JGR Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2024JH000235

**Key Points:**

- Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) is an algorithm that predicts plasma and magnetic field in Earth's magnetosheath using inputs from in-situ monitors at L1
- PRIME-SH accurately predicts the magnetosheath conditions in a statistical sense and its predictions obey conservation laws at the shock
- PRIME-SH can be used to easily assemble continuous maps of the magnetosheath, addressing spatial limitations of in situ data

**Correspondence to:**
C. O'Brien,
obrienco@bu.edu

**Citation:**
O'Brien, C., Walsh, B. M., Zou, Y., Qudsi, R., Tasnim, S., Zhang, H., & Sibeck, D. G. (2024). PRIME-SH: A data-driven probabilistic model of Earth's magnetosheath. *Journal of Geophysical Research: Machine Learning and Computation*, *1*, e2024JH000235. https://doi.org/10.1029/2024JH000235

Received 22 APR 2024
Accepted 10 JUL 2024

© 2024 The Author(s). Journal of Geophysical Research: Machine Learning and Computation published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# PRIME-SH: A Data-Driven Probabilistic Model of Earth's Magnetosheath

**C. O'Brien[1], B. M. Walsh[1,2], Y. Zou[3], R. Qudsi[1], S. Tasnim[4], H. Zhang[5], and D. G. Sibeck[6]**

[1]Center for Space Physics, Boston University, Boston, MA, USA, [2]Department of Mechanical Engineering, Boston University, Boston, MA, USA, [3]Johns Hopkins University Applied Physics Lab, Laurel, MD, USA, [4]Institute for Solar-Terrestrial Physics, German Aerospace Center (DLR), Neustrelitz, Germany, [5]Computer Science Department, University of Alabama in Huntsville, Huntsville, AL, USA, [6]Heliophysics Science Division, NASA/GSFC, Greenbelt, MD, USA

**Abstract** A data-driven model of Earth's magnetosheath is developed by training a recurrent neural network (RNN) with probabilistic outputs to reproduce Magnetospheric MultiScale (MMS) measurements of the magnetosheath plasma and magnetic field using measurements from the Wind spacecraft upstream of Earth at the first Earth-Sun Lagrange point (L1). This model, called Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) in reference to its progenitor algorithm PRIME, is shown to predict spacecraft observations of magnetosheath conditions accurately in a statistical sense with a continuous rank probability score of $0.227\sigma$ (dimensionless standard deviation units). PRIME-SH is shown to be more accurate than many current analytical models of the magnetosheath. Furthermore, PRIME-SH is shown to reproduce physics not explicitly enforced during training, such as field line draping, the dayside plasma depletion layer, the magnetosheath flow stagnation point, and the Rankine-Hugoniot MHD shock jump conditions. PRIME-SH has the additional benefits of being computationally inexpensive relative to global MHD simulations, being capable of reproducing difficult-to-model physics such as temperature anisotropy, and being capable of reliably estimating its own uncertainty to within 3.5%.

**Plain Language Summary** As the solar wind encounters Earth's magnetosphere and diverts around it, a shock is formed that heats and compresses the plasma and warps the magnetic field frozen into it. This shocked plasma and magnetic field, known as the magnetosheath, is what drives energy transfer at the magnetopause. Due to orbital constraints there is no continuous in-situ monitor of magnetosheath conditions. Studies of solar wind magnetosphere interaction typically rely on solar wind conditions measured at L1 propagated to Earth by some algorithm, which are then either used directly or used to drive some model of the magnetosheath. This process has numerous points of uncertainty, from the choice of propagation algorithm to the choice of magnetosheath model (or lack thereof). To address these concerns with the traditional approach, this study develops a data-driven model of the magnetosheath that uses data from L1 as its input. This new model, called Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath, adapts a recurrent neural network architecture that is capable of estimating uncertainties for its predictions. This new model is verified to be accurate in a statistical sense, and is also capable of representing physics that is not explicitly incorporated in the model during training.

## 1. Introduction

The region of turbulent, shocked solar wind plasma downstream of Earth's bow shock is known as the magnetosheath. The magnetosheath plasma and magnetic field transfer energy to Earth's magnetosphere via magnetic reconnection and viscous interaction (Axford, 1964; Dungey, 1961). Despite this, the solar wind conditions upstream of the bow shock are frequently taken as the input to the system in studies of solar wind-magnetosphere interaction (e.g., Barker et al., 2005; Borovsky, 2008; Fleetham et al., 2023; X. Li, 2004; Newell et al., 2007). This is largely because of the absence of any continuous in-situ magnetosheath monitor due to orbital constraints. Continuous records of the magnetosheath conditions therefore require modeling the magnetosheath by some method.

Early models of the magnetosheath used gas dynamics as their basis, incorporating some physical assumptions and including limited consideration of the magnetic field outside the magnetopause (Spreiter & Alksne, 1969; Spreiter et al., 1966). These models have matured through the inclusion of additional physics into modern MHD

codes (e.g., Lyon et al., 2004; Powell et al., 1999), that offer spatially and temporally complete model magnetosheaths at the cost of some physical assumptions and increased computational expense. In situations where the computational expense of MHD modeling is prohibitive, some magnetosheath modeling efforts fit analytical expressions derived from physical assumptions to spacecraft measurements of the magnetosheath (Kobel & Flückiger, 1994; Soucek & Escoubet, 2012; Tsyganenko et al., 2023). Others, such as the recent Mshpy23 model (Jung et al., 2024), parameterize the outputs of MHD models to reduce their computational cost but retain some of their accuracy. A shared feature of these approaches is that they all include physical assumptions (discussed in more detail below). While they may often be valid, there remains differences between their outputs and the actual magnetosheath that can limit their representational power. This issue could be addressed by reducing the number of assumptions used to construct the model; for example, hybrid-Vlasov codes capable of simulating the entire magnetosheath have recently come online (Hoilijoki et al., 2016; Von Alfthan et al., 2014) but come with an even higher computational cost than MHD codes.

One possible way of addressing this limitation is the use of neural network codes that do not assume a functional form or simplified physics. Neural networks have been used to assemble models of geophysical quantities for the past few decades since the early relativistic electron flux model of Koons and Gorney (1991), and have continued to be regularly utilized for space physics tasks. The applications of neural networks to the magnetosheath are typically classification tasks, including identifying measurements of the magnetosheath from ion or electron distribution functions (Gilet et al., 2021; Olshevsky et al., 2021), identifying planetary bow shock and magnetopause crossings (Cheng et al., 2022), classifying magnetosheath jets (Raptis et al., 2020), and identifying scientifically interesting data to be downlinked by spacecraft operators (Argall et al., 2020). However, neural networks are also suited to prediction and spatio-temporal inversion tasks in the magnetosheath as they do not require physical assumptions to construct tractable or analytical descriptions of the magnetosheath plasma and magnetic field, and are also computationally inexpensive. In particular, recurrent neural networks designed to produce probabilistic predictions have shown good performance in spatio-temporal inversion tasks such as electron density in the inner magnetosphere (Huang et al., 2022).

A crucial aspect of any prediction algorithm that is typically lacking in magnetospheric physics is uncertainty quantification (Borovsky, 2021). There is growing evidence that uncertainty in solar wind data affects correlation studies of the cross polar cap potential (Sivadas et al., 2022), development of solar wind-magnetosphere coupling functions (Lockwood et al., 2019), and global MHD simulation outputs (Al Shidi et al., 2023); the solar wind data uncertainty and the magnetosheath model uncertainty compound. Since it is the shocked solar wind at the magnetopause rather than the solar wind upstream of the bow shock that interacts with the magnetosphere, this uncertainty has the potential to affect any study that tries to associate solar wind conditions with magnetospheric response in a way that is difficult to account for without a magnetosheath model that estimates this uncertainty.

Another challenge with traditional models aside from their physical assumptions is the fact that they typically use solar wind data that has been propagated from in-situ monitors far from Earth as input. Much like the magnetosheath, there is no continuous in-situ monitor of the solar wind near Earth due to orbital constraints. In order to obtain inputs for each of the previously mentioned models, data from monitors at the L1 position $235R_E$ (1,500,000 km) ahead of Earth need to be propagated to Earth to account for the travel time of the solar wind plasma and interplanetary magnetic field (generally 30–60 min). This propagation task is made difficult by the structure and dynamics of the solar wind (Borovsky, 2018), and a variety of algorithms have been developed in order to propagate measurements between L1 and Earth accurately. One such algorithm, the Probabilistic Regressor for Input to the Magnetosphere Estimation (PRIME) (O'Brien et al., 2023) was recently developed to address some of these difficulties with traditional propagation algorithms, and its architecture is well suited to be adapted to the problem of magnetosheath prediction from L1 inputs (since the physics of solar wind propagation is the first "step" of that task).

Motivated by the limitations of traditional algorithms outlined above, a new algorithm capable of predicting magnetosheath plasma and magnetic field conditions given measurements made by an in-situ monitor at L1 is developed. This algorithm, named PRIME-SH after its progenitor algorithm PRIME (O'Brien et al., 2023) with the "-SH" suffix to denote "magnetosheath," requires a data set of in situ magnetosheath measurements and associated solar wind inputs at L1 (Section 2), and a network architecture adapted from PRIME and optimized for the magnetosheath (Section 3). Outputs from PRIME-SH are validated statistically on a holdout data set. PRIME-



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SH is subjected to additional validation verifying that it reproduces some expected physics (Section 4), after which the results can be summarized and discussed (Section 5).

## 2. Data

### 2.1. MMS Target Data Set

Plasma and magnetic field data from the Magnetospheric Multiscale 1 (MMS-1) spacecraft's (Burch et al., 2016) Fast Plasma Investigation (FPI) (Pollock et al., 2016) and Fluxgate Magnetometer (FGM) (Russell et al., 2016) instruments are utilized as targets for the algorithm to be optimized against. MMS is a constellation of four spacecraft designed to study magnetic reconnection at Earth's magnetopause and neutral sheet. It therefore spends considerable time in Earth's magnetosheath and carries instruments particularly designed to measure the plasma and magnetic field there, making data it collects highly suitable for use as targets to optimize PRIME-SH. The large volume of data produced by MMS-1's instruments have motivated the development of automated classification, identification, and segmentation tools for MMS data that allow rapid selection of large amounts of data with particular features or from particular plasma regimes. The coordinate systems used in this study are geocentric solar ecliptic (GSE) and geocentric solar magnetic (GSM) coordinates. Both systems are centered on the Earth with their $x$-axes oriented toward the Sun. GSE coordinates have a $z$-axis oriented normal to the ecliptic plane, while GSM coordinates have a $z$-axis oriented parallel to Earth's dipole moment.

To assemble a solar wind data set using MMS, an automatic tool developed by Olshevsky et al. (2021) is used to classify all MMS-1 FPI 3D ion distributions from 2 September 2015 to 1 January 2023. The classifier is capable of discriminating between magnetospheric, magnetosheath, non-foreshock solar wind, and foreshock plasma using the shape of the ion distribution function, and outputs a normalized probability that a given distribution belongs to each class. Periods of time where MMS-1 is in the magnetosheath with probability $p > 0.7$ are found using the classifier; all other time periods are removed thereby removing the magnetosphere, solar wind, foreshock, and ambiguous classifications from the data set. Remaining FGM magnetic field and FPI ion moments are averaged in 100 s bins to match the input data from the Wind spacecraft (See Section 2.2). Since the classifier is trained only on data from dayside orbits, any data on the nightside (GSE/GSM $X < 0$) are removed. It is also worth noting that the Olshevsky et al. (2021) classifier was trained on 2 months of MMS data during solar minimum. Such small training sets could favor identifying the magnetosheath preferentially for certain upstream solar wind as some conditions (e.g., fast solar wind) may not be present or may be underrepresented in such a short interval.

This procedure is applied to all MMS-1 data from 2 September 2015 to 1 January 2023. 117,427 100 s-averaged data points consisting of $\vec{B}$ (GSM coordinates), $\vec{V}$ (GSE coordinates), $n_i$, $T_{i\|}$, and $T_{i\perp}$ result. The full spatial distribution of the magnetosheath data are shown in Figure 1. Only 100 s periods when MMS-1 is in the magnetosheath with probability $p > 0.7$ for all FPI moments (roughly 22 at FPI's 4.5 s cadence) are included. Visible in Figure 1 are the different orbits corresponding to different MMS mission phases. This is worth noting because the inner orbit (apogee $12R_E$) corresponding to mission phase 1 (September 2015 to January 2017) is potentially biased toward magnetosheath measurements during high dynamic pressure solar wind. Since there is more data after MMS's apogee is raised (May 2017 to January 2023), the biases in the full data set were verified to be small.

### 2.2. Input Data Set

The input solar wind data at L1 comes from the Magnetic Field Investigation (MFI) (Lepping et al., 1995) and Solar Wind Experiment (SWE) (Ogilvie et al., 1995) aboard the Wind spacecraft. Wind was selected for this study because it had the best coverage over the time period of the MMS-1 data set used here (2 September 2015 to 1 January 2023). "Key parameter" (moments) data are utilized, resulting in time series of plasma flow velocity $\vec{v}$ (GSE coordinates), ion density $n_{ion}$, ion thermal speed $v_{\perp th}$, and IMF $\vec{B}$ (GSM coordinates) at a 100 s cadence. Due to the difficulty involved with spacecraft intercalibration, data from other L1 monitors are not included in this study (King & Papitashvili, 2005). To give PRIME-SH information about the spatial separation of the input and target spacecraft and the location in the sheath at which the prediction is being made, the positions of Wind and MMS-1 in GSE coordinates are included in the input data.

