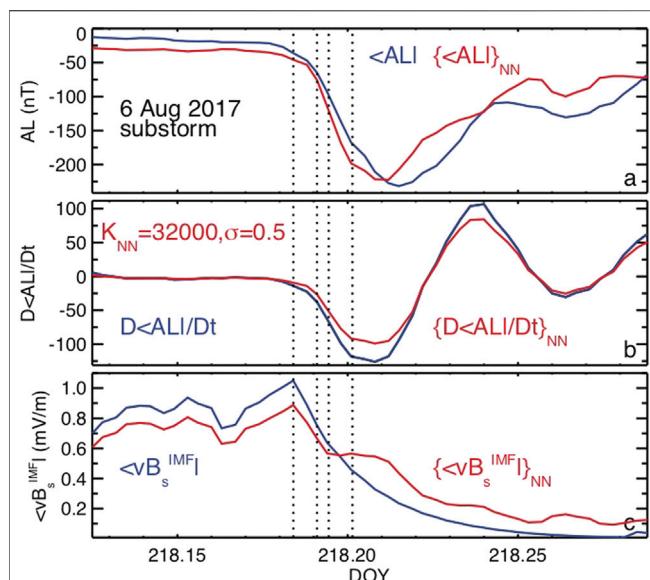
so far to study the kinetic dissipation in relatively slow moving DFs.

In spite of these caveats, simulation and observation results presented in **Figure 14** have interesting similarities. In particular, both simulations and data show the accumulation of positive  $P_i - D$  values for electrons after integration (over  $x$  in simulations and in time in observations) seen in **Figures 14B–G**. Both in simulations and in observations (**Figures 14B,F**) the electron dissipation builds up behind the DF, upstream of the ion dissipation buildup, in the regions with relatively small values of the magnetic field, while for ions the dissipation starts accumulating at or even before the DF.

## 6 DISCUSSION

### 6.1 Error Analysis of Empirical Reconstructions

In this study we provided a DM reconstruction of magnetic reconnection in the Earth's magnetotail associated with its dipolarizations during substorms. A direct validation of this reconstruction can only be provided using a limited number of in-situ observations available at the moment of interest. This is an unavoidable feature of the DM method as a data discovery



**FIGURE 15 |** The substorm part of the binning parameters (4)–(6)  $\langle AL \rangle$ ,  $D\langle AL \rangle / Dt$  and  $\langle vB_s^{\text{IMF}} \rangle$  (blue lines) and their means over 32,000 NN bins ( $\langle \dots \rangle_{\text{NN}}$ , red lines) for the 6 August 2017 substorm. Parameters  $\langle AL \rangle$  and  $D\langle AL \rangle / Dt$  are normalized by the corresponding unit convolution integrals  $\langle 1 \rangle$ . Vertical dashed lines mark two dipolarization intervals considered in **Section 4**.

tool, which extracts from data the information (e.g., on the global structure of the magnetotail), which cannot be obtained by other methods. We simply have no real constellations of  $\sim 3 \cdot 10^4$  probes to comprehensively validate our results. Still, the 13 February 2008 substorms were validated by all five THEMIS probes (Figs. S6–S7 in [27]), while for the 6 August 2017 event, the MMS1 validation results are presented in **Figure 5**. Moreover, the uncertainty of the DM method caused by averaging over the NN bins can be quantified by comparing the original values of the parameters  $G_1$ – $G_5$  with their NN means. For the 13 February 2008 reconstruction such information was provided in Fig. 19 of [27]. For the 6 August 2017 substorm we provide it in **Figure 15**. This figure shows in particular that during the dipolarization intervals considered in **Section 4** and shown by vertical dashed lines, the maximum deviation of the binning parameters averaged over their NN bins from their original values defined by Eqs. 4–6 does not exceed  $\sim 10\%$  (the largest deviation is seen for  $\langle AL \rangle$  at the end of the second dipolarization interval). This means that statistical errors of the presented reconstruction of the magnetic field during this substorm are much smaller compared to major variations of the binning parameters. Therefore the presented DM-based picture of magnetotail reconfigurations should indeed reflect the characteristic features of magnetic reconnection during substorms.

Consistent with the analysis of the 13 February 2008 substorms [27], we have found that the relatively strong deviations of the binning parameters from their means over NNs take place for the solar wind parameter  $\langle vB_s^{\text{IMF}} \rangle$  and the  $AL$  index in the recovery phase. This suggests that the solar wind and the magnetosphere after substorms are less coherent

(perhaps turbulent) and hence less reproducible, compared to the evolution of the magnetosphere during growth and expansion phases.

An important source of uncertainty in the present NN approach may be the instrument errors and combining probes from different epochs. Fortunately, the accuracy of magnetic field measurements critical for our investigation (with a few nT accuracy necessary to resolve the  $B_z$  magnetic field in the tail) was sufficiently high. In particular, the IMP8 magnetometer was good to 0.3 nT [116] and later missions had largely better instruments (e.g. [117–119]) with a few caveats. Significant errors (up to 7 nT) were found for some geosynchronous missions [120] and they were mitigated using inter-spacecraft calibration. The errors in the external magnetic field (difference between the measured and dipole magnetic field values) may also be large in the inner magnetosphere because of the spacecraft attitude uncertainty and large values of the dipole field there [121]. However, this is not an issue in the magnetotail.

## 6.2 Implications for Local Reconnection Models and Tearing Stability

The concept of magnetic reconnection was introduced to explain explosive energy release and rapid changes of magnetic field topology associated with solar flares [2, 3], magnetospheric substorms [7, 108, 122] and laboratory current disruptions ([123], and refs. therein). But its theory turned out to be built mainly on models of steady-state reconnection regimes ([63, 124–127], and refs. therein). The few exceptions include the tearing instability theory [87, 108, 122, 128], and catastrophe models of coronal mass ejections and solar flares [129, 130].

At the same time, the description of transition from the slow evolution of the tail to its rapid reconfiguration has long been complicated by the almost universal tearing stability of the tail current sheet provided by magnetization of electrons due to nonzero northward magnetic field  $B_z$  [131, 132]. As a result, the tail can be unstable when electrons become unmagnetized, under the condition  $B_z/B_0 \lesssim k\rho_{0e}$ , where  $B_0$  is the field outside CS,  $k$  is the wave vector and  $\rho_{0e}$  is the thermal electron gyroradius in the field  $B_0$  [73, 122, 133]. The resulting electron tearing instability is enabled by the free energy of the mutual attraction of the parallel electric current filaments and the electron Landau dissipation of unmagnetized electrons. In PIC simulations, the corresponding electron-demagnetization mediated reconnection (EDMR) onset used to be reproduced due to stretching and thinning of a CS by the external electric field [73, 101, 134]. It is important that after the electron tearing instability phase (or in its absence in simulations with spatially localized driving [84, 85]) the reconnection process becomes quasi-steady ([83], and refs. therein), consistent with regimes found earlier in kinetic simulations with non-self-consistent setups using 1-D CS equilibria with an imposed X-line perturbation ([127], and refs. therein).

In 1974 Schindler [108] hypothesized that the tail could become unstable even with magnetized electrons if the CS is sufficiently thin to demagnetize ions and provide their Landau dissipation. The corresponding tearing instability must be much

faster compared to the electron tearing. However, later it was found [135] that magnetized electrons change the free energy of the tearing mode, and eventually Lembege and Pellat [131] showed that the corresponding sufficient stability condition coincides with the Wentzel–Kramers–Brillouin (WKB) approximation  $\pi(B_z/B_0) \lesssim kL_z$ , which allows one to consider stability neglecting the CS variations along the tail with the scale  $L_x \sim L_z(B_0/B_z)$  ( $L_z$  is the CS half-thickness) making the ion tearing impossible.

A missing key for ion tearing destabilization was found relatively recently when it was discovered [77] that the stability condition derived by Lembege and Pellat [131] is only valid for constant  $B_z$  values. If  $B_z$  changes along the tail, the stability condition takes the form  $\pi(B_z/B_0)C_d^2 \lesssim kL_z$ , where the parameter  $C_d = VB_z/(\pi L_z)$  is determined by the flux tube volume per unit magnetic flux  $V = \int dl/B$ . In particular, in the presence of a flux accumulation area with the tailward gradient of  $B_z$ , the parameter  $C_d > 1$  and a room for instability arises. The corresponding instability had indeed been found in PIC simulations with ad hoc configurations having  $B_z(x)$  profiles with a hump [72, 88, 89]. Since electrons remained initially magnetized by the field  $B_z$ , the instability was essentially the ion tearing. It first led to the formation of an earthward-moving dipolarization front (DF), in whose wake new X-lines formed due to the flux starvation process [89]. The resulting ion-demagnetization dominated reconnection (IDMR) onset did not require any external driving and could be considered as spontaneous or “internally driven” by the DF formation and evolution processes.

Despite this clarity in the tearing stability theory and consistent simulation results, until now, the role of EDMR and IDMR regimes in the actual magnetotail dynamics remained unclear. In particular, it is unknown if/when the driving (ultimately due to the solar wind) is sufficiently strong to squeeze the CS down to electron scales and to provide EDMR with the subsequent steady reconnection, and when (if any)  $B_z$  humps form to provide IDMR.

The present study provides interesting implications for the magnetotail stability and reconnection onset mechanisms. Our DM reconstructions suggest that both steady and unsteady reconnection regimes are possible in the magnetotail during substorms. At the same time, our PIC simulations guided by empirical reconstructions suggest that both IDMR and EDMR regimes are possible in the tail. Moreover, the former resembles the unsteady reconnection, while the latter becomes eventually steady, consistent with the classical fast and steady reconnection models ([62] and refs. therein).

### 6.3 Role of Thin Current Sheets

The use in **Section 5** of isotropic plasma equilibria with shifted Maxwellian distributions for ions and electrons, inherited from the 1962 Harris solution [79], to explain the reconnection features found in our DM reconstructions may be questioned in view of another discovery in the DM analysis of substorms, namely the buildup of extended TCSs in the substorm growth phase and their decay in the expansion phase [26, 27] (see also **Figures 10A,B, 11A,B** of the present study).

The analysis of 2-D isotropic equilibrium models [136] suggests that they require strongly stretched magnetic field configurations (with sufficiently large values of the ratio  $B_0/B_z$ ) to explain the formation of the ion-scale TCS sufficiently far from the Earth. Large values of  $B_0/B_z$  are required to maintain the force balance between the magnetic field line tension and the pressure gradient  $1/L_x \sim (B_z/B_0)/L_z$ , where  $L_x$  is the inhomogeneity scale of the TCS,  $L_z$  is its half-thickness and  $B_0$  is the lobe field [137]. Modeling TCSs with  $L_x \gg L_z(B_0/B_z)$  might require more sophisticated equilibria with anisotropic and agyrotropic particle distributions (e.g. [136], and refs. therein).

Indeed, three of four substorm events on 13 February 2008 considered in [27] had relatively small values of  $B_0/B_z \sim 10$  (according to their Figs. 15b–15d), whereas their aspect ratios  $L_x/L_z$  often exceeded 50 (Fig. 16 in [27]). That finding was consistent with signatures of the multiscale structure of the magnetotail inferred from local observations of the pre-onset CSs [19, 20, 23, 51].

However, this is not the case for the event considered in **Section 3**, whose specific features (the  $B_z$  hump and the ion-scale TCS earthward of it) guided our PIC simulations. In that first substorm of the 13 February 2008 series, the ratio  $B_0/B_z$  reaches 70 in the late growth phase (yellow line in Fig. 15a, corresponding to 02:25 UT). Thus, the specific substorm event, considered in **Section 3** of our DM analysis is close to the isotropic force balance state and it can be consistently described by 2-D isotropic CS equilibrium models of the class [78]. Moreover, the specific parameters used in our simulations correspond to  $L_x/L_z \sim B_0/B_z \approx 33$  and  $D_{TCS} \sim 0.1R_E$  and they are quite close to similar TCS parameters of the first substorm in the 13 February 2008 series in its late growth phase (02:25 UT):  $L_x/L_z \sim 25$ ,  $B_0/B_z = 20 - 70$  and  $L_z = 0.2R_E$ .

One can also provide more general arguments why the isotropic 2-D models can still be used in the local stability analysis of the realistic magnetotail. First, statistical studies show that the tail plasmas away from the dipole region are weakly anisotropic [138, 139]. At the same time, the DM reconstructions demonstrate that the current of the embedded TCS in the late growth phase may be small, compared to the total current, as is seen, for instance, from **Figure 5L** (this is the case for all four 13 February 2008 events as is seen from Fig. 8f in [27]). This suggests that the embedded TCS features and underlying non-isotropic plasma properties may only serve to provide the formation of the ion-scale TCSs sufficiently far from Earth, where their local stability properties can still be realistically reproduced by PIC simulations with isotropic equilibria and open x-boundaries. This is consistent with the results of statistical studies based on Geotail data [140], which suggest that the near-Earth X-line mainly forms near the tailward edge of the TCS. This appears to be the case during the second dipolarization in the 6 August 2017 event (**Figure 11**), although this is likely not the case during the first dipolarization when the near-Earth X-line forms in the middle of a very long TCS (**Figure 10**). Besides, even if the initial TCS is relatively short because of the corresponding force balance, the simulations performed in **Section 5** suggest that it becomes more stretched and closer to empirical TCS reconstructions due to the external driving. To conclude, while

some substorm dipolarizations certainly require a generalization of the isotropic plasma approximation, as it was outlined in [136], others can still be described using the conventional class of isotropic CS models [78, 79].

## 7 CONCLUSION

In this study, we investigated for the first time the magnetotail reconnection picture using modern data-mining methods, which allow us to employ for the reconstruction not only the magnetic field measurements available at the moment of interest but also other events in the historical database when the magnetosphere was in similar global states (substorm phases). The DM reconstruction revealed two distinctly different regions of magnetic reconnection with weak and strong changes of the magnetic field geometry. For both the 13 February 2008 and the 6 August 2017 substorms considered in our study the near-Earth X-line appears near  $x = -20R_E$  at the substorm onset, which is defined in our work as a transition to the  $AL(t)$  index evolution with a strong negative slope (dashed vertical orange line in **Figure 5G–M**). This result is consistent with the original conjecture of Hones [7], later single- and multi-probe studies of the near-Earth X-lines [57, 58], as well as with the plasmoid statistics [141]. In both events, the near-Earth X-line first appears in the pre-midnight sector (**Figures 2, 6B**), which is consistent with the earlier statistical investigations using Geotail [142] and Cluster [143] data.

In addition to earlier investigations, our DM reconstruction reveals that the near-Earth X-line ( $X_n$ ) often co-exists with another more distant midtail X-line ( $X_m$ ) located at  $x \approx -30R_E$ . In spite of the fact that its location is near the edge of the main cloud of historical magnetometer measurements [44], the analysis of data in the corresponding NN bins (**Figures 2, 8**) shows that the selected NN subsets provide sufficiently broad radial coverage of data to resolve both X-lines. The finding of the midtail X-line is consistent with another group of earlier observations suggesting persistent reconnection in the midtail around  $30R_E$ , which was inferred from THEMIS and ARTEMIS statistics of traveling compression regions [59, 60]. However, the coexistence of near-Earth and midtail X-lines has never been demonstrated before.

Moreover, the DM analysis shows that reconnection regimes at near-Earth and midtail X-lines are different. The near-Earth X-line appears at the substorm onset and then disappears from that region or reappears in another near-Earth region, e.g., in the postmidnight sector (compare **Figures 7A,B** or **Figures 2, 10** in [27]). In contrast, the midtail X-line, after its appearance within the reconstruction validity region (here  $R < 32R_E$ ) in the late growth phase remains relatively stable and only gradually approaching the Earth (**Figures 6A–C, 7A–C**). Furthermore, the analysis of the magnetic field changes in the meridional plane (**Figures 3, 4**), which according to the Faraday's law 8) quantifies the steadiness of the reconnection process, suggests that the latter is relatively steady near  $X_m$  and transient at  $X_n$ .

To understand the physical mechanisms of the formation of several X-lines in the magnetotail and their different reconnection regimes, we performed 3-D PIC simulations of a relatively long ( $L_x = 80d_i$ ) tail CS region with open boundaries in the Sun-Earth direction. A new aspect of simulations was the combination of the initial TCS configuration having a region of

the flux accumulation ( $B_z$  hump) with a relatively weak and homogeneous external driving. The formation of the flux accumulation regions prior to unsteady reconnection in the near-Earth tail is found in the DM reconstruction of both substorm events (Fig. 8h in [27], as well as **Figures 6A, 7A**), consistent with earlier statistical results [144, 145]. Recently, it has been inferred from remote-sensing observations of 30–100 keV energy electrons precipitating from the tail CS during the substorm growth phase [146]. This feature is also interesting because the corresponding region with the tailward  $B_z$  gradient (earthward of the  $B_z$  hump) has been found in the tail stability theory [77] to be the only mechanism of destabilization of the ion tearing mode [108]. The second feature, the external driving was used before to reproduce the tail reconnection onset through the electron tearing instability [73]. It was also used in forced reconnection models [84, 85].

The reconnection picture in PIC simulations, guided by the DM reconstructions, is found to be surprisingly consistent with the empirical picture of the magnetotail reconnection. It also reveals two reconnection areas with distinctly different reconnection regimes, whose steadiness can now be checked using the explicit distributions of the electric field in the meridional plane (**Figure 13A**). It is found that farther in the tail, the reconnection process is steady and it reveals many signatures of the sustained collisionless reconnection process with the region of agyrotropic electron motion in its center. The corresponding dusk component of the electric field is broadly distributed in the meridional plane and hence it becomes effectively a global parameter of this reconnection regime. Its value  $E_y \approx 0.1$  matches earlier theoretical estimates for this regime supported by local PIC simulations [61–63]. At the same time, the evolution of the  $B_z$  hump is found to result in an unsteady reconnection process with the peak electric field near the dipolarization front exceeding the steady reconnection rate limit by more than an order of magnitude, the result, which is consistent with earlier PIC simulations of local unsteady reconnection regions [82, 92]. The analysis of kinetic dissipation parameters in the unsteady reconnection region shows that the ion dissipation parameter  $Pi - D^{(i)}$  peaks near the DF and it is further accumulated upstream of the propagating front. The electron dissipation is largely accumulated behind the DF near new X- and O-lines. Similar ion and electron dissipation parameters are inferred from MMS observations.

Both empirical and first-principle pictures of magnetotail reconnection still need further refinement. The present DM approach provides an empirical picture on the magnetotail on the time scales greater than 5 min and on the spatial scales larger than  $\sim 0.2R_E$  for the TCS thickness and a few  $R_E$  in the equatorial plane. On these scales, the magnetic field dipolarization is likely a cumulative effect of the smaller-scale processes, such as multiple DFs (e.g. [58, 66, 147]). These cumulative effects are not yet reproduced in PIC simulations. On the other hand, the midtail X-lines are found close to the gap region  $31R_E < R < 55R_E$  in historical data [44, 148]. Thus, a better resolution of the midtail reconnection picture requires more measurements in that gap region. PIC simulations were made in a relatively thin CS, whose non-Harris properties, such as its negative charging and

multiscale structure, are only partially captured now due to the external driving. In simulations with thicker CSs and broader  $B_z$  humps, as well as more realistic values of the parameters  $m_i/m_e$  and  $c/v_A$ , one can expect stronger negative charging effects, slower growth of DFs and subsequent reconnection, as well as weaker electron dissipation. A further improvement of the tail reconnection and stability picture is also required to better reproduce less stretched embedded TCS.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://doi.org/10.5281/zenodo.4383387>.

## AUTHOR CONTRIBUTIONS

Conceptualization, MSi; methodology, MSi, GS, TM and MSw; software, MSi, GS, TM and MSw; formal analysis, MSi, GS and TM; investigation, MSi, GS and TM; resources, MSi, GS and TM; data curation, MSi, GS and TM; writing—original draft preparation, MSi; writing—review and editing, MSi, GS, TM and MSw; visualization, MSi, GS and TM; supervision, MSi; project administration, MSi; funding acquisition, MSi. All authors have read and agreed to the published version of the manuscript.

## REFERENCES

- Cassak PA. Inside the black box: magnetic reconnection and the magnetospheric multiscale mission. *Space Weather* (2016) 14:186–97. doi:10.1002/2015SW001313
- Giovanelli RG. A theory of chromospheric flares. *Nature* (1946) 158:81–2. doi:10.1038/158081a0
- Sturrock PA. Model of the high-energy phase of solar flares. *Nature* (1966) 211:695–7. doi:10.1038/211695a0
- Forbes TG. A review on the genesis of coronal mass ejections. *J Geophys Res Space Phys* (2000) 105:23153–65. doi:10.1029/2000JA000005
- Dungey JW. Interplanetary magnetic field and the auroral zones. *Phys Rev Lett* (1961) 6:47–8. doi:10.1103/PhysRevLett.6.47
- Akasofu SI. The development of the auroral substorm. *Planet Space Sci* (1964) 12:273–82. doi:10.1016/0032-0633(64)90151-5
- Hones EW. Transient phenomena in the magnetotail and their relation to substorms. *Space Sci Rev* (1979) 23:393–410. doi:10.1007/BF00172247
- Ren Y, Yamada M, Ji H, Dorfman S, Gerhardt SP, and Kulsrud R. Experimental study of the hall effect and electron diffusion region during magnetic reconnection in a laboratory plasma. *Phys Plasmas* (2008) 15: 082113. doi:10.1063/1.2936269
- Nishida A, Uesugi K, Nakatani I, Mukai T, Fairfield DH, and Acuna MH. Geotail mission to explore earth's magnetotail. *Eos, Trans Am Geophys Union* (1992) 73:425–9. doi:10.1029/91EO000314
- Sharma AS, Nakamura R, Runov A, Grigorenko EE, Hasegawa H, Hoshino M, et al. Transient and localized processes in the magnetotail: a review. *Ann Geophysicae* (2008) 26:955–1006. doi:10.5194/angeo-26-955-2008
- Angelopoulos V. The THEMIS mission. *Space Sci Rev* (2008) 141:5. doi:10.1007/s11214-008-9336-1
- Burch JL, Moore TE, Torbert RB, and Giles BL. Magnetospheric multiscale overview and science objectives. *Space Sci Rev* (2016) 199:5–21. doi:10.1007/s11214-015-0164-9
- Sergeev VA, Pellinen RJ, and Pulkkinen TI. Steady magnetospheric convection: a review of recent results. *Space Sci Rev* (1996) 75:551–604. doi:10.1007/BF00833344
- Birdsall CK, and Langdon AB. *Plasma physics via computer simulation*. New York: Taylor & Francis (2005).
- Nakamura TKM, Genestreti KJ, Liu YH, Nakamura R, Teh WL, Hasegawa H, et al. Measurement of the magnetic reconnection rate in the Earth's magnetotail. *J Geophys Res Space Phys* (2018) 123:9150–68. doi:10.1029/2018JA025713
- Lapenta G, El Alaoui M, Berchem J, and Walker R. Multiscale mhd-kinetic PIC study of energy fluxes caused by reconnection. *J Geophys Res Space Phys* (2020) 125:e2019JA027276. doi:10.1029/2019JA027276
- Sergeev VA, Mitchell DG, Russell CT, and Williams DJ. Structure of the tail plasma/current sheet at and its changes in the course of a substorm. *J Geophys Res Space Phys* (1993) 98:17345–65. doi:10.1029/93JA01151
- Nakamura R, Baumjohann W, Runov A, Volwerk M, Zhang TL, Klecker B, et al. Fast flow during current sheet thinning. *Geophys Res Lett* (2002) 29: 55–1–55–4. doi:10.1029/2002GL016200
- Runov A, Sergeev VA, Nakamura R, Baumjohann W, Apatenkov S, Asano Y, et al. Local structure of the magnetotail current sheet: 2001 Cluster observations. *Ann Geophysicae* (2006) 24:247–62. doi:10.5194/angeo-24-247-2006
- Petrukovich AA, Artemyev AV, Malova HV, Popov VY, Nakamura R, and Zelenyi LM. Embedded current sheets in the Earth's magnetotail. *J Geophys Res Space Phys* (2011) 116. doi:10.1029/2010JA015749
- Sergeev VA, Angelopoulos V, Kubyshkina M, Donovan E, Zhou XZ, Runov A, et al. Substorm growth and expansion onset as observed with ideal ground-spacecraft THEMIS coverage. *J Geophys Res* (2011) 116:A00126. doi:10.1029/2010JA015689
- Hoshino M, Nishida A, Mukai T, Saito Y, Yamamoto T, and Kokubun S. Structure of plasma sheet in magnetotail: double-peaked electric current sheet. *J Geophys Res Space Phys* (1996) 101:24775–86. doi:10.1029/96JA02313
- Asano Y, Mukai T, Hoshino M, Saito Y, Hayakawa H, and Nagai T. Evolution of the thin current sheet in a substorm observed by Geotail. *J Geophys Res Space Phys* (2003) 108. doi:10.1029/2002JA009785
- Asano Y, Nakamura R, Baumjohann W, Runov A, Vörös Z, Volwerk M, et al. How typical are atypical current sheets? *Geophys Res Lett* (2005) 32. doi:10.1029/2004GL021834
- Sitnov MI, Ukhorskiy AY, and Stephens GK. Forecasting of global data-binning parameters for high-resolution empirical geomagnetic field models. *Space Weather* (2012) 10. doi:10.1029/2012SW000783

## FUNDING

This work was funded by NASA grants 80NSSC19K0074, 80NSSC19K0847, 80NSSC20K1271, 80NSSC19K0396 and 80NSSC20K1787, as well as NSF grants AGS-1702147 and AGS-1744269.

## ACKNOWLEDGMENTS

We acknowledge interesting and useful discussions of the obtained results with Nikolai Tsytgenenko, Rumi Nakamura, Ferdinand Plaschke, Slava Merkin and Shin Ohtani. We thank the many spacecraft and instrument teams and their PIs who produced the data sets we used in this study, including the Cluster, Geotail, Polar, IMP-8, GOES, THEMIS, Van Allen Probes and MMS, particularly their magnetometer teams. We also thank the SPDF for the OMNI database for solar wind values, which is composed of data sets from the IMP-8, ACE, WIND, and Geotail missions, and also the WDC in Kyoto for the Geomagnetic indices. Simulations were made possible by the NCAR's Computational and Information Systems Laboratory (doi:10.5065/D6RX99HX), supported by the NSF, as well as the NASA High-End Computing Program through the NASA Advanced Supercomputing Division at Ames Research Center.

