

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

C. O'Brien¹, B. M. Walsh¹, Y. Zou², R. Qudsi¹, S. Tasnim³, H. Zhang⁴, D.G.
Sibeck⁵

¹Center for Space Physics, Boston University, Boston, MA, USA

²Johns Hopkins University Applied Physics Lab, Laurel, MD, USA

³Institute for Solar-Terrestrial Physics, German Aerospace Center (DLR), Neustrelitz, Germany

⁴Computer Science Department, University of Alabama in Huntsville, Huntsville, AL, USA

⁵Heliophysics Science Division, NASA/GSFC, Greenbelt, MD, USA

Key Points:

- PRIME-SH is an algorithm that predicts plasma and magnetic field in Earth's magnetosheath using inputs from the Wind spacecraft at L1.
- PRIME-SH accurately predicts the magnetosheath conditions in a statistical sense and obeys mass, momentum, and energy conservation.
- PRIME-SH can be used to easily assemble spatial maps of the magnetosheath and to study physics not captured by MHD simulations.

Corresponding author: Connor O'Brien, obrienco@bu.edu

Abstract

A data-driven model of the Earth’s magnetosheath is developed by training a Bayesian recurrent neural network to reproduce Magnetospheric MultiScale (MMS) measurements of the magnetosheath plasma and magnetic field using measurements from the Wind spacecraft upstream of the Earth at the first Earth-Sun Lagrange point (L1). This model, called PRIME-SH in reference to its progenitor algorithm PRIME (Probabilistic Regressor for Input to the Magnetosphere Estimation), is shown to predict spacecraft observations of magnetosheath conditions accurately in a statistical sense with a continuous rank probability score (CRPS) of 0.227σ and more accurately than analytical models of the magnetosheath. Furthermore, PRIME-SH is shown to reproduce physics not explicitly enforced during training, such as field line draping, quasi-perpendicular and quasi-parallel magnetohydrodynamic (MHD) shock behavior, 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 kinetic physics such as temperature anisotropy, and being capable of reliably estimating its own uncertainty.

Plain Language Summary

As the solar wind encounters the 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 PRIME-SH, adapts a Bayesian 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 the 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 (Dungey, 1961; Axford, 1964). 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. This is because there is no continuous in-situ magnetosheath monitor due to orbital constraints. In order to know the conditions in the magnetosheath during periods of time or locations in space where there is no in-situ monitor in the magnetosheath available, it is therefore necessary to model the magnetosheath.

Early models of the magnetosheath were gas dynamic in nature, incorporating some physical assumptions and including limited consideration of the magnetic field outside the magnetopause (Spreiter et al., 1966; Spreiter & Alksne, 1969). These models have matured through the inclusion of additional physics into modern MHD codes (e.g. Powell et al. (1999); Lyon et al. (2004)), which offer spatially and temporally complete model magnetosheaths at the cost of some physical assumptions and increased computational expense. To overcome the computational expense of MHD modeling, 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 their accuracy. A shared feature of these approaches is that they all include physical assumptions. 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; hybrid-Vlasov codes capable of simulating the entire magnetosheath have recently come online (Von Alfthan et al., 2014) but come with an even higher computational cost than MHD codes and still simplify some physics.

One possible way of addressing this limitation is the use of neural network codes that do not assume a functional form, or any physics for that matter. 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. These algorithms do not require physical assumptions to construct tractable or analytical descriptions of the magnetosheath plasma and magnetic field, and are also computationally lightweight. In particular, new Bayesian recurrent neural network architectures have shown good performance in spatio-temporal inversion tasks such as electron density in the inner magnetosphere (Huang et al., 2022). These algorithms have the benefit of quantifying the uncertainty of their predictions, a crucial aspect of any prediction algorithm and one that is typically lacking in magnetospheric physics (Borovsky, 2021).

Another challenge with these 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 the Earth as input. Much like the magnetosheath, there is no continuous in-situ monitor of the solar wind near the 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 the Earth need to be propagated to the Earth to account for the travel time of the solar wind plasma and interplanetary magnetic field (generally 30-60 minutes). This propagation task is made difficult by the structure and dynamics of the solar wind (Borovsky, 2018), and a great deal of varied algorithms have been developed in order to propagate between L1 and the 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 Bayesian recurrent neural network 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).

In this study, a new algorithm capable of predicting magnetosheath plasma and magnetic field conditions given measurements made by the Wind spacecraft at L1 is developed. This algorithm, named PRIME-SH after its progenitor algorithm PRIME (O’Brien et al., 2023), requires a dataset of in-situ magnetosheath measurements and associated solar wind inputs at L1 (Section 2), a network architecture adapted from PRIME and optimized for the magnetosheath (Section 3). Outputs from PRIME-SH are validated in a statistical sense and by demonstrating PRIME-SH reproduces some expected physics (Section 4), after which the results can be summarized and discussed (Section 5).

2 Data

2.1 MMS Target Dataset

Plasma and magnetic field data from the Magnetospheric Multi Scale 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 the 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.

To assemble a solar wind dataset using MMS, an automatic tool developed by Olshevsky et al. (2021) is used to classify all MMS-1 FPI 3D ion distributions from September 2nd 2015 to January 1st 2023. The classifier is capable of discriminating between magnetospheric, magnetosheath, non-foreshock solar wind, and foreshock plasma through 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 dataset. Remaining FGM magnetic field and FPI ion moments are averaged in 100 second bins. Since the classifier is trained only on data from dayside orbits, any data on the nightside ($GSE\ X < 0$) are removed. The full spatial distribution of the magnetosheath data are shown in Figure 1.

2.2 Wind Input Dataset

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 dataset used here (September 2nd 2015 to January 1st 2023). Key parameter (KP) 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 second cadence. Due to the difficulty involved with spacecraft intercalibration data from other L1 monitors are not included in this study (King, 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. Missing data is linearly interpolated and flagged so they can be excluded if necessary. The precise windows of time in the Wind dataset that are used as input to predict each MMS target heavily influence the performance of the optimized algorithm; therefore these and other parameters pertaining to the exact construction of the dataset are optimized through hyperparameter search (see Section 3.2).