There are data gaps ranging from a few minutes to a little over a day in the wind data set. Data gaps are linearly interpolated over and flagged so that the amount of interpolated data in any given timeseries can be tracked. For



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

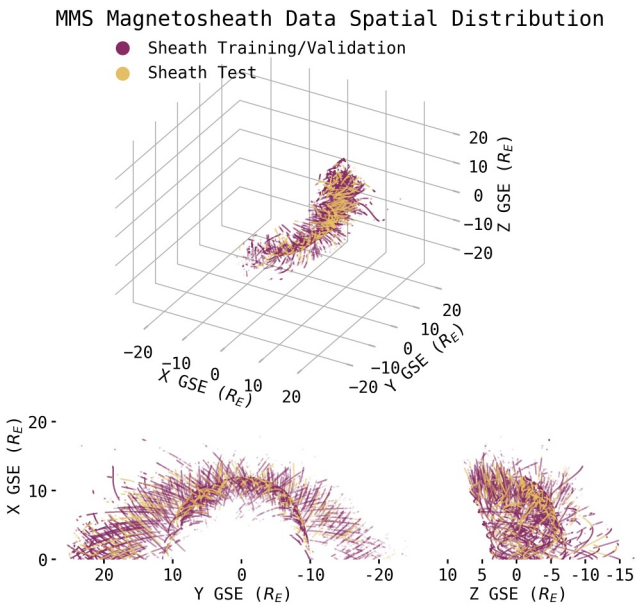


**Figure 1.** 3D spatial distribution of the 117,427 magnetosheath Magnetospheric Multiscale 1 100 s-averaged data points split into 80% training/validation (purple) and 20% test (yellow) subsets. Data consists of $\vec{B}_{GSM}$, $\vec{V}_{GSE}$, $n_i$, $T_{i\parallel}$, and $T_{i\perp}$ from 2 September 2015 to 1 January 2023. Train/validation/test split is as used in the optimized data set (see Section 3.2).

each target MMS-1 observation, a subset of the input timeseries with a defined length and lead time before the observation is used as input. The length of the input timeseries, its lead time, and the permissible amount of data in the timeseries that can be interpolated all affect the performance of the final algorithm. Finding the optimal values for these three quantities is therefore best thought of in the context of the model hyperparameter optimization and is discussed in detail in Section 3.2. To investigate whether the possible biases in the Olshevsky et al. (2021) classifier affected the input data set by preferentially identifying the magnetosheath for certain upstream conditions, the distribution of input solar wind conditions in the final Wind data set were compared to all solar wind data from Wind during the time period of this study. While some small biases toward higher solar wind density were observed, the biases were not significant enough to indicate a systematic bias in the classification algorithm (especially since no bias was observed in solar wind velocity, the primary indicator of fast or slow solar wind).

## 3. Algorithm Methodology

### 3.1. Network Architecture

The overall architecture selected for the algorithm is similar to that utilized to construct PRIME, an algorithm that predicts the solar wind near Earth using data from the Wind spacecraft at L1 (O'Brien et al., 2023). The neural network architecture developed for PRIME is well suited to be adapted to the task of magnetosheath prediction for several reasons. First, it is capable of incorporating information about the time history of solar wind at L1 into its predictions which is important for predicting the solar wind and the evolution of the magnetosheath. Second, it is capable of assigning uncertainties to its predictions which is crucial in the frequently turbulent environment in the magnetosheath. Third, it has proven to be accurate when applied to the task of solar wind propagation, which is essentially the first step of the task undertaken by PRIME-SH.

The overall form of PRIME-SH is shown in Figure 2. Like PRIME, PRIME-SH utilizes a Gated Recurrent Unit (GRU) sequence (see Cho et al., 2014) that is evaluated sequentially over the input timeseries $x$ (length 55, 14 features). GRUs are a class of recurrent neural network (RNN) that show good performance relative to other recurrent network architectures (Chung et al., 2014), and are chosen here as they are suited for timeseries prediction tasks. The GRU's history vector $h$ is then fed into three fully connected neural network layers (see Bebis & Georgiopoulos, 1994). These layers reduce the dimensionality of the GRU's history vector. The hidden layers feed into an output layer of neurons ($y$, length 18) which are taken to be the mean and variance of a Gaussian probability distribution for each parameter rather than single scalar values (Lakshminarayanan et al., 2017; Nix & Weigend, 1994). The input feature size is 14, and the output feature size is 9. The algorithm is implemented in the Keras high-level API for tensorflow (https://keras.io/api/). Details of the architecture such as the length of the input time series and the size of each layer do not have optimal values that can be determined a priori. Instead, they are chosen via hyperparameter tuning (See Section 3.2).

The loss criterion used to optimize the algorithm during training is chosen to be the continuous rank probability score (CRPS) (Hersbach, 2000; Matheson & Winkler, 1976). The CRPS is a common scoring metric used to compare probabilistic forecasts for weather prediction (Zamo & Naveau, 2018). For a detailed description of the CRPS see Section 2 of Camporeale and Carè (2021) or Section 3.1 of O'Brien et al. (2023). Briefly, the CRPS is given by

$$CRPS = \int_{-\infty}^{\infty} \left[F(y) - H(y - y_{obs})\right]^2 dy \tag{1}$$

where $F(y)$ is the cumulative distribution function of a probabilistic prediction for some observation $y_{obs}$ and $H(y)$ is the Heaviside step function (Wilks, 2011). The CRPS is desirable as a loss function because it more



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

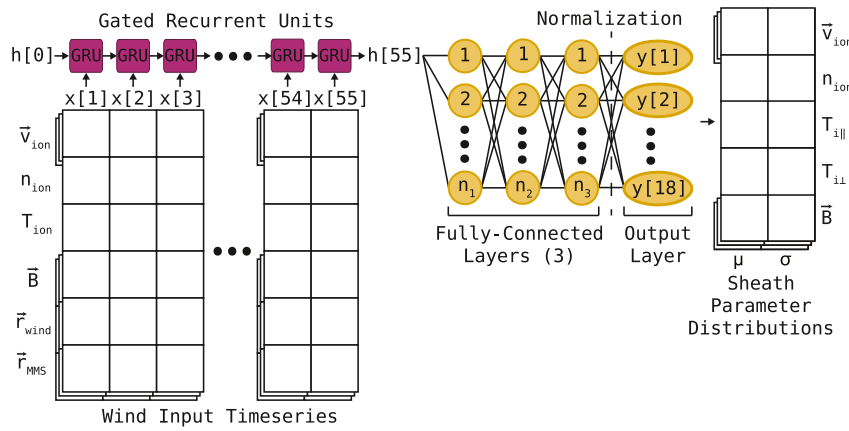


**Figure 2.** Schematic of Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath's (PRIME-SH's) neural network architecture, based on the architecture of PRIME (O'Brien et al., 2023). Note that the Gated Recurrent Unit (GRU) sequence feeds into a fully connected neural network in order to output a mean and variance for each desired parameter instead of a single value (See Section 3.1 for explanations of these architectures). x[i] are the *i*th elements of input timeseries x, h[i] are the *i*th states of the GRU history vector h, and the output layer is y. Vector quantities such as magnetic field, flow velocity, and spacecraft position are stacked to show that they constitute three units in the input/output but describe one physical vector quantity. Exact layer size and additional regularization features (see Table 1) chosen via hyperparameter search.

symmetrically punishes over and under confident predictions than the negative log probability density (the most commonly used score for probabilistic predictions) (Camporeale & Carè, 2021). A side benefit is that the CRPS has the same unit as the variable of interest, making it more intuitively human-readable. In the case of Gaussian predictions with mean $\mu$ and variance $\sigma^2$ the CRPS is given by

$$CRPS(y_{obs},\mu,\sigma) = \sigma\left[\frac{y_{obs}-\mu}{\sigma} erf\left(\frac{y_{obs}-\mu}{\sqrt{2}\,\sigma}\right) + \sqrt{\frac{2}{\pi}}\, e^{-\frac{(y_{obs}-\mu)^2}{2\sigma^2}} - \frac{1}{\sqrt{\pi}}\right] \tag{2}$$

(Gneiting et al., 2005). Since PRIME-SH outputs Gaussian probability distributions, and since CRPS is negatively oriented, Equation 2 is used as a loss function during training. The 18 output units in PRIME-SH's last layer are taken to be the means ($\mu$s) and variances ($\sigma$s) defining a Gaussian probability distribution for each parameter. During training the CRPS for each target parameter is assigned equal weight when assessing PRIME-SH's performance.

The primary limitation of the CRPS as a loss function training probabilistic algorithms is the fact that it does not explicitly enforce reliability of the algorithm's predicted uncertainties (Camporeale et al., 2019). Reliability is measure of the degree to which a probabilistic forecast's uncertainties are statistically consistent with the observed probabilities of the events the forecast seeks to predict (Anderson, 1996). It has been shown that accuracy and reliability are competing metrics that must be balanced, and that simply minimizing the CRPS does not necessarily mean that the resulting model is reliable (Camporeale & Carè, 2021). Since reliability is not explicitly enforced, the reliability of PRIME-SH's uncertainties must be verified after training (See Section 4.1) (Tasistro-Hart et al., 2021).

### 3.2. Algorithm Optimization

Optimization of PRIME-SH follows a three step process. First, the optimal length, lead time, and composition of the input timeseries data set is determined (the data set hyperparameter search). Then the algorithm hyperparameters are systematically varied using the Hyperband tournament-style algorithm (L. Li et al., 2018) in order to find the optimal algorithm, then finally the optimal algorithm is instantiated and trained. This algorithm then becomes the canonical version of PRIME-SH.



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

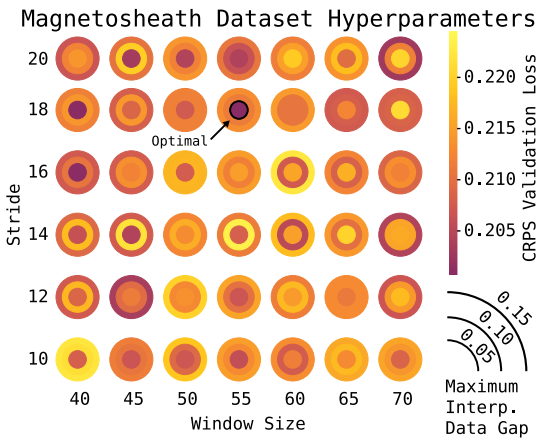


**Figure 3.** Results from data set optimization trials over timeseries window (length), stride (lead time), and permitted fraction of interpolated data. Units for window and stride are 100 s (the Wind KP data cadence). The optimal set (window 55, stride 18, largest interp. fraction ≤5%) is shown in darkest purple and labeled "optimal." Loss is given in dimensionless units of parameter interquartile range to ensure comparability of continuous rank probability score for each parameter.

Given a particular time when a prediction of the magnetosheath conditions is desired, it is difficult to say a priori what time period of Wind data from L1 contains the necessary information to make that prediction (especially given the flexible nature of neural network algorithms). Since the solar wind typically takes 30–60 min to get from L1 to Earth, there is likely only so much time history that can be incorporated before including more yields diminishing returns in terms of accuracy. Similarly, it is likely that including conditions at L1 right up until the time the sheath prediction is desired is not necessary, since the solar wind at that time has not had sufficient time to get to Earth. To find the optimal start and stop times of the timeseries used to make each prediction, a range of start and stop times are tested by optimizing a test version of PRIME-SH using different input time series lengths (windows) and lead times before each prediction (strides). It is also likely that large data gaps that are filled with interpolated data can affect the algorithm's performance, therefore a range of permissible data gap sizes are also tested (expressed in terms of fractions of the window size). Whichever parameters produce a model that can achieve the best results on the validation data set before overfitting are taken as optimal. When training these test models and for any time a model is trained, the input/target data sets are split into 60% training, 20% validation, and 20% test subsets. Since temporally adjacent entries in the input data set are almost entirely overlapping, randomly assigning input/target pairs to each subset would result in significant data leakage. To avoid this, the full data set is split into independent blocks containing an amount of MMS data four times the length of the timeseries window used as input (i.e., for a window size of 55 measurements/∼1 hr 32 min, the data set is split into chunks of length 220 measurements/∼6 hr 8 min) and those blocks are then assigned to each subset in order to achieve a 60%–20%–20% train-validation-test split. The 6 hours of MMS data in each block need not be continuous, as 6 hours is the minimum block size that ensures the input data in the training, validation, and test data sets do not overlap. To ensure that no parameter dominates others due to their absolute relative values, each subset is rescaled to the interquartile range of the training set in order to account for outliers without leaking information about the validation/test sets during training. Results on the validation data set from the search are shown in Figure 3.

Whichever set from Figure 3 has the lowest CRPS is taken to be optimal. The optimal window size is 55 measurements (∼5,500 s, ∼1 hr 32 min), the optimal stride/lead time is 18 measurements (∼1,800 s, ∼30 min). That is to say, for an MMS measurement at time $t$, the input timeseries from Wind runs from time $t - 5{,}500\ \mathrm{s} - 1{,}800\ \mathrm{s} \approx t - 122$ min to time $t - 1{,}800\ \mathrm{s} \approx t - 30$ min. The largest data gap that can be interpolated over is 4.6 min (≤5% of the input window).

Once the optimal data set structure is found, the optimal model configuration can be determined via hyperparameter search. The nine hyperparameters that are optimized are listed in Table 1, along with the values used for determining the optimal data set, the optimal values used for the canonical version of PRIME-SH, and the search range for each hyperparameter. The hyperparameter search is conducted using the Hyperband tournament bracket style algorithm (L. Li et al., 2018) implemented in the KerasTuner API (O'Malley et al., 2019). The meaning of each hyperparameter is described in the following paragraph. After the optimal model configuration is determined, the canonical version of PRIME-SH is optimized on the training data set until its loss on the validation dataset does not improve for five successive epochs (20 epochs total in this case).