26. Stephens GK, Sitnov MI, Korth H, Tsyganenko NA, Ohtani S, Gkioulidou M, et al. Global empirical picture of magnetospheric substorms inferred from multimission magnetometer data. *J Geophys Res Space Phys* (2019) 124. doi:10.1029/2018JA025843
27. Sitnov MI, Stephens GK, Tsyganenko NA, Miyashita Y, Merkin VG, Motoba T, et al. Signatures of nonideal plasma evolution during substorms obtained by mining multimission magnetometer data. *J Geophys Res Space Phys* (2019) 124:8427–56. doi:10.1029/2019JA027037
28. Cover T, and Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theor* (1967) 13:21–7. doi:10.1109/TIT.1967.1053964
29. Mitchell T. *Machine learning*. New York: McGraw-Hill (1997).
30. Tsyganenko NA. Effects of the solar wind conditions on the global magnetospheric configurations as deduced from data-based field models. Proceedings of the Third International Conference on Substorms (ICS-3) (1996), ESA SP-389, Versailles, France (1996).
31. Partamies N, Jususola L, Tanskanen E, and Kauristie K. Statistical properties of substorms during different storm and solar cycle phases. *Ann Geophysicae* (2013) 31:349–58. doi:10.5194/angeo-31-349-2013
32. Davis TN, and Sugiura M. Auroral electrojet activity index AE and its universal time variations. *J Geophys Res* (1896–1977) 71:785–801. doi:10.1029/JZ071i003p00785
33. Hsu TS, and McPherron RL. The characteristics of storm-time substorms and non-storm substorms. In: A Wilson, editor. *Fifth international conference on substorms*, 443. Noordwijk, Netherlands: ESA Special Publication (2000). p. 439.
34. Surjalal Sharma A. Assessing the magnetosphere’s nonlinear behavior: its dimension is low, its predictability, high. *Rev Geophys* (1995) 33:645–50. doi:10.1029/95RG00495
35. Sitnov MI, Sharma AS, Papadopoulos K, Vassiliadis D, Valdivia JA, Klimas AJ, et al. Phase transition-like behavior of the magnetosphere during substorms. *J Geophys Res Space Phys* (2000) 105:12955–74. doi:10.1029/1999JA000279
36. Sitnov MI, Tsyganenko NA, Ukhorskiy AY, and Brandt PC. Dynamical data-based modeling of the storm-time geomagnetic field with enhanced spatial resolution. *J Geophys Res Space Phys* (2008) 113. doi:10.1029/2007JA013003
37. Burton RK, McPherron RL, and Russell CT. An empirical relationship between interplanetary conditions and Dst. *J Geophys Res* (1896–1977) 80:4204–14. doi:10.1029/JA080i031p04204
38. Runge J, Balasis G, Daglis IA, Papadimitriou C, and Donner RV. Common solar wind drivers behind magnetic storm-magnetospheric substorm dependency. *Scientific Rep* (2018) 8:16987. doi:10.1038/s41598-018-35250-5
39. Sitnov MI, Sharma AS, Papadopoulos K, and Vassiliadis D. Modeling substorm dynamics of the magnetosphere: from self-organization and self-organized criticality to nonequilibrium phase transitions. *Phys Rev E* (2001) 65:016116. doi:10.1103/PhysRevE.65.016116
40. Ukhorskiy AY, Sitnov MI, Sharma AS, and Papadopoulos K. Global and multi-scale features of solar wind-magnetosphere coupling: from modeling to forecasting. *Geophys Res Lett* (2004) 31. doi:10.1029/2003GL018932
41. Pulkkinen TI, Baker DN, Fairfield DH, Pellinen RJ, Murphree JS, Elphinstone RD, et al. Modeling the growth phase of a substorm using the Tsyganenko model and multi-spacecraft observations: CDAW-9. *Geophys Res Lett* (1991) 18:1963–6. doi:10.1029/91GL02002
42. Kubyskina MV, Sergeev VA, and Pulkkinen TI. Hybrid input algorithm: an event-oriented magnetospheric model. *J Geophys Res Space Phys* (1999) 104: 24977–93. doi:10.1029/1999JA900222
43. Tsyganenko NA, and Andreeva VA. An empirical RBF model of the magnetosphere parameterized by interplanetary and ground-based drivers. *J Geophys Res Space Phys* (2016) 121:10,786–10,802. doi:10.1002/2016JA023217
44. Tsyganenko NA, and Sitnov MI. Magnetospheric configurations from a high-resolution data-based magnetic field model. *J Geophys Res Space Phys* (2007) 112. doi:10.1029/2007JA012260
45. Sitnov MI, Stephens GK, Tsyganenko NA, Ukhorskiy AY, Wing S, Korth H, et al. *Spatial structure and asymmetries of magnetospheric currents inferred from high-resolution empirical geomagnetic field models*. Washington, D.C: American Geophysical Union (2017). p. 199–212. chap. 15. doi:10.1002/9781119216346.ch15
46. Iijima T, and Potemra TA. The amplitude distribution of field-aligned currents at northern high latitudes observed by Triad. *J Geophys Res* (1896–1977) 81:2165–74. doi:10.1029/JA081i013p02165
47. Harang L. The mean field of disturbance of polar geomagnetic storms. *Terrrestrial Magnetism Atmos Electricity* (1946) 51:353–80. doi:10.1029/TE051i003p00353
48. Tsyganenko NA. A model of the near magnetosphere with a dawn-dusk asymmetry 1. Mathematical structure. *J Geophys Res Space Phys* (2002) 107: SMP 12–1–SMP 12–5. doi:10.1029/2001JA000219
49. McPherron RL, Russell CT, and Aubrey MP. Satellite studies of magnetospheric substorms on August 15, 1968: 9. Phenomenological model for substorms. *J Geophys Res* (1896–1977) 78:3131–49. doi:10.1029/JA078i016p03131
50. Camporeale E. The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather* (2019) 17:1166–207. doi:10.1029/2018SW002061
51. Runov A, Angelopoulos V, and Zhou XZ. Multipoint observations of dipolarization front formation by magnetotail reconnection. *J Geophys Res Space Phys* (2012) 117. doi:10.1029/2011JA017361
52. Kubat M. *An introduction to machine learning*. 1st ed. Incorporated: Springer Publishing Company (2015).
53. Nagai T, Fujimoto M, Nakamura R, Baumjohann W, Ieda A, Shinohara I, et al. Solar wind control of the radial distance of the magnetic reconnection site in the magnetotail. *J Geophys Res Space Phys* (2005) 110. doi:10.1029/2005JA011207
54. Nagai T, Shinohara I, and Zenitani S. The dawn-dusk length of the x line in the near-earth magnetotail: Geotail survey in 1994–2014. *J Geophys Res Space Phys* (2015) 120:8762–73. doi:10.1002/2015JA021606
55. Hasegawa H, Sonnerup BUO, Denton RE, Phan TD, Nakamura TKM, Giles BL, et al. Reconstruction of the electron diffusion region observed by the Magnetospheric Multiscale Spacecraft: First results. *Geophys Res Lett* (2017) 44:4566–74. doi:10.1002/2017GL073163
56. Denton RE, Torbert RB, Hasegawa H, Dors I, Genestreti KJ, Argall MR, et al. Polynomial reconstruction of the reconnection magnetic field observed by multiple spacecraft. *J Geophys Res Space Phys* (2020) 125:e2019JA027481. doi:10.1029/2019JA027481
57. Baker DN, Pulkkinen TI, Angelopoulos V, Baumjohann W, and McPherron RL. Neutral line model of substorms: Past results and present view. *J Geophys Res Space Phys* (1996) 101:12975–3010. doi:10.1029/95JA03753
58. Angelopoulos V, Runov A, Zhou XZ, Turner DL, Kiehas SA, Li SS, et al. Electromagnetic energy conversion at reconnection fronts. *Science* (2013) 341:1478–82. doi:10.1126/science.1236992
59. Imber SM, Slavin JA, Auster HU, and Angelopoulos V. A THEMIS survey of flux ropes and traveling compression regions: location of the near-Earth reconnection site during solar minimum. *J Geophys Res Space Phys* (2011) 116. doi:10.1029/2010JA016026
60. Zhao S, Tian A, Shi Q, Xiao C, Fu S, Zong Q, et al. Statistical study of magnetotail flux ropes near the lunar orbit. *Sci China Technol Sci* (2016) 59: 1591–6. doi:10.1007/s11431-015-0962-3
61. Shay MA, and Drake JF. The role of electron dissipation on the rate of collisionless magnetic reconnection. *Geophys Res Lett* (1998) 25:3759–62. doi:10.1029/1998GL900036
62. Cassak PA, Liu YH, and Shay M. A review of the 0.1 reconnection rate problem. *J Plasma Phys* (2017) 83:715830501. doi:10.1017/S0022377817000666
63. Liu YH, Hesse M, Guo F, Daughton W, Li H, Cassak PA, et al. Why does steady-state magnetic reconnection have a maximum local rate of order 0.1? *Phys Rev Lett* (2017) 118:085101. doi:10.1103/PhysRevLett.118.085101
64. Sitnov MI, and Swisdak M. Onset of collisionless magnetic reconnection in two-dimensional current sheets and formation of dipolarization fronts. *J Geophys Res* (2011) 116:12216. doi:10.1029/2011JA016920
65. Nakamura R, Baumjohann W, Klecker B, Bogdanova Y, Balogh A, Rème H, et al. Motion of the dipolarization front during a flow burst event observed by Cluster. *Geophys Res Lett* (2002) 29:3–1–3–4. doi:10.1029/2002GL015763

66. Runov A, Angelopoulos V, Sitnov MI, Sergeev VA, Bonnell J, McFadden JP, et al. Themis observations of an earthward-propagating dipolarization front. *Geophys Res Lett* (2009) 36. doi:10.1029/2009GL038980
67. Sitnov MI, Swisdak M, and Divin AV. Dipolarization fronts as a signature of transient reconnection in the magnetotail. *J Geophys Res Space Phys* (2009) 114. doi:10.1029/2008JA013980
68. Huang SY, Fu HS, Yuan ZG, Zhou M, Fu S, Deng XH, et al. Electromagnetic energy conversion at dipolarization fronts: Multispacecraft results. *J Geophys Res Space Phys* (2015) 120:4496–502. doi:10.1002/2015JA021083
69. Runov A, Angelopoulos V, Zhou XZ, Zhang XJ, Li S, Plaschke F, et al. A THEMIS multicase study of dipolarization fronts in the magnetotail plasma sheet. *J Geophys Res Space Phys* (2011) 116. doi:10.1029/2010JA016316
70. Sitnov MI, Stephens GK, Tsyganenko NA, Korth H, Roelof EC, Brandt PC, et al. Reconstruction of extreme geomagnetic storms: Breaking the data paucity curse. *Space Weather* (2020) 18:e2020SW002561. doi:10.1029/2020SW002561
71. Galeev AA. Spontaneous reconnection of magnetic field lines in a collisionless plasma. In: AA Galeev and RN Sudan, editors. *Handbook of plasma physics*, Vol. 2. New York): North-Holland (1984). p. 305–35.
72. Sitnov MI, Buzulukova N, Swisdak M, Merkin VG, and Moore TE. Spontaneous formation of dipolarization fronts and reconnection onset in the magnetotail. *Geophys Res Lett* (2013) 40:22–7. doi:10.1029/2012GL054701
73. Liu YH, Birn J, Daughton W, Hesse M, and Schindler K. Onset of reconnection in the near magnetotail: PIC simulations. *J Geophys Res Space Phys* (2014) 119:9773–89. doi:10.1002/2014JA020492
74. Shukhtina MA, Dmitrieva NP, and Sergeev VA. On the conditions preceding sudden magnetotail magnetic flux unloading. *Geophys Res Lett* (2014) 41: 1093–9. doi:10.1002/2014GL059290
75. Artemyev AV, Angelopoulos V, Runov A, and Petrokovich AA. Properties of current sheet thinning at  $x$  -10 to -12  $R_E$ . *J Geophys Res Space Phys* (2016) 121: 6718–31. doi:10.1002/2016JA022779
76. Sun WJ, Fu SY, Wei Y, Yao ZH, Rong ZJ, Zhou XZ, et al. Plasma sheet pressure variations in the near-Earth magnetotail during substorm growth phase: themis observations. *J Geophys Res Space Phys* (2017) 122: 12,212–12,228. doi:10.1002/2017JA024603
77. Sitnov MI, and Schindler K. Tearing stability of a multiscale magnetotail current sheet. *Geophys Res Lett* (2010) 37:08102. doi:10.1029/2010GL042961
78. Schindler K. A selfconsistent theory of the tail of the magnetosphere. In: BM McCormac, editor. *Earth's magnetospheric processes*. Dordrecht-Holland: D. Reidel (1972). p. 200–9.
79. Harris E. On a plasma sheath separating regions of oppositely directed magnetic field. *Il Nuovo Cimento Ser* (1962) 23:115–21. doi:10.1007/BF02733547
80. Divin AV, Sitnov MI, Swisdak M, and Drake JF. Reconnection onset in the magnetotail: Particle simulations with open boundary conditions. *Geophys Res Lett* (2007) 34:L09109. doi:10.1029/2007GL029292
81. Zeiler A, Biskamp D, Drake JF, Rogers BN, Shay MA, and Scholer M. Three-dimensional particle simulations of collisionless magnetic reconnection. *J Geophys Res* (2002) 107:1230. doi:10.1029/2001JA000287
82. Sitnov MI, Merkin VG, Swisdak M, Motoba T, Buzulukova N, Moore TE, et al. Magnetic reconnection, buoyancy, and flapping motions in magnetotail explosions. *J Geophys Res Space Phys* (2014) 119:7151–68. doi:10.1002/2014JA020205
83. Sitnov M, Birn J, Ferdousi B, Gordeev E, Khotyaintsev Y, Merkin V, et al. Explosive magnetotail activity. *Space Sci Rev* (2019) 215:31. doi:10.1007/s11214-019-0599-5
84. Pritchett PL. Externally driven magnetic reconnection in the presence of a normal magnetic field. *J Geophys Res Space Phys* (2005) 110. doi:10.1029/2004JA010948
85. Pritchett PL. Onset of magnetic reconnection in the presence of a normal magnetic field: Realistic ion to electron mass ratio. *J Geophys Res Space Phys* (2010) 115. doi:10.1029/2010JA015371
86. Nishimura Y, and Lyons LR. Localized reconnection in the magnetotail driven by lobe flow channels: Global MHD simulation. *J Geophys Res Space Phys* (2016) 121:1327–38. doi:10.1002/2015JA022128
87. Loureiro NF, Schekochihin AA, and Cowley SC. Instability of current sheets and formation of plasmoid chains. *Phys Plasmas* (2007) 14:100703. doi:10.1063/1.2783986
88. Bessho N, and Bhattacharjee A. Instability of the current sheet in the Earth's magnetotail with normal magnetic field. *Phys Plasmas* (2014) 21: 102905.
89. Pritchett PL. Instability of current sheets with a localized accumulation of magnetic flux. *Phys Plasmas* (2015) 22:062102. doi:10.1063/1.4921666
90. Sitnov MI, Merkin VG, Pritchett PL, and Swisdak M. Distinctive features of internally driven magnetotail reconnection. *Geophys Res Lett* (2017) 44:3028–37. doi:10.1002/2017GL072784
91. Swisdak M. Quantifying gyrotropy in magnetic reconnection. *Geophys Res Lett* (2016) 43:43–9. doi:10.1002/2015GL066980
92. Pritchett PL, and Coroniti FV. Structure and consequences of the kinetic ballooning/interchange instability in the magnetotail. *J Geophys Res Space Phys* (2013) 118:146–59. doi:10.1029/2012JA018143
93. Runov A, Sergeev VA, Baumjohann W, Nakamura R, Apatenkov S, Asano Y, et al. Electric current and magnetic field geometry in flapping magnetotail current sheets. *Ann Geophysicae* (2005) 23:1391–403. doi:10.5194/angeo-23-1391-2005
94. Petrukovich AA, Zhang T, Baumjohann W, Nakamura R, Runov A, Balogh A, et al. Oscillatory magnetic flux tube slippage in the plasma sheet. *Ann Geophysicae* (2006) 24:1695–704. doi:10.5194/angeo-24-1695-2006
95. Sergeev VA, Sormakov DA, Apatenkov SV, Baumjohann W, Nakamura R, Runov AV, et al. Survey of large-amplitude flapping motions in the midtail current sheet. *Ann Geophys* (2006) 24:2015–24. doi:10.5194/angeo-24-2015-2006
96. Panov EV, Sergeev VA, Pritchett PL, Coroniti FV, Nakamura R, Baumjohann W, et al. Observations of kinetic ballooning/interchange instability signatures in the magnetotail. *Geophys Res Lett* (2012) 39. doi:10.1029/2012GL051668
97. Gao JW, Rong ZJ, Cai YH, Lui ATY, Petrukovich AA, Shen C, et al. The distribution of two flapping types of magnetotail current sheet: implication for the flapping mechanism. *J Geophys Res Space Phys* (2018) 123:7413–23. doi:10.1029/2018JA025695
98. Sergeev VA, Angelopoulos V, and Nakamura R. Recent advances in understanding substorm dynamics. *Geophys Res Lett* (2012) 39. doi:10.1029/2012GL050859
99. Pritchett PL, and Coroniti FV. Formation of thin current sheets during plasma sheet convection. *J Geophys Res Space Phys* (1995) 100:23551–65. doi:10.1029/95JA02540
100. Hesse M, Winske D, and Birn J. On the ion-scale structure of thin current sheets in the magnetotail. *Physica Scripta* (1998) T74:63–6. doi:10.1088/0031-8949/1998/t74/012
101. Birn J, and Hesse M. Forced reconnection in the near magnetotail: onset and energy conversion in PIC and MHD simulations. *J Geophys Res* (2014) 119: 290–309. doi:10.1002/2013JA019354
102. Lu S, Lin Y, Angelopoulos V, Artemyev AV, Pritchett PL, Lu Q, et al. Hall effect control of magnetotail dawn-dusk asymmetry: A three-dimensional global hybrid simulation. *J Geophys Res Space Phys* (2016) 121:11,882–11,895. doi:10.1002/2016JA023325
103. 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 Sci Rev* (2016) 199:631–50. doi:10.1007/s11214-015-0150-2
104. 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) 362:1391–5. doi:10.1126/science.aat2998
105. Cozzani G, Retinò A, Califano F, Alexandrova A, Le Contel O, Khotyaintsev Y, et al. *In situ* spacecraft observations of a structured electron diffusion

- region during magnetopause reconnection. *Phys Rev E* (2019) 99:043204. doi:10.1103/PhysRevE.99.043204
106. Merkin VG, Sitnov MI, and Lyon JG. Evolution of generalized two-dimensional magnetotail equilibria in ideal and resistive MHD. *J Geophys Res Space Phys* (2015) . doi:10.1002/2014JA020651
  107. Birn J, Merkin VG, Sitnov MI, and Otto A. MHD stability of magnetotail configurations with a hump. *J Geophys Res Space Phys* (2018) 123. doi:10.1029/2018JA025290
  108. Schindler K. A theory of the substorm mechanism. *J Geophys Res* (1974) 79: 2803–10. doi:10.1029/JA079i019p02803
  109. Sitnov MI, Merkin VG, Roytershteyn V, and Swisdak M. Kinetic dissipation around a dipolarization front. *Geophys Res Lett* (2018) 45:4639–47. doi:10.1029/2018GL077874
  110. Yang Y, Matthaeus WH, Parashar TN, Wu P, Wan M, Shi Y, et al. Energy transfer channels and turbulence cascade in Vlasov-Maxwell turbulence. *Phys Rev E* (2017) 95:061201. doi:10.1103/PhysRevE.95.061201
  111. Nakamura R, Amm O, Laakso H, Draper NC, Lester M, Grocott A, et al. Localized fast flow disturbance observed in the plasma sheet and in the ionosphere. *Ann Geophysicae* (2005) 23:553–66. doi:10.5194/angeo-23-553-2005
  112. Zhou XZ, Angelopoulos V, Sergeev VA, and Runov A. On the nature of precursor flows upstream of advancing dipolarization fronts. *J Geophys Res Space Phys* (2011) 116. doi:10.1029/2010JA016165
  113. Eastwood JP, Goldman MV, Hietala H, Newman DL, Mistry R, and Lapenta G. Ion reflection and acceleration near magnetotail dipolarization fronts associated with magnetic reconnection. *J Geophys Res Space Phys* (2015) 120:511–25. doi:10.1002/2014JA020516
  114. Liu CM, Vaivads A, Graham DB, Khotyaintsev YV, Fu HS, Johlander A, et al. Ion-beam-driven intense electrostatic solitary waves in reconnection jet. *Geophys Res Lett* (2019) 46:12702–10. doi:10.1029/2019GL085419
  115. Pollock C, Moore T, Jacques A, Burch J, Gliese U, Saito Y, et al. Fast plasma investigation for Magnetospheric Multiscale. *Space Sci Rev* (2016) 199: 331–406. doi:10.1007/s11214-016-0245-4
  116. Frank LA, Ackerson KL, and Lepping RP. On hot tenuous plasmas, fireballs, and boundary layers in the Earth's magnetotail. *J Geophys Res* (1896–1977) 81:5859–81. doi:10.1029/JA081i034p05859
  117. Balogh A, Carr CM, Acuña MH, Dunlop MW, Beek TJ, Brown P, et al. The Cluster magnetic field investigation: Overview of in-flight performance and initial results. *Ann Geophysicae* (2001) 19:1207–17. doi:10.5194/angeo-19-1207-2001
  118. Fröhlauff D, Plaschke F, and Glassmeier KH. Spin axis offset calibration on THEMIS using mirror modes. *Ann Geophysicae* (2017) 35:117–21. doi:10.5194/angeo-35-117-2017
  119. Russell CT, Anderson BJ, Baumjohann W, Bromund KR, Dearborn D, Fischer D, et al. The magnetospheric multiscale magnetometers. *Space Sci Rev* (2016) 199:189–256. doi:10.1007/s11214-014-0057-3
  120. Tsyganenko NA, Singer HJ, and Kasper JC. Storm-time distortion of the inner magnetosphere: How severe can it get? *J Geophys Res Space Phys* (2003) 108. doi:10.1029/2002JA009808
  121. Stephens GK, Sitnov MI, Ukhorskiy AY, Roelof EC, Tsyganenko NA, and Le G. Empirical modeling of the storm time innermost magnetosphere using Van Allen Probes and THEMIS data: Eastward and banana currents. *J Geophys Res Space Phys* (2016) 121:157–70. doi:10.1002/2015JA021700
  122. Coppi B, Laval G, and Pellat R. Dynamics of the geomagnetic tail. *Phys Rev Lett* (1966) 16:1207–10. doi:10.1103/PhysRevLett.16.1207
  123. Yamada M. Review of controlled laboratory experiments on physics of magnetic reconnection. *J Geophys Res Space Phys* (1999) 104:14529–41. doi:10.1029/1998JA900169
  124. Sweet PA. 14. The neutral point theory of solar flares. *Symp - Int Astronomical Union* (1958) 6:123–34. doi:10.1017/S0074180900237704
  125. Parker EN. Sweet's mechanism for merging magnetic fields in conducting fluids. *J Geophys Res* (1896–1977) 62:509–20. doi:10.1029/JZ062i004p00509
  126. Petschek HE. *Magnetic Field Annihilation*. AAS-NASA Symposium on Physics of Solar Flares. NASA Spec. Publ. (1964) 50:425.
  127. Birn J, Drake JF, Shay MA, Rogers BN, Denton RE, Hesse M, et al. Geospace Environmental Modeling (GEM) magnetic reconnection challenge. *J Geophys Res Space Phys* (2001) 106:3715–9. doi:10.1029/1999JA900449
  128. Pucci F, and Velli M. Reconnection of quasi-singular current sheets: the ideal tearing mode. *ApJL* (2014) 780:L19. doi:10.1088/2041-8205/780/2/l19
  129. Forbes TG, and Isenberg PA. A catastrophe mechanism for coronal mass ejections. *Astrophysical J* (1991) 373:294. doi:10.1086/170051
  130. Cassak PA, Shay MA, and Drake JF. Catastrophe model for fast magnetic reconnection onset. *Phys Rev Lett* (2005) 95:235002. doi:10.1103/PhysRevLett.95.235002
  131. Lembege B, and Pellat R. Stability of a thick two-dimensional quasineutral sheet. *Phys Fluids* (1982) 25:1995–2004. doi:10.1063/1.863677
  132. Pellat R, Coroniti FV, and Pritchett PL. Does ion tearing exist? *Geophys Res Lett* (1991) 18:143–6. doi:10.1029/91GL00123
  133. Daughton W. The unstable eigenmodes of a neutral sheet. *Phys Plasmas* (1999) 6:1329–43. doi:10.1063/1.873374
  134. Hesse M, and Schindler K. The onset of magnetic reconnection in the magnetotail. *Earth, Planets and Space* (2001) 53:645–53.
  135. Galeev AA, and Zelenyi LM. Tearing instability in plasma configurations. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki* (1976) 70:2133–51.
  136. Sitnov MI, and Merkin VG. Generalized magnetotail equilibria: Effects of the dipole field, thin current sheets, and magnetic flux accumulation. *J Geophys Res Space Phys* (2016) 121:7664–83. doi:10.1002/2016JA023001
  137. Rich FJ, Vasiliunas VM, and Wolf RA. On the balance of stresses in the plasma sheet. *J Geophys Res* (1972) 77:4670–6. doi:10.1029/JA077i025p04670
  138. Walsh AP, Owen CJ, Fazakerley AN, Forsyth C, and Dandouras I. Average magnetotail electron and proton pitch angle distributions from Cluster PEACE and CIS observations. *Geophys Res Lett* (2011) 38. doi:10.1029/2011GL046770
  139. Wang CP, Zaharia SG, Lyons LR, and Angelopoulos V. Spatial distributions of ion pitch angle anisotropy in the near-Earth magnetosphere and tail plasma sheet. *J Geophys Res Space Phys* (2013) 118:244–55. doi:10.1029/2012JA018275
  140. Asano Y, Mukai T, Hoshino M, Saito Y, Hayakawa H, and Nagai T. Statistical study of thin current sheet evolution around substorm onset. *J Geophys Res Space Phys* (2004) 109. doi:10.1029/2004JA010413
  141. Ieda A, Machida S, Mukai T, Saito Y, Yamamoto T, Nishida A, et al. Statistical analysis of the plasmoid evolution with Geotail observations. *J Geophys Res Space Phys* (1998) 103:4453–65. doi:10.1029/97JA03240
  142. Nagai T, Shinohara I, Zenitani S, Nakamura R, Nakamura TKM, Fujimoto M, et al. Three-dimensional structure of magnetic reconnection in the magnetotail from Geotail observations. *J Geophys Res Space Phys* (2013) 118:1667–78. doi:10.1002/jgra.50247
  143. Genestreti K, Fuselier S, Goldstein J, Nagai T, and Eastwood J. The location and rate of occurrence of near-earth magnetotail reconnection as observed by Cluster and Geotail. *J Atmos Solar-Terrestrial Phys* (2014) 121:98–109. doi:10.1016/j.jastp.2014.10.005
  144. Wang CP, Lyons LR, Nagai T, and Samson JC. Midnight radial profiles of the quiet and growth-phase plasma sheet: The Geotail observations. *J Geophys Res Space Phys* (2004) 109:12201. doi:10.1029/2004JA010590
  145. Machida S, Miyashita Y, Ieda A, Nosé M, Nagata D, Liou K, et al. Statistical visualization of the Earth's magnetotail based on Geotail data and the implied substorm model. *Ann Geophys* (2009) 27:1035–46. doi:10.5194/angeo-27-1035-2009

146. Sergeev VA, Gordeev EI, Merkin VG, and Sitnov MI. Does a local B-minimum appear in the tail current sheet during a substorm growth phase? *Geophys Res Lett* (2018) 0. doi:10.1002/2018GL077183
147. Liu J, Angelopoulos V, Runov A, and Zhou XZ. On the current sheets surrounding dipolarizing flux bundles in the magnetotail: the case for wedgelets. *J Geophys Res Space Phys* (2013) 118:2000–2020. doi:10.1002/jgra.50092
148. Kiehas SA, Runov A, Angelopoulos V, Hietala H, and Korovinskiy D. Magnetotail fast flow occurrence rate and dawn-dusk asymmetry at  $X_{GSM} \sim 60R_E$ . *J Geophys Res Space Phys* (2018) 123:1767–78. doi:10.1002/2017JA024776

**Conflict of Interest:** 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.

Copyright © 2021 Sitnov, Stephens, Motoba and Swisdak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## NOMENCLATURE

The following abbreviations are used in this manuscript:

**CME** Coronal Mass Ejection

**CS** Current Sheet

**DM** Data Mining

**EDR** Electron Diffusion Region

**EDMR** Electron Demagnetization Mediated Reconnection

**FAC** Field Aligned Current system

**GSM** Geocentric Solar Magnetospheric coordinate system

**IDMR** Ion Demagnetization Mediated Reconnection

**KNN** K Nearest Neighbors method

**PIC** Particle-In-Cell simulation method

**R1,2** Region 1,2 field-aligned current

**SMC** Steady Magnetospheric Convection

**TCS** Thin Current Sheet

**UT** Universal Time

**WKB** Wentzel–Kramers–Brillouin approximation



# Exploring Three Recurrent Neural Network Architectures for Geomagnetic Predictions

Peter Wintoft \* and Magnus Wik

Swedish Institute of Space Physics, Lund, Sweden

## OPEN ACCESS

### Edited by:

Enrico Camporeale,  
University of Colorado Boulder,  
United States

### Reviewed by:

Jannis Teunissen,  
Centrum Wiskunde and Informatica,  
Netherlands  
Shiyong Huang,  
Wuhan University, China

### \*Correspondence:

Peter Wintoft  
peter@lund.irf.se

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
*Frontiers in Astronomy and  
Space Sciences*

**Received:** 05 February 2021

**Accepted:** 22 April 2021

**Published:** 12 May 2021

### Citation:

Wintoft P and Wik M (2021) Exploring Three Recurrent Neural Network Architectures for Geomagnetic Predictions. *Front. Astron. Space Sci.* 8:664483.  
doi: 10.3389/fspas.2021.664483

Three different recurrent neural network (RNN) architectures are studied for the prediction of geomagnetic activity. The RNNs studied are the Elman, gated recurrent unit (GRU), and long short-term memory (LSTM). The RNNs take solar wind data as inputs to predict the Dst index. The Dst index summarizes complex geomagnetic processes into a single time series. The models are trained and tested using five-fold cross-validation based on the hourly resolution OMNI dataset using data from the years 1995–2015. The inputs are solar wind plasma (particle density and speed), vector magnetic fields, time of year, and time of day. The RNNs are regularized using early stopping and dropout. We find that both the gated recurrent unit and long short-term memory models perform better than the Elman model; however, we see no significant difference in performance between GRU and LSTM. RNNs with dropout require more weights to reach the same validation error as networks without dropout. However, the gap between training error and validation error becomes smaller when dropout is applied, reducing over-fitting and improving generalization. Another advantage in using dropout is that it can be applied during prediction to provide confidence limits on the predictions. The confidence limits increase with increasing Dst magnitude: a consequence of the less populated input-target space for events with large Dst values, thereby increasing the uncertainty in the estimates. The best RNNs have test set RMSE of 8.8 nT, bias close to zero, and linear correlation of 0.90.

**Keywords:** space weather, recurrent neural net, cross-validation, solar wind-magnetosphere–ionosphere coupling, prediction, dropout

## 1 INTRODUCTION

In this work we explore recurrent neural networks (RNNs) for the prediction of geomagnetic activity using solar wind data. An RNN can learn input–output mappings that are temporally correlated. Many solar terrestrial relations exhibit such behavior that contains both directly driven processes and dynamic processes that depend on time. The geomagnetic Dst index has been addressed in numerous studies and serves as a parameter for general space weather summary and space weather models. The Dst index is derived from magnetic field measurements at four near-equatorial stations and primarily indicates the strength of the equatorial ring current and the magnetopause current (Mayaud, 1980). The Dst index has attained a lot of attention over the years, both for understanding solar terrestrial relations and for use in space weather.

An early attempt to predict the Dst index from the solar wind made use of a linear filter (Burton et al., 1975) derived from the differential equation containing a source term (the solar wind driver) and a decay term. After removing the variation in Dst that is controlled by the solar wind dynamic

pressure, one arrives at the pressure-corrected  $Dst^*$  index (O'Brien and McPherron, 2000) which is modeled as

$$\frac{dDst^*}{dt} = Q - \frac{Dst^*}{\tau}, \quad (1)$$

where  $Q$  is the source term that depends on the solar wind and  $t$  is the decay time of the ring current. The decay time  $\tau$  may be a constant, but it may also vary with the solar wind. (see, e.g., the AK1 (constant  $\tau$ ) and AK2 (variable  $\tau$ ) models in O'Brien and McPherron (2000)). As the functional form of  $Q$  is not known, the equation is numerically solved by

$$Dst^*(t + \Delta t) = Dst^*(t) + \left( Q(t) - \frac{Dst^*(t)}{\tau(t)} \right) \Delta t. \quad (2)$$

Based on observed solar wind data, for hourly sampled data, the time step is  $\Delta t = 1$  hour. The source term  $Q$  is a nonlinear function of the solar wind parameters, and different forms have been suggested. The AK1 model defines the source term as

$$Q(V, B_s) = aVB_s \text{ nT/h}, \quad (3)$$

where  $a = -2.47$  is a constant,  $V$  is the solar wind speed (km/s), and  $B_s$  is

$$B_s = \begin{cases} 0 & , B_z \geq 0 \\ -B_z & , B_z < 0 \end{cases} \text{ nT}. \quad (4)$$

$B_z$  is the vertical solar wind magnetic field component. Thus, as long as  $B_z < 0$ , the  $Dst$  index will be driven to increasing negative values; for example, if  $\tau$  is a constant and the solar wind conditions are constant with negative  $B_z$ , then  $Dst^*$  will asymptotically approach  $Q \cdot \tau$ . With  $\tau = 17$  hours (AK1),  $V = 600$  km/s and  $B_z = -20$  nT give  $VB_s = 12$  mV/m and  $Q \cdot \tau = -500$  nT. The AK1 model has been further extended by letting  $\tau$  be a function of  $Dst$  and adding components for the diurnal and seasonal variation that are present in  $Dst$  (O'Brien and McPherron, 2002).