3 Algorithm Methodology

3.1 Network Architecture

The overall architecture selected for the algorithm is similar to that utilized to construct PRIME (Probabilistic Regressor for Input to the Magnetosphere Estimation), an algorithm that predicts the solar wind near the Earth using data from the Wind spacecraft at L1 (O’Brien et al., 2023). The Bayesian recurrent 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 so-

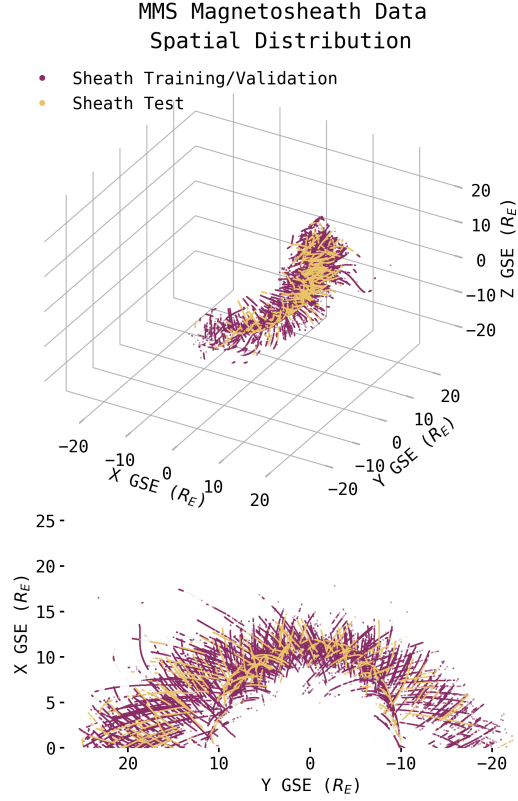


Figure 1. 3D spatial distribution of the 117,427 magnetosheath MMS-1 data points split into 80% training/validation (purple) and 20% test (yellow) subsets. Data consists of \vec{B}_{GSM} , \vec{V}_{GSE} , n_i , $T_{i\parallel}$, and $T_{i\perp}$ from September 2nd 2015 to January 1st 2023. Train/validation/test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

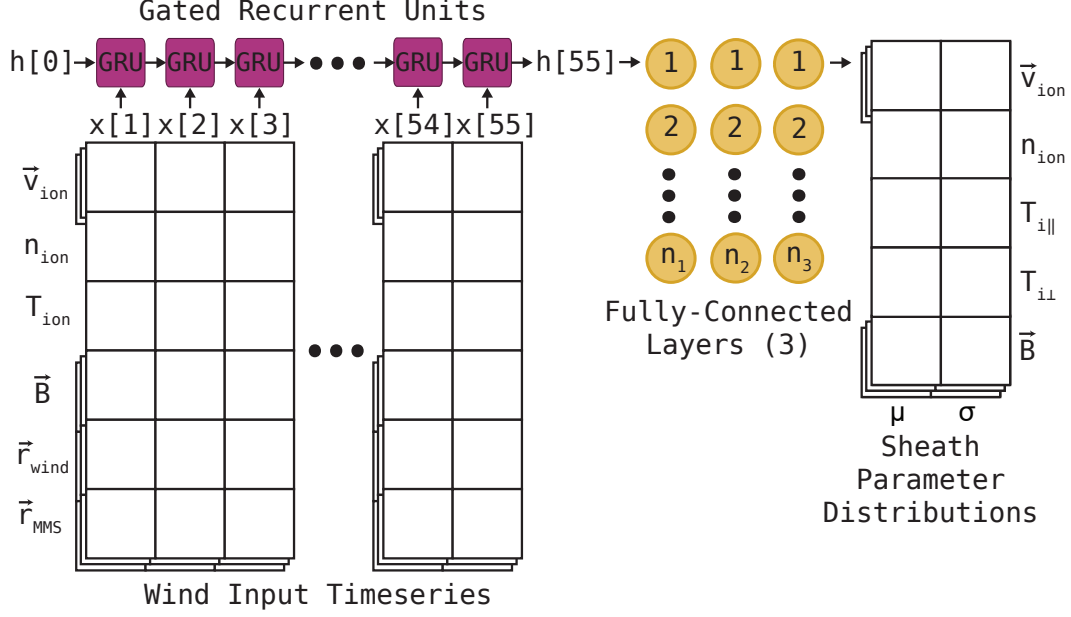


Figure 2. Schematic of PRIME-SH’s neural network architecture, based on the architecture of PRIME (O’Brien et al., 2023). Note that the Gated Recurrent Unit (GRU) sequence feeds into a Fully Connected Neural Network (FCNN) in order to output a mean and variance for each desired parameter instead of a single value. 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.

lar 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 fed into fully connected neural network (FCNN) layers (See Bebis and Georgiopoulos (1994)). The last layer of neurons are taken to be the mean and variance of a Gaussian probability distribution for each parameter rather than single scalar values (Nix & Weigend, 1994; Lakshminarayanan et al., 2017). 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) (Matheson & Winkler, 1976; Hersbach, 2000). 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 continuous rank probability score is given by

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

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 continuous rank probability score is desirable as a loss function because it more 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 μ and variance σ^2 the CRPS is given by

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

(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 (μ s) and variances (σ s) defining a Gaussian probability distribution for each parameter. During training the CRPS over all nine parameters in the target dataset is averaged with equal weight assigned to all parameters.

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, unlike the two step process often utilized for training neural networks. First, the optimal length, lead time, and composition of the input timeseries dataset is determined (the dataset hyperparameter search). Then the algorithm hyperparameters are systematically varied 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.

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 to 60 minutes to get from L1 to the 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 the 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 dataset before overfitting is taken as optimal. When training these test models and for any time a model is trained, the input/target datasets are split into 60% training, 20% validation, and 20% test subsets. Since temporally adjacent entries in the input dataset are almost

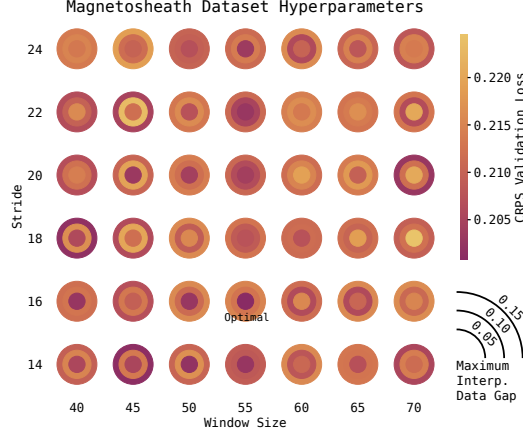


Figure 3. Results from dataset optimization trials over timeseries window (length), stride (lead time), and permitted fraction of interpolated data. Units for window and stride are 100s (the Wind KP data cadence). The optimal set (window 55, stride 16, largest interp. fraction $\leq 5\%$) is shown in darkest green and labelled “optimal”. Loss is given in dimensionless units of parameter interquartile range to ensure comparability of CRPS for each parameter.

entirely overlapping, randomly assigning input/target pairs to each subset results in significant data leakage. To avoid this, the full dataset is split into independent blocks four times the length of the timeseries window used as input (i.e. for a window size of 50 measurements/ ~ 1 hour 20 minutes, the dataset is split into chunks of length 200 measurements/ ~ 5 hours 20 minutes) and those blocks are then assigned to each subset in order to achieve a 60%-20%-20% train-validation-test split. 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 dataset from the search are shown in Figure 3.

The optimal window size is 55 measurements ($\sim 5,500$ seconds, ~ 1 hour 32 minutes), the optimal stride/lead time is 16 measurements ($\sim 1,600$ seconds, ~ 27 minutes). That is to say, for an MMS measurement at time t , the input timeseries from Wind runs from time $t-5,500s-1,600s \approx t-120min$ to time $t-1,600s \approx t-27min$. The largest data gap that can be interpolated over is 4.6 minutes ($\leq 5\%$ of the input window).

Once the optimal dataset 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 dataset, 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 (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 dataset for 20 epochs (the maximum before the loss on the validation dataset starts to increase).