The nine hyperparameters are as follows (see also Table 1). The first four are the node sizes of the GRU layer and the following three fully-connected layers. The fifth is where in the algorithm sequence to perform a layer normalization step, which stabilizes neural networks during optimization to reduce the time it takes to optimize them (Ba et al., 2016). Layer normalization normalizes a given layer's output vector before passing it to the next layer, which speeds up the convergence of the algorithm used to optimize the weights and biases of the algorithm by reducing the extent to which the gradients with respect to the weights in one layer covary with the outputs of the previous layer. The sixth and seventh hyperparameters are the dropout locations and rate used during training. Dropout is a technique to mitigate overfitting that involves randomly removing some percentage of the units from the network every training epoch. This prevents units from co-adapting which can lead to overfitting (Srivastava



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**Table 1**
*Detailed Layer Sizes and Architecture Parameters for the Test Version of Probabilistic Regressor for Input to the Magnetosphere Estimation-Magnetosheath (PRIME-SH) Used to Optimize the Data Set Parameters (Left Column), the Canonical Version of PRIME-SH Determined by Hyperparameter Search (Middle Column), and the Range of Each Parameter for Which the Hyperparameter Search was Conducted (Right Column)*

| | Data set HP test | Canonical algorithm | HP range |
|---|---|---|---|
| GRU Layer | 192 | 416 | 128–640 |
| FCNN Layer 1 | 352 | 352 | 128–640 |
| FCNN Layer 2 | 48 | 32 | 16–128 |
| FCNN Layer 3 | N/A | 64 | 16–128 |
| Normalization | Last Layer | Last Layer | Any Combination |
| Dropout Location | Last Layer | Last Layer | Any Combination |
| Dropout Rate | 20% | 35% | 20%–50% |
| Optimizer | Adamax | Adam | Adam, Adamax, Adagrad |
| Learning Rate | $10^{-4}$ | $10^{-4}$ | $10^{-3}$, $10^{-4}$, $10^{-5}$ |

et al., 2014). The eighth and ninth hyperparameters are the optimization algorithm used to update the weights and biases in the network and that algorithm's learning rate. Included in the search are the adaptive gradient descent algorithms Adam, Adamax, and Adagrad. An adaptive gradient descent algorithm changes the step size it uses to update parameter weights during optimization to avoid getting stuck in local minima or skipping over minima. Adam updates parameters according to estimates of first order and second moments and has been shown to be suitable for optimizing large algorithms (Kingma & Ba, 2017), Adamax updates parameters according to first order moments and the infinity norm and has been shown to be suitable for recurrent networks (Kingma & Ba, 2017), and Adagrad updates its gradient descent step size per parameter based on the number of updates the parameter receives during training making it suitable for sparse gradients (Duchi et al., 2011). Since each of these conditions could apply to PRIME-SH and the data set used to optimize it, these three algorithms were included.

## 4. Results

### 4.1. Statistical Performance

PRIME-SH's performance is evaluated on the test data set (not seen by the algorithm at any point during training) by calculating the CRPS between its predictions and the test data set. Additionally, the mean absolute error (MAE) and Pearson's $r$ correlation coefficient are calculated between the means of PRIME-SH's predicted probability distributions and the MMS test set thereby ignoring the uncertainty information. To gain a better sense of the accuracy of PRIME-SH's predictions in a statistical sense, its outputs are compared to several analytical models and a parameterization of a popular MHD code for the same MMS-1 test data set (Figures 4–6, Tables 2 and 3). For magnetic field, the model derived in Cooling et al. (2001) is utilized. The Cooling et al. (2001) model essentially "drapes" the interplanetary magnetic field over the Shue et al. (1998) axisymmetric conic section magnetosheath model (based on Kobel & Flückiger, 1994). For magnetosheath flow, the model derived in Soucek and Escoubet (2012) is utilized. The Soucek and Escoubet (2012) model is partially based on Génot et al. (2011) and Kobel and Flückiger (1994), but extends those works to additional magnetopause and bow shock shapes. For density and temperature, the model derived in Spreiter et al. (1966) is utilized. The Spreiter et al. (1966) model is a gas dynamic model that assumes a nondissipative, ideal, compressible, steady flow. Additionally, PRIME-SH is compared to a parameterization of the OpenGGCM MHD code (Raeder et al., 2001, 2008) developed in Jung et al. (2024). This parameterization cannot capture small-scale structure in the MHD code's outputs, but has been shown to be accurate when compared to observations and is importantly computationally inexpensive enough to enable the statistical comparison in this study. The Soucek and Escoubet (2012) and Spreiter et al. (1966) models are implemented in the Mshpy23 package (Jung et al., 2024) and accept 1 min resolution OMNI data as input (King & Papitashvili, 2020). The Spreiter et al. (1966) and OpenGGCM models produce isotropic temperatures, therefore their temperatures are compared to the average temperature measured by MMS $T_{iAV} = (2T_{i\perp} + T_{i\parallel})/3$. None of the models PRIME-SH is compared to have uncertainty information, therefore the MAE and CRPS



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

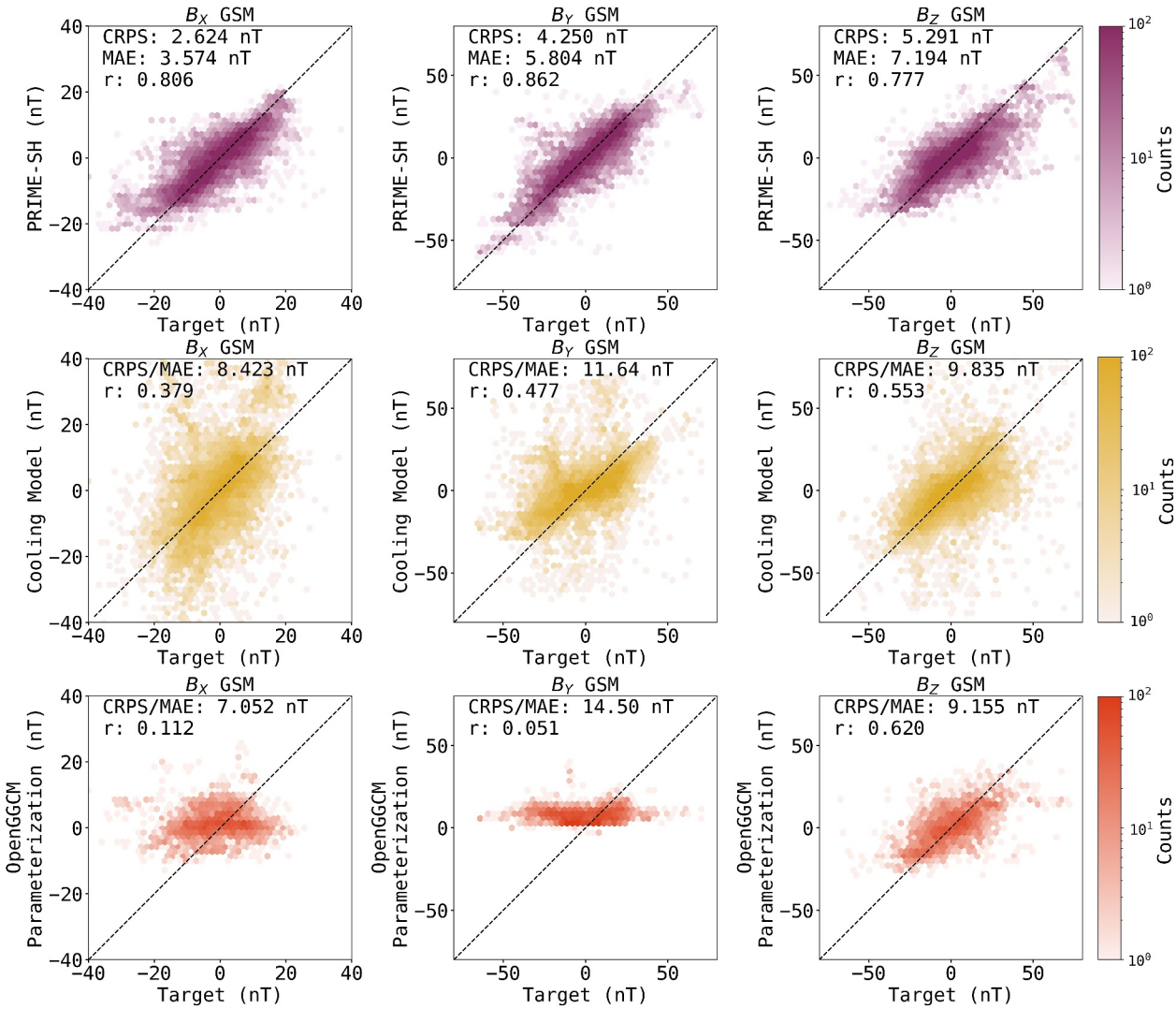


**Figure 4.** Joint distributions of Magnetospheric Multiscale 1 magnetic field data ($x$ axis) with predicted parameters from Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) (purple, top), the Cooling et al. (2001) (yellow, middle), and a parameterization of the OpenGGCM MHD code (orange, bottom). Continuous rank probability score, the mean absolute error (MAE), and Pearson's r correlation coefficient for each parameter shown in the top left of each distribution. The MAE is calculated between the peaks of PRIME-SH's predicted distributions and each MMS observation (thereby throwing away uncertainty information). A perfect prediction corresponds to the line $y = x$, plotted overtop of each distribution for convenience.

reduce to the same form and number (Hersbach, 2000); both metrics are provided for PRIME-SH's outputs so that all comparisons can be made.

On average, PRIME-SH predicts plasma parameters ($\vec{v}$, $n_i$, $T_{i\perp}$, and $T_{i\|}$) slightly more accurately than magnetic field parameters. This is possibly due to the fact that fluctuations in magnetic field happen more quickly than those in the plasma, and neural networks tend to have more difficulty representing smaller scale variations than larger scale ones whether temporal or spatial in nature. PRIME-SH has a Pearson's $r$ higher than 0.75 for every parameter. There are no strong biases or systematic errors visible in Figures 4–6, only some small degree of underprediction in the most extreme values of $V_X$ and $n_i$ (and therefore in $P_{dyn}$ as well). The highest values of $V_X$ occur in the flanks, which indicates that the inaccuracies are due to either the relative rarity of measurements toward the flanks with high speed or due to ambiguous classifications between magnetosheath and solar wind. It also indicates that the high speed observations are not magnetosheath jets, which tend to occur closer to the nose (Raptis et al., 2020). The largest values of $n_i$ occur close to the subsolar point. Density enhancements close to the nose could be due to solar wind transient phenomena such as pressure pulses or coronal mass ejections, or due to



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

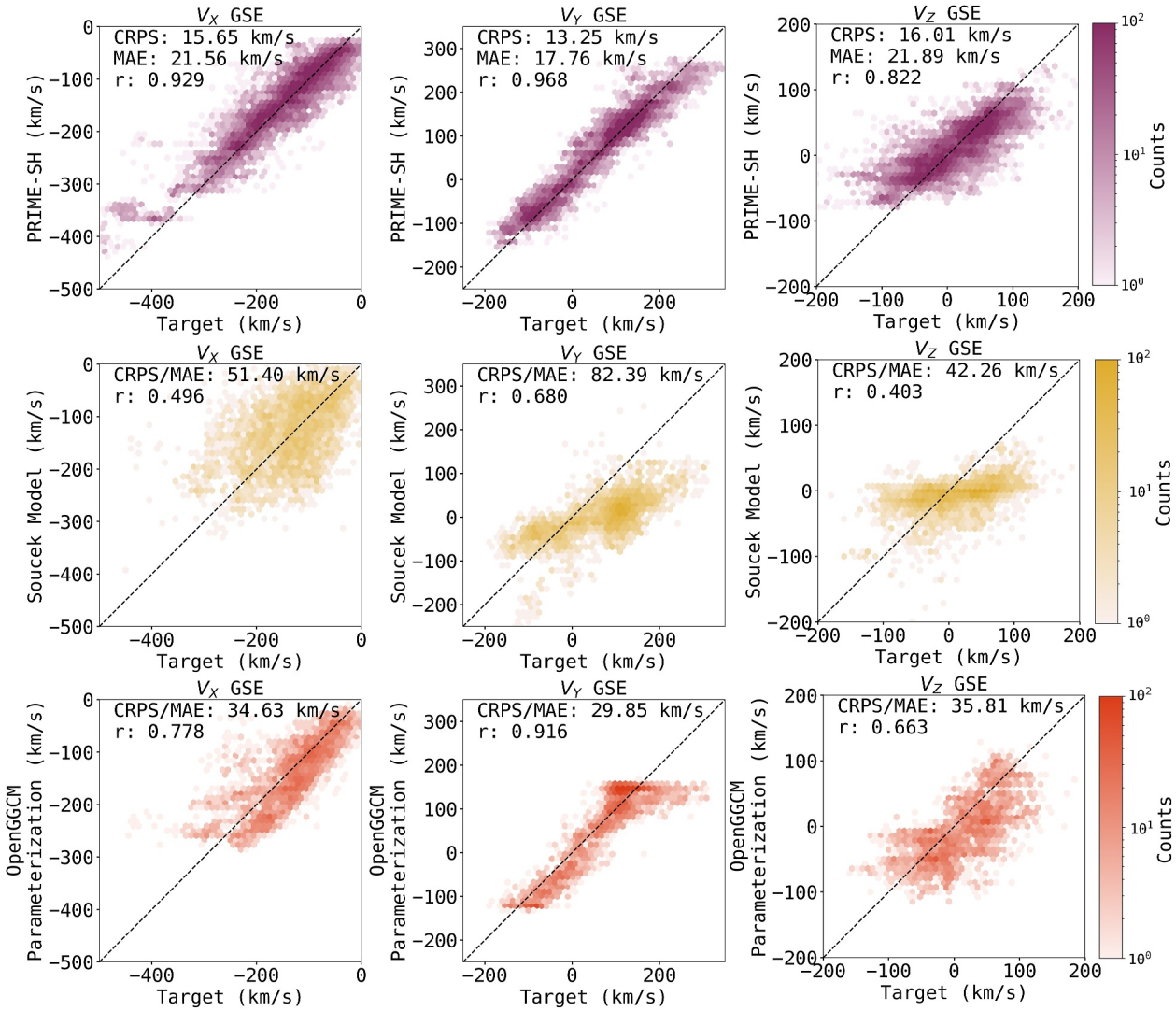


**Figure 5.** Joint distributions of Magnetospheric Multiscale 1 ion velocity data ($x$ axis) with predicted parameters from Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) (purple, top), the Soucek and Escoubet (2012) model (yellow, middle), and a parameterization of the OpenGGCM MHD code (orange, bottom). Continuous rank probability score, the mean absolute error (MAE), and Pearson's r correlation coefficient for each parameter shown in the top left of each distribution. The MAE is calculated between the peaks of PRIME-SH's predicted distributions and each MMS observation (thereby throwing away uncertainty information). A perfect prediction corresponds to the line $y = x$, plotted overtop of each distribution for convenience.

mesoscale foreshock transients, all of which are relatively rare during the time period of the training data set. Interestingly, PRIME-SH predicts magnetosheath conditions almost as accurately as its progenitor algorithm PRIME predicts solar wind conditions given the same type of input data from L1 (PRIME-SH's average CRPS of $0.221\sigma$ and PRIME's average CRPS of $0.214\sigma$), despite that it has to represent not only the physics of the solar wind's propagation from L1 to Earth but the physics of the bow shock as well. This could be due to the fact that MMS's FPI instrument was designed to study the magnetosheath rather than the solar wind (Roberts et al., 2021).