The machine learning (ML) approach could be viewed as a set of more general algorithms that can model complex functions. The development of an ML model is more involved and time consuming. For the prediction of the  $Dst$  index, many ML methods have been applied, and we here list some examples using different approaches: neural network with input time delays (Lundstedt and Wintoft, 1994; Gleisner et al., 1996; Watanabe et al., 2002), recurrent neural network (Wu and Lundstedt, 1997; Lundstedt et al., 2002; Pallocchia et al., 2006; Gruet et al., 2018), ARMA (Vassiliadis et al., 1999), and NARMAX (Boaghe et al., 2001; Boynton et al., 2011).

An RNN models dynamical behavior through internal states so that the output depends on both the inputs and the internal state (see, e.g., Goodfellow et al. (2016)) for an overview. Thus, structures that are temporally correlated can be modeled without explicitly parameterizing the temporal dependence; instead, the weights in the hidden layer that connects to the internal state units are adjusted during the training phase. An early RNN was the Elman network (Elman, 1990) which was applied to geomagnetic predictions (Wu and Lundstedt, 1997) and later

implemented for real-time operation (Lundstedt et al., 2002). The Elman RNN can model complex dynamical behavior; however, it was realized that it could be difficult to learn dynamics for systems with long-range memory (Bengio et al., 1994). To overcome that limitation, other RNN architectures were suggested, such as the GRU (Cho et al., 2014) and LSTM (Hochreiter and Schmidhuber, 1997). The LSTM has been applied to geomagnetic predictions of the  $K_p$  (Tan et al., 2018) and  $Dst$  indices (Gruet et al., 2018; Xu et al., 2020). It should be noted that Elman RNN is less complex and has the shortest training times of the three architectures and may be suited for certain problems, and that it is not clear whether there is a general advantage of using GRU or LSTM (Chung et al., 2014; Goodfellow et al., 2016).

In this work, the main goal is to compare the three RNNs: Elman, GRU, and LSTM. The geomagnetic  $Dst$  index is chosen as target as it captures several interesting features of the geomagnetic storm with different temporal dynamics. The initial phase is marked by an increase in  $Dst$  caused by a directly driven pressure increase in the solar wind; the main phase is marked by a sudden decrease in  $Dst$  when solar wind energy enters the magnetosphere through mainly reconnection with southward  $B_z$ , and later, the storm enters the recovery phase when energy is dissipated by internal processes not related to the solar wind condition.

The inputs to the RNNs are solar wind, local time, and time of year. Specifically, past values of  $Dst$  are not used as inputs, although the autocorrelation is very strong (0.98). Clearly, all statistical measures of performance will improve for short lead time predictions when past values of  $Dst$  are used. However, as the solar wind controls the initial and main phases of the storm, the strong autocorrelation is mainly a result of quiet time variation and the relatively slow increase in  $Dst$  during the recovery phase. Another aspect is that for real-time predictions, the variable lead time given by the solar wind must be matched against available real-time  $Dst$  if it is used as input. Also, any errors in real-time  $Dst$  will affect the predictions, and as an example, during the period June–September 2020, the real-time  $Dst$  was offset by about  $-30$  nT. It is also interesting to note that in a recent  $Dst$  prediction competition<sup>1</sup> hosted by NOAA, it was stated that the models “may not take  $Dst$  as an input.”

As the idea here is to compare three RNN architectures that map from solar wind to  $Dst$ , the prediction lead time is not explored. The solar wind data used have been propagated to a location close to Earth, and no further lead time is added; thus, propagated solar wind at time  $t$  is mapped to  $Dst(t)$ . Clearly, possibilities to increase the lead time are of great interest, and many attempts have been made with models driven by measured solar wind (e.g., Gruet et al. (2018); Xu et al. (2020)). However, without any information other than L1 solar wind measurements, the initial phase cannot be predicted with any additional lead time, except that given by the L1-Earth solar wind propagation time, while the main phase may be predicted with possibly up to an additional hour due to magnetospheric processes. The effect of

<sup>1</sup><https://www.drivendata.org/competitions/73/noaa-magnetic-forecasting/page/278/>

forcing models driven by measured solar wind to predict  $K_p$  and  $Dst$  with different lead times was studied in Wintoft and Wik (2018).

## 2 MODELS AND ANALYSIS

### 2.1 Models

A neural network performs a sequence of transforms by multiplying its inputs with a set of coefficients (weights) and applying nonlinear functions to provide the output. It has been shown that the neural network can approximate any continuous function (Cybenko, 1989). For a supervised network, the weights are adjusted to produce a desired output, given the inputs, known as the training phase. The training phase requires known input and target values, a cost function, and an optimization algorithm that minimizes the cost.

The Elman RNN was first developed for  $Dst$  predictions by Wu and Lundstedt (1997) and later implemented for real-time operation (Lundstedt et al., 2002). In this work, we use the term Elman network, but it is the same as the simple RNN used in the Tensorflow package that we use (Abadi et al., 2015). Using linear units at the output layer, the Elman network at time  $t$  is described by

$$y_t = b + \sum_{j=1}^J v_j h_t^j = \mathbf{V}\mathbf{h}_t, \quad (5)$$

$$h_t^j = f \left( a_j + \sum_{i=1}^I w_{ji} x_t^i + \sum_{k=1}^J u_{jk} h_{t-1}^k \right) = f(\mathbf{W}\mathbf{x}_t + \mathbf{U}\mathbf{h}_{t-1})^j, \quad (6)$$

with the output layer bias  $b$ ,  $J$  hidden weights  $v_j$ , nonlinear activation function  $f$ ,  $J$  hidden layer biases  $a_j$ ,  $J \times I$  weights  $w_{ji}$ , and  $J \times J$  recurrent weights  $u_{jk}$ . Note that we use superscripts  $i, j$ , and  $k$  as indices, not powers. The equations can be written more condensed using weight matrices  $\mathbf{W}$  and  $\mathbf{U}$ , where the bias terms ( $a_j, b$ ) have been collected into the matrices and increasing the lengths of  $\mathbf{x}_t$  and  $\mathbf{h}_t$  by adding a constant set to one. For example, in Lundstedt et al. (2002), there are  $I = 3$  inputs ( $B_z, n, V$ ) and  $J = 4$  hidden units.

A minimalistic Elman network can be constructed by using only one input unit and one linear unit in the hidden layer, thus  $b = 0$ ,  $v_1 = 1$ , leading to

$$y_t = a_1 + w_{11}x_t^1 + u_{11}y_{t-1}, \quad (7)$$

which after some rearranging can be written as

$$y_t = y_{t-1} + a_1 + w_{11}x_t^1 - (1 - u_{11})y_{t-1}, \quad (8)$$

which is identical to Eq. 2 for  $\tau = \text{const}$  and  $\Delta t = 1$ , and by letting  $a_1 = 0$ ,  $w_{11}x_t^1 = Q(t)$ , and  $1 - u_{11} = 1/\tau$ . This simple network is trained using the pressure-corrected  $Dst$  index as the target. As the weights in the network are initiated with random values before training begins, there will be some variation in the final weight values if the training is repeated. We find typical values of  $w_{11}$  and  $u_{11}$  corresponding to  $a \in [-2.4, -2.7]$  (Eq. 3) and  $\tau \in [14, 16]$  hours, which are close to the values used by O'Brien and McPherron (2000). However, the algorithm can

get stuck in local minima that results in quite different values. We provide code on Github<sup>2</sup> for the minimalistic Elman network (see Model005.py).

The gated recurrent unit (GRU) neural network (Cho et al., 2014) has a more complex architecture than the Elman network. We implement a single GRU layer, and the output from the network is as before, given by  $y_t = \mathbf{V}\mathbf{h}_t$  (Eq. 5). The GRU layer output at unit  $j$  is

$$h_t^j = z_t^j h_{t-1}^j + (1 - z_t^j) \tilde{h}_t^j, \quad (9)$$

where  $z_t^j$  is the update gate and  $\tilde{h}_t^j$  is the candidate activation. The update gate is defined as

$$z_t^j = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1})^j, \quad (10)$$

where  $\sigma$  is the sigmoid function with output range 0–1. The weight matrix  $\mathbf{W}_z$  operates on the input vector  $\mathbf{x}_t$ , and the matrix  $\mathbf{U}_z$  operates on the past activation  $\mathbf{h}_{t-1}$ . The candidate activation is defined as

$$\tilde{h}_t^j = f(\mathbf{W}\mathbf{x}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1}))^j, \quad (11)$$

where  $f$  is a nonlinear function with two additional weight matrices  $\mathbf{W}$  and  $\mathbf{U}$ . The  $\mathbf{U}$  matrix operates on the past activation weighted by the reset gate

$$r_t^j = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1})^j, \quad (12)$$

which has a further set of weights matrices  $\mathbf{W}_r$  and  $\mathbf{U}_r$ . Clearly, the GRU network is more complex than the Elman network, and it has approximately 3 times more weights than the Elman network for the same number of units. As the update and reset gates have outputs between 0 and 1, we see that when both produce ones [ $(z_t^j, r_t^j) = (1, 1)$ ], the GRU network simplifies to the Elman network. On the other hand, when  $z_t^j = 0$ , no information of the current input  $\mathbf{x}_t$  is used, only the past state  $\mathbf{h}_{t-1}$ . Finally, when  $r_t^j = 0$ , no information of past states goes through the candidate activation (Eq. 11); information on past states only goes through Eq. 9 and is weighted by  $1 - z_t^j$ .

The long short-term memory (LSTM) neural network (Hochreiter and Schmidhuber, 1997) was introduced before GRU and has further complexity with the number of weights approximately four times that of the Elman network, given the same number of units. We will not repeat the equations here but instead refer to for example, Chung et al. (2014). The LSTM has three gating functions, instead of GRU's two, that control the flow of information: the output gate, the forget gate, and the input gate. When they have values 1, 0, and 1, respectively, the LSTM simplifies to the Elman network.

Given a network with a sufficient number of weights, it can be trained to reach zero MSE; however, such a network will have poor generalizing capabilities; that is, it will have large errors on predictions on samples not included in the training data. Different strategies exist to prevent over-fitting (Goodfellow

<sup>2</sup>[https://github.com/spacedr/dst\\_rnn](https://github.com/spacedr/dst_rnn)

**TABLE 1 |** Summary of the five subsets showing the years, number of samples, the mean (nT), standard deviation (nT), and minimum  $Dst$  (nT).

	<b>Years</b>	<b>Count</b>	<b>Mean</b>	<b>Std</b>	<b>Min</b>
1	1995, 2003, 2006, 2010	34,890	-15.0	19.8	-422
2	2001, 2002, 2009, 2011	35,021	-13.2	23.4	-387
3	1998, 2004, 2008, 2012	35,089	-11.3	20.9	-374
4	1996, 2000, 2013, 2015	35,059	-13.0	21.0	-301
5	1997, 1999, 2005, 2014	34,754	-12.7	19.2	-247

et al., 2016). We apply early stopping and dropout (Srivastava et al., 2014; Gal and Ghahramani, 2016).

In order to make a robust estimation of the performance of the networks, we apply  $k$ -fold cross-validation (Goodfellow et al., 2016). During a training session, one subset is held out for testing and the remaining  $k - 1$  subsets are used for training and validation, and out of the  $k - 1$  subsets, one is used for validation and the remaining  $k - 2$  subsets for the training. During training, the validation mean squared error (MSE) is monitored, and the network with lowest validation MSE is chosen (early stopping). In practice, to know that the minimum validation MSE has been reached, the training is continued for a number of epochs after the lowest MSE has been reached. The final evaluation of the models is performed on the  $k$  different test sets (see Section 2.2 to know how the different sets are selected).

## 2.2 Data Sets

The hourly solar wind data and  $Dst$  index are obtained from the OMNI dataset (King and Papitashvili, 2005). The inputs are the solar wind magnetic field magnitude  $B$ , the  $y$ - and  $z$ -components ( $B_y, B_z$ ) in the geocentric solar magnetospheric (GSM) coordinate system, the particle density  $n$ , and speed  $V$ . To provide information on diurnal and seasonal variations (O'Brien and McPherron, 2002), four additional variables are added: the day of year parameterized as sine and cosine of  $2\pi DOY/365$ , and local time as sine and cosine of  $2\pi UT/24$ . Thus, in total, nine input variables. Several previous geomagnetic prediction models also use diurnal and seasonal inputs (e.g., Temerin and Li (2006); Wintoft et al. (2017); Wintoft and Wik, (2018)). Many different coupling functions ( $Q$ ) for the dayside reconnection rate have been suggested and investigated (Borovsky and Birn, 2014), but as the neural network can approximate any function, the exact function does not have to be specified as long as the relevant inputs are available.

The target variable ( $Dst$ ) depends on both the solar wind and past states of the system, where past states can be described by past values of  $Dst$  itself or by past values of the solar wind. We choose to only include solar wind, thereby not relying on past observed or predicted values of  $Dst$ . For the RNN training algorithm, the data are organized so that the past  $T$  solar wind observations are presented at each time step. The input data are thus collected into a  $N \times T \times 9$  tensor, and the target data have  $N \times 1$  dimension, where  $N$  is the number of samples in the set. The input history should be long enough to capture typical storm dynamics, and we found that validation errors leveled

**TABLE 2 |** Selection of subsets for the different cross-validation (CV) sets.

<b>CV</b>	<b>Training</b>	<b>Validation</b>	<b>Test</b>
1	2, 4, 5	3	1
2	1, 4, 5	3	2
3	1, 4, 5	2	3
4	1, 3, 5	2	4
5	1, 3, 4	2	5

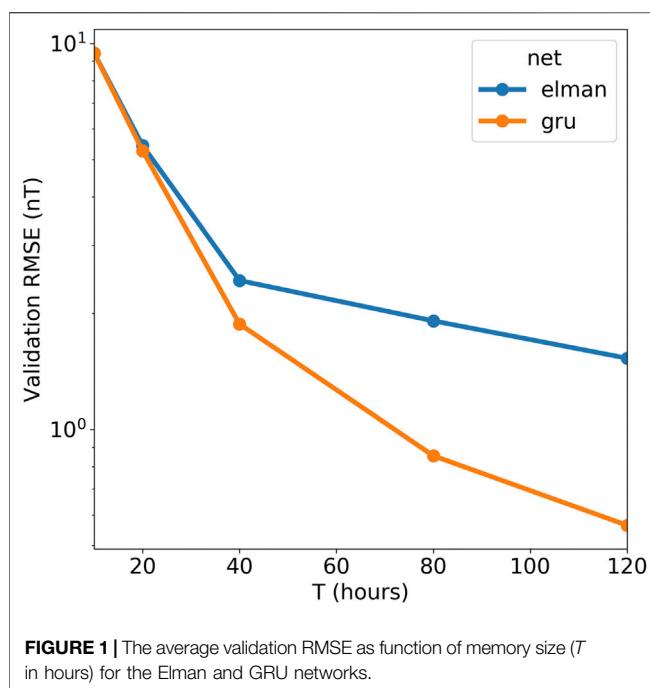
out at  $T = 120$  hours (see also the results regarding  $T$  in Sections 2.3 and 2.4).

To implement the  $k$ -fold cross-validation (CV), the dataset must be partitioned into subsets; we perform a five-fold CV. We choose the five sets to each have similar target ( $Dst$ ) mean and standard deviation so that training, validation, and testing are based on comparable data. If a more blind approach were done, then there is a high risk that training is performed on data dominated by storms, while testing is performed on more quiet conditions. Further, the samples in a subset cannot be randomly selected because there will be considerable temporal overlap between samples of different subsets due to the  $T = 120$  hour window. Instead, we build the subsets from data covering complete years. The data we use cover the years 1995 to 2015, extending over almost two solar cycles and with few solar wind data gaps. We define five subsets based on the data for the years shown in Table 1. The datasets used for training, validation, and test are selected by cycling through the subsets. For the first CV (CV-1) subset, one is selected as test set; subsets two, four, and five for training; and subset three for validation. The process is repeated according to Table 2.

The input and target values span very different numerical ranges, whereas the training algorithm should receive input-target data that have similar numerical ranges. Therefore, the input and target data are normalized, where the normalization coefficients are found from the training set. By subtracting the mean and dividing with the standard deviation for each variable separately, the training set will have zero mean and one standard deviation on all its inputs and target variables. However, as the distributions for each variable are highly skewed, they result in several normalized values with magnitudes much larger than one. Another way to normalize is to instead rescale the minimum and maximum values to the range  $[-1, 1]$ . This guarantees that there will be no values outside this range for the training set. We found that the min-max normalization gave slightly better results, especially at the large values.

## 2.3 Hyperparameters

There are a number of hyperparameters (HPs) that control the model complexity and training algorithm that need to be tuned, but it is not feasible to make an exhaustive search. Initially, a number of different combinations of HP values were manually tested to provide a basic insight into reasonable choices and how the training and validation MSEs vary with epochs. In this initial exploration, we found the Tensorboard (Abadi et al., 2015) software valuable in monitoring the MSE.



**FIGURE 1 |** The average validation RMSE as function of memory size ( $T$  in hours) for the Elman and GRU networks.

The Adam learning algorithm (Kingma and Lei, 2015), which is a stochastic gradient descent method, has three parameters: learning rate  $\epsilon$  and two decay rates for the moment estimates ( $\beta_1, \beta_2$ ). We fix the latter to the suggested values  $\beta_1, \beta_2 = 0.9, 0.999$  and vary  $\epsilon \in [5 \cdot 10^{-4}, 1 \cdot 10^{-3}, 2 \cdot 10^{-3}]$ .

The learning algorithm updates the weights in batches of samples from the training set, where the number of samples in each batch  $N_B$  is much smaller than the total number of training samples ( $N_B \ll N$ ). We test batch sizes of  $N_B \in [32, 64, 128]$ . One training epoch includes approximately  $N/N_B$  training iterations in which the weights are updated at each iteration.

The model capacity is determined by the number of weights and the network architecture. In this work, we have one input layer, a recurrent layer (hidden layer), and a single output. Thus, the capacity is determined by the network type ( $M \in [\text{Elman, GRU, LSTM}]$ ) and the number of hidden units ( $N_H$ ).

The current state ( $\mathbf{h}_t$ ) in the RNN depends on both its inputs ( $\mathbf{x}_t$ ) and the past state ( $\mathbf{h}_{t-1}$ ). For computational performance reasons, past states are not kept indefinitely; instead, there is a limit  $T$  on the length of the memory. We explored  $T \in [48, 72, 96, 120]$  hours and found that the validation MSE decreased with increasing  $T$ , but that it leveled out for large  $T$ . We therefore set  $T = 120$  hours. This also means that any dynamical processes extending past 120 h cannot be modeled internally by the RNN. The choice of  $T$  for the Elman and GRU networks is studied on simulated  $Dst$  data in the next section.

The dropout is controlled by parameters that specify the fraction of network units in a layer that are randomly selected per epoch and temporarily disregarded. The dropout can be applied to all layers: the input layer ( $d_i$ ), the recurrent layer ( $d_r$ ), and the hidden layer ( $d_h$ ). The dropout is a number between 0 and 1, where 0 means all units are included and 1 that all units are unused.

For each combination of HP that we explore, we train three networks initiated with different random weight values as there is no guarantee that the training algorithm will find a good local minimum. The network with the lowest validation error is selected. Note that here validation refers to the split into training and validation sets used during training, which is different from the cross-validation sets that make up the independent test set.

## 2.4 Training Network on Simulated Data

It is interesting to study the RNNs on data generated from a known function relating solar wind to  $Dst$ , and for this purpose, we use the AK1 model (O'Brien and McPherron, 2000). Using the datasets defined in the previous section, we apply the AK1 model to the solar wind inputs and create the target data. Thus, there exists an exact relation between input and output, and the learning process of the RNN will only be limited by the amount of data, network structure (type of RNNs), and network capacity (size of RNNs). We showed that the minimalistic Elman network (Eq. 7) can model the pressure-corrected  $Dst$ . The AK1 model also includes the pressure term, and its inputs are  $B_z$ ,  $n$ , and  $V$ . The five-fold CV is applied to Elman and GRU networks, and we vary the time window  $T$  and the network size  $N_H$ .

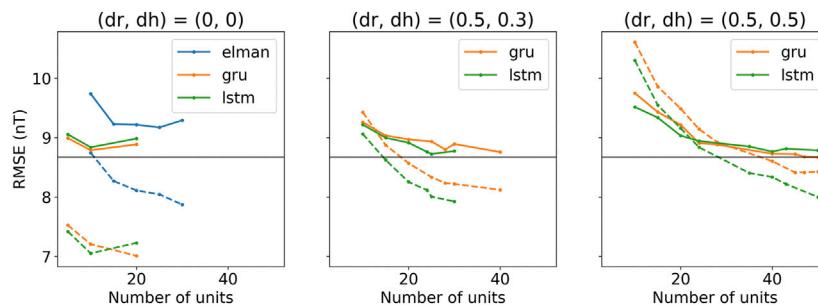
In Figure 1, the validation errors as function of  $T$  are shown for the Elman and GRU networks. At each  $T$ , the optimal networks with respect to  $N_H$  are used. We see that for small  $T$ , the RMSE is large, but it is similar for the two network types. At small  $T$ , only part of the storm recovery phase can be modeled. But as  $T$  is increased, the RMSE becomes much smaller for the GRU network than for the Elman network. It is likely that the Elman network suffers from the vanishing gradient problem (Bengio et al., 1994): the reason for introducing GRU and LSTM networks. We also see that the GRU network reaches an RMSE of lower than 0.6 nT, which could be further decreased by increasing  $T$ . Thus, the GRU network can learn the AK1 model using the observed solar wind data.

## 2.5 Result for the $Dst$ Index

As described in Section 2.2, the inputs to the  $Dst$  model are solar wind magnetic field ( $B, B_y, B_z$ ), density ( $n$ ), and speed ( $V$ ); the day of year parameterized as sine and cosine of  $2\pi\text{DOY}/365$ ; and local time as sine and cosine of  $2\pi\text{UT}/24$ . The DOY and UT are added to model the seasonal and diurnal variations in  $Dst$  (O'Brien and McPherron, 2002).

We perform a search in the hyperparameter space as described above and conclude that training is not very sensitive on the learning rate ( $\epsilon$ ) or batch size ( $N_B$ ), and therefore fix them at  $(\epsilon, N_B) = (10^{-3}, 128)$ .

For each of the five splits, we select the corresponding training set (Tables 1,2), and RNNs are trained with different number of hidden units ( $N_H$ ) and different dropout rates ( $d_i, d_r, d_h$ ). For each combination of  $(N_H, d_i, d_r, d_h)$ , three networks are trained starting from different random initial weights. During training, the validation error is monitored, and the network with lowest validation error is selected. The training is stopped 20 epochs after the minimum validation error has been reached, and the



**FIGURE 2** | The average training (dashed lines) and validation (solid lines) RMSE as function of the number of hidden units ( $N_H$ ) for the Elman, GRU, and LSTM networks. The panels show errors, from left to right, when no dropout is applied ( $d_r = 0, d_h = 0$ ), dropout  $d_r = 0.5, d_h = 0.3$ , and dropout  $d_r = 0.5, d_h = 0.5$ . Note that the dropout on the inputs is  $d_i = 0$ . Gray horizontal line marks the minimum validation RMSE.

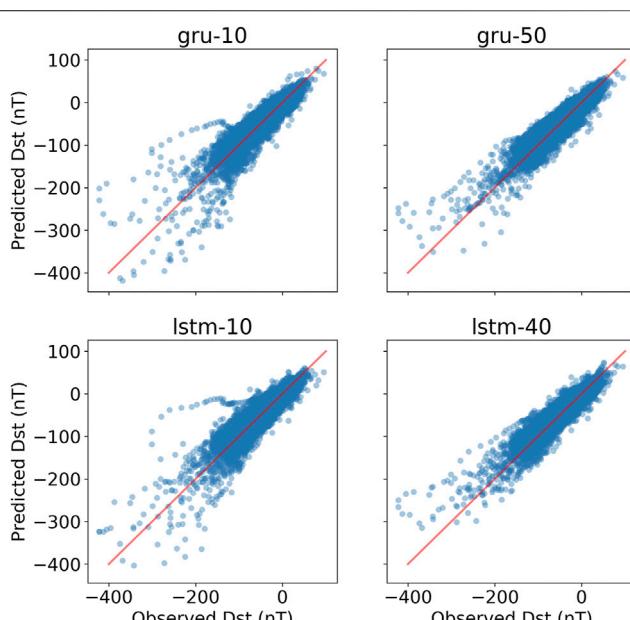
network at minimum validation error is saved. Typically, the minimum validation RMSE is found after 40 to 80 epochs. This results in five different networks for each HP combination that can be tested using the CV approach. **See Appendix for software used and typical training times.**

The coupling function from solar wind to observed  $Dst$  is subject to a number of uncertainties, and to provide a few examples: the solar wind data have been measured at different locations upstream of Earth, mostly from orbit around the L1 location, and then shifted to a common location closer to Earth (King and Papitashvili, 2005); We rely on a point measurement; there may be both systematic and random errors in the derived  $Dst$  index. The uncertainties introduce errors in the input–output mapping, and to reduce their effect and improve generalization,

we apply dropout. From a search of different combinations of  $(N_H, d_i, d_r, d_h)$ , it was found that dropout on the inputs ( $d_i$ ) always led to poor performance, which can be understood as several inputs individually are critical, for example  $B_z$ . Therefore, we set  $d_i = 0$ . The performance improved when dropout was applied on the recurrent and hidden layers. **Figure 2** shows the training and validation RMSE as a function of  $N_H$  for different dropouts. In the case with no dropout (left panel), it is seen that the GRU and LSTM validation errors are similar and significantly below the Elman validation errors. There is also a large gap between the training and validation errors, indicating over-fitting on the training set. When dropout is introduced (middle and right panels), the network size must be increased to reach similar validation RMSE as when no dropout is applied, which is expected as only a fraction of units are active at any one time. But we also see that the gap between training and validation errors decreases. We also applied dropout on the Elman network, but the validation errors became large when  $d_r > 0$ ; therefore, the results are not included in the middle and right panels. When  $d_r = 0.5$  and  $d_h = 0.5$  (right panel), the optimal GRU and LSTM networks have  $N_H = 50$  and  $N_H = 40$ , respectively. In terms of the number of weights, they are of similar sizes, 9,051 and 8,041, respectively. When dropout is applied, the number of active weights drops to 2,651 and 2,421.

For each CV set, we select the GRU and LSTM networks with minimum validation RMSE with and without dropout, and run them and collect the 5 CV sets into one set. We thereby get an estimate of the generalization performance for the whole 1995 to 2015 period. **Figure 3** shows scatterplots of predicted  $Dst$  vs. observed  $Dst$  on the test sets for different networks. **Table 3** summarizes the performance on the training, validation, and test sets. The 95% confidence intervals have been estimated by both assuming independent data points and taking into account the autocorrelation (Zwiers and von Storch, 1995). It is clear that using dropout significantly improves the generalization capability. We also see that there is no significant difference between the GRU and LSTM networks. The bias (mean of errors) and linear correlation coefficient are computed on the test set and shown in **Table 4**.

The performance of the networks varies with the level of  $Dst$ ; the errors have a tendency to increase with the magnitude of  $Dst$ .



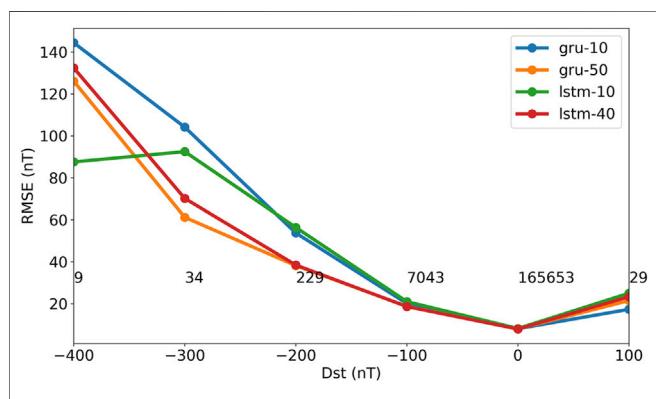
**FIGURE 3** | Scatterplots of predicted vs. observed  $Dst$  based on the five CV test sets. The left panels show predictions without dropout using GRU and LSTM networks (gru-10 and lstm-10), while the right panels are predictions based on networks trained using dropout of  $(d_r, d_h) = (0.5, 0.5)$  (gru-50 and lstm-40).

**TABLE 3 |** Training, validation, and test RMSE (nT) for networks with and without dropout. The training and validation RMSE are averages over the five CV splits, while the test RMSE is computed from the combined five CV test sets. Networks with  $N_H = 10$  have no dropout and the larger networks have dropout  $(d_r, d_h) = (0.5, 0.5)$ . The 95% RMSE confidence interval is approximately  $\pm 0.03$  nT assuming independent errors but increases to  $\pm 0.17$  nT if the autocorrelation is taken into account.