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 pass-

| | Dataset HP Test | Canonical Algorithm | HP Range |
|------------------|-----------------|---------------------|-----------------------------------|
| GRU Layer | 192 | 544 | 128-640 |
| FCNN Layer 1 | 352 | 224 | 128-640 |
| FCNN Layer 2 | 48 | 64 | 16-128 |
| FCNN Layer 3 | N/A | 80 | 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 | Adamax | Adam, Adamax, Adagrad |
| Learning Rate | 10^{-4} | 10^{-4} | 10^{-3} , 10^{-4} , 10^{-5} |

Table 1. Detailed layer sizes and architecture parameters for the test version of PRIME-SH used to optimize the dataset 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).

ing 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 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 dataset used to optimize it, these three algorithms were included.

4 Results

4.1 Statistical Performance

PRIME-SH’s performance is evaluated on the test dataset (not seen by the algorithm at any point during training) by calculating the CRPS between its predictions and the test dataset. 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. Joint distributions of the MMS test dataset with PRIME-SH’s predictions are shown in Figure 4 and Table 2, along with the dynamic pressure calculated from PRIME-SH’s outputs.

PRIME-SH on average predicts plasma parameters (\vec{v} , n_i , $T_{i\perp}$, and $T_{i\parallel}$) slightly more accurately than magnetic field parameters. This is possibly due to the fact that fluctuations in magnetic field happen at a shorter timescale than those in the plasma,

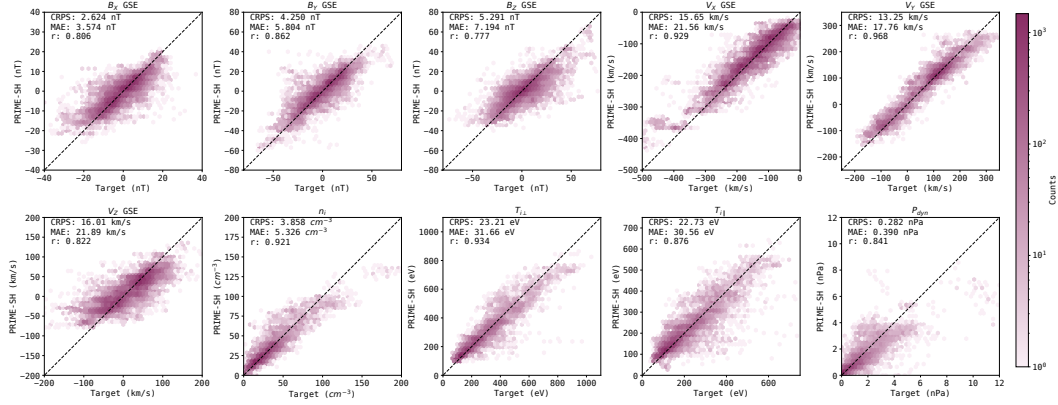


Figure 4. Joint distributions of MMS-1 data (x axis) with predicted parameters from PRIME-SH (y axis). CRPS, the mean absolute error (MAE), and Pearson’s r correlation coefficient for each parameter shown in the top left of each distribution. The MAE is calculated between the peaks of its predicted distributions and each MMS observation (thereby throwing away uncertainty information). A perfect prediction corresponds to the line $y = x$, plotted over-top of each distribution for convenience.

| Parameter | PRIME-SH CRPS | PRIME-SH MAE | PRIME-SH r |
|------------------|--|---|------------|
| B_x GSM | 2.62nT (0.293 σ) | 3.57nT (0.400 σ) | 0.806 |
| B_y GSM | 4.25nT (0.249 σ) | 5.80nT (0.340 σ) | 0.862 |
| B_z GSM | 5.29nT (0.329 σ) | 7.19nT (0.447 σ) | 0.777 |
| V_x GSE | 15.65km/s (0.203 σ) | 21.56km/s (0.280 σ) | 0.929 |
| V_y GSE | 13.25km/s (0.128 σ) | 17.76km/s (0.171 σ) | 0.968 |
| V_z GSE | 16.01km/s (0.304 σ) | 21.89km/s (0.416 σ) | 0.822 |
| n_i | 3.86cm ⁻³ (0.180 σ) | 5.326cm ⁻³ (0.248 σ) | 0.921 |
| $T_{i\perp}$ | 23.21eV (0.154 σ) | 31.66eV (0.210 σ) | 0.934 |
| $T_{i\parallel}$ | 22.73eV (0.198 σ) | 30.56eV (0.266 σ) | 0.876 |
| P_{dyn} | 0.282nPa (0.249 σ) | 0.390nPa (0.345 σ) | 0.841 |

Table 2. Performance of PRIME-SH on the MMS test dataset across continuous rank probability score (CRPS, Equation 1), mean absolute error (MAE), and Pearson’s r correlation coefficient (also shown in Figure 4). CRPS is given in the units of each parameter as well as dimensionless units of standard deviations of each parameter in the MMS training dataset to facilitate comparison between each parameter.

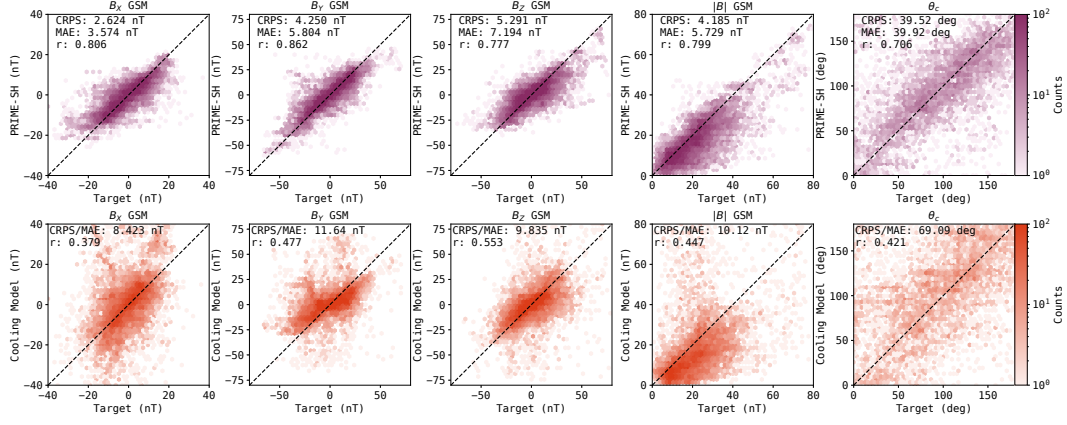


Figure 5. Joint distributions of MMS-1 data (x axis) with predicted parameters from PRIME-SH (top, purple) and the Cooling et al. (2001) model (bottom, orange). CRPS, the mean absolute error (MAE), and Pearson’s r correlation coefficient for each parameter shown in the top left of each distribution. For PRIME-SH, the MAE is calculated between the peaks of its predicted distributions and each MMS observation (thereby throwing away uncertainty information). For the Cooling et al. (2001) model, the CRPS and MAE are equivalent. A perfect prediction corresponds to the line $y = x$, plotted overtop of each distribution for convenience.

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 Figure 4, only some amount of regression to the mean in the most extreme values of V_X and n_i (and therefore in P_{dyn} as well). 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 0.227σ and PRIME’s average CRPS of 0.214σ), 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.