For all parameters PRIME-SH outperforms all of the analytical models considered here with respect to MAE and CRPS. For each component in the magnetic field, PRIME-SH predicts MMS-1 observations more accurately than the Cooling et al. (2001) model. Specifically, PRIME-SH's CRPS and MAE are both lower than the Cooling et al. (2001) model's MAE, and PRIME-SH's Pearson's $r$ is higher than the Cooling et al. (2001) model's Pearson's $r$. There appears to be some systematic overprediction in the Cooling et al. (2001) model's outputs for $B_X$. This means that PRIME-SH reproduces the actual magnetic field in the magnetosheath given upstream conditions more accurately than the Cooling et al. (2001) model, but whether it produces a physically accurate draped field



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

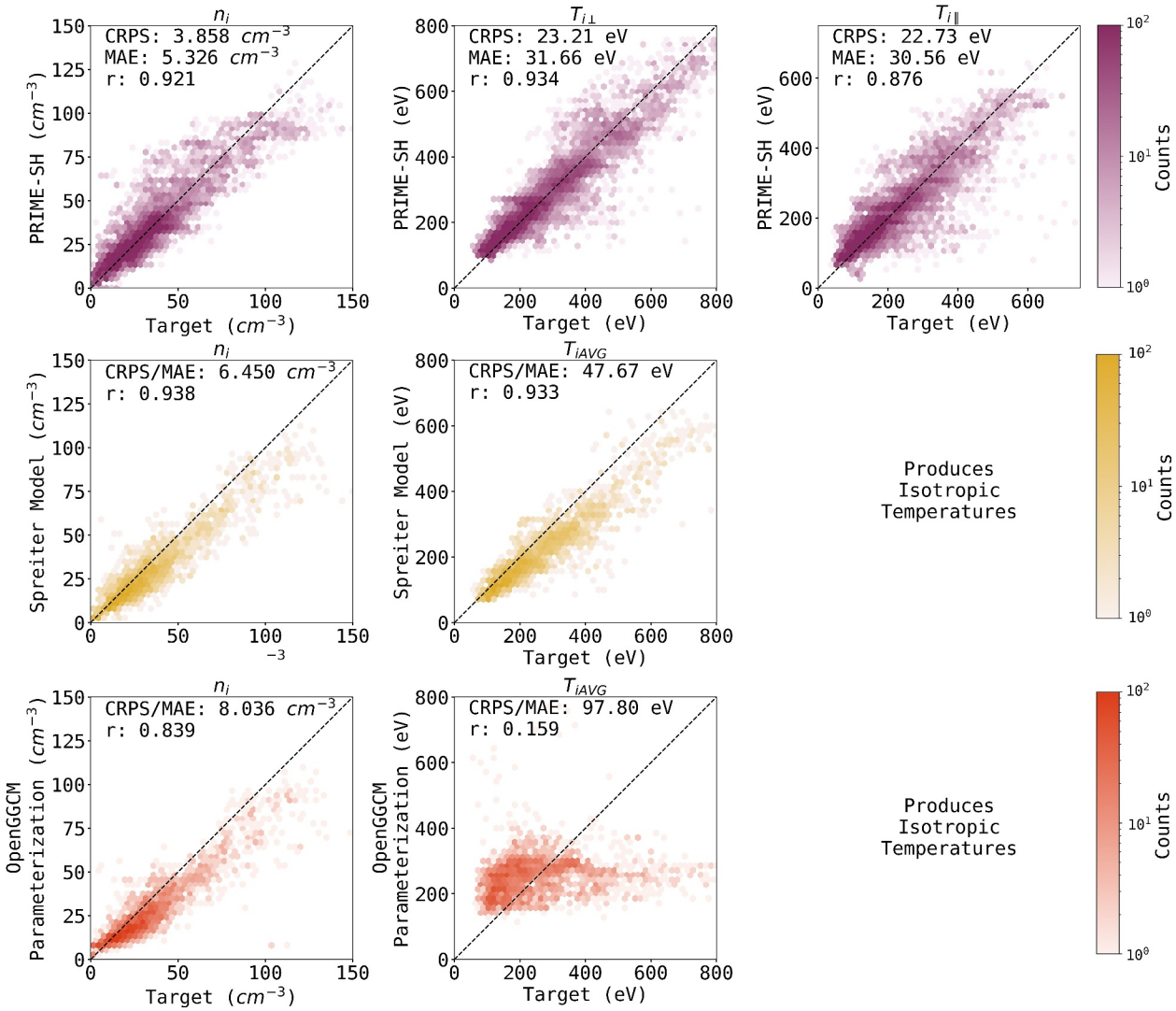


**Figure 6.** Joint distributions of Magnetospheric Multiscale 1 ion density and temperature data ($x$ axis) with predicted parameters from Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) (purple, top), the Spreiter et al. (1966) model (yellow, middle), and a parameterization of the OpenGGCM MHD code (orange, bottom). Continuous rank probability score, the mean absolute error (MAE), and Pearson's r correlation coefficient for each parameter shown in the top left of each distribution. The MAE is calculated between the peaks of PRIME-SH's predicted distributions and each MMS observation (thereby throwing away uncertainty information). A perfect prediction corresponds to the line $y = x$, plotted overtop of each distribution for convenience.

must be separately validated in Section 4.2.1. The Soucek and Escoubet (2012) model has a large variance in $V_X$ and does not reproduce fast flows (>300 km/s) as accurately as PRIME-SH does. It also underpredicts $V_Y$ and $V_Z$, all of which indicate the model's outputs could be too "smooth." The Spreiter et al. (1966) comes the closest to outperforming PRIME-SH of any model considered here, but still does not predict $n_i$ or $T_{iAVG}$ more accurately than PRIME-SH.

It is worth noting that PRIME-SH's training data set includes observations of the magnetosheath for quasiparallel and quasiperpendicular shock conditions (Eastwood et al., 2005). The plasma and magnetic field conditions are often very different downstream of these shock conditions, in particular during quasiparallel shock conditions when high speed magnetosheath jets (Plaschke et al., 2018) and foreshock transients (Zhang et al., 2022) are observed. These effects can significantly effect magnetosheath properties (e.g., Karlsson et al., 2021), and can last longer than the 100 s averaging applied to MMS-1 data in PRIME-SH's target data set. Evaluating quasi-perpendicular and quasi-parallel shock geometries using temperature anisotropy (from the test data set) as an indication of each geometry showed little variation in the accuracy of PRIME-SH. While this indicates that both



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**Table 2**
*Performance of Probabilistic Regressor for Input to the Magnetosphere Estimation-Magnetosheath on the Magnetospheric MultiScale (MMS) Test Data Set Across Continuous Rank Probability Score (CRPS, Equation 1), Mean Absolute Error, and Pearson's r Correlation Coefficient (Also Shown in Figures 4–6)*

| Parameter | CRPS | MAE | $r$ |
|---|---|---|---|
| $B_{XGSM}$ | 2.65 nT (0.296$\sigma$) | 3.61 nT (0.403$\sigma$) | 0.800 |
| $B_{YGSM}$ | 4.18 nT (0.245$\sigma$) | 5.65 nT (0.331$\sigma$) | 0.864 |
| $B_{ZGSM}$ | 5.19 nT (0.323$\sigma$) | 7.08 nT (0.440$\sigma$) | 0.779 |
| $V_{XGSE}$ | 14.03 km/s (0.182$\sigma$) | 19.25 km/s (0.250$\sigma$) | 0.945 |
| $V_{YGSE}$ | 13.22 km/s (0.127$\sigma$) | 17.95 km/s (0.173$\sigma$) | 0.969 |
| $V_{ZGSE}$ | 15.35 km/s (0.291$\sigma$) | 21.04 km/s (0.399$\sigma$) | 0.838 |
| $n_i$ | 3.63 cm$^{-3}$ (0.169$\sigma$) | 4.96 cm$^{-3}$ (0.231$\sigma$) | 0.929 |
| $T_{i\perp}$ | 23.76 eV (0.158$\sigma$) | 32.58 eV (0.216$\sigma$) | 0.936 |
| $T_{i\parallel}$ | 22.67 eV (0.198$\sigma$) | 30.70 eV (0.268$\sigma$) | 0.881 |

*Note.* CRPS is given in the units of each parameter as well as dimensionless units of standard deviations of each parameter in the MMS training data set to facilitate comparison between each parameter.

geometries are accurately modeled by PRIME-SH, further work is needed to fully evaluate the effect of kinetic processes at the foreshock. These could play a role in shaping small scale and mesoscale variability of particle moments and magnetic fields that could have an effect on the model performance. It is also worth noting that there was no significant difference in the occurrence of quasiparallel and quasiperpendicular shock conditions between the training and test data.

Compared to the parameterized MHD model, PRIME-SH has higher representational power and therefore higher accuracy across the parameters. For $B_X$ and $B_Y$ the parameterized MHD model does not vary by much (both have Pearson's $r < 0.12$), which could be consistent with the results presented in Jung et al. (2024) Figures 2–4. For plasma flow velocity, the parameterized MHD model clearly reaches the bounds of its parameterization (most visible for $V_Y < -120$ km/s and $V_Y > 160$ km/s). The shape of the distribution for $T_{iAVG}$ is also consistent with results presented in Jung et al. (2024) Figure 2. The MHD model is more accurate than the associated analytical model for all parameters except $B_Y$, $n_i$, $T_{iAVG}$, and $P_{dyn}$, but is not more accurate than PRIME-SH for any of the parameters it is capable of predicting.

PRIME-SH is a 3D model, and its outputs are valid over any regions covered by MMS-1's orbit on the dayside (GSE $X > 0R_E$, GSE $|Y| < 5R_E$). Since the magnetosheath conditions vary significantly across its extent, PRIME-SH's accuracy evaluated against the test set is displayed in GSE coordinates in Figure 7. In general, PRIME-SH's outputs are generally less accurate on predictions closer to the Earth than on those further from the Earth. This suggests that PRIME-SH is less accurate during periods where the magnetosheath is highly compressed or when it makes predictions close to the magnetopause. These periods are rare relative to nominal conditions in the training data set, so PRIME-SH being somewhat less accurate under these conditions is expected and should be taken into account when using PRIME-SH. It has been noted in prior studies that predictions of the magnetosheath using solar wind values propagated from L1 are less accurate in the flanks as compared to the subsolar magnetosheath (Raptis et al., 2020). Since no such trend is observed in Figure 7, it seems that the more sophisticated structure of PRIME-SH has overcome this difficulty. It is worth noting that PRIME-SH has not been trained outside of the areas shown in Figure 7 and thus its predictions outside of those areas are likely to be inaccurate or unphysical due to its nature as a neural network algorithm.