Net	$N_H$	Train	Val	Test
GRU	10	7.21	8.79	9.24
GRU	50	8.43	8.67	8.85
LSTM	10	7.06	8.84	9.37
LSTM	40	8.34	8.77	8.81

**TABLE 4 |** BIAS, RMSE, and CORR for the GRU and LSTM models on the test set. BIAS and RMSE are in units of nT. The 95% confidence intervals are  $\pm 0.04$  to  $\pm 0.2$  nT for BIAS,  $\pm 0.03$  to  $\pm 0.17$  nT for RMSE (same as **Table 3**), and  $\pm 0.001$  to  $\pm 0.01$  for CORR. The lower limits assume independence and the higher limits take into account the autocorrelations.

	Gru-10	Gru-50	Lstm-10	Lstm-40
BIAS	-0.41	-0.10	-0.59	0.16
RMSE	9.24	8.85	9.37	8.81
CORR	0.89	0.90	0.89	0.90



**FIGURE 4 |** The test RMSE binned by observed  $Dst$ . The RMSE is computed on the five CV test sets. Bins are 100 nT wide, and the numbers show the number of samples in each bin. Legend: GRU (gru-10) and LSTM (lstm-10) networks without dropout, and GRU (gru-50) and LSTM (lstm-40) networks with dropout  $d_r = 0.5, d_h = 0.5$ .

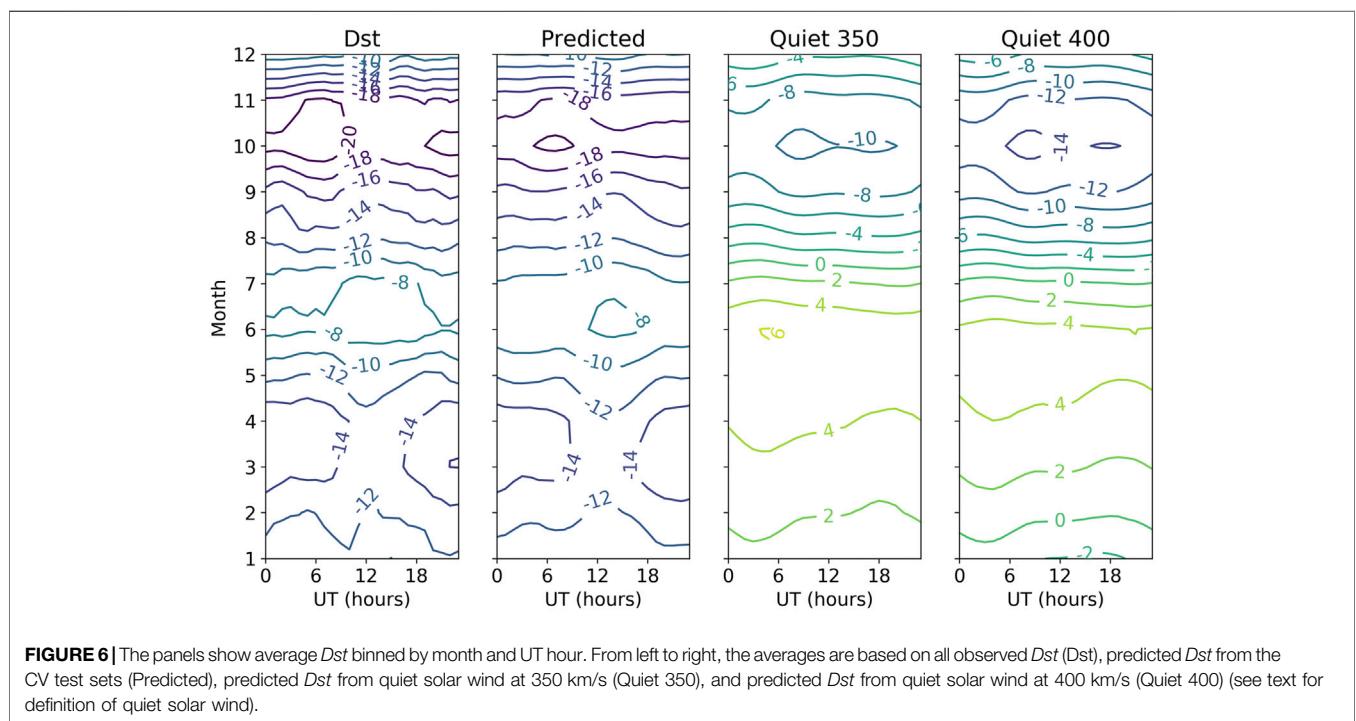
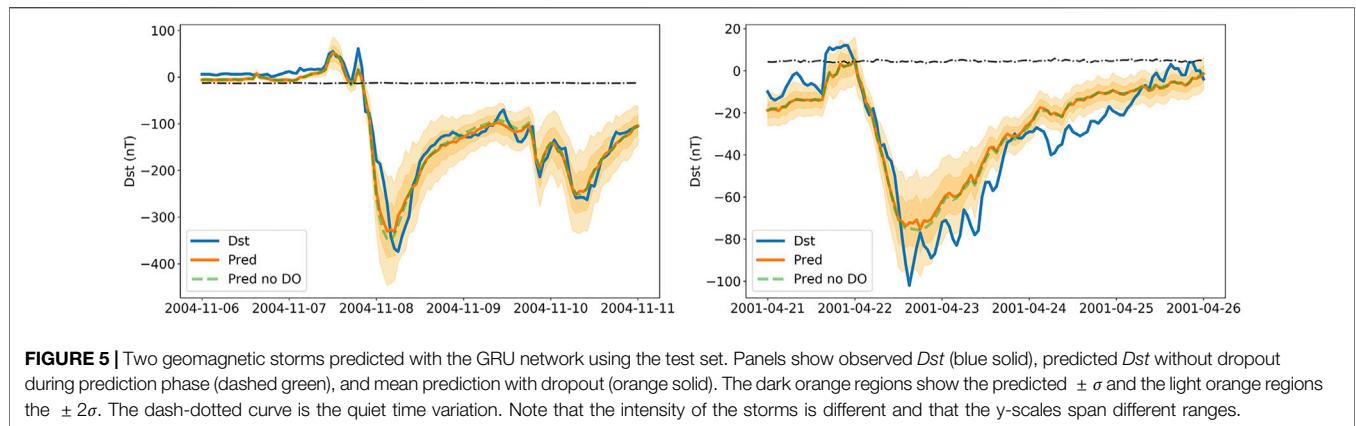
**Figure 4** shows the RMSE binned by observed  $Dst$ . Down to  $Dst = -300$  nT, the networks with dropout have the lowest RMSE. The bin at  $Dst = -400$  nT has too few samples to be interpretable. The main reason the errors increase with the magnitude of  $Dst$  is that there are very few samples around the extremes; thus, the uncertainty of the function estimation will be large. In Wintoft et al. (2017), this problem was addressed by using an ensemble of networks; the predictions from several networks with different weights were averaged. In this work, we study the use of dropout not only in the training phase but also in the prediction phase. The algorithm that temporarily cancels

units at random during training can also be applied during prediction. This means that there is practically an indefinitely number of weight combinations that can be used to produce an arbitrary number of predictions at each time step. For the GRU network with  $(N_H, d_r, d_h) = (50, 0.5, 0.5)$ , there are more than  $10^{28}$  possible combinations. There is a Bayesian interpretation of dropout as the estimation of model uncertainty (Gal and Ghahramani, 2016b). The idea is that the weights are random variables leading to a distribution of predictions for fixed inputs. For each sample, a large number of predictions can be generated, which randomly use different combinations of network units. For each sample, we generate 100 predictions and compute the mean and standard deviation. **Figure 5** shows two examples, the first a severe geomagnetic storm and the second a major storm. The mean predictions with dropout come close to the predictions without dropout. The prediction uncertainty is small during quiet times ( $Dst$  close to zero) and increases with storm magnitude. Again, this is a result of the greater uncertainty in parameter estimates in regions which are poorly sampled.

As time of day and season are included in the inputs, the network can model diurnal and seasonal variations in  $Dst$ . These variations are not strong, and the left panel in **Figure 6** shows  $Dst$  for all years averaged over month and UT hour. Running the GRU networks on the test data from the five CV sets reveals a very similar pattern (second panel from left). Thus, the network shows similar long-term statistics considering that it is driven only by solar wind and time information. The two left panels contain contributions from all levels of  $Dst$  from quiet conditions to storm conditions. But we may now simulate solar wind conditions that we can define as quiet conditions. The two right panels show predicted  $Dst$ , assuming solar wind flowing out from the Sun (GSEQ system) along the Parker spiral with a  $45^\circ$  angle at L1 at two different speeds, 350 km/s and 400 km/s, respectively. In this configuration,  $B_z = 0$  in the solar coordinate system, but via geometric effects (Sun's and Earth's tilts with respect to the ecliptic and Earth's dipole tilt),  $B_z$  will be nonzero in the GSM system showing diurnal and seasonal variations (Lockwood et al., 2020).

### 3 DISCUSSION AND CONCLUSION

There is a close correspondence between Elman networks and models expressed in terms of the differential equation for the  $Dst$  index. A minimalistic Elman network trained on simulated data from the pressure-corrected  $Dst$  index (**Eq. 1**) results in weights that translate to values around  $a = 2.45$  and  $\tau = 15$ , close to those used in **Eqs 2,3**. However, using solar wind data from the years 1995–2015 and targeting simulated  $Dst$  from the AK1 model, we find that the RMSE for Elman network basically levels out for temporal history of  $T \gtrsim 40$  hours. This is not the case for the GRU network, which has similar RMSE up to  $\lesssim 20$  hours but continues to improve for  $T > 20$  hours. We interpret this as an effect of the vanishing gradient problem (Bengio et al., 1994) that is solved in the GRU and LSTM networks. It should be noted that the Elman network takes less time to train, and if the dynamics of the system can be captured in less than about 20 time steps, then the Elman



network could be sufficient. In the future, it would be interesting to perform similar experiments for other solar terrestrial variables, for example, other geomagnetic indices with different temporal dynamics. Another line of experimenting could be to separate processes with different dynamics in the construction of the RNN.

The GRU (Cho et al., 2014) and LSTM (Hochreiter and Schmidhuber, 1997) networks include gating units that control information flow through time. However, it is not clear if one architecture is better than the other (Chung et al., 2014). In order to reliably study the differences between the two RNNs, we applied five-fold cross-validation. Further, it was also essential to apply dropout (Gal and Ghahramani, 2016) to reduce over-fitting and achieve consistent results. Using solar wind data and observed  $Dst$  from 1995 to 2015, we see no significant difference between the two

architectures. However, the GRU network is slightly less complex than the LSTM and will therefore have shorter training times.

An interesting effect of using dropout is that it can also be applied during the prediction phase as a way of capturing model uncertainty (Gal and Ghahramani, 2016b). Using dropout during prediction is similar to ensemble prediction based on a collection of networks with identical architectures but different specific weights (Wintoft et al., 2017), but with the great advantage that the predictions can be based on, in principle, unlimited number of models. However, it is different from using an ensemble of different types of models like in Xu et al. (2020). We illustrated the prediction uncertainty using dropout for a couple of storms from the test set. Estimating the prediction uncertainty is important and was addressed by Gruet et al. (2018) using a combination of LSTM network and a Gaussian process (GP) model. In that case, the LSTM network provides the mean function to

the GP model from which a distribution of prediction can be made. For future work, it will be interesting to further study the use of dropout for estimating model uncertainty.

Predictions based on the test sets using the GRU networks show very good agreement with observed *Dst* when averaged over month and UT (**Figure 6**). The semiannual variation (Lockwood et al., 2020) is clear, with a deeper minimum in autumn than in spring and a weak UT variation. It is a combination of geometrical effects that cause the asymmetric semiannual variation leading to a modulation of the  $B_z$  component in the GSM system, and, together with the nonlinear solar wind–magnetosphere coupling, gives rise to the variation in *Dst*. The two rightmost panels in **Figure 6** show predictions based on simulated data with  $B_z = 0$  in the GSEQ system using two different speeds. In these cases, the semiannual variation is only caused by geometrical effects, while the two panels to the left also contain storms caused by different solar wind disturbances like coronal mass ejections. We also see that the difference between the spring and autumn minima is about 6 nT for both observed and predicted *Dst*, while the difference is about 14–18 nT for quiet time-simulated *Dst*. In this work, we only showed that the semiannual variation is reproduced by the

simulations, but for the future, other types of simulations that contain CME structures could be performed to provide further insights into the semiannual variations.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: <https://omniweb.gsfc.nasa.gov/ow.html>.

## AUTHOR CONTRIBUTIONS

PW and MW have carried out this work with main contribution from PW.

## ACKNOWLEDGMENTS

We acknowledge the use of NASA/GSFC's Space Physics Data Facility's OMNIWeb (or CDAWeb or ftp) service and OMNI data.

## REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.
- Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* 5 157–166. doi:10.1109/72.279181
- Boaghe, O. M., Balikhin, M. A., Billings, S. A., and Alleyne, H. (2001). Identification of Nonlinear Processes in the Magnetospheric Dynamics and Forecasting of *Dst* Index. *J. Geophys. Res.* 106, 30047–30066. doi:10.1029/2000ja900162
- Borovsky, J. E., and Birn, J. (2014). The Solar Wind Electric Field Does Not Control the Dayside Reconnection Rate. *J. Geophys. Res. Space Phys.* 119 751–760. doi:10.1002/2013JA019193
- Boynton, R. J., Balikhin, M. A., Billings, S. A., Sharma, A. S., and Amarutei, O. A. (2011). Data Derived Narmax *Dst* Model. *Ann. Geophys.* 29 965–971. doi:10.5194/angeo-29-965-2011
- Burton, R. K., McPherron, R. L., and Russell, C. T. (1975). An Empirical Relationship between Interplanetary Conditions and *Dst*. *J. Geophys. Res.* 80 4204–4214. doi:10.1029/ja080i031p04204
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). “Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (Doha, Qatar: Association for Computational Linguistics) 1724–1734. doi:10.3115/v1/D14-1179
- Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *NIPS 2014 Workshop on Deep Learning*, December 2014.
- Cybenko, G. (1989). Approximation by Superposition of a Sigmoidal Function. *Math. Control Signals Syst.* 2 303–314. doi:10.1007/bf02551274
- Elman, J. L. (1990). Finding Structure in Time. *Cogn. Sci.* 14 179–211. doi:10.1207/s15516709cog1402\_1
- Gal, Y., and Ghahramani, Z. (2016a). “A Theoretically Grounded Application of Dropout in Recurrent Neural Networks,” in 30th Conference on Neural Information Processing Systems (Barcelona, Spain: NIPS).
- Gal, Y., and Ghahramani, Z. (2016b). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of the 33rd International Conference on Machine Learning*, New York, NY, United States, 2016, (. JMLR: W&CP), 48.
- Gleisner, H., Lundstedt, H., and Wintoft, P. (1996). Predicting Geomagnetic Storms from Solar-Wind Data Using Time-Delay Neural Networks. *Ann. Geophys.* 14 679–686. doi:10.1007/s00585-996-0679-1
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning* (MIT Press)
- Gruet, M. A., Chandorkar, M., Sicard, A., and Camporeale, E. (2018). Multiple-hour-ahead Forecast of the *Dst* Index Using a Combination of Long Short-Term Memory Neural Network and Gaussian Process. *Space Weather* 16, 1882, 1896. doi:10.1029/2018SW001898
- Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* 9 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Hunter, J. D. (2007). Matplotlib: A 2d Graphics Environment. *Comput. Sci. Eng.* 9 90–95. doi:10.1109/MCSE.2007.55
- King, J. H., and Papitashvili, N. E. (2005). Solar Wind Spatial Scales in and Comparisons of Hourly Wind and Ace Plasma and Magnetic Field Data. *J. Geophys. Res.* 110, A02104. doi:10.1029/2004JA010649
- Kingma, D. P., and Lei, Ba. J. (2015). “Adam: A Method for Stochastic Optimization,” in The 3rd International Conference on Learning Representations (ICLR), arXiv:1412.6980.
- Lockwood, M., Owens, M. J., Barnard, L. A., Haines, C., Scott, C. J., McWilliams, K. A., et al. (2020). Semi-annual, Annual and Universal Time Variations in the Magnetosphere and in Geomagnetic Activity: 1. Geomagnetic Data. *J. Space Weather Space Clim.* 10 . doi:10.1051/swsc/2020023
- Lundstedt, H., Gleisner, H., and Wintoft, P. (2002). Operational Forecasts of the geomagnetic *Dst* index. *Geophys. Res. Lett.* 29, 34. doi:10.1029/2002GL016151
- Lundstedt, H., and Wintoft, P. (1994). Prediction of Geomagnetic Storms from Solar Wind Data with the Use of a Neural Network. *Ann. Geophys.* 12, 19–24. doi:10.1007/s00585-994-0019-2
- Mayaud, P. N. (1980). Derivation, Meaning, and Use of Geomagnetic Indices, *Geophysical Monograph*, 22 (American Geophysical Union).
- O’Brien, T. P., and McPherron, R. L. (2000). Forecasting the Ring Current *Dst* in Real Time. *J. Atmos. Solar-Terrestrial Phys.* 62 1295–1299.
- O’Brien, T. P., and McPherron, R. L. (2002). Seasonal and Diurnal Variation of *Dst* Dynamics. *J. Geophys. Res.* 107, 1341. doi:10.1029/2002JA009435
- Pallocchia, G., Amata, E., Consolini, G., Marcucci, M. F., and Bertello, I. (2006). Geomagnetic *Dst* Index Forecast Based on IMF Data Only. *Ann. Geophys.* 24, 989–999. doi:10.5194/angeo-24-989-2006

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a Simple Way to Prevent Neural Networks from Overfitting. *J. Machine Learn. Res.* 15, 1929–1958.
- Tan, Y., Hu, Q., Wang, Z., and Zhong, Q. (2018). Geomagnetic Index Kp Forecasting with Lstm. *Space Weather* 16, 406–416. doi:10.1002/2017SW001764
- Temerin, M., and Li, X. (2006). Dst Model for 1995–2002. *J. Geophys. Res.* 111, A04221. doi:10.1029/2005JA011257
- T. pandas-dev/pandas: Pandas (2020). pandas development team. doi:10.5281/zenodo.3509134[Dataset]
- Vassiliadis, D., Klimas, A. J., Valdivia, J. A., and Baker, D. N. (1999). The Dst Geomagnetic Response as Function of Storm Phase and Amplitude and the Solar Wind Electric Field. *J. Geophys. Res.* 104, 957–976. doi:10.1029/1999ja900185
- Watanabe, S., Sagawa, E., Ohtaka, K., and Shimazu, H. (2002). Prediction of the Dst Index from Solar Wind Parameters by a Neural Network Method. *Earth Planets Space* 54, 1263–1275. doi:10.1186/bf03352454
- Wintoft, P., and Wik, M. (2018). Evaluation of Kp and Dst Predictions Using Ace and Dscovr Solar Wind Data. *Space Weather* 16, 1972–1983. doi:10.1029/2018SW001994
- Wintoft, P., Wik, M., Matzka, J., and Shprits, Y. (2017). Forecasting Kp from Solar Wind Data: Input Parameter Study Using 3-hour Averages and 3-hour Range Values. *J. Space Weather Space Clim.* 7, A29. doi:10.1051/swsc/2017027
- Wu, J. G., and Lundstedt, H. (1997). Neural Network Modeling of Solar Wind-Magnetosphere Interaction. *J. Geophys. Res.* 102, 14457–14466. doi:10.1029/97ja01081
- Xu, S. B., Huang, S. Y., Yuan, Z. G., Deng, X. H., and Jiang, K. (2020). Prediction of the Dst Index with Bagging Ensemble-Learning Algorithm. *Astrophysical J. Suppl. Ser.* 248. doi:10.3847/1538-4365/ab880e
- Zwiers, F. W., and von Storch, H. (1995). Taking Serial Correlation into Account in Tests of the Mean. *J. Clim.* 8, 336–351. doi:10.1175/1520-0442(1995)008<0336:tsciai>2.0.co;2

**Conflict of Interest:** 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.

Copyright © 2021 Wintoft and Wik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX: SOFTWARE AND HARDWARE

The code has been written in Python where we rely on several software packages: Pandas for data analysis (T. pandas-dev/pandas: Pandas, 2020); Matplotlib for plotting (Hunter, 2007); and TensorFlow and TensorBoard for RNN training (Dataset] Abadi et al., 2015).

The simulations have been run on an Intel Core i9-7960X CPU at 4.2 GHz with 64 GB memory. In total, 32 threads can be

run in parallel. Typical training time for one Elman network with 30 hidden units for 50 epochs ranges between 5 and 15 min, where the shorter time is due to that the process could be distributed on multiple threads. We noted that one training process could be distributed over four threads, when the overall load was low. A GRU network with 10 hidden units could take between 30 min and slightly more than 1 h for 50 epochs. A 10-hidden unit LSTM network ranged between 50 min and 1.5 h.



# Concurrent Empirical Magnetic Reconstruction of Storm and Substorm Spatial Scales Using Data Mining and Virtual Spacecraft

Grant K. Stephens\* and Mikhail I. Sitnov

Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, United States

## OPEN ACCESS

### Edited by:

Enrico Camporeale,  
University of Colorado Boulder,  
United States

### Reviewed by:

Xiangning Chu,  
University of Colorado Boulder,  
United States

Jacob Bortnik,

University of California, Los Angeles,  
United States

### \*Correspondence:

Grant K. Stephens  
grant.stephens@jhuapl.edu

### Specialty section:

This article was submitted to  
Space Physics,  
a section of the journal  
Frontiers in Physics

Received: 13 January 2021

Accepted: 06 April 2021

Published: 13 May 2021

### Citation:

Stephens GK and Sitnov MI (2021)  
Concurrent Empirical Magnetic  
Reconstruction of Storm and  
Substorm Spatial Scales Using Data  
Mining and Virtual Spacecraft.  
*Front. Phys.* 9:653111.  
doi: 10.3389/fphy.2021.653111

Data mining (DM) has ushered in a new era of empirical magnetic reconstructions of the magnetosphere via application of the k-nearest neighbors (kNN) method. In this approach, the combined magnetosphere storm-substorm state is characterized by the *Sym-H* and *AL* indices, their time derivatives, and the solar wind electric field  $vB_z^{IMF}$ . However, using the DM reconstructions to account for the substorm contributions to the ring current as well as describing storm-time substorms remains a problem. The inner region  $r \leq 12R_E$ , where the ring current develops, has a much higher density of data than the tail region  $12R_E \leq r \leq 22R_E$ , where substorms operate. This results in two models inconsistent in their scales dictated by the corresponding data densities. The inner model reconstructs storm time dynamics, including the formation of the westward and eastward ring current and pressure distributions. The outer model captures substorm features, including the thinning and rapid dipolarization of the tail sheet during the growth and expansion phases, respectively. However, the substorm model is insufficient to reconstruct the eastward ring current while the storm model cannot fully reproduce substorm effects because it overfits in the tail region. This issue is addressed by constructing a hybrid model which is fit using virtual magnetic field observations generated by sampling the other two models. The resulting merged resolution model concurrently captures the spatial scales associated with both storms in the inner region and substorms in the near-tail region. Hence it is particularly useful for investigation of the storm-substorm relationship, including storm-time substorms and the impact of individual substorm injections to the buildup of the storm-time ring current.

**Keywords:** substorm, storm, data mining, empirical model, magnetotail, ring current

## 1. INTRODUCTION

Storms and substorms represent two major modes of magnetospheric activity and the resulting space weather (e.g., [1, 2]), which is reflected in the corresponding low- and high-latitude geomagnetic indices, e.g., *Sym-H* [3] and *AL* [4]. Substorms frequently occur during the development of storm main phases and to a lesser extent during the recovery phase [5]. Although the original idea of substorms viewed them as building blocks of storms [6], which is reflected in the names of these phenomena, that view is now strongly modified [7, 8], as it is clear that storms and substorms are related because of their common driver, the solar wind [9, 10]. Despite this,

storm-time *in-situ* observations during substorm expansions show charged particles injected deep into the inner magnetosphere [11, 12]. Furthermore, statistical analyses of particle measurements from the RBPSICE and HOPE instruments on the Van Allen Probes identified an energization of the ring current during substorm expansions [13, 14]. Additionally, empirical magnetic field models found that the substorm expansion phase was correlated with an increase in the amount of current contained within the symmetric ring current (SRC) during both storm and non-storm substorms [15]. As both types of disturbances have dramatic impacts on the global configuration of the geomagnetic field and associated current systems, a concurrent description of storms and substorms and their phases is needed.

Since the development of the earliest empirical magnetic field models [16], the storm state of the magnetosphere was considered in their construction by binning the magnetometer data by storm activity level. Over time, empirical models became more complex in their description of the storm state, by incorporating additional current systems and making them functionally dependent on storm activity indices and solar wind conditions [17–20]. Data mining (DM) ushered in an entirely different approach to the empirical modeling of storms, the first of which was termed the TS07D model [21, 22]. The DM technique refits the model for each snapshot in time to a small subset of the entire magnetometer database. This subset of data is identified by mining the whole database for other time intervals when the magnetosphere was presumably in a similar storm state configuration, characterized using the storm index *Sym-H*, its derivative, and the solar wind electric field parameter  $vB_z^{IMF}$ . The DM algorithm employed in TS07D is the k-Nearest Neighbor (kNN) method [23], which is described in section 2.2. The first applications of the TS07D model were to investigate the global magnetic field and current system configuration of storms driven by different solar wind phenomena [22, 24].

In contrast to the storm state, until recently, inclusion of the description of the substorm state of the magnetosphere within empirical magnetic field models has been limited. Event-oriented approaches sought to describe the substorm growth phase by incorporating a magnetotail thin current sheet (TCS) into storm models and then adjusting the TCS to match particular event observations [25, 26]. Wire-models attempted to describe the substorm expansion by hand tailoring the substorm current wedge (SCW) [27–29]. Showcasing the flexibility of the DM approach, TS07D was customized to picture geomagnetic substorms [15, 30] by including the substorm index *AL* and its time derivative as binning parameters and by incorporating a TCS into the model structure. Termed the SST19 model, it was successful in reconstructing the primary substorm features of the magnetosphere including the stretching of the magnetotail associated with the enhancement of the TCS in the near-tail during the growth phase and the rapid dipolarization of the magnetotail along with the formation of the substorm current wedge during the expansion phase. It also revealed a connection between the substorm expansion phase and the ring current enhancement, seen as  $\approx 1$  MA increase in the dayside ring current during a non-storm substorm.

However, the resolution of the SST19 model was insufficient to fully resolve the innermost magnetosphere, in particular, it was unable to reconstruct the eastward component of the ring current. Earthward ( $r \lesssim 4R_E$ ) of the dominant westward ring current, the azimuthal component of the ring current changes sign becoming eastwardly oriented [31]. Assuming force-balance and pressure isotropy, the boundary between the eastward and westward oriented ring current identifies the location of the plasma pressure peak [32].

An advantage of the DM approach, which reflects a postulate of machine learning techniques, is that the addition of more high quality data sets allows for increasing model complexity. Thus, an obvious remedy to enable the model to more fully resolve the inner magnetosphere is to increase the resolution of the equatorial field, taking advantage of the numerous spacecraft missions that fly through this region (e.g., Van Allen Probes, THEMIS, Polar, Cluster). Indeed, several applications of the TS07D family of models have demonstrated this [33–36], but unlike SST19, the focus of these studies was largely limited to the inner magnetosphere. As this study reveals, the higher equatorial resolution used for those inner magnetosphere investigations overfits the near-tail, introducing numerous artifacts in the reconstruction of the magnetic fields and electric currents. On one hand, this is a key advantage of the TS07D approach, that is, the resolution of the model can be customized to the particular region of interest, but on the other hand, it is also a shortcoming in that no single resolution is adequate for the entirety of the spatial domain of the magnetosphere. This dilemma is discussed in detail in section 2.4.

In section 3 we present a simple yet effective solution for how to concurrently reconstruct both the inner magnetosphere necessary for storm features and the near-tail needed for the description of substorms. Two separate models are constructed with varying equatorial resolutions, one customized for the inner and the other for the near-tail regions. Both models are then sampled in their respective regions to synthesize a distribution of virtual magnetic field observations dense enough to fit a third model, effectively merging the two other models while maintaining the divergenceless property of the magnetic and current density fields. The focus of this study will be a pair of non-storm time substorms that occurred on 8 March 2009. These were previously analyzed by Stephens et al. [15] and as such represent a good test case for comparing the merged model to the SS19 model, as is described in section 3.2.

## 2. USING DATA MINING TO EMPIRICALLY PICTURE THE MAGNETOSPHERE

### 2.1. Magnetic Field Architecture

Empirical magnetic field models are designed with two primary considerations: the spatial structure of the current systems and their dynamical evolution over time. For describing the spatial structure, it is useful to model each current system individually as a sub-model termed a module [20]. The total magnetic field  $\mathbf{B}_{tot}$  is then the sum of magnetic field of each module along with the internal field, e.g.,  $\mathbf{B}_{tot} = \mathbf{B}_{FAC} + \mathbf{B}_{PRC} + \mathbf{B}_{SRC} +$

$\mathbf{B}_{tail} + \mathbf{B}_{MP} + \mathbf{B}_{int}$  corresponding to the magnetic field from the field-aligned currents (FACs), partial and symmetric ring current (PRC and SRC), the cross-tail current, and the magnetopause (or Chapman-Ferraro) currents, respectively. The internal field,  $\mathbf{B}_{int}$ , is readily determined with ground magnetometers and as such is not in the scope of this research and the IGRF model [37] is used to represent it. Earlier models sought to define these modules by hand, crafting a mathematical description based on the theoretical picture of the current system (e.g., [19]). Each of these modules will have non-linear parameters that determine their spatial scales and linear amplitude coefficients controlling their intensity. For example, a magnetic field module could take the form  $\mathbf{B}'(\mathbf{r}) = a_1 \mathbf{B}(\mathbf{r}; \beta_1)$ , where  $a_1$  is the amplitude coefficient and  $\beta_1$  is a non-linear parameter defining the module's mathematical structure (e.g., the current system's spatial scale size or thickness). The total model's set of  $a_i$ s and  $\beta_j$ s are then fit to the available magnetometer data [20]. The dynamical evolution of the current systems can thus be introduced by simply making  $a_i$ s and  $\beta_j$ s functions of time. Some of the earliest models achieved this straightforwardly by binning the magnetometer data by the  $K_p$  storm index and performing separate fits for each bin [16, 38]. The proceeding models instead opted to make  $a_i$ s and  $\beta_j$ s as functions of solar wind conditions and geomagnetic indices [17, 18]. Again, the mathematical structure of the functions was hand-tailored.