To gain a better sense of the accuracy of PRIME-SH’s predictions in a statistical sense, its outputs are compared to those from a popular magnetosheath magnetic field model (Cooling et al., 2001). The model derived in Cooling et al. (2001) is analytical; essentially the interplanetary magnetic field is “draped” over the Shue et al. (1998) axisymmetric conic section magnetosheath model. For each MMS observation in the target dataset, the solar wind conditions from the OMNI dataset (King & Papitashvili, 2020) are used to determine the magnetopause shape, and the IMF from the OMNI database is “draped” across the boundary. The three component magnetic field at MMS’s location is then compared to the actual value measured by MMS-1 FGM. In Figure 5 the magnetic field values, magnitude, and clock angle predicted by PRIME-SH and the Cooling et al. (2001) model are compared to the MMS-1 test dataset to illustrate each model’s statistical accuracy. The Cooling et al. (2001) model does not have uncertainties associated with it, therefore the MAE and CRPS reduce to the same form and number (Hersbach, 2000); both are provided for PRIME-SH’s outputs so that all comparisons can be made.

For each component in the magnetic field, the field magnitude, and the clock angle $\theta_C = \tan^{-1}(B_Y/B_Z)$, 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 system-

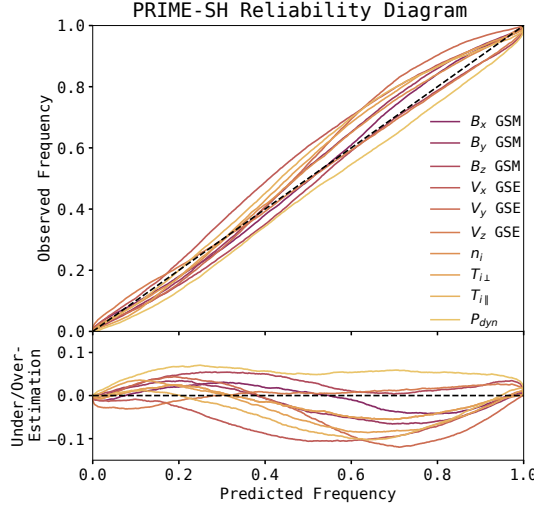


Figure 6. Reliability diagram constructed from PRIME-SH’s outputs on the test dataset for each parameter. Shown versus the predicted frequency of the observation from PRIME-SH are the absolute value of the observed frequency (top) and the deviation from perfect reliability (bottom). For the bottom plot, a given parameter being over (under) the line by an amount corresponds to PRIME-SH over (under) predicting the frequency by that amount.

atic error in the Cooling et al. (2001) model’s outputs for $|B|$, which overpredicts $|B|$ especially when $|B|$ is low. 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 must be separately validated in Section 4.2.1.

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 μ_i, σ_i with $i = 1, \dots, N$ are defined as $\eta_i = (y_{obs,i} - \mu_i)/(\sqrt{2}\sigma_i)$. The probability density of a given Gaussian forecast is therefore $\Phi_i = \frac{1}{2}[erf(\eta_i) + 1]$, allowing the reliability diagram to be constructed from the empirical cumulative distribution of Φ_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 dataset for each parameter and presented in Figure 6.

PRIME-SH is not perfectly reliable (its reliability diagram does not exactly follow the dashed line in Figure 6); it generally tends to overestimate the likelihood of unlikely events, and underestimate the likelihood of likely events. With the exception of V_Z , B_Z , and P_{dyn} , 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 dataset (Kull & Flach, 2015). The largest departures from perfect calibration are observed in V_Y GSE (predicts events that occur with $p = 0.833$ as occurring with $p = 0.713$), V_X GSE (predicts events that occur with $p = 0.662$ as occurring with $p = 0.556$), and T_{\parallel} (predicts events that occur with $p = 0.741$ as occurring with $p = 0.638$). On average PRIME-SH is reliable to within 1.2% with a maximum difference 12% (calculated $p_{obs} -$

p_{pred}). This is more reliable than its progenitor algorithm PRIME and other probabilistic prediction algorithms for other 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 in a statistical sense, it is also important to verify that a model can reproduce expected physics. This is especially important for neural network models, which can relatively easily overfit and reproduce a dataset’s noise rather than the underlying data representation or physics. In the following sections PRIME-SH’s outputs for real and 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 diverts 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 the same input data. The grid is chosen to lie in the GSE X-Y plane with a grid scale of $0.1R_E$. All grid cells inside the Shue 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 input data is chosen to be a 400km/s solar wind only in the GSE X direction with otherwise average solar wind conditions from the Wing L1 dataset: $|B| = 5.34\text{nT}$, $V_X = -400\text{km/s}$, $V_Y = 0\text{km/s}$, $V_Z = 0\text{km/s}$, $n_i = 7.12\text{cm}^{-3}$, and $v_{th} = 34.9\text{km/s}$. In order to investigate whether PRIME-SH is capable of draping, conditions on the grid are calculated for four different IMF orientations: one radial toward the Earth (cone angle 0°), one downward (cone angle -90°), one duskward (cone angle $+90^\circ$) and one radial away from the Earth (cone angle 180°). Shown in Figure 7 are these four grids, with the sheath magnetic field streamlines plotted in black arrows and the uncertainty in B_Z PRIME-SH predicts in each cell in color.

As can be seen in Figure 7, 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 $\pm 90^\circ$ the magnetic field piles up at the nose of the magnetopause, more than it does for radial IMF. The predicted uncertainty in B_Z is plotted in color to investigate whether PRIME-SH captures the physics of quasiparallel shocks. In collisionless shocks, if the upstream magnetic field is roughly parallel to the shock normal particles can be reflected at the shock boundary (Kennel, 1986). This causes “bubbles” to form upstream of the shock in a region called the foreshock, and causes turbulence in the plasma and magnetic field downstream of the shock (Burgess et al., 2005). Downstream of the quasiparallel shock in Figure 7 (cone angle 0° and 180°), the uncertainty in the magnetic field predicted by PRIME-SH is larger than the uncertainty it predicts downstream of the quasiperpendicular shock. This is likely due to the fact that it becomes more difficult to accurately predict the magnetic field in a turbulent environment, as the magnetometer is sampling from a noisy distribution. This is reflected in PRIME-SH’s predicted probability distributions broadening downstream of the quasiparallel shock.