**Table 3**
*Performance of the Analytical Models Cooling et al. (2001), Soucek and Escoubet (2012), and Spreiter et al. (1966) and the MHD Parameterization of OpenGGCM Derived by Jung et al. (2024) on the Magnetospheric MultiScale (MMS) Test Data Set Across Mean Absolute Error and Pearson's r Correlation Coefficient (Also Shown in Figures 4–6)*

| | Analytical models | | | MHD parameterization | |
|---|---|---|---|---|---|
| Parameter | Model | MAE | $r$ | MAE | $r$ |
| $B_{XGSM}$ | Cooling et al. (2001) | 8.44 nT (0.943$\sigma$) | 0.374 | 7.01 nT (0.783$\sigma$) | 0.111 |
| $B_{YGSM}$ | Cooling et al. (2001) | 11.6 nT (0.680$\sigma$) | 0.479 | 14.5 nT (0.850$\sigma$) | 0.055 |
| $B_{ZGSM}$ | Cooling et al. (2001) | 9.89 nT (0.616$\sigma$) | 0.546 | 9.16 nT (0.570$\sigma$) | 0.617 |
| $V_{XGSE}$ | Soucek and Escoubet (2012) | 51.6 km/s (0.670$\sigma$) | 0.487 | 34.8 km/s (0.451$\sigma$) | 0.773 |
| $V_{YGSE}$ | Soucek and Escoubet (2012) | 82.2 km/s (0.790$\sigma$) | 0.682 | 30.0 km/s (0.288$\sigma$) | 0.914 |
| $V_{ZGSE}$ | Soucek and Escoubet (2012) | 42.2 km/s (0.800$\sigma$) | 0.404 | 35.3 km/s (0.670$\sigma$) | 0.672 |
| $n_i$ | Spreiter et al. (1966) | 6.55 cm$^{-3}$ (0.305$\sigma$) | 0.938 | 8.15 cm$^{-3}$ (0.380$\sigma$) | 0.830 |
| $T_{iAVG}$ | Spreiter et al. (1966) | 47.5 eV (0.316$\sigma$) | 0.932 | 97.5 eV (0.648$\sigma$) | 0.170 |

*Note.* Since MAE is given in the units of each parameter as well as dimensionless units of standard deviations of each parameter in the MMS training data set to facilitate comparison between each parameter. Since these models output isotropic temperatures, their temperatures are compared to $T_{iAV} = (2T_{i\perp} + T_{i\parallel})/3$.



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

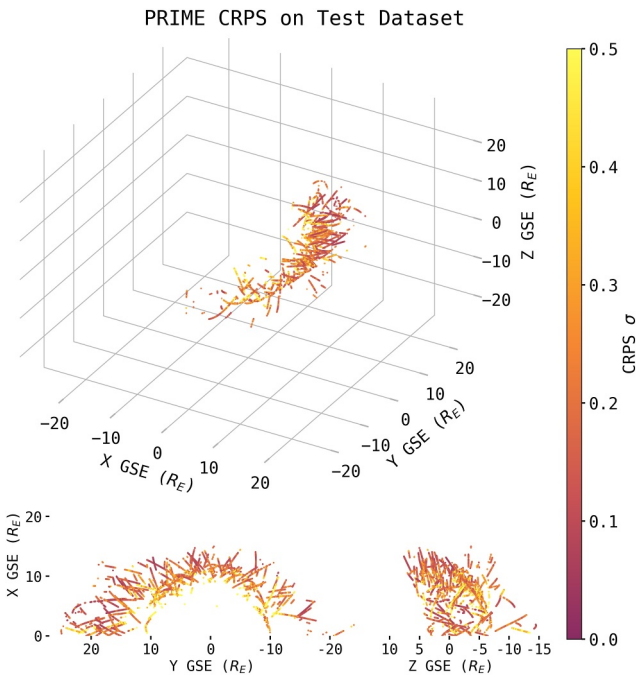


**Figure 7.** Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath's accuracy on the test data set averaged across all nine target parameters in dimensionless standard deviation units ($\sigma$). Targets arranged spatially in 3D (top), the geocentric solar ecliptic (GSE) X-Y plane (bottom left), and the GSE X-Z plane (bottom right).

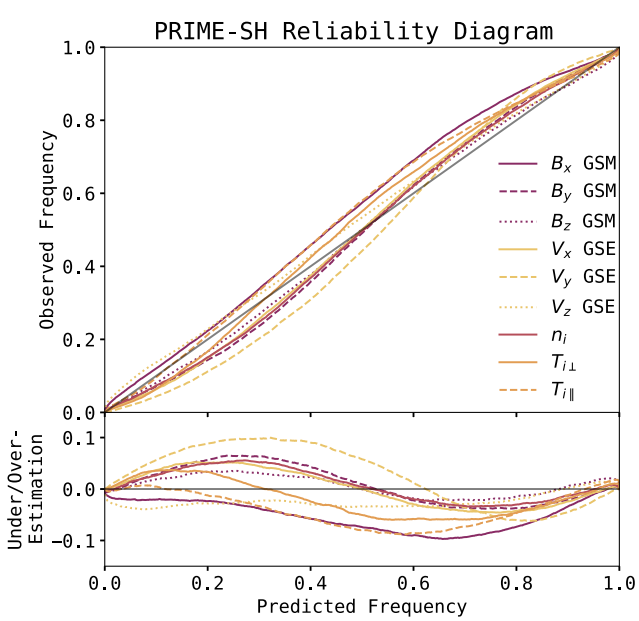


**Figure 8.** Reliability diagram constructed from Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath's (PRIME-SH's) outputs on the test data set for each parameter following the procedure of Camporeale et al. (2019). Shown versus the predicted frequency of the observation from PRIME-SH are the value of the observed frequency (top) and the deviation from perfect reliability (bottom). For the bottom plot, a given parameter being over (under) the line by an amount corresponds to PRIME-SH over (under) predicting the frequency by that amount.

Since reliability is not enforced by the CRPS loss function during training, PRIME-SH's output uncertainties must be validated quantitatively through the use of a reliability diagram (Hamill, 1997, 2001). Following the procedure in Camporeale et al. (2019) and Camporeale and Carè (2021), the standardized errors associated with prediction $\mu_i$, $\sigma_i$ with $i = 1, \ldots, N$ are defined as $\eta_i = (y_{obs,i} - \mu_i)/\left(\sqrt{2}\,\sigma_i\right)$. The probability density of a given Gaussian forecast is therefore $\Phi_i = \frac{1}{2}\left[erf(\eta_i) + 1\right]$, allowing the reliability diagram to be constructed from the empirical cumulative distribution of $\Phi_i$ given by $C(\phi) = \frac{1}{N}\sum_{i=1}^{N} H(\phi - \Phi_i)$ (with $H$ being the Heaviside step function). $C(\phi)$ is the observed frequency as a function of the predicted frequency, the same as reliability diagrams of forecasts of discrete events (e.g., those in Hamill (1997)). This method has the benefit of not requiring binning, which has been shown to affect the results of reliability diagrams of discrete events (Bröcker & Smith, 2007). $C(\phi)$ is calculated for all observations in the test data set for each parameter and presented in Figure 8.

PRIME-SH is not perfectly reliable (its reliability diagram does not exactly follow the dashed line in Figure 8); it generally tends to overestimate the likelihood of unlikely events, and underestimate the likelihood of likely events. With the exception of $V_Z$, $B_X$, and $T_\parallel$, PRIME-SH tends to be conservative. This is not unexpected, as even models perfectly calibrated on training data can suffer calibration loss on the test data set (Kull & Flach, 2015). The largest departures from perfect calibration are observed in $V_Y$ GSE (predicts events that occur with $p = 0.221$ as occurring with $p = 0.320$), $B_X$ GSM (predicts events that occur with $p = 0.754$ as occurring with $p = 0.657$), and $T_\parallel$ (predicts events that occur with $p = 0.674$ as occurring with $p = 0.586$). On average PRIME-SH is reliable to within 3.5% with a maximum difference 10% (calculated $p_{obs} - p_{pred}$). This is roughly as reliable as its progenitor algorithm PRIME and other probabilistic prediction algorithms for space weather tasks (e.g., Tasistro-Hart et al., 2021), but less reliable than those that use loss functions that enforce reliability explicitly (e.g., Hu et al., 2022).

### 4.2. Physical Validation

While a model's accuracy and reliability are important to quantify statistically, it is also important to verify that a model can reproduce expected physics. This is especially important for neural network models that can relatively easily overfit and reproduce a data set's noise rather than the underlying data representation or physics. In the following sections PRIME-SH's outputs for synthetic data are investigated to ensure that it can reproduce magnetic field and plasma physics in the magnetosheath.

#### 4.2.1. Field Line Draping and Uncertainty

Since the interplanetary magnetic field is frozen into the solar wind plasma, as the plasma is shocked and diverted around the magnetopause the magnetic field "drapes" over the obstacle forming a tangential discontinuity at the magnetopause (Crooker et al., 1985). In order to verify that PRIME-SH captures this feature of the magnetosheath, outputs are generated on a grid of points for synthetic input data. The grid is chosen to lie in the GSE X-Y or GSE X-Z plane (depending on IMF orientation) with a grid scale of $0.1R_E$. Since PRIME-SH does not output the location of the bow shock and magnetopause directly, these boundaries must be imposed somehow (a limitation of PRIME-SH). For this study all grid cells inside the Shue



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

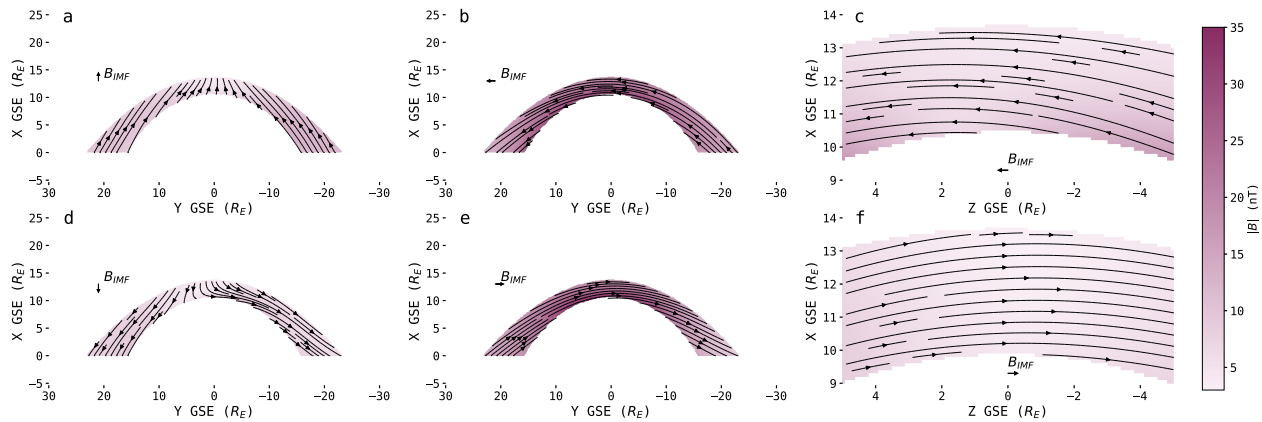


**Figure 9.** Magnetosheath conditions output by Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath (PRIME-SH) using synthetic data for six different IMF orientations (shown with arrows in top left or bottom). Plasma conditions are average conditions from the input data set, magnetic field magnitude is 5.34 nT (the average magnitude from the input data set). Shown in color is the magnitude of B, and the arrows are $B_X$ and $B_Y$ geocentric solar magnetic (GSM) field lines (for the left four plots) or the $B_X$ and $B_Z$ GSM field lines (for the right two plots).

et al. (1998) or outside the Jelínek et al. (2012) bow shock (calculated using the conditions at L1 used as inputs for PRIME-SH) are left unused. The Jelínek et al. (2012) model uses an aberrated GSE coordinate system with an $x$ axis aligned with the solar wind flow. For synthetic data with solar wind velocity parallel to the GSE $x$-axis, differences in the coordinate systems are minor. Only grid cells in regions well sampled by the MMS training data are included, hence the $Z$ extent is restricted to $\pm 5R_E$ away from the ecliptic and the nightside is not included (see Figure 1). The input data are chosen to be a 400 km/s solar wind only in the GSE X direction with otherwise average solar wind conditions from the Wind L1 data set: $|B| = 5.34$ nT, $V_X = -400$ km/s, $V_Y = 0$ km/s, $V_Z = 0$ km/s, $n_i = 7.12$ cm$^{-3}$, and $v_{th} = 34.9$ km/s. In order to investigate whether PRIME-SH is capable of draping, conditions on the grid are calculated for six different IMF orientations: one radial toward Earth (cone angle 0°), one dawnward (cone angle −90°), one duskward (cone angle +90°) one radial away from Earth (cone angle 180°), one purely northward (clock angle 0°), and one purely southward (clock angle 180°). Shown in Figure 9 are these six grids, with the sheath magnetic field streamlines plotted in black arrows and the magnitude of B in each cell in color.

As can be seen in Figure 9, PRIME-SH reproduces the draping of the magnetic field in the magnetosheath well despite the frozen in condition not being enforced during training. For cone angles of ±90° the magnetic field piles up at the nose of the magnetopause, much more than it does for radial IMF. This can be seen in the magnitude of the magnetic field, which is higher at the nose than the flanks for cone angles of ±90°. For cone angles of 0° or 180°, the flanks have a relatively higher magnetic field than the nose (though it is not as strong as the field at the nose in the cone angle ±90° case). For northward IMF, somewhat more magnetic field pileup is observed at the northern and southern flanks than for the southern IMF case. The magnetic field magnitude is also slightly higher overall in the northward IMF case than in the southward IMF case. These maps suggest that a lower reconnection rate for northward IMF at the nose causes magnetic field pileup and rearrangement in the sheath (Phan & Paschmann, 1996), sometimes referred to as the plasma depletion layer (Cummings & Coleman, 1968). This phenomenon is investigated in more detail in Section 4.2.4. Comparing to the magnetic field model of Tsyganenko et al. (2023), PRIME-SH predicts qualitatively similar field line orientations for radial IMF and IMF with cone angle ±90°. The increase in flank magnetic field for radial IMF and the increase in magnetic pileup at the nose for IMF with cone angle ±90° have almost the same magnitude and location in Figure 9 as shown in Tsyganenko et al. (2023) Figure 10. The increase in magnetic field magnitude at the nose for northward IMF is also consistent with Tsyganenko et al. (2023) Figure 6. Additionally, $|\nabla \cdot \vec{B}| < 0.005 nT/m$ across the six magnetic field configurations shown in Figure 9. While the condition $\nabla \cdot \vec{B} = 0$ is not explicitly enforced on PRIME-SH's outputs during training, the algorithm has been trained to represent the physics of the magnetosheath such that the divergence the magnetic field in its outputs is close to zero for these cases.