The TS07D [22] and derivative models [15, 33–35] utilized a wholly different approach that sought to eliminate many of the hand tailored elements, motivated by the principle that the data should dictate the model instead. First, all the equatorial field modules (SRC, PRC, and tail current) were replaced by a single regular expansion that had no predefined azimuthal or radial structure derived from the general magnetic vector potential solution of a thin current sheet in the cylindrical coordinate system [21] taking the form:

$$\mathbf{B}_{sheet}(\rho, \phi, z) = \sum_{n=1}^N a_{0n}^{(s)} \mathbf{B}_{0n}^{(s)} + \sum_{m=1}^M \sum_{n=1}^N (a_{mn}^{(o)} \mathbf{B}_{mn}^{(o)} + a_{mn}^{(e)} \mathbf{B}_{mn}^{(e)}) \quad (1)$$

where  $\mathbf{B}_{0n}^{(s)}$ ,  $\mathbf{B}_{mn}^{(o)}$ , and  $\mathbf{B}_{mn}^{(e)}$  are the basis functions with azimuthally symmetric, odd (sine) symmetry, and even (cosine) symmetry, respectively; while  $a_{mn}^{(s)}$ ,  $a_{mn}^{(o)}$ , and  $a_{mn}^{(e)}$  are the corresponding amplitude coefficients determined in the fitting procedure.  $M$  represents the number of azimuthal harmonics (odd/even pairs) and  $N$  determines the number of radial (Bessel functions) harmonics used in the expansion. The thickness of the current sheet comes about by substituting  $z$  with  $\zeta = \sqrt{z^2 + D^2}$  in the magnetic vector potential solution, introducing  $D$  as the characteristic half-thickness parameter. The SST19 model expanded upon this approach by including two such systems, one for the thick current sheet and one for the TCS, giving:

$$\mathbf{B}^{(eq)}(\rho, \phi, z) = \mathbf{B}^{(eq)}(\rho, \phi, z; D) + \mathbf{B}^{(eq)}(\rho, \phi, z; D_{TCS}) \quad (2)$$

where  $D_{TCS}$  is the half-thickness for the TCS. This TCS system is key for reconstructing the enhancement of a thin

cross tail current sheet which acts to thin and stretch the magnetotail during the substorm growth phase [15]. Later storm investigations also found the TCS facilitated in reconstructing the eastward ring current during quiet and weak storm times [35].

The FAC module is similarly mathematically described using a regular expansion, in this case a Fourier series in local time [39] which is duplicated and initialized to different latitudes to mimic an expansion [34] in both local time and latitude. The SST19 configuration is used here, which employs four different latitudes with the first four Fourier harmonics, totaling 16 total basis functions which describe the FAC field structure [15]. This factor of four increase over the original TS07D configuration proved critical in reconstructing a realistic FAC morphology associated with substorms (e.g., [40]). The FAC current sheets are bent to flow along approximately dipolar field lines [41] and are allowed to expand and contract by introducing two global rescaling factors  $\kappa_{R1}$  and  $\kappa_{R2}$ .  $\kappa_{R1}$  applies to the 8 basis functions at higher latitudes (region-1 or R1 FACs) while the  $\kappa_{R2}$  corresponds to the 8 at lower latitudes (region-2 or R2 FACs).

Each current system along with  $\mathbf{B}_{int}$  is given a complementary shielding field together represented as  $\mathbf{B}_{MP}$  which acts to contain  $\mathbf{B}_{tot}$  within the magnetopause boundary [20]:  $\mathbf{B}_{tot} \cdot \mathbf{n}|_S = 0$ , where  $S$  is the modeled magnetopause boundary [42].

## 2.2. kNN Method Application

The second key element of the TS07D model is the DM approach. DM identifies time intervals when the state of the magnetosphere was in a similar global configuration as the moment of interest. These time intervals are intersected with the historical magnetometer database to form a subset of data that is then used to fit the model's  $a_i$ s and  $\beta_j$ s. The procedure is repeated for each step in time giving the  $a_i$ s and  $\beta_j$ s their time dependence, thus fulfilling the motivating principle that the data should dictate the model's structure and dynamic evolution. The DM algorithm employed is k-nearest neighbors (kNN), where the global configuration of the magnetosphere and its dynamics are assumed to be represented by some finite dimensional state-space constructed from global parameters, such as geomagnetic indices or solar wind conditions and their time derivatives [43].

As the magnetosphere evolves in time, it traces curves, represented by the state-vector  $\mathbf{G}(t)$ , within state-space. Here, as with previous studies, the state-vector is discretized to a 5 min cadence, forming a cloud of points. Similar dynamical events (such as storms and substorms) will trace similar curves in this state-space, meaning for a moment of interest  $\mathbf{G}(t = t')$  there will be other points in proximity to it, termed nearest-neighbors (NNs). The number of NNs used ( $K_{NN}$ ) is much larger than unity but much smaller than the total number of points in the state-space database ( $K_{DB}$ ):  $1 \ll K_{NN} \ll K_{DB}$ . The distance between an NN point  $\mathbf{G}^{(i)}$  and the moment of interest  $\mathbf{G}(t = t') = \mathbf{G}^{(q)}$  is determined using the standard Euclidean distance metric:

$$R_q^{(i)} = \sqrt{\sum_{k=1}^5 (G_k^{(i)} - G_k^{(q)})^2 / \sigma_{G_k}^2} \quad (3)$$

where each component is standardized by dividing by its standard deviation  $\sigma_{G_k}$ . Here, the combined storm-substorm 5D state-space from SST19 is utilized:

$$G_1^{(sst)}(t) = \langle Sym-H^* \rangle \propto \int_{-\Pi_{st}/2}^0 Sym-H^*(t+\tau) \cos(\pi\tau/\Pi_{st}) d\tau \quad (4)$$

$$\begin{aligned} G_2^{(sst)}(t) &= D\langle Sym-H^* \rangle / Dt \\ &\propto \int_{-\Pi_{st}/2}^0 Sym-H^*(t+\tau) \cos(2\pi\tau/\Pi_{st}) d\tau \end{aligned} \quad (5)$$

$$G_3^{(sst)}(t) = \langle AL \rangle \propto \int_{-\Pi_{sst}/2}^0 AL(t+\tau) \cos(\pi\tau/\Pi_{sst}) d\tau \quad (6)$$

$$G_4^{(sst)}(t) = D\langle AL \rangle / Dt \propto \int_{-\Pi_{sst}/2}^0 AL(t+\tau) \cos(2\pi\tau/\Pi_{sst}) d\tau \quad (7)$$

$$G_5^{(sst)}(t) = \langle vB_s^{IMF} \rangle \propto \int_0^{\tau_\infty} vB_s^{IMF}(t-\tau_\infty+\tau) \exp[(\tau-\tau_\infty)/\tau_0] d\tau \quad (8)$$

where  $Sym-H^*$  and  $AL$  are common indices used to measure storm and substorm intensities, respectively [3, 4, 44]. The correction  $Sym-H^* = A \cdot Sym-H - B \cdot \sqrt{P_{dyn}}$  is performed to approximately isolate the ring current contributions to this index (by removing magnetopause and induction fields) [45]. The half-wave rectified solar wind electric field value  $vB_s^{IMF}$  (where  $B_s^{IMF} = -B_z^{IMF}$  when  $B_z^{IMF} < 0$  and  $B_s^{IMF} = 0$  otherwise) is correlated with both storm [46] and substorm [47] activity. The integrals notated by  $\langle \dots \rangle$  in Equations (4), (6), and (8) act to smooth the inputs. The storm intensity parameter  $\langle Sym-H^* \rangle$  and the substorm intensity parameter  $\langle AL \rangle$  use half-cosine smoothing windows sized to the characteristic time scales of storms and substorms with  $\Pi_{st} = 12$  h and  $\Pi_{sst} = 2$  h, respectively.  $vB_s^{IMF}$  is instead smoothed using the more responsive exponential smoothing window with  $\tau_0 = 0.5$  h and  $\tau_\infty = 6\tau_0$ , which better captures the substorm growth phase. Also included in the state-space are the smoothed time derivatives of  $Sym-H^*$  and  $AL$  constructed using derivative windows (5) and (7) represented by the notation  $D\langle \dots \rangle / Dt$  [48]. It is critical to include the derivatives in the state-space as they act to differentiate between storm/substorm phases (main/expansion vs. recovery phases). Each of the components are standardized by dividing them by their standard deviations computed over the entire state-space.

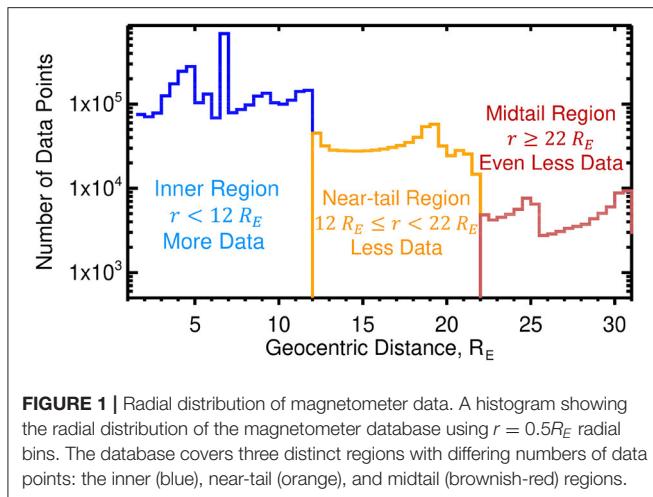
Not all the hand tailored elements were removed from the TS07D model, in particular, as with earlier models, some dynamical features are explicitly built into the model structure, including the contraction/expansion of the magnetosphere in response to the changes in the solar wind dynamic pressure  $P_{dyn}$  and the warping of the current sheets due to dipole tilt angle effects [21]. The contraction/expansion can readily be modeled by assuming all the current systems change in a self-similar way,

that is, by using a simple spatial rescaling:  $\mathbf{r}' \propto P_{Dyn}^{-\kappa} \mathbf{r}$  (e.g., [20]). Meanwhile, near the Earth ( $r \lesssim 4R_E$ ), the geometry of the current systems tends to be oriented with respect to the geodipole axis, while further down the tail ( $r \gtrsim 8R_E$ ), the current geometry is controlled by the solar wind flow direction. These dipole tilt angle effects are accounted for in the model by application of the general deformation technique [49, 50]. For example the flat current sheet described by Equation (1) is warped to account for the “bowl-shaped” deformation when the dipole tilt angle is non-zero [51], introducing three additional non-linear parameters (the hinge distance  $R_H$ , the warping parameter  $G$ , and the twisting parameter  $TW$ ), yielding  $\mathbf{B}^{(eq)}(\rho, \phi, z)$  [35].

It is important to mention here that the kNN DM technique is an instance-based machine learning method [52] which drastically differs both from model-based ML methods, such as ANNs [53] and classical Tsyganenko models [17, 19]. In the case of Tsyganenko models, tuning the parameters other than the linear regression coefficients and selected non-linear parameters is redundant, because the resulting model is universal and its architecture is custom-made and fixed from the outset. In case of ANNs, it is usually possible to split the model into training and validation sets and to use the latter to further optimize the model architecture [parameters, like  $K_{NN}$ , or  $M$  and  $N$  in Equation (1) that are termed hyper-parameters]. This can also be done when kNN is used for data classification or when the binning and fitting spaces are the same. However, this is almost impossible in our case, when the binning and fitting procedures are made in different spaces, the global index state-space (4)–(8) and the real space, which is extremely sparsely filled with data, with no more than a dozen of probes available for validation at any given moment.

At the same time, the selection of the kNN binning space (4)–(8) can be optimized in our DM method using important physics constraints and dynamic systems theory. First, we take explicitly into account the storm and substorm states of the magnetosphere that are known to be assessed by indices  $Sym-H$  [3] and  $AL$  [4], whereas their trends are described by the corresponding time derivatives (5) and (7). The latter can be extended to higher time derivatives following the idea of the time delay embedding in the non-linear time series analysis [48]. The averaging time scales in (4)–(8) are also physics-based and they are consistent with the observed characteristic times for storms [19, 46, 54] and substorms [55]. This physics-based optimization makes our kNN method similar to gray-box models (e.g., [56]).

Further selection of the hyper-parameters is done as follows. For a given set of the magnetic field model complexity (parameters  $M$ ,  $N$ , and the number of the FAC modules NFAC) there is usually an optimal range of  $K_{NN}$  values where the reconstruction is stable (no overfitting) and yet resolves important storm or substorm features, such as the westward and eastward ring currents, thinning and dipolarization of the magnetotail. Then validation tests are performed within that optimal range (e.g., SI in [30] to quantify the reconstruction fidelity). These tests are discussed below in section 3.4. At the same time, the uncertainty of the kNN binning is independently quantified by comparing the original binning parameters (4)–(8) with their means and standard deviations within the bin



**FIGURE 1 |** Radial distribution of magnetometer data. A histogram showing the radial distribution of the magnetometer database using  $r = 0.5R_E$  radial bins. The database covers three distinct regions with differing numbers of data points: the inner (blue), near-tail (orange), and midtail (brownish-red) regions.

as is also detailed in section 3.4. The optimal choice of  $K_{NN}$  minimizes the bias and standard deviation of the NN means but at the same time avoids overfitting with the model structure parameters ( $M$ ,  $N$ , NFAC, and others), i.e.,  $K_{NN}$  still needs to be large enough to resolve the corresponding storm and substorm structures and dynamics.

### 2.3. Magnetometer Database

The spacecraft magnetometer database used in this study is also the same as [30] spanning the years 1995 through 2018. It covers three distinct regions with different densities of magnetometer measurements as shown by the radial histogram in **Figure 1**. The inner magnetosphere region  $r \leq 12R_E$  has ample spatial coverage with the THEMIS [57] (five probes) and Van Allen [58] (two probes) missions sampling the inner equatorial magnetosphere, including the vicinity of the eastward current system ( $2R_E \leq r \leq 4R_E$ ) which is crucial to resolving the peak in the plasma pressure. The geosynchronous orbiting ( $r \sim 6.6R_E$ ) GOES (08, 09, 10, and 12) spacecraft reside within the ring current region. The THEMIS orbits have changed over the years but one of the primary apogees for the three inner probes has been  $r = 12R_E$ , providing good coverage throughout the *Inner Region*.

The *Near-tail Region*  $12R_E \leq r \leq 22R_E$  has a noticeable drop in data coverage, as only two outer THEMIS probes were ever located here and only for about 2 years before then were moved into a lunar orbit becoming the ARTEMIS mission [59]. With an apogee of  $r \approx 18R_E$ , the Cluster mission (four probes) helps populate this region, however, as a polar orbiting spacecraft, they spent a limited amount of time in the equatorial region.

Beyond  $22R_E$  the data density drops off by nearly an order of magnitude as the only spacecraft in the database that spent a considerable time in this region was Geotail. The near-earth reconnection sites are expected to be located here [60]. This motivated including the 2016–2017 MMS data. With an apogee of  $r \sim 26R_E$ , it nearly doubled the amount of data between  $22R_E \leq r \leq 26R_E$ . The only other spacecraft included in the database is IMP-8, however, it comprises a relatively small amount of the total dataset.

For each time step, the model is fit to the identified subset of magnetometer data by minimizing the root mean square of the difference between the model  $\mathbf{B}^{(mod)}$  and the observed  $\mathbf{B}_i^{j,obs}$  magnetic field vectors:

$$M_{err}^{(NN)} = \sqrt{\sum_{j \in S_{NN}} \sum_{i=x,y,z} w_j w_{(0)}(r) \left[ B_i^{(mod)}(\mathbf{r}^{(j)}) - B_i^{j,obs} \right]^2}, \quad (9)$$

where  $S_{NN}$  is the number of data points in the magnetometer subset identified through the kNN technique. The model is evaluated at the spacecraft's position  $\mathbf{r}$ . Two weight factors are incorporated into the objective function, one based on the data point's position in physical space  $w_{(0)}(r)$  and the other on its position in state-space  $w_j$ . The first,  $w_{(0)}(r)$ , lowers the weight factor in regions of the magnetosphere with a high density of data and was introduced to limit the bias of the fit toward these regions, in particular, to decrease the influence of the GOES satellites which are all located at the same radial distance [21]. The other weighting factor  $w_j$  gives higher weights toward observations which correspond to NNs that are closer to the moment of interest in state-space and will be described in the next section. The set of non-linear parameters ( $D$ ,  $D_{TCS}$ ,  $\kappa_{R1}$ ,  $\kappa_{R2}$ ,  $R_H$ ,  $G$ , and  $TW$ ) are found by minimizing Equation (9) using the downhill simplex method [61], while the amplitude coefficients are solved using the singular value decomposition (SVD) pseudo-inversion method [62, 63] also by minimizing (9).

### 2.4. Model Resolution Dilemma

The kNN approach has shortcomings caused by both the disparate density of NNs in the state-space and also the disparate density of magnetometer observations within the magnetosphere. The cause of the former is that, like many observed outputs of complex natural systems, geomagnetic indices, such as the substorm (*AL*) and storm (e.g., *Sym-H\**) indices, tend to follow lognormal distributions [64–66], meaning the distribution contains many weaker storms and substorms than stronger ones. The result is that the NNs are inhomogeneous within the state-space, biasing the kNN method toward weaker events, which is especially problematic for strong and extreme events. Decreasing  $K_{NN}$  helps resolve this problem but introduces another, overfitting. This problem was addressed in a pair of studies [35, 36] that demonstrated introducing a simple distance-weighting of the NNs could significantly reduce this bias while keeping the  $K_{NN}$  large enough to temper overfitting. The NNs are weighted using a Gaussian function of the form

$$w_j = \exp \left[ - \left( R_q^{(j)} / \sigma R_{NN} \right)^2 / 2 \right], \quad (10)$$

where  $R_q$  is the distance of each NN to the query point and  $R_{NN}$  is the radius of the NN n-sphere, which is the maximum  $R_q$  (the distance between the query point and the most distant NN).  $\sigma$  determines the narrowness of the Gaussian and the value of  $\sigma = 0.3$  is utilized here. These weights are then attached to the magnetometer datapoints when the model is fit by

minimizing Equation (9). However, the second problem that was left unaddressed in those studies was the disparate data density in real/physical space caused by the distribution of spacecraft as shown in **Figure 1**.

The choice of  $K_{NN}$  in the kNN approach is a tradeoff. A small number ( $K_{NN} \sim 1$ ) is akin to event-oriented modeling but requires a similarly small number of degrees of freedom in the model, i.e., the combined number of scaling coefficients  $a_i$  and non-linear parameters  $\beta_j$  are also on the order of unity [25, 26]. Increasing  $K_{NN}$  permits more complex models with a larger number of degrees of freedom (more  $a_i$ s and  $\beta_j$ s), but when  $K_{NN}$  approaches the size of the database the model becomes a universal statistical fit similar to classical Tsyganenko models (e.g., [17, 19, 38]) with a weaker sensitivity to storm and substorm phases. During the development of the SST19 model, it was found that an equatorial model resolution of  $(M, N) = (6, 8)$  using  $K_{NN} = 32,000$  was sufficient for reconstructing the primary substorm configuration of the magnetotail while avoiding overfitting. For this study, the model described in section 2 using these SST19 values of  $(M, N) = (6, 8)$  and  $K_{NN} = 32,000$  will be labeled the *Tail Model*. **Figure 2** (right panels) displays the *Tail Model*'s 2D equatorial distributions of the magnetic, electric current, and pressure fields during the late growth phase of a substorm (described below in section 3.2). For the sake of simplicity, these equatorial slices ignore dipole tilt and twisting effects, thus aligning the magnetic equator with the equatorial plane. Current densities are determined by numerically evaluating Ampere's law  $\mathbf{j} = \frac{1}{\mu_0} \nabla \times \mathbf{B}$ . The pressure is computed by integrating  $\mathbf{j} \times \mathbf{B}$  radially inward starting at the boundary  $r = 10R_E$  in the manner detailed in [35]:

$$P(r) - P(r_0) = \int_{r=10R_E}^r [-\mathbf{j} \times \mathbf{B}]_r dr' \quad (11)$$

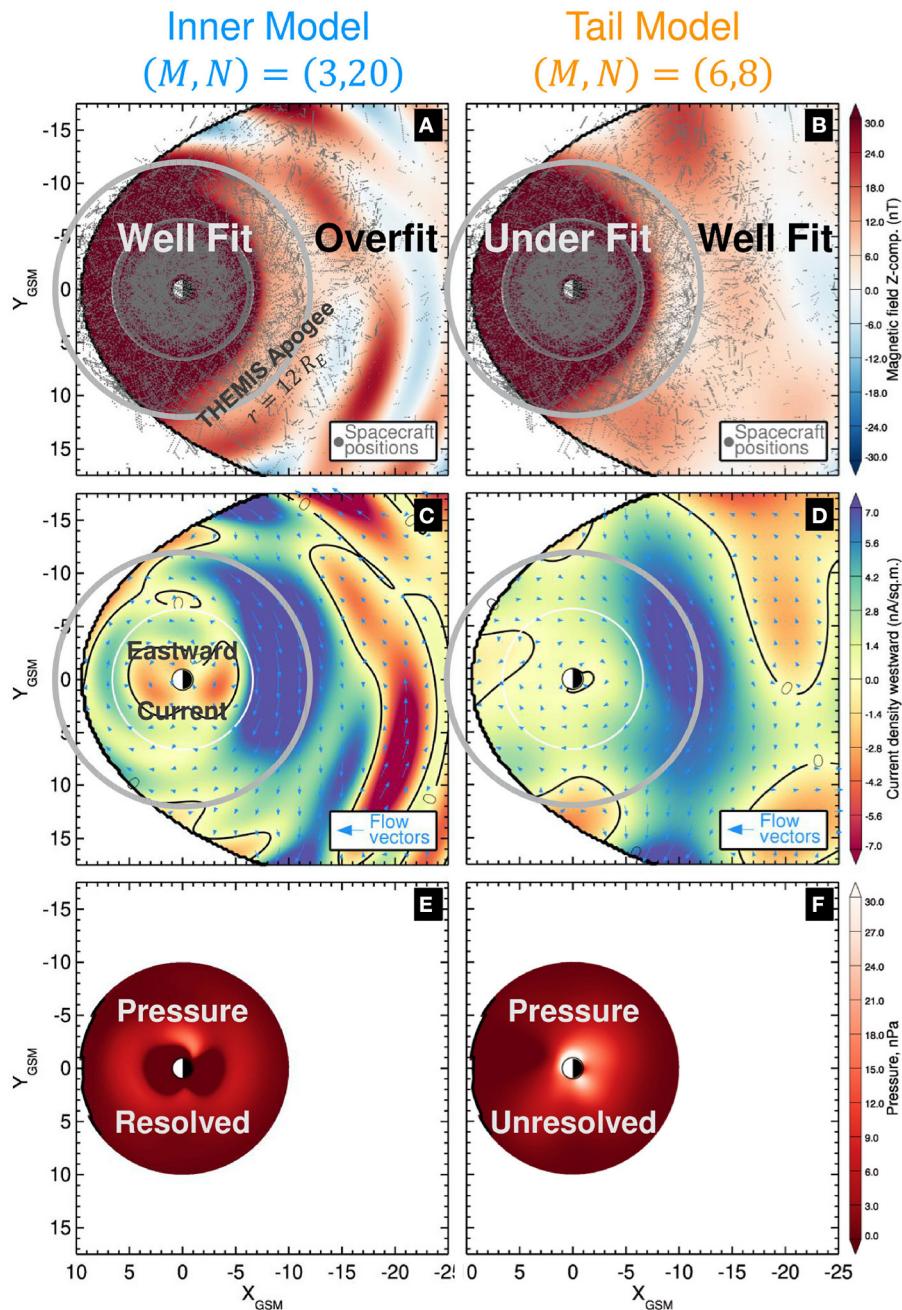
where the boundary pressure  $P(r_0)$  is assumed to be small and set to zero. **Figure 2** demonstrates how the *Tail Model* performs well throughout the near-tail and midtail regions and thus is suitable for substorm studies. However, within the predominant THEMIS inner probes apogee (gray circle) the current densities appear underresolved. Specifically, the eastward currents are almost entirely absent (**Figure 2D**). Without an eastward current, the integrand  $[-\mathbf{j} \times \mathbf{B}]_r$  monotonically increases earthward from the boundary  $r = 10R_E$  meaning the pressure also increases monotonically. Thus, the location of the pressure peak is absent and the pressure radial gradients are also missing (**Figure 2F**).

In contrast, storm-time empirical reconstructions of the inner region ( $r \leq 12R_E$ ) were able to reconstruct the eastward current and thus the pressure, its peak, and gradient with the essential difference in the configuration of these models being a larger radial expansion number  $N = 20$  used in the equatorial current module [33, 35, 36, 67]. Furthermore, the inclusion of the TCS particularly helped in resolving the eastward current when the storm state of the magnetosphere was relatively quiet [35]. However, the addition of the TCS in the model resulted

in reconstructed pressures with rather large azimuthal gradients. This was mitigated by halving their number from  $M = 6$  to  $M = 3$ , yielding a smoother pressure distribution, resulting in an equatorial resolution of  $(M, N) = (3, 20)$ . Thus, the *Inner Model* is defined using this resolution of  $(M, N) = (3, 20)$  with all other model configurations being the same as the *Tail Model*. The analogous reconstruction of the substorm growth phase using this *Inner Model* are plotted in the left-hand panels of **Figure 2** demonstrating that it does indeed capture the eastward ring current (**Figure 2C**) and appropriate pressure distributions (**Figure 2E**). However, the magnetic and current density reconstructions now contain numerous artifacts throughout the near-tail region (**Figures 2A,C**) indicating overfitting there.

Herein lies the dilemma; no single resolution is capable of adequately reconstructing both the inner and near-tail regions using  $K_{NN} = 32,000$  (the amount required to reconstruct the near-tail). Higher equatorial resolutions are necessary to describe the eastward current systems but will overfit the near-tail due to the lesser amount of data there and vice versa; lower equatorial resolutions perform well in the near-tail but miss the eastward currents in the inner region. This results in a bifurcation of models: the *Inner Model* (higher resolution) is suitable for storm spatial-scale reconstructions and the *Tail Model* (lower resolution) is applicable to substorm scales. A potential solution is to dramatically increase the value of  $K_{NN}$ , however, due to the disparate density of data in state-space, this would weaken the model's sensitivity to the event of interest and would begin to resemble statistical modeling instead of the DM approach sought. A simple solution would utilize a piecewise field, that is, to evaluate the *Inner Model* in the inner magnetosphere region and the *Tail Model* in the near-tail region. However, the equatorial current sheet described by Equations (1) and (2) ensures a divergenceless  $\mathbf{B}$  and  $\mathbf{j}$  fields. Such a piecewise field would introduce discontinuities which would violate these conditions and would also introduce infinitely thin current sheets. The resultant question is how to transition between these regions in a way that maintains  $\nabla \cdot \mathbf{B} = 0$  and  $\nabla \cdot \mathbf{j} = 0$ ? The next section 3 presents a simple resolution to this dilemma that smoothly transitions between the two regions while maintaining divergenceless fields.

It must be stressed that this resolution dilemma is not just a technical issue related to the sparse data distribution in space and the need to improve it is some regions. It reflects different physics processes associated with magnetic storms in the inner magnetosphere and substorms that reconfigure the magnetotail and create new FAC systems. In the earlier DM algorithms both storm and substorm descriptions used fleets of synthetic probes mined in the history of the magnetosphere. But these descriptions did not exchange the information gained from inner and outer magnetosphere description. So the fitting challenge resembles the first-principles model problem of the concurrent description of magnetic storms using the dedicated ring current models [68–70] and global MHD models [71–73] that often properly describe the outer magnetosphere. The resolution of that physics-based problem is similarly offered



**FIGURE 2 |** The model resolution dilemma. 2D equatorial distributions of the late growth phase (11:25) of a substorm on 8 March 2009 using two different equatorial resolutions. **(A,B)** The equatorial distribution of the total modeled magnetic field. The location of the spacecraft magnetic field observations used to fit the model are overplotted with gray dots. The predominant apogee of the inner probes of the THEMIS mission  $r = 12R_E$  is represented by the gray circle. **(B,C)** The equatorial distribution of the current density. The overplotted arrows show the direction and magnitude of the current density vectors. **(E,F)** The equatorial distribution of the pressure computed by integrating  $\mathbf{j} \times \mathbf{B}$  according to Equation (11). The *Inner Model* is used for the left hand panels **(A,C,E)** which is fit using an equatorial resolution of  $(M, N) = (3, 20)$ . It performs well in the inner region but overfits the near-tail region. The *Tail Model* is used for the right hand panels **(B,D,F)** which is fit using an equatorial resolution of  $(M, N) = (6, 8)$ . It performs well in the near-tail region but under resolves the inner region.

in the form of coupled models [74–77], when the information on the plasma conditions at some boundary separating inner and outer magnetosphere is transferred from MHD to ring

current models. At the same time, the latter are used to adjust the equation of state in MHD taking storm effects into account.