4.2.2 Stagnation Point

As the solar wind plasma diverts and is slowed around the magnetopause, a region known as the stagnation point is thought to develop 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,

B Field Draping and Uncertainty for Four IMF Cone Angles

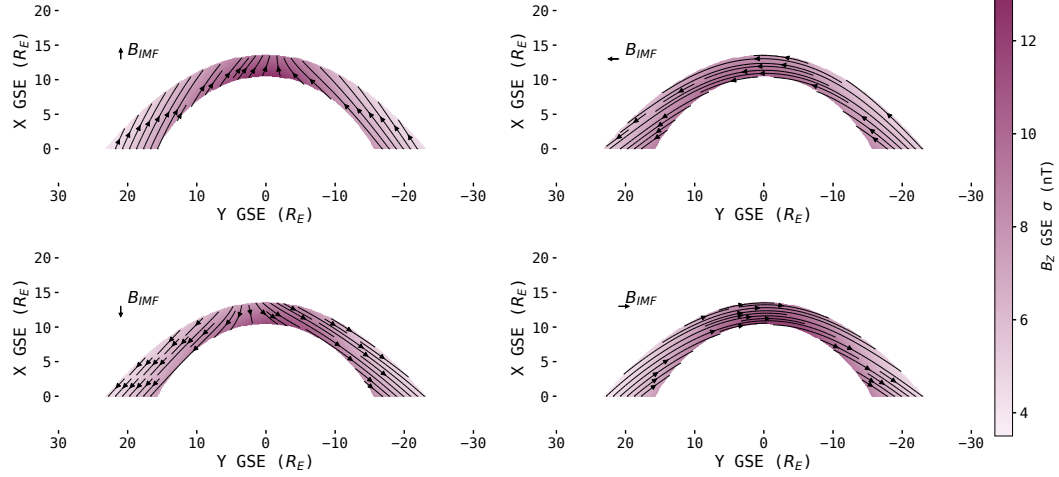


Figure 7. Magnetosheath conditions output by PRIME-SH using synthetic data at four different IMF cone angles (Shown with arrows in top left). Plasma conditions are average conditions from the input dataset, magnetic field magnitude is $5.34nT$ (the average magnitude from the input dataset). Shown in color is the uncertainty in B_z GSM, and the arrows are B_x and B_y GSM field lines.

but MHD theory predicts that the stagnation point should deflect downward for solar wind flows with low Alfvén Mach numbers (Russell et al., 1981). Here PRIME-SH is used to assemble predictions on more grids of the same configuration as Section 4.2.1, however this time the Alfvén Mach number of the synthetic dataset is varied from $M_A = 4$ to $M_A = 16$ (the solar wind typically has $M_A \approx 10$). The density and velocity are held the same ($n_i = 7.12$, $V_X = -400km/s$) and the magnetic field is kept at a 45° Parker spiral cone angle as its magnitude is decreased in steps from $12nT$ to $2.4nT$ to yield the four Alfvén Mach numbers. Shown in Figure 8 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 7, 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 downward as predicted by MHD theory and simulations. 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 dawn-dusk asymmetry in the flow velocity maps produced by PRIME-SH, visible especially at higher Alfvén Mach number. This could be due to biases in MMS-1’s orbit showing up in PRIME-SH’ outputs, as the asymmetry does not appear in MHD simulations of the magnetosheath.

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 \vec{n} , the plasma flow velocity to be \vec{v} , the plasma mass density to be ρ , the thermal pressure to be P , the specific heat

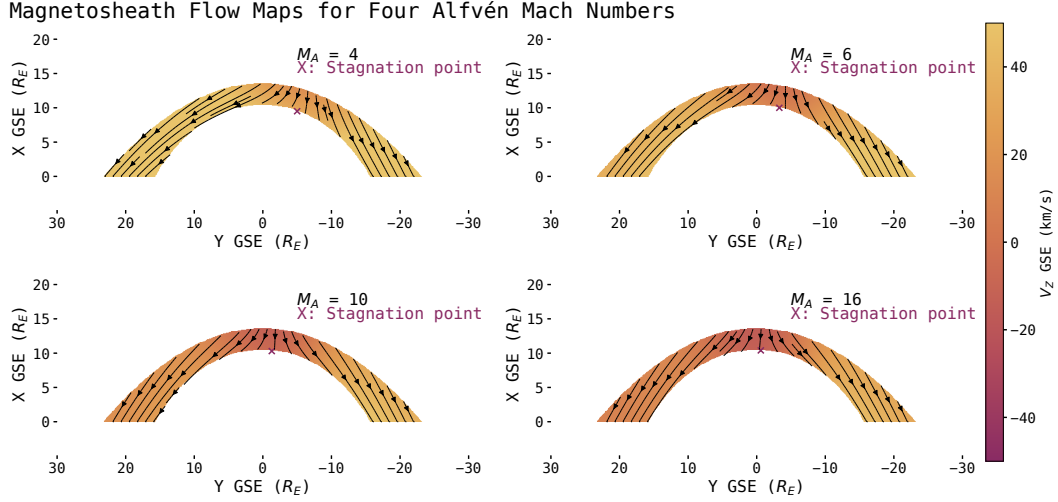


Figure 8. Magnetosheath conditions output by PRIME-SH using synthetic data at four different Alfvén Mach numbers ($M_A = 4, 6, 10, 16$). Plasma conditions are average conditions from the input dataset, magnetic field is a typical Parker spiral orientation whose magnitude is changed to result in the four Alfvén Mach numbers. Shown in color is the Z 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.

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

$$[\rho \vec{u} \cdot \vec{n}] = 0 \quad (3)$$

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

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

Energy conservation can be written:

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

(Kallenrode, 2010).

None of these conditions are explicitly enforced during training, but they are part of the underlying physics PRIME-SH ought to be representing. To validate that PRIME-SH obeys these conservation laws, a range of synthetic solar wind conditions with densities ranging from $1cm^{-3}$ to $50cm^{-3}$ with $V_{GSE} = -400km/s$, $\vec{B} = (-4nT)\vec{x} + (-4nT)\vec{y}$, and $v_{th} = 34.9km/s$ are initialized and used to generate predictions just behind the Jelínek et al. (2012) bow shock nose. This range was chosen to reflect the full range of densities from the input dataset, which results in better coverage of the range of the three upstream fluxes observed than varying other conditions such as velocity. Equations 3, 4, and 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, 4, and 5 to obtain uncertainties for the downstream fluxes as well. The downstream fluxes are plotted as a function of upstream fluxes in Figure 9.

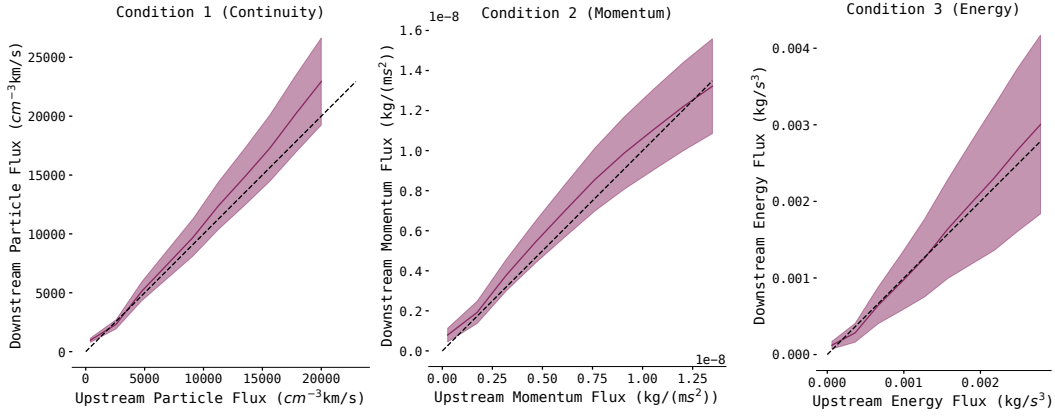


Figure 9. Particle, momentum, and energy fluxes calculated across a range of synthetic input conditions roughly corresponding to the range of the training dataset. Fluxes are calculated just upstream of the bow shock nose (using the input data) and just downstream (using PRIME-SH’s outputs), and uncertainties 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.