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

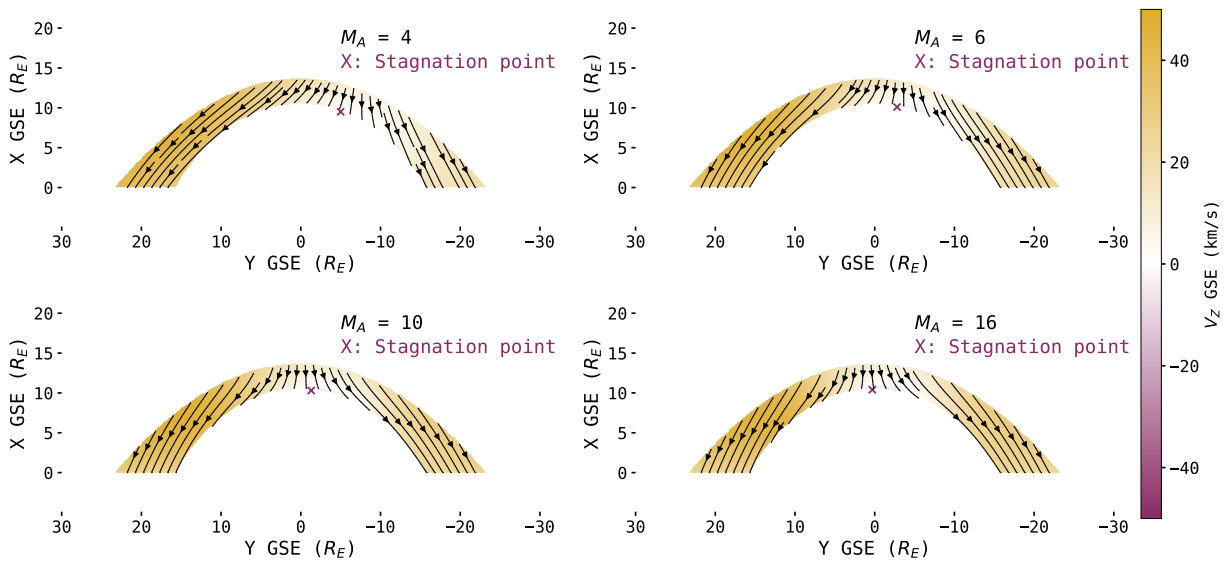


**Figure 10.** Magnetosheath conditions output by Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath using synthetic data at four different Alfvén Mach numbers ($M_A = 4, 6, 10, 16$). Flow velocity is 400 km/s with $V_Y = V_Z = 0$, magnetic field is a Parker spiral orientation whose magnitude is varied for each case to result in the four Alfvén Mach numbers. Shown in color is the Z geocentric solar ecliptic (GSE) velocity, and the arrows are the X and Y GSE velocity. The point of minimum velocity in the sheath (the stagnation point) is marked with the purple X.

#### 4.2.2. Stagnation Point

As the solar wind plasma diverts and is slowed around the magnetopause, a region known as the stagnation point develops where there is very little to no plasma flow (Spreiter et al., 1966). For radial flow and typical Parker spiral magnetic field orientation, this point is thought to be roughly located at the nose of the magnetopause (with slight aberration from Earth's ≈ 30 km/s motion in the negative GSE $Y$ direction). MHD theory predicts that for a Parker spiral IMF, the stagnation point should deflect dawnward for solar wind flows with low Alfvén Mach numbers (Russell et al., 1981). Here PRIME-SH is used to assemble predictions using the same gridded magnetosheath configuration as Section 4.2.1, however this time the Alfvén Mach number of the synthetic data set is varied from $M_A = 4$ (a) to $M_A = 16$ (d) (the solar wind typically has $M_A \approx 10$). The density and velocity are held the same ($n_i = 7.12$, $V_X = -400$ km/s) and the magnetic field is kept at a 45° Parker spiral as its magnitude is decreased in steps from 12 to 2.4 nT to yield the four Alfvén Mach numbers. Shown in Figure 10 are these four grids, with the X and Y GSE plasma flow velocity depicted with black arrows and the Z GSE flow velocity in color. Also depicted is the stagnation point, marked with a purple X.

As can be seen in Figure 10, PRIME-SH produces continuous flow maps that divert around the magnetopause for all four Alfvén mach numbers. Additionally, as the Alfvén Mach number decreases the stagnation point is observed to move dawnward as predicted by MHD theory and simulations (compare the location of the stagnation point in typical conditions in Figure 10c to its deflection in Figures 10a and 10b). This feature is hard to observe using in-situ instruments, but here through what is essentially a spatio-temporal inversion the feature is shown to occur in reality.

One interesting feature is that there appears to be some weak dawn-dusk asymmetry in the flow velocity maps produced by PRIME-SH. This could be due to biases in MMS-1's orbit showing up in PRIME-SH's output. Other experimental work has also found dawn-dusk asymmetries in the magnetosheath properties (Dimmock & Nykyri, 2013; Walsh et al., 2012), an effect not typically reproduced by MHD simulations or other physics-based simulations of flow around the magnetosphere.

#### 4.2.3. Shock Jump Conditions

Shocks, whether they are collisional or collisionless, conserve mass, momentum, and energy. The Rankine-Hugoniot shock jump conditions are formulations of each of these conservation laws in terms of the conditions upstream and downstream of the shock. For an MHD shock, define the shock normal direction to be $\hat{n}$, the plasma flow velocity to be $\vec{v}$, the plasma mass density to be $\rho$, the thermal pressure to be $P$, the specific heat ratio



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

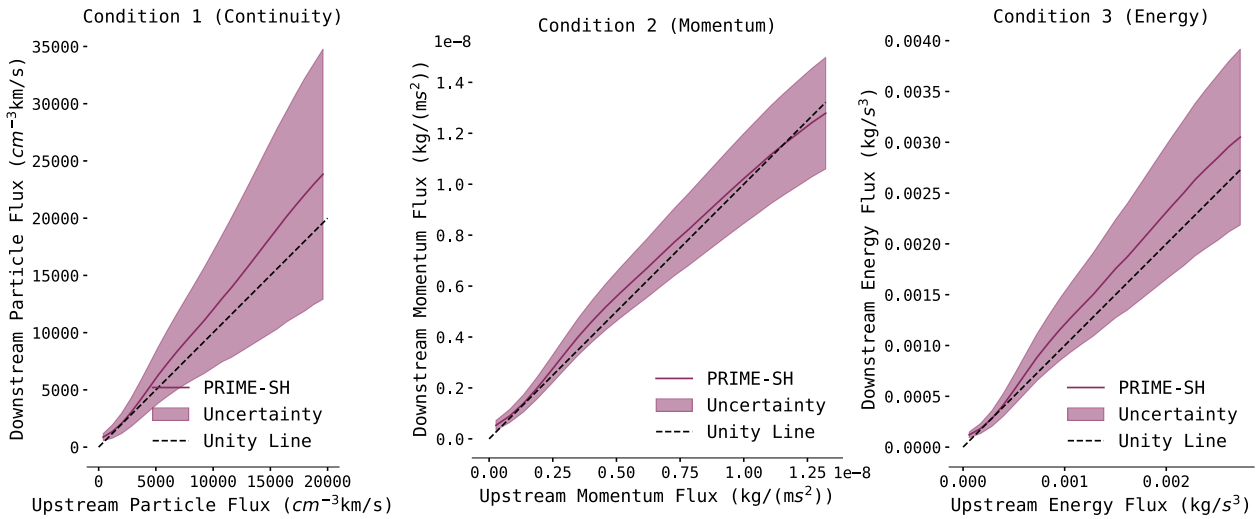


**Figure 11.** Particle, momentum, and energy fluxes calculated across a range of synthetic input conditions roughly corresponding to the range of the training data set. Fluxes are calculated just upstream of the bow shock nose (using the input data) and just downstream (using Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath's [PRIME-SH's] outputs), and $1\sigma$ uncertainties (shaded region) are calculated by propagating PRIME-SH's predicted uncertainties through the MHD shock jump condition equations. Within PRIME-SH's predicted uncertainties the three Rankine-Hugoniot MHD jump conditions are obeyed.

to be $\gamma$, and the magnetic field to be $\vec{B}$. For some quantity $\vec{X}$ upstream and downstream of the shock, define the notation $\vec{X}_{up} - \vec{X}_{down} = \left[\vec{X}\right]$. Mass conservation upstream and downstream of the shock can then be written:

$$\left[\rho\vec{u}\cdot\hat{n}\right] = 0 \tag{3}$$

Momentum conservation (with magnetic pressure included) can be written:

$$\left[\rho\vec{u}\left(\vec{u}\cdot\hat{n}\right) + \left(P + \frac{\vec{B}^2}{2\mu_0}\right)\hat{n} - \frac{\left(\vec{B}\cdot\hat{n}\right)\vec{B}}{\mu_0}\right] = 0 \tag{4}$$

Energy conservation can be written:

$$\left[\vec{u}\cdot\hat{n}\left(\rho\frac{\vec{u}^2}{2} + \frac{\gamma}{\gamma-1}P + \frac{\vec{B}^2}{\mu_0}\right) - \frac{\left(\vec{B}\cdot\hat{n}\right)\left(\vec{B}\cdot\vec{u}\right)}{\mu_0}\right] = 0 \tag{5}$$

(Kallenrode, 2010).

None of these conditions are explicitly enforced during training, but they are part of the underlying physics PRIME-SH should be representing. To validate that PRIME-SH reproduces these conservation laws, a range of synthetic solar wind conditions with densities ranging from 1 to 50 cm$^{-3}$ with $V_{GSE} = -400$ km/s, $\vec{B} = (-4nT)\hat{x} + (-4nT)\hat{y}$, and $v_{th} = 30$ km/s are initialized and used to generate predictions just behind the Jelínek et al. (2012) bow shock nose along the Sun-Earth line. This range was chosen to reflect the full range of densities from the input data set, which results in better coverage of the range of the three upstream fluxes observed than varying other conditions such as velocity. Equations 3–5 are used to calculate the particle, momentum, and energy flux from the synthetic input data (upstream) and from PRIME-SH's outputs (downstream). The uncertainties predicted by PRIME-SH can be propagated through Equations 3–5 to obtain uncertainties for the downstream fluxes as well. Only magnetosheath conditions on the Sun-Earth line just behind the Jelínek et al. (2012) bow shock nose are included so it can be assumed that $\hat{n} = \hat{x}$ (although this does introduce some dependence on the choice of analytical bow shock model to the results). The downstream fluxes are plotted as a function of upstream fluxes in Figure 11.



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Perfect conservation of each flux is represented by the dashed lines in Figure 11. As can be seen, while the quantities predicted by PRIME-SH do not perfectly conserve mass/particles, momentum, and energy, it does conserve them within the $1\sigma$ uncertainty bounds (shaded region) for each quantity. One contribution to this uncertainty is an experimental one. Although the instruments on Wind and MMS have been carefully calibrated, they were not calibrated together. Previous studies have found mismatches when comparing plasma and magnetic field parameters from different missions, even those with very similar instruments (King & Papitashvili, 2005; Roberts et al., 2021). The points of largest *fractional* difference between upstream and downstream fluxes occur for the smallest fluxes (when $n_{up} = 1$ cm$^{-3}$), which happens relatively infrequently in the input data set. Another source of uncertainty in this figure is the fact that the form of these conservation laws are derived from fluid equations, whereas PRIME-SH is based on observations of the magnetosheath. Kinetic effects at the bow shock can cause apparent violations of the simple fluid form conservation laws when in reality the quantity is conserved (e.g., particles reflected at the quasiparallel shock can cause downstream flux in the magnetosheath to appear lower than the incident solar wind flux). Despite the fact that mass/particle conservation, momentum conservation, and energy conservation were not explicitly enforced during training, PRIME-SH has been optimized such that it successfully represents the underlying physics to a degree that the three quantities are conserved.

#### 4.2.4. Plasma Depletion Layer

The plasma depletion layer is a transient region of the subsolar magnetosheath characterized by decreased density and increased magnetic field strength. This layer exists when the reconnection rate at the magnetopause is insufficient to prevent "pile-up" of magnetic flux, and as such is typically observed during periods of northward IMF (although it can sometimes be observed during periods of southward IMF). This "pile-up" can modify the local reconnection rate, and could even enable reconnection at the subsolar magnetopause for northward IMF (Anderson, 1996). It has also been shown that the plasma depletion layer has stronger temperature anisotropy than the rest of the magnetosheath, although it is currently unclear whether this is a formation mechanism of the region or simply a consequence of the flux pile-up (Phan & Paschmann, 1996). Despite the fact that the plasma depletion layer has been observed by in-situ spacecraft for many years (Cummings & Coleman, 1968), the dynamics and global geometry of the region is difficult to determine from observations due to their spatio-temporal ambiguity (Wang et al., 2004).

Both to verify PRIME-SH has been properly trained to replicate solar wind flow around the magnetosphere and to overcome the spatio-temporal ambiguity of in situ observations, PRIME-SH is used to assemble predictions using the same gridded magnetosheath configuration as Section 4.2.1 for northward $\left(\vec{B} = 5nT\hat{z}\right)$ and southward $\left(\vec{B} = -5nT\hat{z}\right)$ IMF. Plasma conditions are the same between each run ($V_{GSE} = -400$ km/s, $n = 5$ cm$^{-3}$, and $v_{th} = 30$ km/s, Alfvén Mach number 8). The magnetic field magnitude, density, and temperature anisotropy ($T_\perp/T_\parallel$) are shown for each configuration in Figure 12 in the ecliptic and in cuts along the Sun-Earth line.