### 3. MERGED MODEL

#### 3.1. Merged Model Algorithm

In this section, a simple method for concurrent reconstruction of both substorm and storm spatial scales is proposed, thus addressing the model resolution dilemma described above. The general approach is to introduce a third model, termed here the *Merged Model*, constructed using an equatorial resolution of  $(M, N) = (6, 20)$ , the lowest common resolution between the *Tail* ( $M, N) = (6, 8)$  and *Inner* ( $M, N) = (3, 20$ ) models. However, instead of fitting the *Merged Model* to actual spacecraft magnetometer data points, it is instead fit to virtual data, which are simulated by randomly sampling the equatorial regions of the *Tail* and *Inner* models outside and inside the merging boundary, respectively. Because the *Tail* and *Inner* models can be sampled at any location, the introduction of these virtual datapoints allows the *Merged Model* to be fit with an arbitrary density of points. The result is a *Merged model* (given a proper distribution of virtual data points) that tends to reflect the *Inner Model* within the merging boundary and the *Tail Model* beyond it with a smooth transition between the two in the vicinity of the boundary. Also, because the *Merged Model's* equatorial field is still described by Equations (1) and (2), by construction  $\nabla \cdot \mathbf{B} = 0$  and  $\nabla \cdot \mathbf{j} = 0$  is ensured.

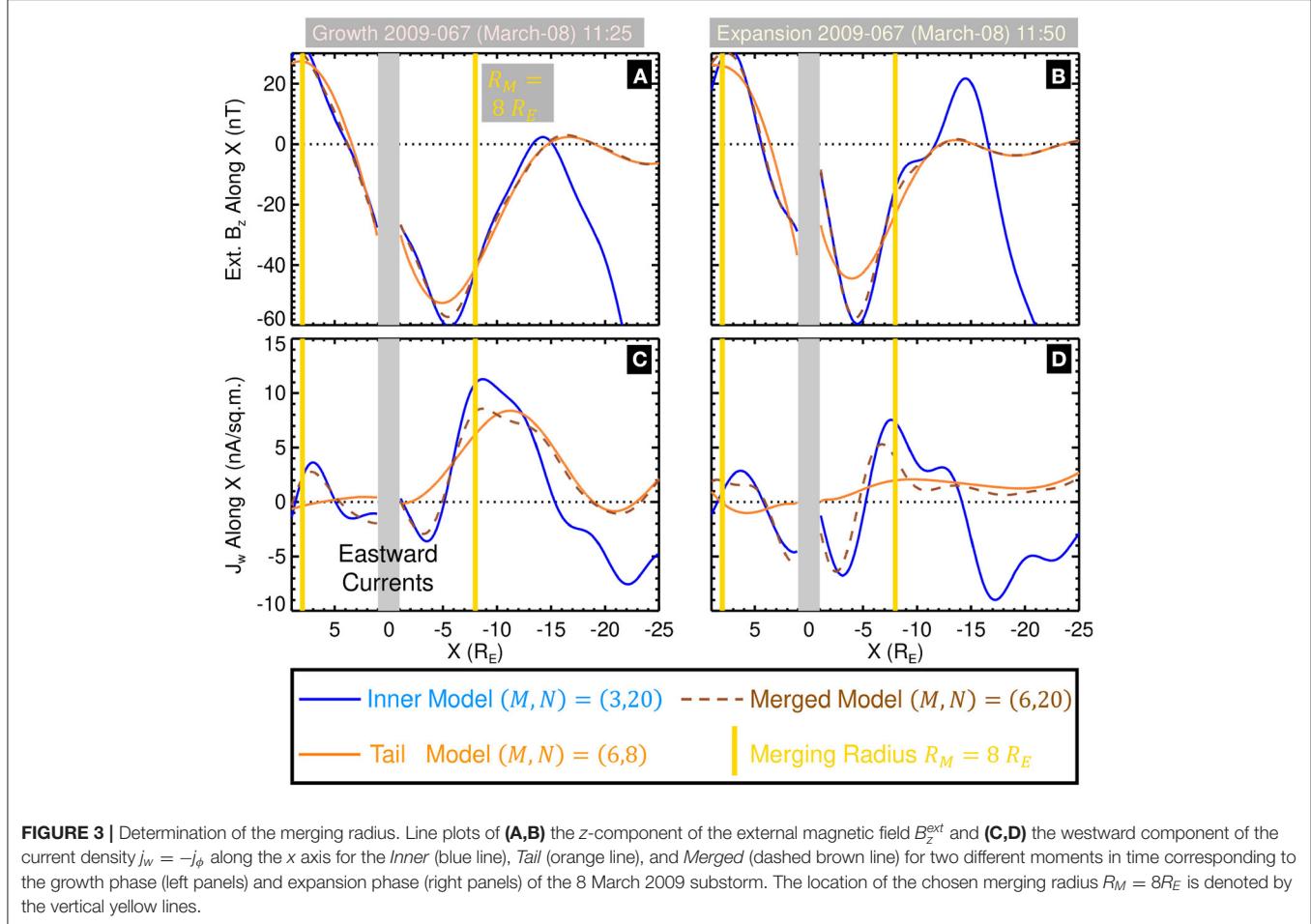
For simplicity, a cylindrical boundary is used identified by the merging radius  $R_M$ . Initially,  $R_M$  was defined to be commensurate with the demarcation between the inner and near-tail regions as characterized by the predominant THEMIS inner probes apogee ( $r = 12R_E$  from **Figure 1**), though, this yielded unsatisfying results. **Figure 3** shows 1D plots of the external magnetic field  $B_z^{ext}$  (**Figures 3A,B**) and the westward component of the current density  $j_w = -j_\phi$  (**Figures 3C,D**) for the *Inner* (blue lines) and *Tail* (orange lines) models along the  $X$  axis. Two different moments are shown corresponding to the late growth (left panels) and expansion (right panels) phases, respectively. Ideally, there would exist a clear boundary where  $B_z^{ext}$  and  $j_w$  from the two models intersect. Instead, the intersections (between the blue and orange lines) vary between  $\sim 6R_E$ – $10R_E$ . This justifies moving  $R_M$  earthward of the THEMIS apogee of  $12R_E$ . Ultimately, the center of this region, that is, a cylindrical radius of  $R_M = 8R_E$  was chosen, which will be used as a constant merging boundary throughout the rest of the study. Although this is a rather arbitrary value, as shown below, this choice of  $R_M$  enables the *Merged Model* to fully capture the deep  $B_z^{IMF}$  minimum and eastward currents of the *Inner Model* while also reconstructing the substorm scale features of the SST19 and *Tail Model*. The merging algorithm can be enhanced in future studies by incorporating a dynamical merging boundary that minimizes the differences of  $B_z^{ext}$  and  $j_w$  between the *Tail* and *Inner* models. The use of elliptic cylindrical boundaries should also be pursued as **Figure 3** indicates that day-night asymmetries exist in the optimal merging location.

To create the *Merged Model*, the non-linear parameters ( $D$ ,  $D_{TCS}$ ,  $\kappa_{R1}$ ,  $\kappa_{R2}$ ,  $R_H$ ,  $G$ , and  $TW$ ) are taken to be the average of their values from the *Inner* and *Tail* models and are not included in the fit; while the linear amplitude coefficients for the equatorial and FAC systems are fit in the manner described before, that is by

using the SVD least-squares method to minimize Equation (9). However, instead of using real magnetometer data points, the  $B_i^{j,obs}$  terms in (9) are populated with virtual measurements constructed by evaluating the *Inner* and *Tail* models. The *Inner Model* was randomly sampled to achieve a data point density of  $1R_E^{-3}$  within the cylindrical volume  $1.1R_E \leq r \leq R_M$ ,  $-6R_E \leq z \leq +6R_E$ . The value of  $\pm 6R_E$  was chosen as it is twice the typical value of the thick current sheet half thickness  $D \approx 3R_E$ . Meanwhile, given the lower radial resolution of the *Tail Model*, it was only sampled at a data point density of  $0.2R_E^{-3}$  within the cylindrical volume  $R_M \leq r \leq 31R_E$ ,  $-6R_E \leq z \leq +6R_E$ . While this worked well for resolving the thick current sheet, due to the small height scale size of the TCS, additional samples were needed from the region  $-1R_E \leq z \leq +1R_E$ , where  $\pm 1R_E$  is twice the typical TCS half-thickness  $D_{TCS} \approx 0.5R_E$  now using six times the densities from before,  $6R_E^{-3}$  and  $1.2R_E^{-3}$  for the *Inner Model* and *Tail Model*, respectively. The reason for using a random distribution of points in contrast to a regular grid, is that fitting a regular grid of points with a magnetic field described by the regular expansion in Equation (1) may introduce artifacts caused by the model aligning to the grid instead of the underlying magnetic field distribution. In total, 18,260 virtual spacecraft datapoints were included in fitting the amplitude coefficients of the *Merged Model*.

The output of the *Merged Model* is overplotted in **Figure 3** as brown dashed lines. Earthward of  $R_M$  the *Merged Model* generally tracks the *Inner Model* (blue line) including the deep minimums in  $B_z^{ext}$  (**Figures 3A,B**) and the eastward currents (**Figures 3C,D**), both of which are key characteristics needed to reconstruct the storm-time dynamics of the inner magnetosphere [33]. Tailward of  $R_M$  the *Merged Model* closely matches the *Tail Model* (orange line). This indicates that the *Merged Model* should also reconstruct the primary substorm scale features, such as the thinning and stretching of the magnetotail during the growth phase and its rapid dipolarization during the expansion phase, as will be shown in the next section 3.2. In the  $\sim \pm 2R_E$  region bounding  $R_M$ , the *Merged Model* smoothly transitions between the *Inner* and *Tail* models.

In order to qualitatively compare the *Merged Model* to the *Inner* and *Tail* models, 2D equatorial slices of the magnetic and current density fields are plotted for the late growth phase in **Figure 4**. Inside the merging radius  $R_M = 8R_E$  (yellow circles) the *Merged Model* qualitatively resembles the *Inner Model*, and importantly for the evaluation of the pressure, resolves the eastward currents (orangish-red region  $r \lesssim 5R_E$  in **Figures 4D,F**). Outside of the  $R_M$ , the *Merged Model* (**Figures 4C,F**) is nearly indistinguishable from the *Tail Model* (**Figures 4B,E**). Notably, the *Merged Model* resolves the  $B_z$  minimum formation in the tail at  $10R_E \lesssim r \lesssim 13R_E$  and the enhancement of the cross tail current at  $7R_E \lesssim r \lesssim 16R_E$ . The distribution of the virtual spacecraft projected into the equatorial plane used to fit the *Merged Model* is overplotted in **Figure 4C**, showing how random sampling of the other two models evenly fills in data gaps present in the near-tail and midtail regions, particularly beyond the Cluster apogee ( $r \geq 19R_E$ ). A note, as can be seen in **Figure 4**, the *Inner Model* is now only being



fit using data within the predominant THEMIS inner probes apogee of  $r \leq 12R_E$ . Including data beyond this serves no purpose because the *Inner Model* is only being sampled within  $r \leq R_M$  when creating the *Merged Model* and their inclusion may bias its reconstructions.

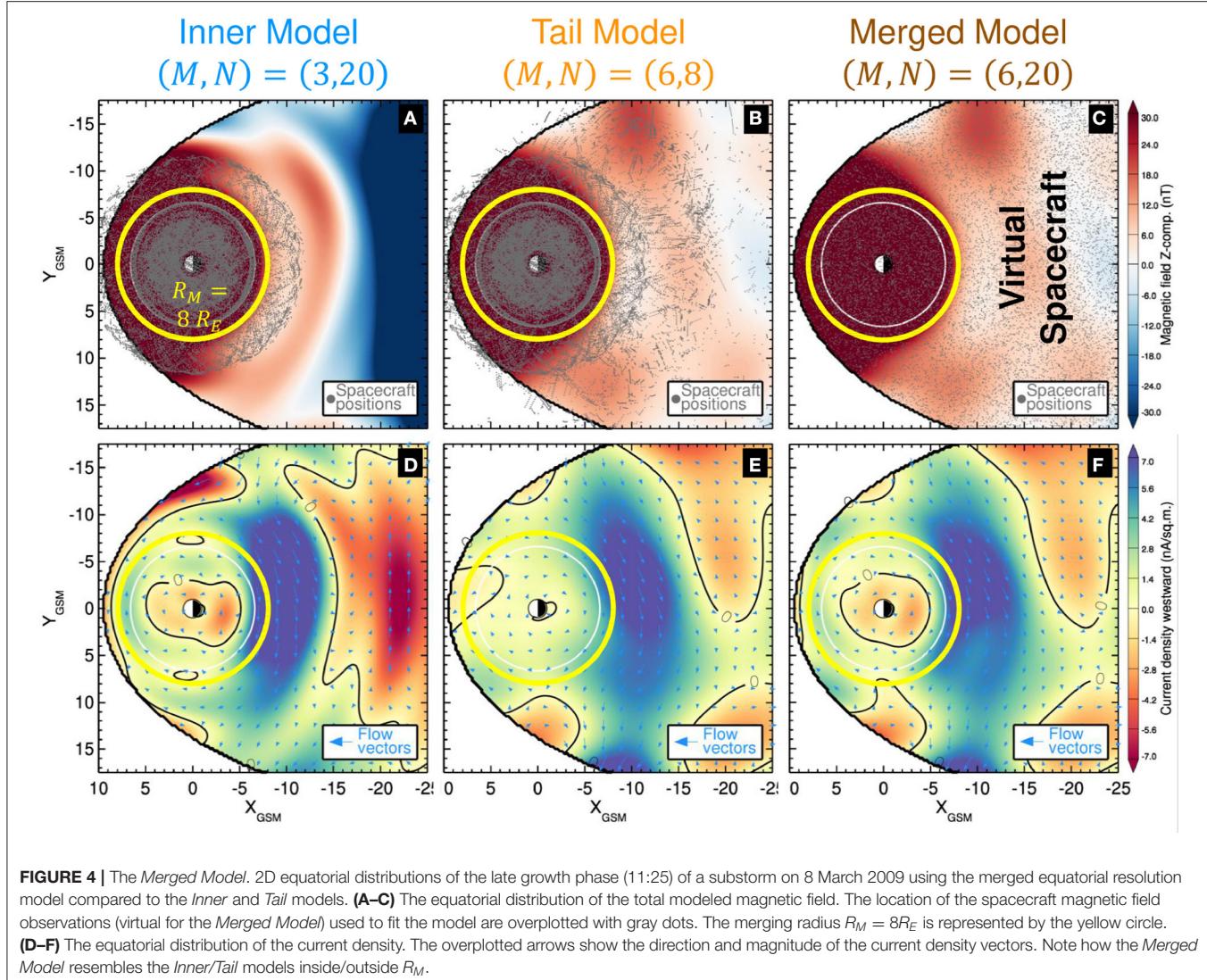
The differences between the *Inner/Tail* and *Merged* models from the panels in **Figure 4** are displayed in **Figure 5**. Within geosynchronous orbit, the differences between *Inner* and *Merged* models are relatively small (**Figures 5A,C**), with maximum values of  $B = 1.9$  nT and  $j = 3.6$  nA/sq.m and mean differences of  $B = 0.66$  nT and  $j = 0.9$  nA/sq.m. In this same region, the differences between the *Tail* and *Merged* models are about two to three times larger, with maximum values of  $B = 3.9$  nT and  $j = 6.8$  nA/sq.m and means of  $B = 1.8$  nT and  $j = 2.5$  nA/sq.m. The comparison between the *Tail* and *Merged* models shows negligible differences beyond  $\sim 10R_E$  (**Figures 5B,D**), while the equivalents for the *Inner Model* display large differences there. This confirms that the *Merged Model* largely mimics the *Inner Model* in the inner magnetosphere and the *Tail Model* in the near-tail. However, of interest are the differences in the vicinity of the merging boundary. Within the  $\pm 2R_E$  region bounding  $R_M$ , the mean differences are  $B \approx 1$  nT and  $j \approx 2$  nA/sq.m but show rather large maximum differences;  $B = 5$  nT,  $j = 10$

nA/sq.m for the *Inner Model* and  $B = 3$  nT and  $j = 6.5$  nA/sq.m for the *Tail Model*. The cause of this being the relatively large mismatch between the *Inner* and *Tail* models as is evident in the **Figures 3, 4**. This further supports additional investigation into a more optimal merging boundary in future works.

### 3.2. Merged Model Reconstruction of 8 March 2009 Substorms

Global reconstructions for the second 8 March 2009 substorm are displayed in **Figures 6, 7**, the first corresponding to the late growth phase at approximately the time of the substorm onset and the second is 25 min later during the expansion phase as indicated by the *AL* index (**Figure 6a**). These reconstructions highly resemble the reconstructions of the same event using the SST model [15] (**Figures 3, 4** from that work). This confirms that the *Merged Model* is mostly analogous to the SST19 model throughout the near-tail region as was expected from the analysis shown in **Figures 3, 4**. However, the *Merged Model* now reconstructs the eastward currents as can be seen near the planet in **Figures 6e, 7e**.

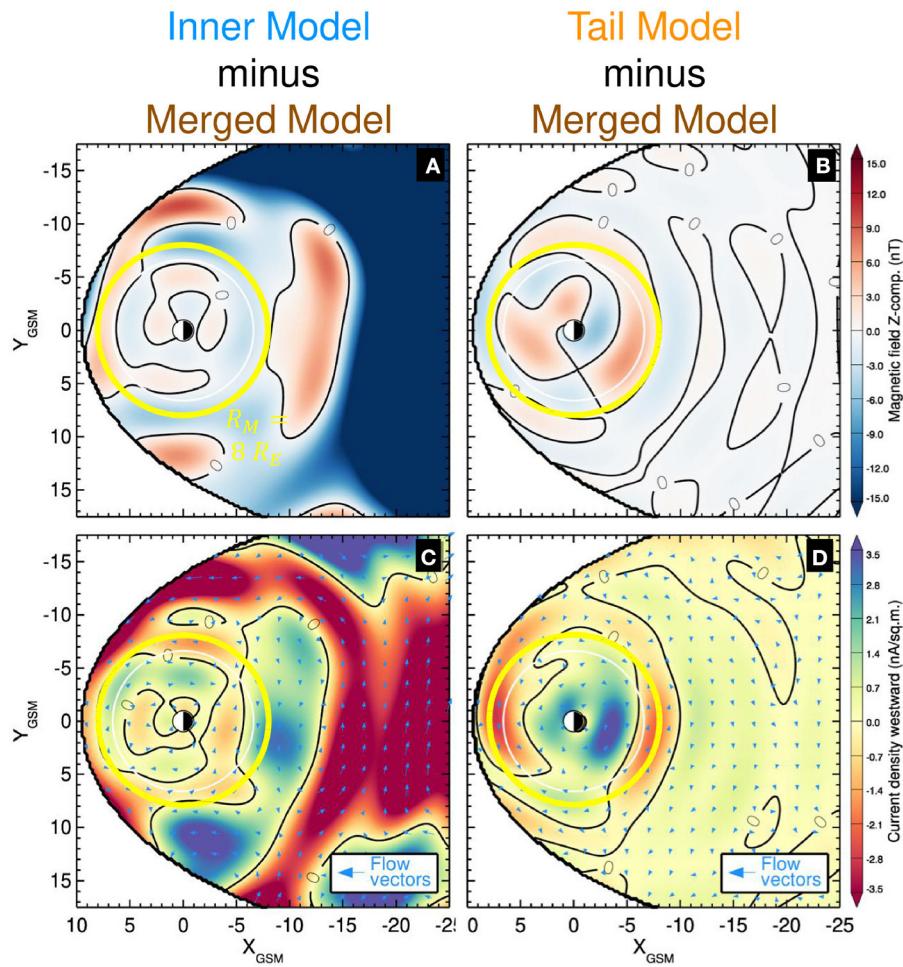
During the growth phase, the magnetotail is stretched (**Figure 6e**), particularly in the region between  $\sim 8R_E$  and  $15R_E$  corresponding to the  $B_z$  minimum region (**Figure 6d**). The



TCS is especially intense on the night side in this same region (**Figure 6e**). The 66 and  $68^\circ$  field lines cross the magnetic equator at  $12R_E$  and  $18R_E$ , respectively while the  $68^\circ$  line is open. The height integrated thick current sheet is plotted (**Figure 6f**) as it is a better proxy for the content of the ring current compared to the equatorial slice of the current density (as is shown in **Figures 2, 4**) because the current density of the TCS tends to be the dominant source of current density along the magnetic equator. This panel shows a modest westward ring current within geosynchronous orbit  $r \lesssim 6.6R_E$  as well as a strong outflow to the magnetopause at the dusk terminator. The pressure during the growth phase is relatively azimuthally symmetric, with a pressure peak at  $4\text{--}5R_E$ .

During the expansion phase, which takes only 25 min, the global configuration of the magnetosphere is drastically altered (**Figure 7**). The magnetotail becomes much more dipolar. The  $66^\circ$  field line (**Figure 7e**) now crosses the magnetic equator at  $7.5R_E$  (compared to  $12R_E$  in the growth phase) and the  $68^\circ$  line

crosses at  $10R_E$  (compared to  $18R_E$ ). The previously open  $70^\circ$  field line now crosses at  $14R_E$ , indicating the conversion of open to closed flux presumably from reconnection. This dipolarization is congruous with the formation of the magnetic flux pileup in this region (**Figure 7d**). There is a strong enhancement of an westwardly directed thick current all across the night side taking the appearance of a PRC (**Figure 7f**). However, the SST19 analysis [15] along with magnetohydrodynamic (MHD) simulations [78] indicated that the enhancement of this PRC is associated with closure through the substorm current wedge. The substorm current wedge manifests as a eastwardly directed TCS in the vicinity of the magnetic equator, but its closure out of the plane is via a westwardly directed thick current, some of which closes through the ionosphere as R2 currents. Another point of agreement between the Merged Model and the SST19 model is the strengthening of the dayside thick current sheet within geosynchronous orbit ( $r \leq 6.6R_E$ ), which was interpreted as an intensification of the symmetric ring current.



**FIGURE 5 |** The *Merged Model* differences. The difference between *Inner/Tail* and the *Merged Model* 2D equatorial distributions from **Figure 4**. The format is the same as **Figure 4** but the range on the color bars has been halved to emphasize the differences.

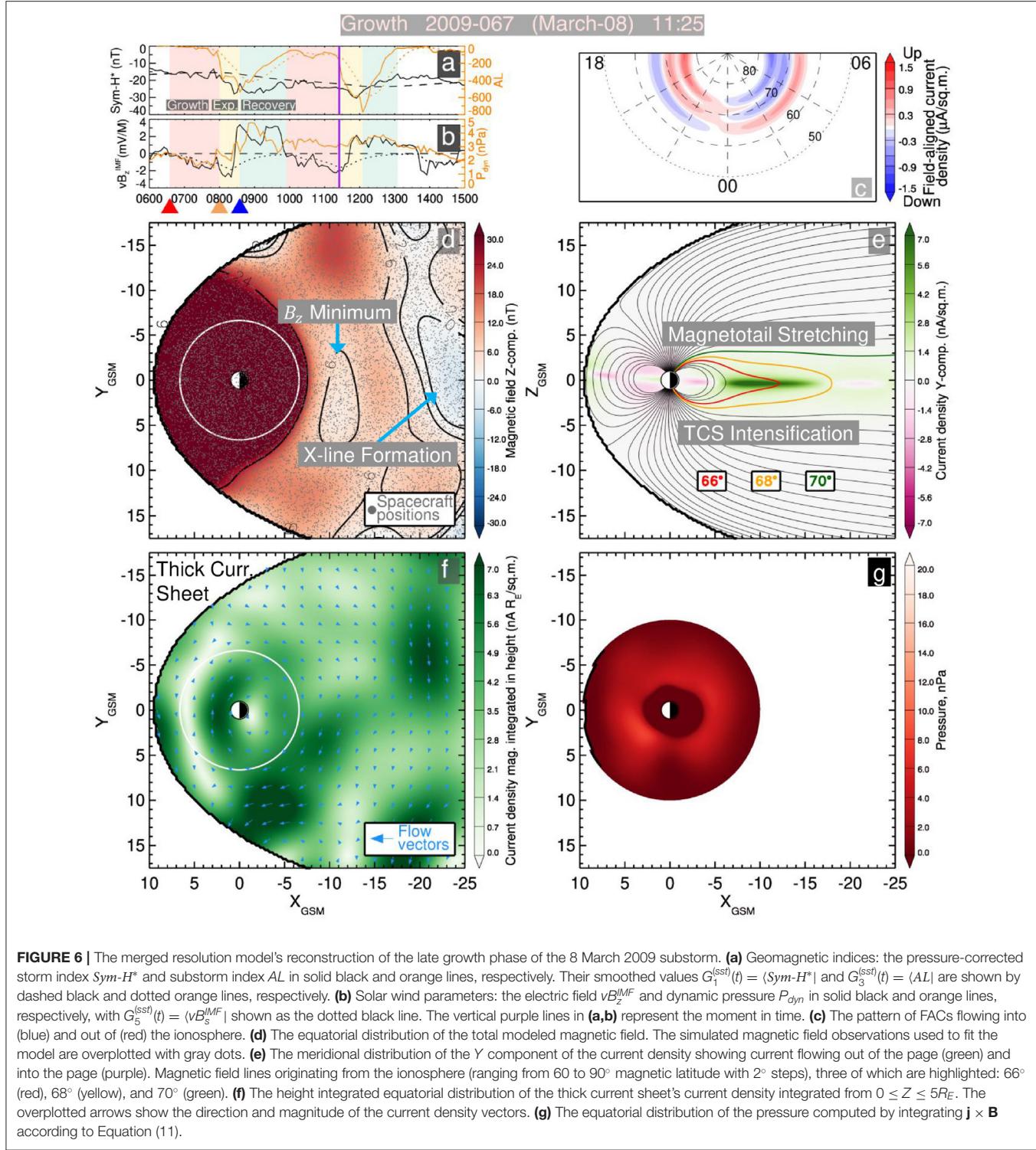
### 3.3. Merged Picture vs. Earlier Substorm Reconstruction

The quantitative analysis performed using the SST19 model for the 8 March 2009 substorms [15] is now recreated using the *Merged Model*, in particular their Figure 6 is recreated and is shown as **Figure 8**. Compared to that substorm-focused reconstruction, the *Merged Model* reveals more profound variations of the R2 FAC amplitudes (**Figure 8d**), stronger buildups and decays of the TCS integrated current (orange line in **Figure 8f**) with significant negative values implying bifurcated current structures. One such bifurcated current structure is well seen in the meridional current distribution presented in **Figure 7e**.

The merged resolution picture reveals stronger variations of the  $B_z$  magnetic field at the Van Allen Probes inner probes apogee (red line in **Figure 8g**). These stronger variations reflect the impact of distance-weighting the NNs along with a better resolution of the inner magnetosphere region resulting eventually in the resolution of the eastward current and the

plasma pressure shown in **Figures 6g, 7g**. At the same time, it resolves the formation of the near-Earth X-line around the substorm onset (red and orange lines in **Figures 8h,i**), which is also seen in **Figure 6d**. Its resolution becomes possible not only due to the solved overfitting problem for the tail region, but also due to the use of the MMS data and the distance-weighted kNN algorithm as is elaborated in detail in [30].

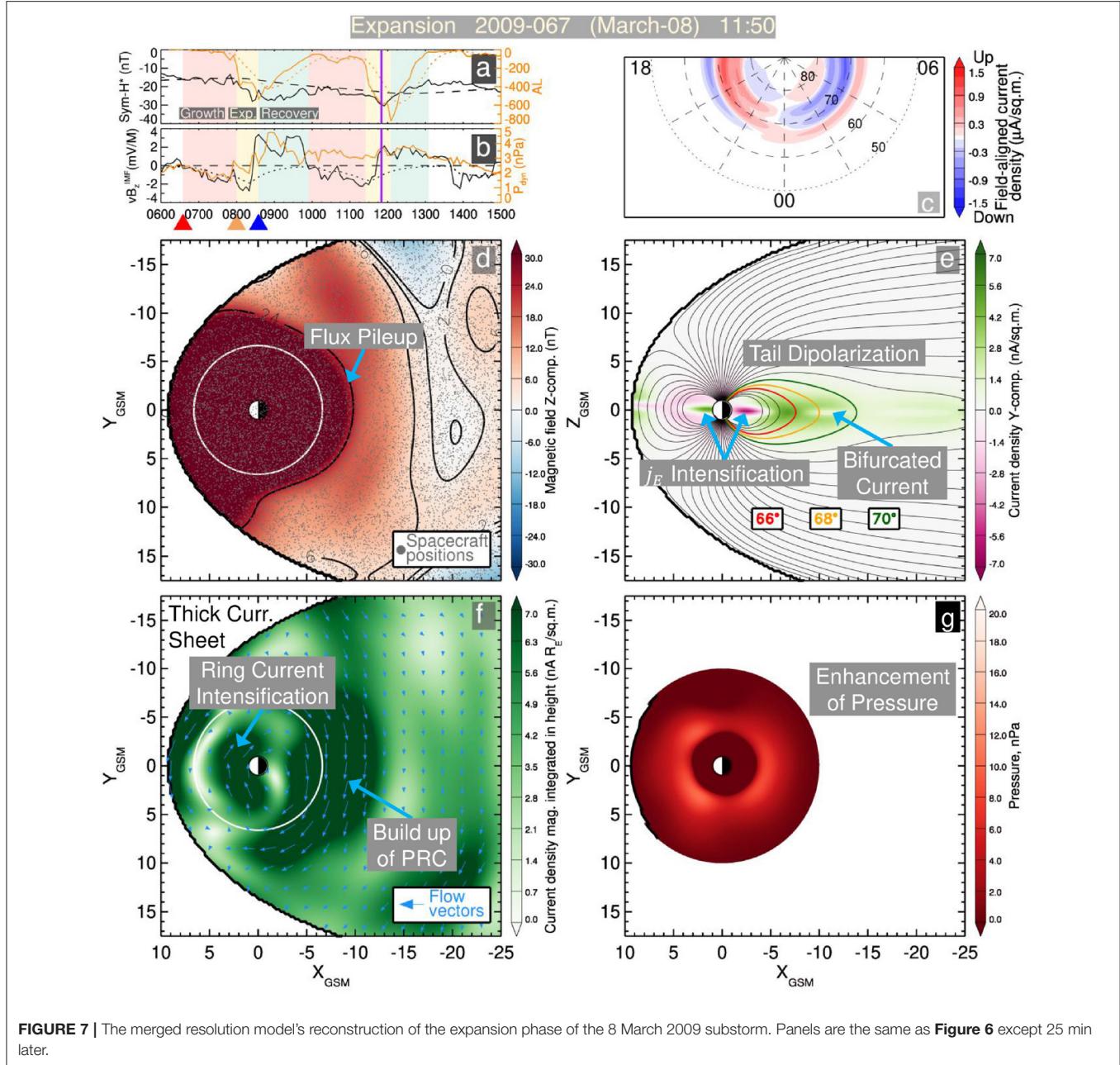
The comparison also reveals some pitfalls of the present merging algorithm. In particular, unlike the SST19 model, the merged resolution model no longer has a clear correlation between the substorm phase and the amplitude of the R1 currents (**Figure 8c**). The probable explanation, is that because the R1 currents do not close through the equatorial plane (in contrast to the R2 currents), relatively few of the virtual spacecraft points sample them. To rectify this in future reconstructions using the *Merged Model*, the amplitude coefficients for the FAC systems should be set in a similar fashion to the non-linear parameters and not included in the fit. This is reasonable since the focus of



the merged modeling approach is to combine two models with different equatorial resolutions.