Perfect conservation of each flux is represented by the dashed lines in Figure 9. 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 the 1σ uncertainty bounds for each quantity. The points of largest fractional difference between upstream and downstream fluxes occur for the smallest fluxes (when $n_{up} = 1\text{cm}^{-3}$), which happens relatively infrequently in the input dataset. 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.

5 Conclusions

A Bayesian recurrent neural network is trained to predict MMS-1 observations of the Earth’s magnetosheath given timeseries input measured by the Wind spacecraft at L1. This algorithm, called PRIME-SH in reference to its 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 dataset of MMS-1 data not used during training (Average CRPS 0.227σ). The uncertainties predicted by PRIME-SH are shown to be reliable to within 1.2% with a maximum difference 12% through a comparison to the test dataset. Additionally, PRIME-SH more accurately predicts the magnetic field magnitude and direction measured by MMS-1 than the frequently used Cooling et al. (2001) magnetic field draping model. 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 for several different orientations of the upstream magnetic field, and that uncertainties produced by PRIME-SH downstream of quasiparallel shock regions are slightly larger than

those downstream of quasiperpendicular shock regions. 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 downward with decreasing Alfvén Mach number as predicted by MHD theory (Russell et al., 1981). PRIME-SH is also shown to conserve particle/mass flux, momentum flux, and energy flux within 1σ uncertainty across the bow shock for the range of input parameters it is trained on. 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 into the magnetosheath.

PRIME-SH’s rapid and straightforward generation of magnetosheath conditions for real or synthetic inputs at L1 enables large scale statistical studies of the bow shock and magnetosheath’s role in mediating energy transport between the solar wind and Earth’s magnetosphere. The fact that it provides uncertainties also allows confidence intervals to be assigned to these studies. For example, it has been argued using 3 month cadence averaged THEMIS data that the solar wind electric field and Poynting flux that can be driven across the bow shock into the magnetosheath saturate at their highest values (Pulkkinen et al., 2016), a conclusion that can be investigated in more detail with a more complete dataset using PRIME-SH. It has been argued that the solar wind alone controls the rate of energy transfer across the magnetopause due to rearrangements of the magnetosheath flow (Lopez, 2016), which can be investigated using PRIME-SH as a kind of spatio-temporal inversion algorithm. In short, PRIME-SH is an accurate and lightweight magnetosheath prediction algorithm that offers functionality no other magnetosheath prediction algorithm does, and enables new statistical and event-based studies of the magnetosheath.

Acknowledgments

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 dataset preparation, algorithm development, and analysis presented in this paper are available at <https://github.com/connor-obrien888/primesh>. 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

- Anderson, J. L. (1996, July). A Method for Producing and Evaluating Probabilistic Forecasts from Ensemble Model Integrations. *Journal of Climate*, 9(7), 1518–1530. Retrieved 2023-07-21, from [http://journals.ametsoc.org/doi/10.1175/1520-0442\(1996\)009<1518:AMFPAE>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2) doi: 10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2
- Axford, W. (1964, January). Viscous interaction between the solar wind and the earth’s magnetosphere. *Planetary and Space Science*, 12(1), 45–53. Retrieved 2021-11-28, from <https://linkinghub.elsevier.com/retrieve/pii/0032063364900674> doi: 10.1016/0032-0633(64)90067-4
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016, July). *Layer Normalization*. arXiv. Retrieved 2023-07-18, from <http://arxiv.org/abs/1607.06450> (arXiv:1607.06450 [cs, stat])
- Bebis, G., & Georgiopoulos, M. (1994, October). Feed-forward neural networks. *IEEE Potentials*, 13(4), 27–31. Retrieved 2023-07-19, from <http://ieeexplore.ieee.org/document/329294/> doi: 10.1109/45.329294
- Borovsky, J. E. (2018, October). 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. Retrieved 2023-03-13, from

- 526 <https://linkinghub.elsevier.com/retrieve/pii/S1364682617300159>
 527 doi: 10.1016/j.jastp.2017.03.014
- 528 Borovsky, J. E. (2021, February). Is Our Understanding of Solar-
 529 Wind/Magnetosphere Coupling Satisfactory? *Frontiers in Astronomy*
 530 *and Space Sciences*, 8, 634073. Retrieved 2023-04-13, from [https://](https://www.frontiersin.org/articles/10.3389/fspas.2021.634073/full)
 531 www.frontiersin.org/articles/10.3389/fspas.2021.634073/full doi:
 532 10.3389/fspas.2021.634073
- 533 Bröcker, J., & Smith, L. A. (2007, June). Increasing the Reliability of Reliability
 534 Diagrams. *Weather and Forecasting*, 22(3), 651–661. Retrieved 2023-07-24,
 535 from <https://journals.ametsoc.org/doi/10.1175/WAF993.1> doi: 10.1175/
 536 WAF993.1
- 537 Burch, J. L., Moore, T. E., Torbert, R. B., & Giles, B. L. (2016, March). Magne-
 538 topheric Multiscale Overview and Science Objectives. *Space Science Reviews*,
 539 199(1-4), 5–21. Retrieved 2021-12-01, from [http://link.springer.com/10](http://link.springer.com/10.1007/s11214-015-0164-9)
 540 [.1007/s11214-015-0164-9](http://link.springer.com/10.1007/s11214-015-0164-9) doi: 10.1007/s11214-015-0164-9
- 541 Burgess, D., Lucek, E. A., Scholer, M., Bale, S. D., Balikhin, M. A., Balogh,
 542 A., ... Walker, S. N. (2005). Quasi-parallel Shock Structure and Pro-
 543 cesses. In G. Paschmann, S. J. Schwartz, C. P. Escoubet, & S. Haaland
 544 (Eds.), *Outer Magnetospheric Boundaries: Cluster Results* (Vol. 20, pp.
 545 205–222). Berlin/Heidelberg: Springer-Verlag. Retrieved 2024-01-24, from
 546 http://link.springer.com/10.1007/1-4020-4582-4_7 (Series Title: Space
 547 Sciences Series of ISSI) doi: 10.1007/1-4020-4582-4_7
- 548 Camporeale, E., & Carè, A. (2021). ACCRUE: ACCURATE AND RELI-
 549 ABLE UNCERTAINTY ESTIMATE IN DETERMINISTIC MODELS.
 550 *International Journal for Uncertainty Quantification*, 11(4), 81–94. Re-
 551 trieved 2023-01-05, from [http://www.dl.begellhouse.com/journals/](http://www.dl.begellhouse.com/journals/52034eb04b657aea,3ec0b84376cfff3d2,1801e97431c5911b.html)
 552 [52034eb04b657aea,3ec0b84376cfff3d2,1801e97431c5911b.html](http://www.dl.begellhouse.com/journals/52034eb04b657aea,3ec0b84376cfff3d2,1801e97431c5911b.html) doi:
 553 10.1615/Int.J.UncertaintyQuantification.2021034623
- 554 Camporeale, E., Chu, X., Agapitov, O. V., & Bortnik, J. (2019, March). On the
 555 Generation of Probabilistic Forecasts From Deterministic Models. *Space*
 556 *Weather*, 17(3), 455–475. Retrieved 2023-07-21, from [https://onlinelibrary](https://onlinelibrary.wiley.com/doi/abs/10.1029/2018SW002026)
 557 [.wiley.com/doi/abs/10.1029/2018SW002026](https://onlinelibrary.wiley.com/doi/abs/10.1029/2018SW002026) doi: 10.1029/2018SW002026
- 558 Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014, October). *On*
 559 *the Properties of Neural Machine Translation: Encoder-Decoder Approaches*.
 560 arXiv. Retrieved 2023-07-19, from <http://arxiv.org/abs/1409.1259>
 561 (arXiv:1409.1259 [cs, stat])
- 562 Cooling, B. M. A., Owen, C. J., & Schwartz, S. J. (2001, September). Role of the
 563 magnetosheath flow in determining the motion of open flux tubes. *Journal of*
 564 *Geophysical Research: Space Physics*, 106(A9), 18763–18775. Retrieved 2023-
 565 07-12, from <http://doi.wiley.com/10.1029/2000JA000455> doi: 10.1029/
 566 2000JA000455
- 567 Crooker, N. U., Luhmann, J. G., Russell, C. T., Smith, E. J., Spreiter, J. R., &
 568 Stahara, S. S. (1985, April). Magnetic field draping against the dayside magne-
 569 topause. *Journal of Geophysical Research: Space Physics*, 90(A4), 3505–3510.
 570 Retrieved 2024-01-05, from [https://agupubs.onlinelibrary.wiley.com/](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/JA090iA04p03505)
 571 [doi/10.1029/JA090iA04p03505](https://agupubs.onlinelibrary.wiley.com/doi/10.1029/JA090iA04p03505) doi: 10.1029/JA090iA04p03505
- 572 Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive Subgradient Methods for Online
 573 Learning and Stochastic Optimization.
- 574 Dungey, J. W. (1961, January). Interplanetary Magnetic Field and the Au-
 575 roral Zones. *Physical Review Letters*, 6(2), 47–48. Retrieved 2021-11-
 576 28, from <https://link.aps.org/doi/10.1103/PhysRevLett.6.47> doi:
 577 10.1103/PhysRevLett.6.47
- 578 Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005, May). Cali-
 579 brated Probabilistic Forecasting Using Ensemble Model Output Statistics and
 580 Minimum CRPS Estimation. *Monthly Weather Review*, 133(5), 1098–1118.