The plasma depletion layer can be identified in Figure 12 as the region of high $|B|$, $T_\perp/T_\parallel$ and low $n$ close to the subsolar point in the northward IMF (Figures 12a–12c) case that is not apparent in the southward IMF case. In the cuts along the Sun-Earth line for each case (Figures 12g–12i), the density can be more readily observed to begin falling off about $1R_E$ from the magnetopause, while at the same time $|B|$ and $T_\perp/T_\parallel$ begin to increase. This is contrasted with the southward case, in which all three parameters increase across the sheath somewhat linearly. This thickness is consistent with reported thicknesses from the literature which range from $0.3R_E$ to $1R_E$ for $M_A = 8$, depending on identification criteria (Wang et al., 2004). It is worth noting that the exact thickness predicted here depends on the subsolar distance predicted by the Shue et al. (1998) magnetopause model used to determine the boundaries in Figure 12. This validates that PRIME-SH has been trained to reproduce magnetic flux pile-up and its effects in the magnetosheath, which are indirect measurements of the dayside magnetic reconnection rate. Unlike numerical simulations, PRIME-SH can generate spatial map of the plasma depletion layer based directly on observations rather than physical assumptions, which can cause deviation between predicted and observed global depletion layer configurations (Southwood & Kivelson, 1995; Zwan & Wolf, 1976).

## 5. Conclusions

A RNN with probabilistic outputs is trained to predict MMS-1 observations of Earth's magnetosheath given timeseries input measured by the Wind spacecraft at L1. This algorithm, called PRIME-SH in reference to its



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2024JH000235

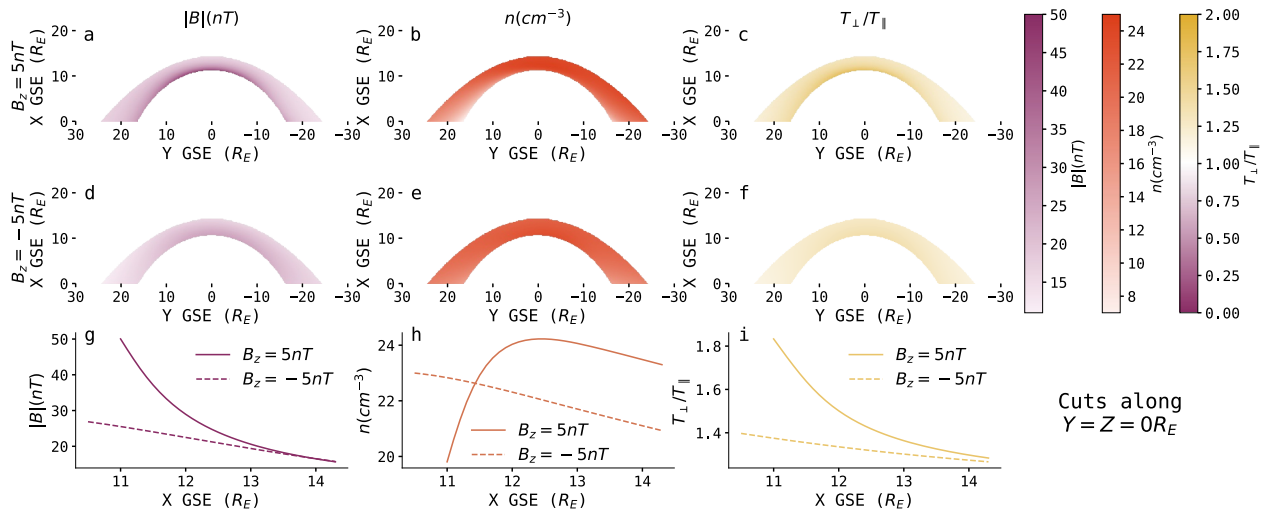


**Figure 12.** Magnetosheath conditions output by Probabilistic Regressor for Input to the Magnetosphere Estimation-magnetosheath for synthetic input conditions with IMF purely northward ($B_Z = 5$ nT, a–c) and purely southward ($B_Z = -5$ nT, d–f). Average plasma conditions from the input data set are used. Top row shows $|B|$, $n_i$, and $T_\perp/T_\parallel$ for $B_Z = 5$ nT, middle row shows the same for $B_Z = -5$ nT, and bottom row shows cuts along $Y = Z = 0R_E$ for both magnetic field orientations for each parameter for ease of comparison.

progenitor algorithm PRIME, incorporates the time history of the solar wind at L1 to generate probability distributions for magnetosheath plasma and magnetic field parameters. These probability distributions can be used to determine the uncertainty associated with PRIME-SH's predictions.

PRIME-SH is shown to have good performance in a statistical sense across a test data set of MMS-1 data not used during training (Average CRPS $0.221\sigma$). The uncertainties predicted by PRIME-SH are shown to be reliable to within 3.5% with a maximum difference 10% through a comparison to the test data set. Additionally, PRIME-SH predicts magnetosheath conditions more accurately than several popular analytical models (Cooling et al., 2001; Kobel & Flückiger, 1994; Soucek & Escoubet, 2012; Spreiter et al., 1966) and a parameterization of the OpenGGCM MHD code (Jung et al., 2024). While statistical validation is important, it is also important to validate that a model is indeed producing physical results. It is verified that the magnetic field values produced by PRIME-SH across a grid of points in the magnetosheath "drape" across the magnetopause in 3D for several different orientations of the upstream magnetic field. Plasma flow velocities output by PRIME-SH across a grid of magnetosheath points divert around the magnetopause as expected, and the point at which the flow stagnates moves dawnward with decreasing Alfvén Mach number as predicted by MHD theory (Russell et al., 1981). PRIME-SH is shown to conserve particle/mass flux, momentum flux, and energy flux within $1\sigma$ uncertainty across the bow shock for the range of input parameters it is trained on. PRIME-SH is also capable of reproducing the plasma depletion layer given input conditions for which the depletion layer is expected to form. From this it may be concluded that PRIME-SH has indeed been optimized to represent the physics of solar wind flow from L1, through the bow shock, and around the magnetopause.

PRIME-SH is not only more accurate in a statistical sense than current analytical models and MHD simulation parameterizations, but it also has additional functionality these other models do not. First, PRIME-SH outputs $T_\perp$ and $T_\parallel$ separately. While it is possible to have anisotropic temperatures in MHD simulations using a few assumptions (Erkaev et al., 1999), most MHD and analytical models currently assume isotropic temperatures. Additionally, PRIME-SH outputs uncertainties for its outputs. These uncertainties were used in this study to assign confidence intervals to fluxes calculated to verify that PRIME-SH conserves particles, mass, and energy. Studies that are enabled by PRIME-SH include investigations of bow shock physics by combining PRIME and PRIME-SH (Pulkkinen et al., 2016), changes in the rate of energy transfer across the magnetopause due to rearrangements of the magnetosheath flow (Lopez, 2016), and any other study that has been previously limited by the lack of continuous in-situ plasma or magnetic field data. In short, PRIME-SH is an accurate and reliable magnetosheath prediction algorithm that offers functionality no other magnetosheath prediction algorithm does, and enables new statistical and event-based studies of the magnetosheath.



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## Data Availability Statement

Magnetospheric Multiscale, Wind, and OMNI data are available through the Coordinated Data Analysis Web (CDAWeb) online portal at https://cdaweb.gsfc.nasa.gov/istp_public/. Codes for data set preparation, algorithm development, and analysis presented in this paper are available at (O'Brien & Qudsi, 2024).

**Acknowledgments**
Authors CO, BMW, and YZ would like to acknowledge support from NASA Grants 80NSSC21K0026 and 80NSSC20K1710. Author ST acknowledges support from the German Aerospace Center (DLR). DGS was supported by NASA's MMS Theory and Modeling program. The authors acknowledge the instrument teams for FPI, FGM, SWE, and MFI, as well as the other MMS and Wind instrument teams whose labor made this study possible.

## References

Al Shidi, Q., Pulkkinen, T., & Welling, D. (2023). Uncertainty quantification in global simulations: Solar wind propagation effects (other). *Oral*. https://doi.org/10.5194/egusphere-egu23-11764

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, *9*(7), 1518–1530. https://doi.org/10.1175/1520-0442(1996)009⟨1518:AMFPAE⟩2.0.CO;2

Argall, M. R., Small, C. R., Piatt, S., Breen, L., Petrik, M., Kokkonen, K., et al. (2020). MMS SITL ground loop: Automating the burst data selection process. *Frontiers in Astronomy and Space Sciences*, *7*, 54. https://doi.org/10.3389/fspas.2020.00054

Axford, W. (1964). Viscous interaction between the solar wind and the earth's magnetosphere. *Planetary and Space Science*, *12*(1), 45–53. https://doi.org/10.1016/0032-0633(64)90067-4

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *arXiv*. Retrieved from http://arxiv.org/abs/1607.06450

Barker, A. B., Li, X., & Selesnick, R. S. (2005). Modeling the radiation belt electrons with radial diffusion driven by the solar wind. *Space Weather*, *3*(10), 2004SW000118. https://doi.org/10.1029/2004SW000118

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), 27–31. https://doi.org/10.1109/45.329294

Borovsky, J. E. (2008). Flux tube texture of the solar wind: Strands of the magnetic carpet at 1 AU? *Journal of Geophysical Research*, *113*(A8), A08110. https://doi.org/10.1029/2007JA012684

Borovsky, J. E. (2018). The spatial structure of the oncoming solar wind at Earth and the shortcomings of a solar-wind monitor at L1. *Journal of Atmospheric and Solar-Terrestrial Physics*, *177*, 2–11. https://doi.org/10.1016/j.jastp.2017.03.014

Borovsky, J. E. (2021). Is our understanding of solar-wind/magnetosphere coupling satisfactory? *Frontiers in Astronomy and Space Sciences*, *8*, 634073. https://doi.org/10.3389/fspas.2021.634073

Bröcker, J., & Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, *22*(3), 651–661. https://doi.org/10.1175/WAF993.1

Burch, J. L., Moore, T. E., Torbert, R. B., & Giles, B. L. (2016). Magnetospheric multiscale overview and science objectives. *Space Science Reviews*, *199*(1–4), 5–21. https://doi.org/10.1007/s11214-015-0164-9

Camporeale, E., & Carè, A. (2021). Accrue: Accurate and reliable uncertainty estimate in deterministic models. *International Journal for Uncertainty Quantification*, *11*(4), 81–94. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2021034623

Camporeale, E., Chu, X., Agapitov, O. V., & Bortnik, J. (2019). On the generation of probabilistic forecasts from deterministic models. *Space Weather*, *17*(3), 455–475. https://doi.org/10.1029/2018SW002026

Cheng, I. K., Achilleos, N., & Smith, A. (2022). Automated bow shock and magnetopause boundary detection with Cassini using threshold and deep learning methods. *Frontiers in Astronomy and Space Sciences*, *9*, 1016453. https://doi.org/10.3389/fspas.2022.1016453

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv*. Retrieved from http://arxiv.org/abs/1409.1259

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv*. Retrieved from http://arxiv.org/abs/1412.3555

Cooling, B. M. A., Owen, C. J., & Schwartz, S. J. (2001). Role of the magnetosheath flow in determining the motion of open flux tubes. *Journal of Geophysical Research*, *106*(A9), 18763–18775. https://doi.org/10.1029/2000JA000455

Crooker, N. U., Luhmann, J. G., Russell, C. T., Smith, E. J., Spreiter, J. R., & Stahara, S. S. (1985). Magnetic field draping against the dayside magnetopause. *Journal of Geophysical Research*, *90*(A4), 3505–3510. https://doi.org/10.1029/JA090iA04p03505

Cummings, W. D., & Coleman, P. J. (1968). Magnetic fields in the magnetopause and vicinity at synchronous altitude. *Journal of Geophysical Research*, *73*(17), 5699–5718. https://doi.org/10.1029/JA073i017p05699

Dimmock, A. P., & Nykyri, K. (2013). The statistical mapping of magnetosheath plasma properties based on THEMIS measurements in the magnetosheath interplanetary medium reference frame: Magnetosheath statistical mapping. *Journal of Geophysical Research: Space Physics*, *118*(8), 4963–4976. https://doi.org/10.1002/jgra.50465

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization.