The primary enhancement of this *Merged Model* over the previous SST19 model is the resolution of the pressure in the inner magnetosphere. In order to quantify the change in pressure

during the course of a substorm, the 2D pressure distributions from **Figures 6g, 7g** are averaged over all local times for a given radial distance  $r$ . The resultant average pressures  $\langle P \rangle$  are plotted as a function of radial distance  $r$  for the second 8 March 2009 substorm in **Figures 9D,E** for several moments during

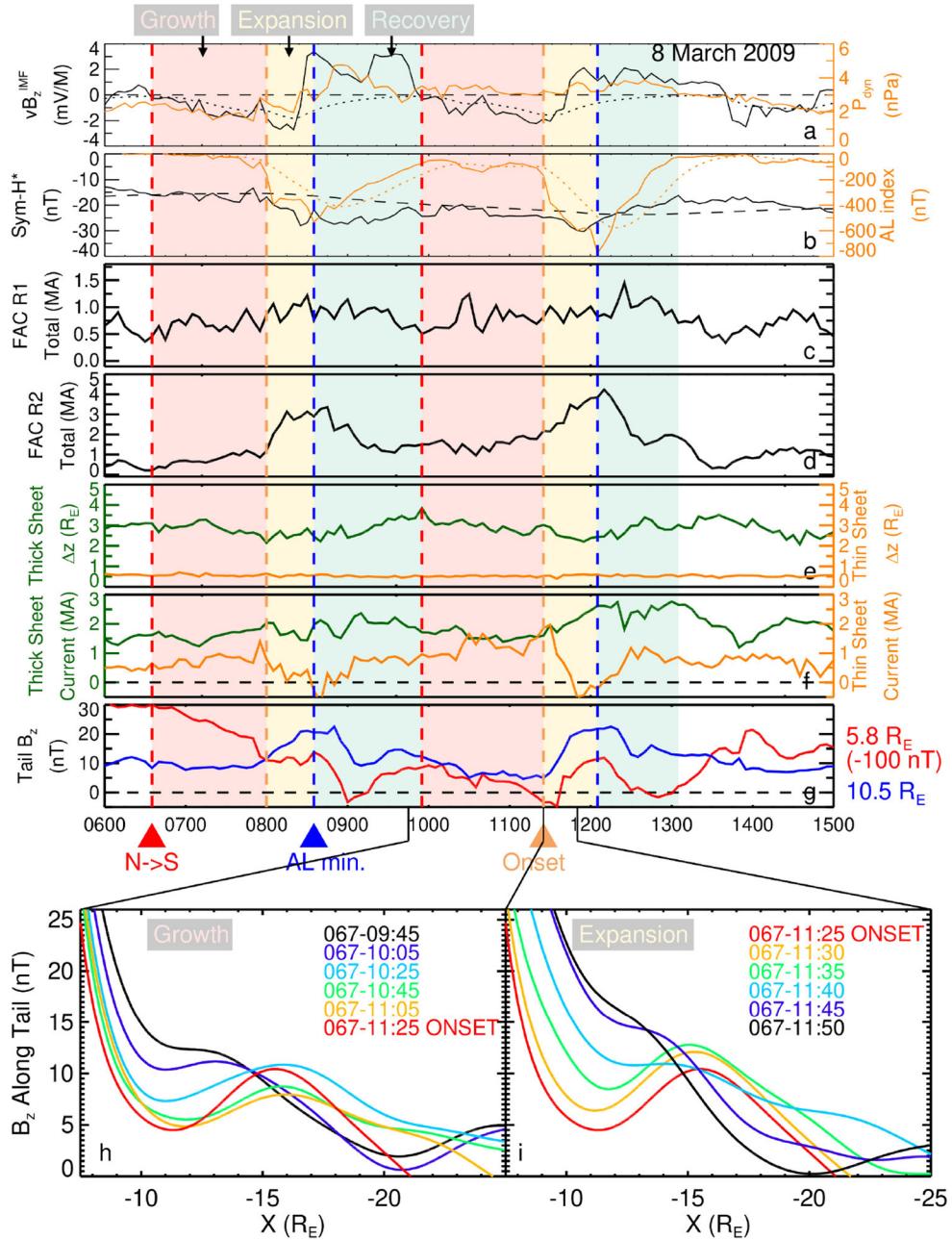


the growth and expansion phases (the solar wind values and geomagnetic indices are included for context in **Figures 9A,B**). During the growth phase, the average pressure stays relatively stable, with a max value between  $\sim 5\text{--}6$  nPa located at  $4.25\text{--}4.5R_E$ . In contrast, during the expansion phase  $\langle P \rangle$  rapidly increases reaching values as high as 10 nPa while the location of the peak stays relatively stable. The time series for  $\langle P \rangle$  at  $4.5R_E$  is then plotted in **Figure 9C** for both substorms, showing that during quiet and growth phases the pressure is stable and low but then rapidly increases during the expansion phase.  $\langle P \rangle$  then decreases during the recovery phase, although for the second stronger substorm, it does not return to the nominal

value until more than 3 h after the substorm onset. Such an enhancement of inner magnetosphere pressure is consistent with *in-situ* observations [12] of particle injections as well as statistical analyses [13] of substorm expansion phases.

### 3.4. Uncertainty Quantification

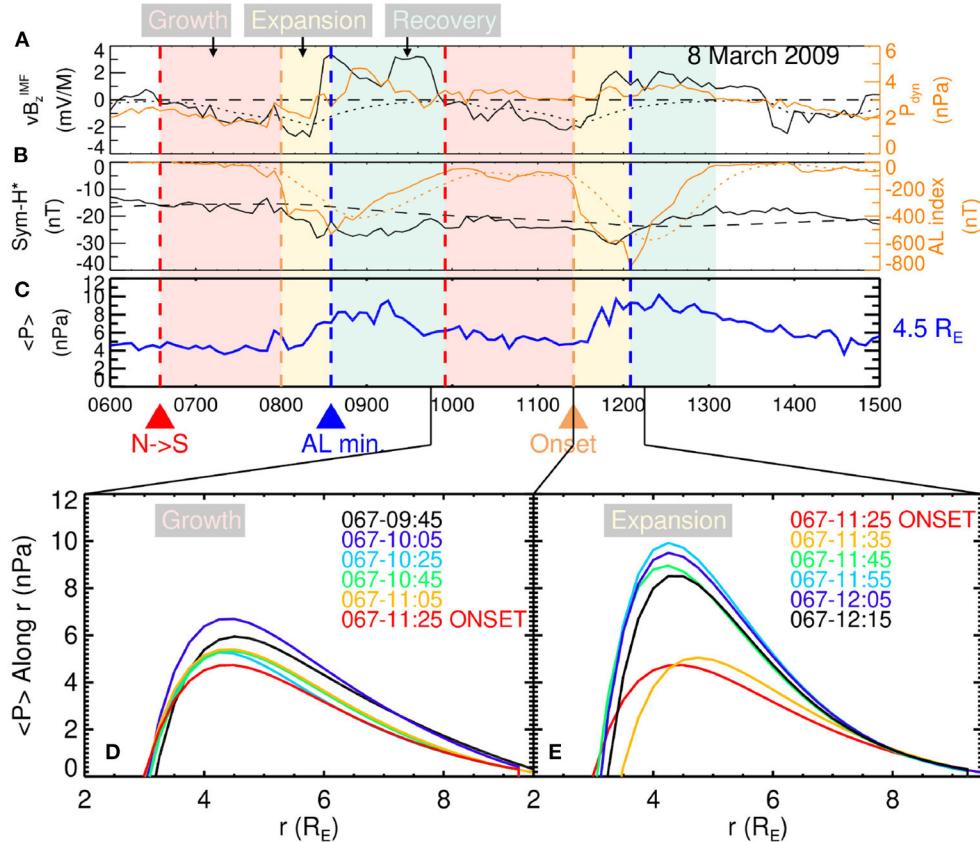
An important aspect of empirical model development is its validation via comparison of the model output to the data that went into its construction. In this section, three different analyses are performed to test the model's fidelity. First, the DM approach is analyzed to ensure that the identified NNs characterize the 8 March 2009 substorm event and next the *in-situ* spacecraft



**FIGURE 8 |** The Merged Model quantitative analysis. **(a,b)** Geomagnetic indices and solar wind parameters in a format similar to **Figures 6a,b**. **(c,d)** The root mean square of the FAC amplitude coefficients for the R1 and R2 FAC modules, respectively. **(e)** The values of the current sheet half thicknesses for the thick sheet (green line) and TCS (orange line), i.e.,  $D$  and  $D_{\text{TCS}}$  from Equation (2). **(f)** A measurement of the amount of current contained in the dayside SRC (green line) and nightside TCS (orange line). The SRC current (green line) is computed by integrating the dayside westward current density of the thick current sheet module flowing within geosynchronous orbit  $1.0R_E \leq x \leq 6.616R_E$  and  $z = \pm 5R_E$ . The TCS current (orange line) is computed by integrating the nightside westward current density of the TCS module flowing within the rectangle  $-16R_E \leq r \leq -6R_E$  and  $z = \pm 1.0R_E$ . **(g)** The z-component of the total magnetic field  $B_z$  sampled at  $x = -5.8R_E$  (red line) and  $x = -10.5R_E$  (blue line). **(h,i)** The z-component of the total magnetic field  $B_z$  along the nightside x-axis at different times during the growth and expansion phases.

magnetometer data is compared to the reconstructed field. Lastly, the three different models are statistically cross-validated using  $\sim 28$  days of model reconstructions to quantitatively assess the model error as a function of activity mode and radial distance.

The DM approach assumes that there are enough similar events in the state-space that the NNs collectively match the event of interest. This assumption can be tested by using the state-space analysis performed in earlier studies [30, 35, 67] in which the



**FIGURE 9** | Substorm enhancement of pressure. **(A,B)** Geomagnetic indices and solar wind parameters in a format similar to **Figures 6a,b**. **(C)** The value of the pressure averaged over all local times at  $r = 4.5R_E$ . **(D,E)** The averaged pressure vs. radial distance at different times during the growth and expansion phases.

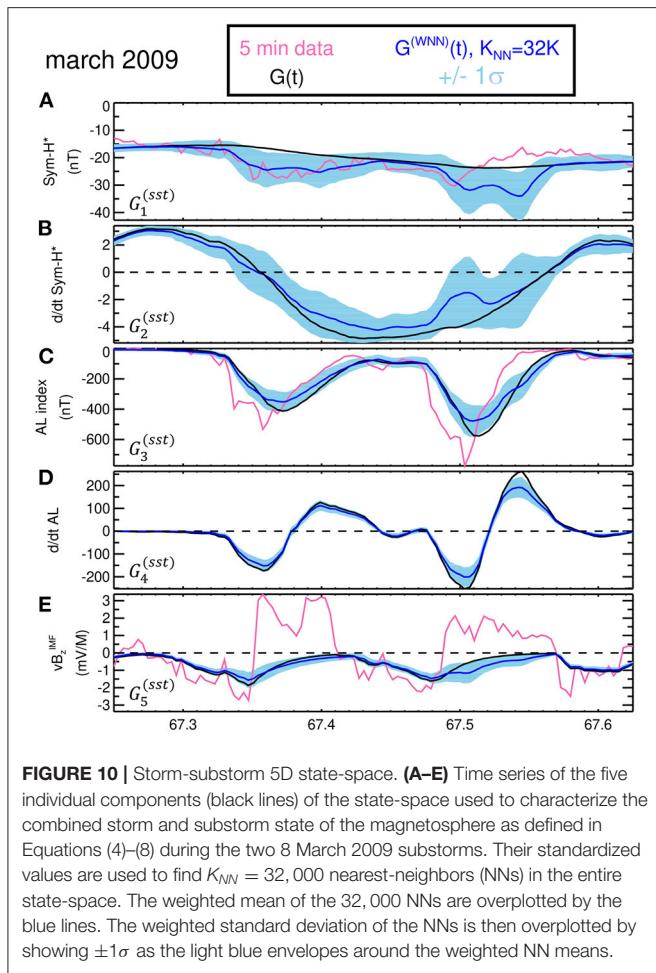
time evolution of the state-space components  $G_i(t)$  (the binning parameters) for the studied event are compared to the mean of those parameters over the NNs  $G_i^{(\text{NN})}(t)$ . Ideally,  $G_i^{(\text{NN})}(t)$  should match  $G_i(t)$ .

**Figure 10** shows the evolution of the five components of the combined storm and substorm state-space (black lines) during the two 8 March 2009 substorms corresponding to the quantities described by Equations (4)–(8). Panels a, c, and e demonstrate how the integral convolutions act to smooth the original data (pink lines). It should be stressed that the smoothing windows are asymmetric in time, that is, they are only computed over previous and not future data points [note the zeros on the upper bounds in the integrals (4)–(8)]. This was introduced to prevent the smoothed parameters from changing prior to the start of an event [48]. For example, without this asymmetry, the model would begin reconstructing storm activity prior to the arrival of the southward IMF in the solar wind. This causes the smoothed values to lag the original (black lines lag behind the pink lines). However, as demonstrated by **Figure 8**, this is not problematic as the time lag is universally applied over the entire dataset. Note how the modeled magnetic field (**Figure 8g**) and model parameters (**Figures 8d–f**) are still largely correlated with the original value

of  $AL$  (**Figure 8a** solid orange line) and not its smoothed value (dotted orange line).

The weighted average of the state-space vector over the closest  $K_{\text{NN}} = 32,000$  NNs is shown by the blue lines:  $\mathbf{G}^{(\text{WNN})} = \sum_j^{K_{\text{NN}}} w_j \mathbf{G}_j / \sum_j^{K_{\text{NN}}} w_j$ , where the weight factor  $w_j$  is computed using the Gaussian distance weighting from Equation (10). In a similar manner, the weighted standard deviations of these parameters over the set of NNs is also computed and the  $\pm 1\sigma$  values are indicated by light blue envelopes that surround  $\mathbf{G}^{(\text{WNN})}$  (blue line) in **Figure 10**. Overall,  $\mathbf{G}^{(\text{WNN})}$  largely stays within  $1\sigma$  for the event of interest (black line).

The most notable inconsistencies appear in  $\langle \text{Sym}-H^* \rangle$  and  $D(\text{Sym}-H^*)$ , which is a result of mixing parameters of different smoothing scales.  $\text{Sym}-H^*$  and its derivative are smoothed over storm time frames of  $\Pi_{st} = 12$  h while the  $AL$  equivalents are instead smoothed over substorm scales  $\Pi_{sst} = 2$  h. This also demonstrates that the pressure-corrected storm indices contain contributions from substorm current systems and is not representative of just a pure ring current as has been discussed in previous studies [79, 80]. Other deviations appear in particularly when  $\langle AL \rangle$  reaches its minimum. The cause of this is the inhomogeneity of the datapoints in state-space as has been extensively discussed in earlier works [35, 36, 67]. This biases



the DM approach toward weaker events as they occur more frequently which is only exacerbated as the events of interest become more extreme [36]. Distance-weighting the NNs can drastically mitigate this issue [35], although, as is shown in the second substorm, it does not entirely correct it. For instance, the minimum value for the second substorm is  $\langle AL \rangle = -576$  nT while the mean over the NNs is only  $\langle AL \rangle^{(WNN)} = -478$  nT.

Next, the reconstructed magnetic field of the *Merged Model* is compared to the available *in-situ* magnetic field measurements from the THEMIS and GOES satellites (**Figure 11**). For this event, all five THEMIS probes are located in the magnetotail, with probes A, D, and E being in a similar orbital configuration situated at about post-midnight and  $r \approx 12R_E$ , while probes B and C are at  $\sim 21$  MLT with  $r \approx 25$  and  $16R_E$ , respectively. As the magnetotail current sheet thins, less magnetic flux threads through it resulting in a decrease in  $B_z$  (e.g., [81]), as is observed by all five THEMIS probes during the growth phases, with  $B_z$  approaching values of  $B_z \sim 1$  nT prior to substorm onset (panels b, e, h, k, and n; black lines). The model captures these thinning signatures for the inner probes A, D, and E (panels b, k, and n; brown lines) as their location places them within the  $B_z$  minimum and TCS region as was shown in **Figures 6d,e**. The

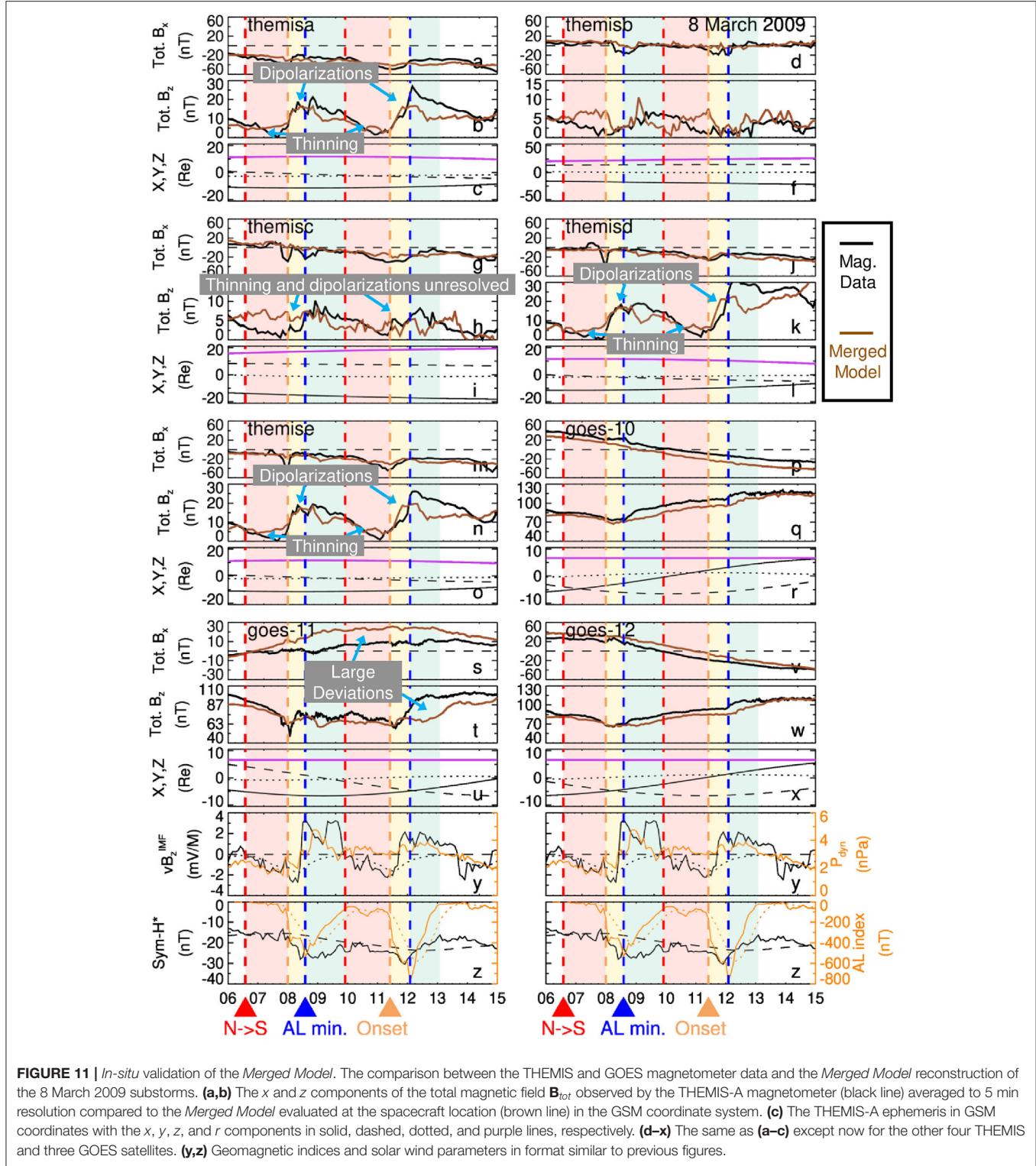
model underestimates the thinning, as the  $B_z$  values only reach  $B_z \sim 4$  nT during the growth phases. One explanation is that the  $\langle vB_s^{IMF} \rangle$  parameter is a suboptimal proxy for the development of the TCS in the magnetotail while another is that the DM approach smears the singularly thin nature of the TCS effectively making it thicker than in reality.

Following substorm onset and throughout the expansion phase, magnetic flux is transported earthward accumulating in the near-tail region and enhancing the value of  $B_z$  there (e.g., [82]), leading to a more dipolar magnetic field configuration. The three inner probes observe this dipolarization as a  $\sim 20$  nT increase in  $B_z$  for the first substorm and a  $\sim 30$  nT increase for the second (panels b, k, n; black lines). As was shown in **Figure 7d**, the model reconstructs this flux pileup across the whole nightside magnetotail to a distance of  $r \approx 16R_E$ . The *in-situ* validation demonstrates that the model nicely captures these  $B_z$  increases for the inner probes (panels b, k, n; black lines). While the model effectively captures the magnitude of the flux enhancement for the first substorm, it underestimates it for the second. This is not unexpected, as the state-space analysis shown in **Figure 10** demonstrated the DM approach underestimated the intensity of the second substorm (**Figure 10c**).

However, the model misses the TCS formation and the dipolarizations at the locations of outer probes B and C (**Figures 11e,h**) which are further down tail and whose MLTs are significantly away from midnight. Further, while the model displays good consistency with the GOES-10 and 12 magnetometers (**Figures 11p,q,v,w**), there are quite large deviations for GOES-11 (**Figure 11t**). This particular pair of substorms appears quite global in nature, showing signs of current sheet thinning as far as  $r \sim 20R_E$  down tail at 21 MLT and large dipolarizations at geosynchronous orbit at 3 MLT. While the  $AL$  index is a good indicator of the overall substorm intensity, it has its limitations. Firstly, it is derived from a rather small (10–12) number of magnetometer stations [44] and secondly, it yields no information about the local time configuration of the substorm currents [83]. As such, the DM approach will tend to reconstruct the typical substorm as is represented by the state-space parameters but will have no specific insight into the event's MLT configuration. This may be the underlying cause of the deviations between the model and the THEMIS B and C as well as the GOES-11 satellites. Some of these issues might be addressed by using another substorm index constructed from a more robust collection of magnetometer stations, such as the *SML* index [84], which furthermore, is computed for differing local time bins [85].

Overall, these results largely mirror to the validation of the SST model as shown in [15] (their Figure 2). This is not unexpected as the *Merged Model* closely matches the *Tail Model* at these spacecraft's locations and given that the *Tail Model* is very similar to the SST19 model. A shortcoming of the spacecraft configuration for this particular event is that none of them are earthward of geosynchronous orbit and according to **Figures 3, 4**, the improvement of the *Merged Model* over the SST19 model should be most apparent in this region. This will be addressed by a quantitative statistical analysis next.

In order to perform a quantitative uncertainty estimate the three models are cross-validated. In this technique, a subset of



the data is reserved, that is it is not included when the model is fit, and instead forms a validation dataset. Statistical analysis is then performed on the independent validation set, and because this set was not used while fitting the model, it represents an out-of-sample test.

For the purposes here, the validation set should include spacecraft which adequately cover both the inner magnetosphere and near tail regions while there is also storm and substorm activity. The year 2015 was chosen to be the validation time interval as both the Van Allen probes and the THEMIS missions

were sampling the inner magnetosphere, while the Cluster and MMS missions observed the near-tail. Further, the maximum sunspot count of solar cycle 24 peaked in April of 2014, so solar activity was still quite high throughout 2015 resulting in an elevated occurrence of geomagnetic storms [86]. Thus, when performing the model cross-validation, the entirety of the spacecraft dataset for the year 2015 was excised from the magnetometer database when fitting the models. That is, only data from spacecraft missions spanning the years 1995–2014 and 2016–2018 are used during fitting, while only data from 2015 are used in this cross-validation. Owing to the computational expense needed to fit the model, it was not feasible in this study to model the entirety of the year 2015. Instead, only times corresponding to storm and substorm activity from the 3 months of June, September, and December were modeled. At least one strong storm ( $Dst \lesssim -100$  nT) occurred during each of these months, and their separation in time allowed the spacecraft to sample a range of different local times. The start of September corresponds to the beginning of the MMS primary science mission phase, and as such is when the MMS magnetometer data becomes available. Here, storm and substorm activity time intervals are defined as when  $\langle Sym-H^* \rangle \leq -50$  nT or  $\langle AL \rangle \leq -300$  nT. To ensure that broader portions of the storm main and recovery phases and substorm growth and recovery phases were also included, the time intervals were expanded by  $\pm 3$  h and  $\pm 30$  min, respectively. In total, the identified storm and substorm activity time intervals across these 3 months spans 673.5 h or about 28 days, corresponding to 8,046 model fits and  $\approx 68,000$  spacecraft magnetic field observations when evaluated at a 5 min cadence. Here, the out-of-sample model error is quantified using the magnitude of the difference between the model evaluated at the spacecraft location and the observed magnetic field there:

$$\epsilon_i = \|\mathbf{B}^{(mod)}(\mathbf{r}_i) - \mathbf{B}_i^{obs}\|. \quad (12)$$

This quantity is similar to the model residuals which are key to its fitting, for example, as seen in the objective function (9), but without the weighting factors. They also have been the primary metric for testing empirical magnetic field model fidelity in previous studies (e.g., [87–89]). Specifically, the average magnitude of the difference between the model and observation  $\langle \epsilon \rangle = \frac{1}{N} \sum \epsilon_i$  is reported in **Table 1**, where N is the number of samples.

The *Inner Model* performs marginally better for the Van Allen Probes data than the *Tail Model* as their orbit is entirely within the inner magnetosphere region. As **Figures 3, 4** indicated, this marginal improvement in  $\mathbf{B}$  is crucial for reconstructing the current density  $\mathbf{j}$  which depends on the spatial derivatives of  $\mathbf{B}$ . In contrast, the *Tail Model* has lower errors for the Cluster spacecraft which spends relatively little time in the inner equatorial magnetosphere. The THEMIS and MMS missions have similar errors for both models. All models display notably lower errors for the less active times, presumably because the global configuration is more regular. Substorm and storm performance is rather similar, although, the storms during the validation interval contain significant substorm activity based on the *AL* index. Further analysis could be performed to separate isolated, multiple, and storm-time substorms, and storm intervals

with and without substorm activity. Importantly, the *Merged Model* generally matches the lower error of the two other models, statistically validating the algorithm discussed in 3.1. Indeed, the *Merged Model* has the smallest error across the entire validation set at  $\langle \epsilon \rangle = 16.228$  nT. To put this in context, running the same model cross-validation for the commonly used T89 and T96 yields errors of  $\langle \epsilon \rangle = 20.006$  and 21.777 nT, respectively, while using no model ( $B_{ext} = 0$ ) gives is 37.458 nT.

To get a better indication of how the model errors change as a function of distance, the errors were collected into different radial bins ( $\Delta r = 0.25R_E$  when  $r < 12R_E$ ;  $\Delta r = 1.0R_E$  when  $r > 12R_E$ , where  $\Delta r$  is the size of the bin) and for the different types of activity levels as are displayed in **Figure 12**. The *Tail Model* (orange lines) has significantly lower errors tailward of  $r = 12R_E$  compared to the *Inner Model* (blue lines), although, recall the later was not fit using data in this region, so it should not be expected to perform well here.

Meanwhile, the *Inner Model* (blue lines) has lower errors within  $3R_E \leq r \leq 6R_E$ , indicating that it does yield a more accurate reconstruction of the ring current region. However, errors are still relatively high here, particularly during storm times (**Figure 12D**), indicating the model still potentially underestimates the ring current intensity. One cause might be, given that there were several strong storms during this validation period (one storm reached a minimum  $Dst \approx -200$  nT while another hit  $Dst \approx -150$  nT), the aforementioned bias toward weaker events. Another, is that strong storms tend to have a complex morphology (e.g., [36]), and the multitude of mesoscale features can simply not be discerned with the DM approach. Also of note, the errors significantly increase earthward of  $r = 2.5R_E$ . This has been observed in *in-situ* comparisons in previous studies and was attributed to attitude uncertainty issues which make it difficult to distinguish between the internal and external fields (e.g., [35]) as the former is very large close to the planet. If these are indeed observational errors, then perhaps these datapoints earthward of  $r = 2.5R_E$  should be excised from the model database in future studies. Importantly for the context of this study, is that the errors of the *Merged Model* (dashed brown line) tend to follow the smaller of the other two across all radial bins. This confirms that the merging algorithm discussed in section 3.1 works as intended.