- Retrieved 2023-05-22, from <http://journals.ametsoc.org/doi/10.1175/MWR2904.1> doi: 10.1175/MWR2904.1
- Hamill, T. M. (1997, December). Reliability Diagrams for Multicategory Probabilistic Forecasts. *Weather and Forecasting*, 12(4), 736–741. Retrieved 2023-07-24, from [http://journals.ametsoc.org/doi/10.1175/1520-0434\(1997\)012<0736:RDFMPF>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2) doi: 10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2
- Hamill, T. M. (2001, March). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, 129(3), 550–560. Retrieved 2023-07-24, from [http://journals.ametsoc.org/doi/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2) doi: 10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2
- Hersbach, H. (2000, October). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting*, 15(5), 559–570. Retrieved 2023-01-23, from [http://journals.ametsoc.org/doi/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2) doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
- Hu, A., Camporeale, E., & Swiger, B. (2022, September). *Multi-Hour Ahead Dst Index Prediction Using Multi-Fidelity Boosted Neural Networks*. arXiv. Retrieved 2023-01-05, from <http://arxiv.org/abs/2209.12571> (arXiv:2209.12571 [physics])
- Huang, S., Li, W., Shen, X., Ma, Q., Chu, X., Ma, D., ... Goldstein, J. (2022, September). Application of Recurrent Neural Network to Modeling Earth's Global Electron Density. *Journal of Geophysical Research: Space Physics*, 127(9). Retrieved 2022-10-25, from <https://onlinelibrary.wiley.com/doi/10.1029/2022JA030695> doi: 10.1029/2022JA030695
- Jelínek, K., Němeček, Z., & Šafránková, J. (2012, May). A new approach to magnetopause and bow shock modeling based on automated region identification. *Journal of Geophysical Research: Space Physics*, 117(A5), 2011JA017252. Retrieved 2024-01-05, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011JA017252> doi: 10.1029/2011JA017252
- Jung, J., K. Connor, H., P. Dimmock, A., Sembay, S., M. Read, A., Soucek, J., ... 5Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic (2024). Mshpy23: a user-friendly, parameterized model of magnetosheath conditions. *Earth and Planetary Physics*, 8(1), 0–0. Retrieved 2023-09-27, from <https://www.eppcgs.org/en/article/doi/10.26464/epp2023065> doi: 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 ed ed.). Berlin Heidelberg: Springer.
- Kennel, C. (1986, January). Quasi-parallel shocks. *Advances in Space Research*, 6(1), 5–16. Retrieved 2024-01-24, from <https://linkinghub.elsevier.com/retrieve/pii/0273117786900037> doi: 10.1016/0273-1177(86)90003-7
- King, J. H. (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. Retrieved 2020-05-26, from <http://doi.wiley.com/10.1029/2004JA010649> doi: 10.1029/2004JA010649
- King, J. H., & Papitashvili, N. E. (2020). *OMNI 1-min Data Set*. NASA Space Physics Data Facility. Retrieved 2023-06-01, from <https://hpde.io/NASA/NumericalData/OMNI/HighResolutionObservations/Version1/PT1M> doi: 10.48322/45BB-8792
- Kingma, D. P., & Ba, J. (2017, January). *Adam: A Method for Stochastic Optimization*. arXiv. Retrieved 2023-07-18, from <http://arxiv.org/abs/1412.6980> (arXiv:1412.6980 [cs])
- Kobel, E., & Flückiger, E. O. (1994, December). A model of the steady