Dungey, J. W. (1961). Interplanetary magnetic field and the auroral zones. *Physical Review Letters*, *6*(2), 47–48. https://doi.org/10.1103/PhysRevLett.6.47

Eastwood, J. P., Lucek, E. A., Mazelle, C., Meziane, K., Narita, Y., Pickett, J., & Treumann, R. A. (2005). The foreshock. *Space Science Reviews*, *118*(1–4), 41–94. https://doi.org/10.1007/s11214-005-3824-3

Erkaev, N. V., Farrugia, C. J., & Biernat, H. K. (1999). Three-dimensional, one-fluid, ideal MHD model of magnetosheath flow with anisotropic pressure. *Journal of Geophysical Research*, *104*(A4), 6877–6887. https://doi.org/10.1029/1998JA900134

Fleetham, A. L., Milan, S. E., Imber, S. M., & Vines, S. K. (2023). Solar wind control of hemispherically-integrated field-aligned currents at earth. *Journal of Geophysical Research: Space Physics*, *128*, e2023JA031540. https://doi.org/10.22541/essoar.168121412.21332677/v1

Génot, V., Broussillou, L., Budnik, E., Hellinger, P., Trávníček, P. M., Lucek, E., & Dandouras, I. (2011). Timing mirror structures observed by Cluster with a magnetosheath flow model. *Annales Geophysicae*, *29*(10), 1849–1860. https://doi.org/10.5194/angeo-29-1849-2011

Gilet, N., De Leon, E., Gallé, R., Vallières, X., Rauch, J., Jegou, K., et al. (2021). Automatic detection of the thermal electron density from the WHISPER experiment onboard CLUSTER-II mission with neural networks. *Journal of Geophysical Research: Space Physics*, *126*(3), e2020JA028901. https://doi.org/10.1029/2020JA028901

Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, *133*(5), 1098–1118. https://doi.org/10.1175/MWR2904.1

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather and Forecasting*, *12*(4), 736–741. https://doi.org/10.1175/1520-0434(1997)012⟨0736:RDFMPF⟩2.0.CO;2

Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, *129*(3), 550–560. https://doi.org/10.1175/1520-0493(2001)129⟨0550:IORHFV⟩2.0.CO;2



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015⟨0559:DOTCRP⟩2.0.CO;2

Hoilijoki, S., Palmroth, M., Walsh, B. M., Pfau-Kempf, Y., von Alfthan, S., Ganse, U., et al. (2016). Mirror modes in the Earth's magnetosheath: Results from a global hybrid-Vlasov simulation. *Journal of Geophysical Research: Space Physics*, *121*(5), 4191–4204. https://doi.org/10.1002/2015JA022026

Hu, A., Camporeale, E., & Swiger, B. (2022). Multi-hour ahead Dst index prediction using multi-fidelity boosted neural networks. *arXiv*. Retrieved from http://arxiv.org/abs/2209.12571

Huang, S., Li, W., Shen, X., Ma, Q., Chu, X., Ma, D., et al. (2022). Application of recurrent neural network to modeling earth's global electron density. *Journal of Geophysical Research: Space Physics*, *127*(9), e2022JA030695. https://doi.org/10.1029/2022JA030695

Jelínek, K., Němeček, Z., & Šafránková, J. (2012). A new approach to magnetopause and bow shock modeling based on automated region identification. *Journal of Geophysical Research*, *117*(A5), 2011JA017252. https://doi.org/10.1029/2011JA017252

Jung, J. K., Connor, H. P., Dimmock, A., Sembay, S., Read, M., Soucek, J., et al. (2024). Mshpy23: A user-friendly, parameterized model of magnetosheath conditions. *Earth and Planetary Physics*, *8*(1), 89–104. https://doi.org/10.26464/epp2023065

Kallenrode, M.-B. (2010). *Space physics: An introduction to plasmas and particles in the heliosphere and magnetospheres; with 12 tables, numerous exercises and problems* (3. ed., paperback edn). Springer.

Karlsson, T., Raptis, S., Trollvik, H., & Nilsson, H. (2021). Classifying the magnetosheath behind the quasi-parallel and quasi-perpendicular bow shock by local measurements. *Journal of Geophysical Research: Space Physics*, *126*(9), e2021JA029269. https://doi.org/10.1029/2021JA029269

King, J. H., & Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field data. *Journal of Geophysical Research*, *110*(A2), A02104. https://doi.org/10.1029/2004JA010649

King, J. H., & Papitashvili, N. E. (2020). OMNI 1-min data set. *NASA Space Physics Data Facility*. https://doi.org/10.48322/45BB-8792

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv*. Retrieved from http://arxiv.org/abs/1412.6980

Kobel, E., & Flückiger, E. O. (1994). A model of the steady state magnetic field in the magnetosheath. *Journal of Geophysical Research*, *99*(A12), 23617–23622. https://doi.org/10.1029/94JA01778

Koons, H. C., & Gorney, D. J. (1991). A neural network model of the relativistic electron flux at geosynchronous orbit. *Journal of Geophysical Research*, *96*(A4), 5549–5556. https://doi.org/10.1029/90JA02380

Kull, M., & Flach, P. (2015). Novel decompositions of proper scoring rules for classification: Score adjustment as precursor to calibration. In A. Appice, P. P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, & A. Jorge (Eds.), *Machine learning and knowledge discovery in databases* (Vol. 9284, pp. 68–85). Springer International Publishing. https://doi.org/10.1007/978-3-319-23528-8_5

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv*. Retrieved from http://arxiv.org/abs/1612.01474

Lepping, R. P., Acũna, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., et al. (1995). The WIND magnetic field investigation. *Space Science Reviews*, *71*(1–4), 207–229. https://doi.org/10.1007/BF00751330

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018). Hyperband: A novel Bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, *18*(185), 1–52. Retrieved from http://jmlr.org/papers/v18/16-558.html

Li, X. (2004). Variations of 0.7–6.0 MeV electrons at geosynchronous orbit as a function of solar wind. *Space Weather*, *2*(3), 2003SW000017. https://doi.org/10.1029/2003SW000017

Lockwood, M., Bentley, S. N., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., & Allanson, O. (2019). The development of a space climatology: 1. Solar wind magnetosphere coupling as a function of timescale and the effect of data gaps. *Space Weather*, *17*(1), 133–156. https://doi.org/10.1029/2018SW001856

Lopez, R. E. (2016). The integrated dayside merging rate is controlled primarily by the solar wind: The solar wind controls dayside merging. *Journal of Geophysical Research: Space Physics*, *121*(5), 4435–4445. https://doi.org/10.1002/2016JA022556

Lyon, J., Fedder, J., & Mobarry, C. (2004). The Lyon–Fedder–Mobarry (LFM) global MHD magnetospheric simulation code. *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(15–16), 1333–1350. https://doi.org/10.1016/j.jastp.2004.03.020

Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096. https://doi.org/10.1287/mnsc.22.10.1087

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables: Universal coupling function. *Journal of Geophysical Research*, *112*(A1), A01206. https://doi.org/10.1029/2006JA012015

Nix, D., & Weigend, A. (1994). Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE international conference on neural networks (ICNN'94)* (Vol. 1, pp. 55–60). IEEE. https://doi.org/10.1109/ICNN.1994.374138

O'Brien, C., & Qudsi, R. (2024). connor-obrien888/primesh: Paper Release. *Zenodo*. https://doi.org/10.5281/ZENODO.11128038

O'Brien, C., Walsh, B. M., Zou, Y., Tasnim, S., Zhang, H., & Sibeck, D. G. (2023). Prime: A probabilistic neural network approach to solar wind propagation from L1. *Frontiers in Astronomy and Space Sciences*, *10*, 1250779. https://doi.org/10.3389/fspas.2023.1250779

Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., et al. (1995). SWE, a comprehensive plasma instrument for the WIND spacecraft. *Space Science Reviews*, *71*(1–4), 55–77. https://doi.org/10.1007/BF00751326

Olshevsky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzen, S., et al. (2021). Automated classification of plasma regions using 3D particle energy distributions. *Journal of Geophysical Research: Space Physics*, *126*(10), e2021JA029620. https://doi.org/10.1029/2021JA029620

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Kerastuner. Retrieved from https://github.com/keras-team/keras-tuner

Phan, T. D., & Paschmann, G. (1996). Low-latitude dayside magnetopause and boundary layer for high magnetic shear: 1. Structure and motion. *Journal of Geophysical Research*, *101*(A4), 7801–7815. https://doi.org/10.1029/95JA03752

Plaschke, F., Hietala, H., Archer, M., Blanco-Cano, X., Kajdič, P., Karlsson, T., et al. (2018). Jets downstream of collisionless shocks. *Space Science Reviews*, *214*(5), 81. https://doi.org/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 Science Reviews*, *199*(1–4), 331–406. https://doi.org/10.1007/s11214-016-0245-4

Powell, K. G., Roe, P. L., Linde, T. J., Gombosi, T. I., & De Zeeuw, D. L. (1999). A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *Journal of Computational Physics*, *154*(2), 284–309. https://doi.org/10.1006/jcph.1999.6299

Pulkkinen, T. I., Dimmock, A. P., Lakka, A., Osmane, A., Kilpua, E., Myllys, M., et al. (2016). Magnetosheath control of solar wind-magnetosphere coupling efficiency. *Journal of Geophysical Research: Space Physics*, *121*(9), 8728–8739. https://doi.org/10.1002/2016JA023011



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Raeder, J., Larson, D., Li, W., Kepko, E. L., & Fuller-Rowell, T. (2008). OpenGGCM simulations for the THEMIS mission. *Space Science Reviews*, *141*(1–4), 535–555. https://doi.org/10.1007/s11214-008-9421-5

Raeder, J., McPherron, R. L., Frank, L. A., Kokubun, S., Lu, G., Mukai, T., et al. (2001). Global simulation of the Geospace Environment Modeling substorm challenge event. *Journal of Geophysical Research*, *106*(A1), 381–395. https://doi.org/10.1029/2000JA000605

Raptis, S., Aminalragia-Giamini, S., Karlsson, T., & Lindberg, M. (2020). Classification of magnetosheath jets using neural networks and high resolution OMNI (HRO) data. *Frontiers in Astronomy and Space Sciences*, *7*, 24. https://doi.org/10.3389/fspas.2020.00024

Roberts, O. W., Nakamura, R., Coffey, V. N., Gershman, D. J., Volwerk, M., Varsani, A., et al. (2021). A study of the solar wind ion and electron measurements from the magnetospheric Multiscale mission's fast plasma investigation. *Journal of Geophysical Research: Space Physics*, *126*(10), e2021JA029784. https://doi.org/10.1029/2021JA029784

Russell, C. T., Anderson, B. J., Baumjohann, W., Bromund, K. R., Dearborn, D., Fischer, D., et al. (2016). The magnetospheric Multiscale magnetometers. *Space Science Reviews*, *199*(1–4), 189–256. https://doi.org/10.1007/s11214-014-0057-3

Russell, C. T., Zhuang, H., Walker, R. J., & Crooker, N. U. (1981). A note on the location of the stagnation point in the magnetosheath flow. *Geophysical Research Letters*, *8*(9), 984–986. https://doi.org/10.1029/GL008i009p00984

Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., et al. (1998). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research*, *103*(A8), 17691–17700. https://doi.org/10.1029/98JA01103

Sivadas, N., Sibeck, D., Subramanyan, V., Walach, M.-T., Murphy, K., & Halford, A. (2022). Uncertainty in solar wind forcing explains polar cap potential saturation. *arXiv*. Retrieved from http://arxiv.org/abs/2201.02137(Number:arXiv:2201.02137arXiv:2201.02137[astro-ph,physics:physics])

Soucek, J., & Escoubet, C. P. (2012). Predictive model of magnetosheath plasma flow and its validation against Cluster and THEMIS data. *Annales Geophysicae*, *30*(6), 973–982. https://doi.org/10.5194/angeo-30-973-2012

Southwood, D. J., & Kivelson, M. G. (1995). Magnetosheath flow near the subsolar magnetopause: Zwan-Wolf and Southwood-Kivelson theories reconciled. *Geophysical Research Letters*, *22*(23), 3275–3278. https://doi.org/10.1029/95GL03131

Spreiter, J. R., & Alksne, A. Y. (1969). Plasma flow around the magnetosphere. *Reviews of Geophysics*, *7*(1–2), 11–50. https://doi.org/10.1029/RG007i001p00011

Spreiter, J. R., Summers, A. L., & Alksne, A. Y. (1966). Hydromagnetic flow around the magnetosphere. *Planetary and Space Science*, *14*(3), 223–253. https://doi.org/10.1016/0032-0633(66)90124-3

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958. Retrieved from http://jmlr.org/papers/v15/srivastava14a.html

Tasistro-Hart, A., Grayver, A., & Kuvshinov, A. (2021). Probabilistic geomagnetic storm forecasting via deep learning. *Journal of Geophysical Research: Space Physics*, *126*(1), e2020JA028228. https://doi.org/10.1029/2020JA028228

Tsyganenko, N. A., Semenov, V. S., & Erkaev, N. V. (2023). Data-based modeling of the magnetosheath magnetic field. *Journal of Geophysical Research: Space Physics*, *128*(11), e2023JA031665. https://doi.org/10.1029/2023JA031665

Von Alfthan, S., Pokhotelov, D., Kempf, Y., Hoilijoki, S., Honkonen, I., Sandroos, A., & Palmroth, M. (2014). Vlasiator: First global hybrid-Vlasov simulations of Earth's foreshock and magnetosheath. *Journal of Atmospheric and Solar-Terrestrial Physics*, *120*, 24–35. https://doi.org/10.1016/j.jastp.2014.08.012

Walsh, B. M., Sibeck, D. G., Wang, Y., & Fairfield, D. H. (2012). Dawn-dusk asymmetries in the earth's magnetosheath: Magnetosheath asymmetries. *Journal of Geophysical Research*, *117*(A12), 12211. https://doi.org/10.1029/2012JA018240

Wang, Y. L., Raeder, J., & Russell, C. T. (2004). Plasma depletion layer: Magnetosheath flow structure and forces. *Annales Geophysicae*, *22*(3), 1001–1017. https://doi.org/10.5194/angeo-22-1001-2004

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences (No. v. 100)* (3rd ed. ed.). Elsevier/Academic Press.

Zamo, M., & Naveau, P. (2018). Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, *50*(2), 209–234. https://doi.org/10.1007/s11004-017-9709-7

Zhang, H., Zong, Q., Connor, H., Delamere, P., Facskó, G., Han, D., et al. (2022). Dayside transient phenomena and their impact on the magnetosphere and ionosphere. *Space Science Reviews*, *218*(5), 40. https://doi.org/10.1007/s11214-021-00865-0

Zwan, B. J., & Wolf, R. A. (1976). Depletion of solar wind plasma near a planetary boundary. *Journal of Geophysical Research*, *81*(10), 1636–1648. https://doi.org/10.1029/JA081i010p01636



29935210, 2024, 3, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024JH000235 by Dtsch Zentrum F. Luft-U. Raum Fahrt In D. Helmholtz Gemein., Wiley Online Library on [02/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License