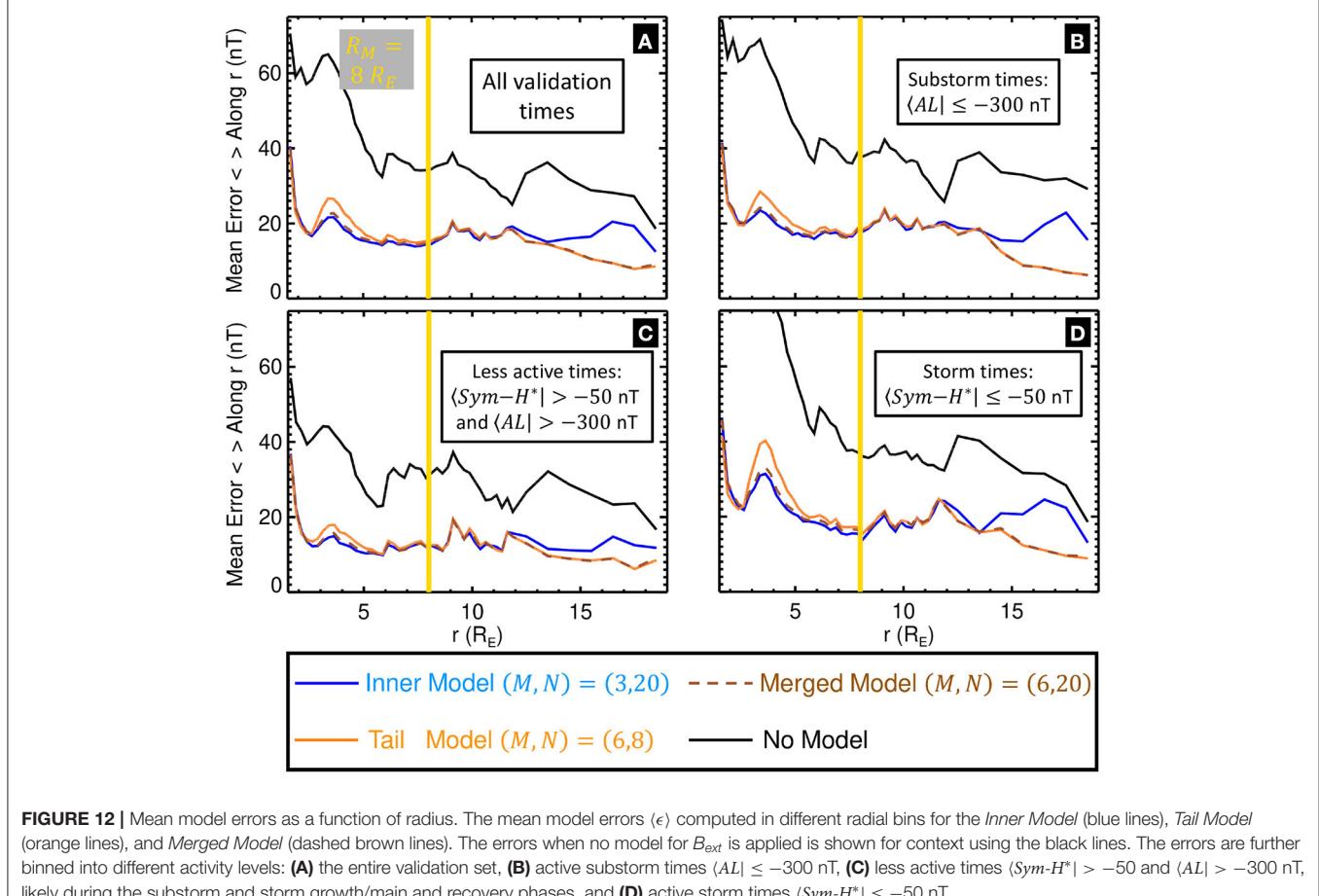
## 4. DISCUSSION AND CONCLUSION

In this paper we presented a new method of the empirical reconstruction of the magnetospheric substorms, which not only resolves the corresponding reconfiguration of the magnetotail but also resolves both westward and eastward currents in the inner magnetosphere that reflect the associated storm-type phenomena. In particular, it becomes possible to resolve and quantitatively evaluate the buildup of the storm-time plasma pressure, which was under-resolved in the previous substorm reconstructions [15]. The key to the solution of such a combined description of the inner magnetosphere is similar to the original kNN DM method [15, 22], where a swarm of many synthetic probes neighboring the event of interest in the binning parameter space was used. Here similar virtual probes were used and

**TABLE 1 |** Out-of-sample model errors ( $\epsilon$ ) for the *Inner*, *Tail*, and *Merged Models* for different spacecraft and activity levels.

	Samples	No model ( $\epsilon$ )(nT)	Inner model ( $\epsilon$ )(nT)	Tail model ( $\epsilon$ )(nT)	Merged model ( $\epsilon$ )(nT)
All	67,801	37.458	17.021	16.891	16.228
Van Allen	14,958	45.520	16.458	18.713	16.862
THEMIS	18,125	35.769	17.753	17.597	17.240
MMS	14,533	34.590	18.721	19.855	19.281
Cluster	20,185	35.093	15.558	12.773	12.649
Storm times:					
$(\text{Sym}-H^*) \leq -50$ nT	28,778	45.757	20.547	20.680	19.709
Substorm times:					
$(AL) \leq -300$ nT	35,111	40.790	19.135	18.801	18.047
Less active times:					
$(\text{Sym}-H^*) > -50$ and $(AL) > -300$ nT	17,265	29.719	13.034	13.010	12.526

The equivalent errors with no external field model is given for reference.



**FIGURE 12 |** Mean model errors as a function of radius. The mean model errors ( $\epsilon$ ) computed in different radial bins for the *Inner Model* (blue lines), *Tail Model* (orange lines), and *Merged Model* (dashed brown lines). The errors when no model for  $B_{\text{ext}}$  is applied is shown for context using the black lines. The errors are further binned into different activity levels: **(A)** the entire validation set, **(B)** active substorm times  $(AL) \leq -300$  nT, **(C)** less active times  $(\text{Sym}-H^*) > -50$  and  $(AL) > -300$  nT, likely during the substorm and storm growth/main and recovery phases, and **(D)** active storm times  $(\text{Sym}-H^*) \leq -50$  nT.

combined from different versions of the SST19 model focused on the inner magnetosphere and the tail region.

This merged resolution approach is also similar to coupled first-principles models of the magnetosphere [74, 75, 77, 90], where the kinetic description of the inner magnetosphere is combined with the magnetohydrodynamic (MHD) description

of the whole magnetosphere in global MHD models. An important advantage of the present empirical method, compared to the aforementioned combinations of the first-principle models, is that the empirical reconstructions weakly depend on the location of the coupling boundary and they do not require any special description of the interaction between

inner and outer magnetosphere models. This approach may be further improved by using a more optimized merging boundary instead of the simple static cylindrical boundary used here. For instance, the merging boundary can be made dynamical, either redetermined for each time step or made a function of storm and substorm activity level. Further, the azimuthally symmetric boundary used here is suboptimal as it does not account for day-night asymmetries. This can be addressed by introducing a shift to the center of the cylinder or by instead using an elliptic cylinder.

In conclusion, the merged modeling technique using virtual observations effectively reconstructs regions of the magnetosphere possessing different spatial scales. This may also have utility in other DM and machine learning applications in which disparate density of data makes it difficult to model the system using a single resolution.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

GS was the primary author writing the original draft, led conceptualization, investigation, and formal analysis including

software, data curation, visualization, and validation. MS assisted in writing by reviewing and editing, developed the methodology presented here including its conceptualization, software, visualizations, validation, managed the project, including project administration, funding acquisition, resources, and supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by NASA Grants 80NSSC19K0074, 80NSSC20K1271, and 80NSSC20K1787, as well as NSF Grants AGS-1702147 and AGS-1744269.

## ACKNOWLEDGMENTS

We acknowledge interesting and useful discussions with Tom Sotirelis and Sasha Ukhorskiy that inspired this concept. We thank the many spacecraft and instrument teams and their PIs who produced the data sets we used in this study, including the Cluster, Geotail, Polar, IMP-8, GOES, THEMIS, Van Allen Probes, and MMS, particularly their magnetometer teams. We also thank the SPDF for the OMNI database for solar wind values, which was composed of data sets from the IMP-8, ACE, WIND, and Geotail missions, and also the WDC in Kyoto for the Geomagnetic indices.

## REFERENCES

- Sharma AS, Baker DN, Grande M, Kamide Y, Lakhina GS, McPherron RM, et al. The storm-substorm relationship: current understanding and outlook. In: Sharma AS, Kamide Y, Lakhina GS, editors. *Disturbances in Geospace: The Storm-Substorm Relationship* (Washington, DC: American Geophysical Union (AGU)) (2003). p. 1–14. doi: 10.1029/142GM01
- Eastwood JP, Nakamura R, Turc L, Mejnertsen L, Hesse M. The Scientific Foundations of forecasting magnetospheric space weather. *Space Sci Rev.* (2017) 212:1221–52. doi: 10.1007/s11214-017-0399-8
- Iyemori T. Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *J Geomagn Geoelectr.* (1990) 42:1249–65. doi: 10.5636/jgg.42.1249
- Davis TN, Sugiura M. Auroral electrojet activity index AE and its universal time variations. *J Geophys Res.* (1966) 71:785–801. doi: 10.1029/JZ071i003p00785
- Lee DY, Min KW. Statistical features of substorm indicators during geomagnetic storms. *J Geophys Res Space Phys.* (2002) 107:SMP 16-1–12. doi: 10.1029/2002JA009243
- Chapman S. Earth storms: retrospect and prospect. *J Phys Soc Jpn Suppl.* (1962) 17:6.
- Kamide Y. Is substorm occurrence a necessary condition for a magnetic storm? *J Geomagn Geoelectr.* (1992) 44:109–17. doi: 10.5636/jgg.44.109
- Kamide Y, Baumjohann W, Daglis IA, Gonzalez WD, Grande M, Joselyn JA, et al. Current understanding of magnetic storms: storm-substorm relationships. *J Geophys Res Space Phys.* (1998) 103:17705–28. doi: 10.1029/98JA01426
- De Michelis P, Consolini G, Materassi M, Tozzi R. An information theory approach to the storm-substorm relationship. *J Geophys Res Space Phys.* (2011) 116:A08225. doi: 10.1029/2011JA016535
- Runge J, Balasis G, Daglis IA, Papadimitriou C, Donner RV. Common solar wind drivers behind magnetic storm-magnetospheric substorm dependency. *Sci Rep.* (2018) 8:16987. doi: 10.1038/s41598-018-35250-5
- Reeves GD, Henderson MG, Skoug RM, Thomsen MF, Borovsky JE, Funsten HO, et al. Image, polar, and geosynchronous observations of substorm and ring current ion injection. *Geophys Monogr Ser.* (2003) 142:91–101. doi: 10.1029/142GM09
- Gkioulidou M, Ukhorskiy AY, Mitchell DG, Sotirelis T, Mauk BH, Lanzerotti LJ. The role of small-scale ion injections in the buildup of Earth's ring current pressure: Van Allen Probes observations of the 17 March 2013 storm. *J Geophys Res Space Phys.* (2014) 119:7327–42. doi: 10.1002/2014JA020096
- Sandhu JK, Rae IJ, Freeman MP, Forsyth C, Gkioulidou M, Reeves GD, et al. Energization of the ring current by substorms. *J Geophys Res Space Phys.* (2018) 123:8131–48. doi: 10.1029/2018JA025766
- Sandhu JK, Rae IJ, Freeman MP, Gkioulidou M, Forsyth C, Reeves GD, et al. Substorm-ring current coupling: a comparison of isolated and compound substorms. *J Geophys Res Space Phys.* (2019) 124:6776–91. doi: 10.1029/2019JA026766
- Stephens GK, Sitnov MI, Korth H, Tsyganenko NA, Ohtani S, Gkioulidou M, et al. Global empirical picture of magnetospheric substorms inferred from multimission magnetometer data. *J Geophys Res Space Phys.* (2019) 124, 1085–110. doi: 10.1029/2018JA025843
- Mead GD, Fairfield DH. A quantitative magnetospheric model derived from spacecraft magnetometer data. *J Geophys Res.* (1975) 80:523–34. doi: 10.1029/JA080i004p00523
- Tsyganenko NA. Modeling the Earth's magnetospheric magnetic field confined within a realistic magnetopause. *J Geophys Res Space Phys.* (1995) 100:5599–612. doi: 10.1029/94JA03193
- Tsyganenko NA. Effects of the solar wind conditions on the global magnetospheric configurations as deduced from data-based field models. In: *Proceedings of the Third International Conference on Substorms (ICS-3), ESA SP-389*, Versailles (1996). p. 181–5.
- Tsyganenko NA, Sitnov MI. Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *J Geophys Res Space Phys.* (2005) 110:A03208. doi: 10.1029/2004JA010798

20. Tsyganenko NA. Data-based modelling of the Earth's dynamic magnetosphere: a review. *Ann Geophys.* (2013) 31:1745–72. doi: 10.5194/angeo-31-1745-2013
21. Tsyganenko NA, Sitnov MI. Magnetospheric configurations from a high-resolution data-based magnetic field model. *J Geophys Res Space Phys.* (2007) 112:A06225. doi: 10.1029/2007JA012260
22. Sitnov MI, Tsyganenko NA, Ukhorskiy AY, Brandt PC. Dynamical data-based modeling of the storm-time geomagnetic field with enhanced spatial resolution. *J Geophys Res Space Phys.* (2008) 113:A07218. doi: 10.1029/2007JA013003
23. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inform Theory.* (1967) 13:21–7. doi: 10.1109/TIT.1967.1053964
24. Sitnov MI, Tsyganenko NA, Ukhorskiy AY, Anderson BJ, Korth H, Lui ATY, et al. Empirical modeling of a CIR-driven magnetic storm. *J Geophys Res Space Phys.* (2010) 115:A07231. doi: 10.1029/2009JA015169
25. Pulkkinen TI, Baker DN, Fairfield DH, Pellinen RJ, Murphree JS, Elphinstone RD, et al. Modeling the growth phase of a substorm using the Tsyganenko Model and multi-spacecraft observations: CDAW-9. *Geophys Res Lett.* (1991) 18:1963–6. doi: 10.1029/91GL02002
26. Kubyshkina MV, Sergeev VA, Pulkkinen TI. Hybrid Input Algorithm: an event-oriented magnetospheric model. *J Geophys Res Space Phys.* (1999) 104:24977–93. doi: 10.1029/1999JA900222
27. Tsyganenko NA. An empirical model of the substorm current wedge. *J Geophys Res Space Phys.* (1997) 102:19935–41. doi: 10.1029/97JA01904
28. Sergeev VA, Tsyganenko NA, Smirnov MV, Nikolaev AV, Singer HJ, Baumjohann W. Magnetic effects of the substorm current wedge in a “spread-out wire” model and their comparison with ground, geosynchronous, and tail lobe data. *J Geophys Res Space Phys.* (2011) 116:A07218. doi: 10.1029/2011JA016471
29. Sergeev VA, Nikolaev AV, Tsyganenko NA, Angelopoulos V, Runov AV, Singer HJ, et al. Testing a two-loop pattern of the substorm current wedge (SCW2L). *J Geophys Res Space Phys.* (2014) 119:947–63. doi: 10.1002/2013JA019629
30. Sitnov MI, Stephens GK, Tsyganenko NA, Miyashita Y, Merkin VG, Motoba T, et al. Signatures of nonideal plasma evolution during substorms obtained by mining multimission magnetometer data. *J Geophys Res Space Phys.* (2019) 124:8427–56. doi: 10.1029/2019JA027037
31. Hoffman RA, Bracken PA. Magnetic effects of the quiet-time proton belt. *J Geophys Res.* (1965) 70:3541–56. doi: 10.1029/JZ070i015p03541
32. Lui ATY, McEntire RW, Krimigis SM. Evolution of the ring current during two geomagnetic storms. *J Geophys Res Space Phys.* (1987) 92:7459–70. doi: 10.1029/JA092iA07p07459
33. Stephens GK, Sitnov MI, Ukhorskiy AY, Roelof EC, Tsyganenko NA, Le G. Empirical modeling of the storm time innermost magnetosphere using Van Allen Probes and THEMIS data: eastward and banana currents. *J Geophys Res Space Phys.* (2016) 121:157–70. doi: 10.1002/2015JA021700
34. Sitnov MI, Stephens GK, Tsyganenko NA, Ukhorskiy AY, Wing S, Korth H, et al. Chapter 15: spatial structure and asymmetries of magnetospheric currents inferred from high-resolution empirical geomagnetic field models. In: Haaland S, Runov A, Forsyth C, editors. *Dawn-Dusk Asymmetries in Planetary Plasma Environments.* Hoboken, NJ; Washington, DC: American Geophysical Union (AGU) (2017). p. 199–212. doi: 10.1002/9781119216346.ch15
35. Stephens GK, Bingham ST, Sitnov MI, Gkioulidou M, Merkin VG, Korth H, et al. Storm-time plasma pressure inferred from multi-mission measurements and its validation using Van Allen probes particle data. *Space Weather.* (2020) 18:e2020SW002583. doi: 10.1029/2020SW002583
36. Sitnov MI, Stephens GK, Tsyganenko NA, Korth H, Roelof EC, Brandt PC, et al. Reconstruction of extreme geomagnetic storms: breaking the data paucity curse. *Space Weather.* (2020) 18:e2020SW002561. doi: 10.1029/2020SW002561
37. Thébault E, Finlay CC, Beggan CD, Alken P, Aubert J, Barrois O, et al. International Geomagnetic Reference Field: the 12th generation. *Earth Planets Space.* (2015) 67:79. doi: 10.1186/s40623-015-0313-0
38. Tsyganenko NA. A magnetospheric magnetic field model with a warped tail current sheet. *Planet Space Sci.* (1989) 37:5–20. doi: 10.1016/0032-0633(89)90066-4
39. Tsyganenko NA. Methods for quantitative modeling of the magnetic field from Birkeland currents. *Planet Space Sci.* (1991) 39:641–54. doi: 10.1016/0032-0633(91)90058-I
40. Murphy KR, Mann IR, Rae IJ, Waters CL, Frey HU, Kale A, et al. The detailed spatial structure of field-aligned currents comprising the substorm current wedge. *J Geophys Res Space Phys.* (2013) 118:7714–27. doi: 10.1002/2013JA018979
41. Tsyganenko NA. A model of the near magnetosphere with a dawn-dusk asymmetry 1. Mathematical structure. *J Geophys Res Space Phys.* (2002) 107:SMP 12-1–15. doi: 10.1029/2001JA000219
42. Shue JH, Song P, Russell CT, Steinberg JT, Chao JK, Zastenker G, et al. Magnetopause location under extreme solar wind conditions. *J Geophys Res Space Phys.* (1998) 103:17691–700. doi: 10.1029/98JA01103
43. Vassiliadis D. Systems theory for geospace plasma dynamics. *Rev Geophys.* (2006) 44:RG2002. doi: 10.1029/2004RG000161
44. Rostoker G. Geomagnetic indices. *Rev Geophys.* (1972) 10:935–50. doi: 10.1029/RG010i004p00935
45. Gonzalez WD, Joselyn JA, Kamide Y, Kroehl HW, Rostoker G, Tsurutani BT, et al. What is a geomagnetic storm? *J Geophys Res Space Phys.* (1994) 99:5771–92. doi: 10.1029/93JA02867
46. Burton RK, McPherron RL, Russell CT. An empirical relationship between interplanetary conditions and Dst. *J Geophys Res.* (1975) 80:4204–14. doi: 10.1029/JA080i031p04204
47. Clauer CR, McPherron RL, Searls C, Kivelson MG. Solar wind control of auroral zone geomagnetic activity. *Geophys Res Lett.* (1981) 8:915–8. doi: 10.1029/GL008i008p00915
48. Sitnov MI, Ukhorskiy AY, Stephens GK. Forecasting of global data-binning parameters for high-resolution empirical geomagnetic field models. *Space Weather.* (2012) 10:S09001. doi: 10.1029/2012SW000783
49. Stern DP. Tail modeling in a stretched magnetosphere: 1. Methods and transformations. *J Geophys Res Space Phys.* (1987) 92:4437–48. doi: 10.1029/JA092iA05p04437
50. Tsyganenko NA. Modeling of twisted/warped magnetospheric configurations using the general deformation method. *J Geophys Res Space Phys.* (1998) 103:23551–63. doi: 10.1029/98JA02296
51. Tsyganenko NA. Data-based modeling of the geomagnetosphere with an IMF-dependent magnetopause. *J Geophys Res Space Phys.* (2014) 119:335–54. doi: 10.1002/2013JA019346
52. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques.* 4th ed. San Francisco, CA: Morgan Kaufmann Publishers Inc. (2016).
53. Bortnik J, Chu X, Ma Q, Li W, Zhang X, Thorne R, et al. Chapter 11–Artificial neural networks for determining magnetospheric conditions. In: Camporeale E, Wing S, Johnson JR, editors. *Machine Learning Techniques for Space Weather.* Elsevier (2018). p. 279–300.
54. O'Brien TP, McPherron RL. An empirical phase space analysis of ring current dynamics: solar wind control of injection and decay. *J Geophys Res Space Phys.* (2000) 105:7707–19. doi: 10.1029/1998JA000437
55. Partamies N, Juusola L, Tanskanen E, Kauristie K. Statistical properties of substorms during different storm and solar cycle phases. *Ann Geophys.* (2013) 31:349–58. doi: 10.5194/angeo-31-349-2013
56. Camporeale E. The challenge of machine learning in space weather: nowcasting and forecasting. *Space Weather.* (2019) 17:1166–207. doi: 10.1029/2018SW002061
57. Angelopoulos V. The THEMIS mission. *Space Sci Rev.* (2008) 141:5. doi: 10.1007/s11214-008-9336-1
58. Mauk BH, Fox NJ, Kanekal SG, Kessel RL, Sibeck DG, Ukhorskiy A. Science objectives and rationales for the radiation belt storm probes mission. *Space Sci Rev.* (2013) 179:3–27. doi: 10.1007/s11214-012-9908-y
59. Angelopoulos V. The ARTEMIS mission. *Space Sci Rev.* (2011) 165:3–25. doi: 10.1007/s11214-010-9687-2
60. Nagai T, Fujimoto M, Saito Y, Machida S, Terasawa T, Nakamura R, et al. Structure and dynamics of magnetic reconnection for substorm onsets with Geotail observations. *J Geophys Res Space Phys.* (1998) 103:4419–40. doi: 10.1029/97JA02190
61. Nelder JA, Mead R. A simplex method for function minimization. *Comput J.* (1965) 7:308–13. doi: 10.1093/comjnl/7.4.308
62. Jackson DD. Interpretation of inaccurate, insufficient and inconsistent data. *Geophys J Int.* (1972) 28:97–109. doi: 10.1111/j.1365-246X.1972.tb06115.x

63. Press WH, Teukolsky SA, Flannery BP, Vetterling WT. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. 2nd ed. New York, NY: Cambridge University Press (1992).
64. Campbell WH. Geomagnetic storms, the DST ring-current myth and lognormal distributions. *J Atmos Terres Phys.* (1996) 58:1171–87. doi: 10.1016/0021-9169(95)00103-4
65. Liou K, Sotirelis T, Richardson I. Substorm occurrence and intensity associated with three types of solar wind structure. *J Geophys Res Space Phys.* (2018) 123:485–96. doi: 10.1002/2017JA024451
66. Kataoka R. Extreme geomagnetic activities: a statistical study. *Earth Planets Space.* (2020) 72:124. doi: 10.1186/s40623-020-01261-8
67. Sitnov MI, Stephens GK, Gkioulidou M, Merkin V, Ukhorskiy AY, Korth H, et al. Chapter 11—empirical modeling of extreme events: storm-time geomagnetic field, electric current, and pressure distributions. In: Buzulukova N, editor. *Extreme Events in Geospace*. Cambridge, MA: Elsevier (2018). p. 259–79. doi: 10.1016/B978-0-12-812700-1.00011-X
68. Wolf RA. The quasi-static (slow-flow) region of the magnetosphere. In: Carovillano RL, Forbes JM, editors. *Solar-Terrestrial Physics*. Dordrecht: Springer Netherlands (1983). p. 303–68. doi: 10.1007/978-94-009-7194-3\_14
69. Jordanova VK, Kozya JU, Nagy AF, Khazanov GV. Kinetic model of the ring current-atmosphere interactions. *J Geophys Res Space Phys.* (1997) 102:14279–91. doi: 10.1029/96JA03699
70. Fok MC, Moore TE, Delcourt DC. Modeling of inner plasma sheet and ring current during substorms. *J Geophys Res Space Phys.* (1999) 104:14557–69. doi: 10.1029/1999JA900014
71. Lyon JG, Brecht SH, Huba JD, Fedder JA, Palmadesso PJ. Computer simulation of a geomagnetic substorm. *Phys Rev Lett.* (1981) 46:1038–41. doi: 10.1103/PhysRevLett.46.1038
72. Powell KG, Roe PL, Linde TJ, Gombosi TI, De Zeeuw DL. A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J Comput Phys.* (1999) 154:284–309. doi: 10.1006/jcph.1999.6299
73. Raeder J, McPherron RL, Frank LA, Kokubun S, Lu G, Mukai T, et al. Global simulation of the geospace environment modeling substorm challenge event. *J Geophys Res Space Phys.* (2001) 106:381–95. doi: 10.1029/2000JA000605
74. De Zeeuw DL, Sazykin S, Wolf RA, Gombosi TI, Ridley AJ, Tóth G. Coupling of a global MHD code and an inner magnetospheric model: initial results. *J Geophys Res Space Phys.* (2004) 109:A12219. doi: 10.1029/2003JA010366
75. Pembroke A, Toftoletto F, Sazykin S, Wiltberger M, Lyon J, Merkin V, et al. Initial results from a dynamic coupled magnetosphere-ionosphere-ring current model. *J Geophys Res Space Phys.* (2012) 117:A02211. doi: 10.1029/2011JA016979
76. Glocer A, Fol M, Meng X, Toth G, Buzulukova N, Chen S, et al. CRCM + BATS-R-US two-way coupling. *J Geophys Res Space Phys.* (2013) 118:1635–50. doi: 10.1002/jgra.50221
77. Cramer WD, Raeder J, Toftoletto FR, Gilson M, Hu B. Plasma sheet injections into the inner magnetosphere: two-way coupled OpenGGCM-RCM model results. *J Geophys Res Space Phys.* (2017) 122:5077–91. doi: 10.1002/2017JA024104
78. Birn J, Hesse M, Haerendel G, Baumjohann W, Shiokawa K. Flow braking and the substorm current wedge. *J Geophys Res Space Phys.* (1999) 104:19895–903. doi: 10.1029/1999JA900173
79. Ohtani S, Nosé M, Rostoker G, Singer H, Lui ATY, Nakamura M. Storm-substorm relationship: contribution of the tail current to DST. *J Geophys Res Space Phys.* (2001) 106:21199–209. doi: 10.1029/2000JA000400
80. Sandhu JK, Rae IJ, Walach MT. Challenging the use of ring current indices during geomagnetic storms. *J Geophys Res Space Phys.* (2021) 126:e2020JA028423. doi: 10.1029/2020JA028423
81. Baker DN, Pulkkinen TI, Angelopoulos V, Baumjohann W, McPherron RL. Neutral line model of substorms: past results and present view. *J Geophys Res Space Phys.* (1996) 101:12975–3010. doi: 10.1029/95JA03753
82. Russell CT, McPherron RL. The magnetotail and substorms. *Space Sci Rev.* (1973) 15:205–66. doi: 10.1007/BF00169321
83. Ahn BH, Kroehl HW, Kamide Y, Kihm EA. Universal time variations of the auroral electrojet indices. *J Geophys Res Space Phys.* (2000) 105:267–75. doi: 10.1029/1999JA900364
84. Newell PT, Gjerloev JW. Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. *J Geophys Res Space Phys.* (2011) 116:A12211. doi: 10.1029/2011JA016779
85. Newell PT, Gjerloev JW. Local geomagnetic indices and the prediction of auroral power. *J Geophys Res Space Phys.* (2014) 119:9790–803. doi: 10.1002/2014JA020524
86. Gerontidou M, Mavromichalaki H, Daglis T. High-speed solar wind streams and geomagnetic storms during solar cycle 24. *Sol Phys.* (2018) 293:131. doi: 10.1007/s11207-018-1348-8
87. McCollough JP, Gannon JL, Baker DN, Gehmeyr M. A statistical comparison of commonly used external magnetic field models. *Space Weather.* (2008) 6:S10001. doi: 10.1029/2008SW000391
88. Brito TV, Morley SK. Improving empirical magnetic field models by fitting to *in situ* data using an optimized parameter approach. *Space Weather.* (2017) 15:1628–48. doi: 10.1002/2017SW001702
89. Kubyshkina M, Sergeev VA, Tsygankova NA, Zheng Y. Testing efficiency of empirical, adaptive, and global MHD magnetospheric models to represent the geomagnetic field in a variety of conditions. *Space Weather.* (2019) 17:672–86. doi: 10.1029/2019SW002157
90. Glocer A, Tóth G, Ma Y, Gombosi T, Zhang JC, Kistler LM. Multifluid block-adaptive-tree solar wind roe-type upwind scheme: magnetospheric composition and dynamics during geomagnetic storms—initial results. *J Geophys Res Space Phys.* (2009) 114:A12203. doi: 10.1029/2009JA014418

**Conflict of Interest:** 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.

Copyright © 2021 Stephens and Sitnov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read for greatest visibility and readership



## FAST PUBLICATION

Around 90 days from submission to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles



## REPRODUCIBILITY OF RESEARCH

Support open data and methods to enhance research reproducibility



## DIGITAL PUBLISHING

Articles designed for optimal readership across devices



FOLLOW US  
@frontiersin



IMPACT METRICS  
Advanced article metrics track visibility across digital media



EXTENSIVE PROMOTION  
Marketing and promotion of impactful research



LOOP RESEARCH NETWORK  
Our network increases your article's readership