- state magnetic field in the magnetosheath. *Journal of Geophysical Research: Space Physics*, 99(A12), 23617–23622. Retrieved 2024-01-05, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/94JA01778> doi: 10.1029/94JA01778
- Koons, H. C., & Gorney, D. J. (1991, April). A neural network model of the relativistic electron flux at geosynchronous orbit. *Journal of Geophysical Research: Space Physics*, 96(A4), 5549–5556. Retrieved 2024-01-16, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/90JA02380> doi: 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). Cham: Springer International Publishing. Retrieved 2023-07-24, from http://link.springer.com/10.1007/978-3-319-23528-8_5 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-319-23528-8_5
- Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017, November). *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*. arXiv. Retrieved 2023-01-05, from <http://arxiv.org/abs/1612.01474> (arXiv:1612.01474 [cs, stat])
- Lepping, R. P., Acuña, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schatten, K. H., ... Worley, E. M. (1995, February). The WIND magnetic field investigation. *Space Science Reviews*, 71(1-4), 207–229. Retrieved 2021-12-01, from <http://link.springer.com/10.1007/BF00751330> doi: 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>
- Lopez, R. E. (2016, May). 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. Retrieved 2021-11-04, from <http://doi.wiley.com/10.1002/2016JA022556> doi: 10.1002/2016JA022556
- Lyon, J., Fedder, J., & Mobarri, C. (2004, October). The Lyon–Fedder–Mobarri (LFM) global MHD magnetospheric simulation code. *Journal of Atmospheric and Solar-Terrestrial Physics*, 66(15-16), 1333–1350. Retrieved 2024-01-24, from <https://linkinghub.elsevier.com/retrieve/pii/S1364682604001439> doi: 10.1016/j.jastp.2004.03.020
- Matheson, J. E., & Winkler, R. L. (1976, June). Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10), 1087–1096. Retrieved 2023-07-18, from <https://pubsonline.informs.org/doi/10.1287/mnsc.22.10.1087> doi: 10.1287/mnsc.22.10.1087
- 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)* (pp. 55–60 vol.1). Orlando, FL, USA: IEEE. Retrieved 2023-07-25, from <http://ieeexplore.ieee.org/document/374138/> doi: 10.1109/ICNN.1994.374138
- Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lobell, J., ... Gergin, E. (1995, February). SWE, a comprehensive plasma instrument for the WIND spacecraft. *Space Science Reviews*, 71(1-4), 55–77. Retrieved 2021-12-01, from <http://link.springer.com/10.1007/BF00751326> doi: 10.1007/BF00751326
- Olshevsky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzen, S., ... Markidis, S. (2021, October). Automated classification of plasma re-

- gions using 3D particle energy distributions. *Journal of Geophysical Research: Space Physics*, 126(10). Retrieved 2022-03-15, from <http://arxiv.org/abs/1908.05715> (arXiv: 1908.05715) doi: 10.1029/2021JA029620
- O'Malley, T., Bursztain, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Keras-tuner. <https://github.com/keras-team/keras-tuner>.
- O'Brien, C., Walsh, B. M., Zou, Y., Tasnim, S., Zhang, H., & Sibeck, D. G. (2023, September). PRIME: a probabilistic neural network approach to solar wind propagation from L1. *Frontiers in Astronomy and Space Sciences*, 10, 1250779. Retrieved 2023-11-13, from <https://www.frontiersin.org/articles/10.3389/fspas.2023.1250779/full> doi: 10.3389/fspas.2023.1250779
- Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., ... Zeuch, M. (2016, March). Fast Plasma Investigation for Magnetospheric Multi-scale. *Space Science Reviews*, 199(1-4), 331–406. Retrieved 2023-02-23, from <http://link.springer.com/10.1007/s11214-016-0245-4> doi: 10.1007/s11214-016-0245-4
- Powell, K. G., Roe, P. L., Linde, T. J., Gombosi, T. I., & De Zeeuw, D. L. (1999, September). A Solution-Adaptive Upwind Scheme for Ideal Magnetohydrodynamics. *Journal of Computational Physics*, 154(2), 284–309. Retrieved 2024-01-24, from <https://linkinghub.elsevier.com/retrieve/pii/S002199919996299X> doi: 10.1006/jcph.1999.6299
- Pulkkinen, T. I., Dimmock, A. P., Lakka, A., Osmane, A., Kilpua, E., Myllys, M., ... Viljanen, A. (2016, September). Magnetosheath control of solar wind-magnetosphere coupling efficiency. *Journal of Geophysical Research: Space Physics*, 121(9), 8728–8739. Retrieved 2021-10-29, from <https://onlinelibrary.wiley.com/doi/abs/10.1002/2016JA023011> doi: 10.1002/2016JA023011
- Russell, C. T., Anderson, B. J., Baumjohann, W., Bromund, K. R., Dearborn, D., Fischer, D., ... Richter, I. (2016, March). The Magnetospheric Multiscale Magnetometers. *Space Science Reviews*, 199(1-4), 189–256. Retrieved 2021-08-25, from <http://link.springer.com/10.1007/s11214-014-0057-3> doi: 10.1007/s11214-014-0057-3
- Russell, C. T., Zhuang, H., Walker, R. J., & Crooker, N. U. (1981, September). A note on the location of the stagnation point in the magnetosheath flow. *Geophysical Research Letters*, 8(9), 984–986. Retrieved 2023-10-19, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/GL008i009p00984> doi: 10.1029/GL008i009p00984
- Shue, J.-H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G., ... Kawano, H. (1998, August). Magnetopause location under extreme solar wind conditions. *Journal of Geophysical Research: Space Physics*, 103(A8), 17691–17700. Retrieved 2020-05-11, from <http://doi.wiley.com/10.1029/98JA01103> doi: 10.1029/98JA01103
- Soucek, J., & Escoubet, C. P. (2012, June). Predictive model of magnetosheath plasma flow and its validation against Cluster and THEMIS data. *Annales Geophysicae*, 30(6), 973–982. Retrieved 2023-08-21, from <https://angeo.copernicus.org/articles/30/973/2012/> doi: 10.5194/angeo-30-973-2012
- Spreiter, J. R., & Alksne, A. Y. (1969). Plasma flow around the magnetosphere. *Reviews of Geophysics*, 7(1, 2), 11. Retrieved 2020-10-22, from <http://doi.wiley.com/10.1029/RG007i001p00011> doi: 10.1029/RG007i001p00011
- Spreiter, J. R., Summers, A. L., & Alksne, A. Y. (1966, March). Hydromagnetic flow around the magnetosphere. *Planetary and Space Science*, 14(3), 223–253. Retrieved 2020-11-09, from <https://linkinghub.elsevier.com/retrieve/pii/0032063366901243> doi: 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, January). Probabilistic Geomagnetic Storm Forecasting via Deep Learning. *Journal of Geophysical Research: Space Physics*, 126(1), e2020JA028228. Retrieved 2023-07-21, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JA028228> doi: 10.1029/2020JA028228
- Tsyganenko, N. A., Semenov, V. S., & Erkaev, N. V. (2023, November). Data-Based Modeling of the Magnetosheath Magnetic Field. *Journal of Geophysical Research: Space Physics*, 128(11), e2023JA031665. Retrieved 2024-01-03, from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023JA031665> doi: 10.1029/2023JA031665
- Von Alfthan, S., Pokhotelov, D., Kempf, Y., Hoilijoki, S., Honkonen, I., Sandroos, A., & Palmroth, M. (2014, December). Vlasiator: First global hybrid-Vlasov simulations of Earth’s foreshock and magnetosheath. *Journal of Atmospheric and Solar-Terrestrial Physics*, 120, 24–35. Retrieved 2024-01-16, from <https://linkinghub.elsevier.com/retrieve/pii/S1364682614001916> doi: 10.1016/j.jastp.2014.08.012
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed ed.) (No. v. 100). Amsterdam ; Boston: Elsevier/Academic Press.
- Zamo, M., & Naveau, P. (2018, February). Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. *Mathematical Geosciences*, 50(2), 209–234. Retrieved 2023-05-31, from <http://link.springer.com/10.1007/s11004-017-9709-7> doi: 10.1007/s11004-017-9709-7