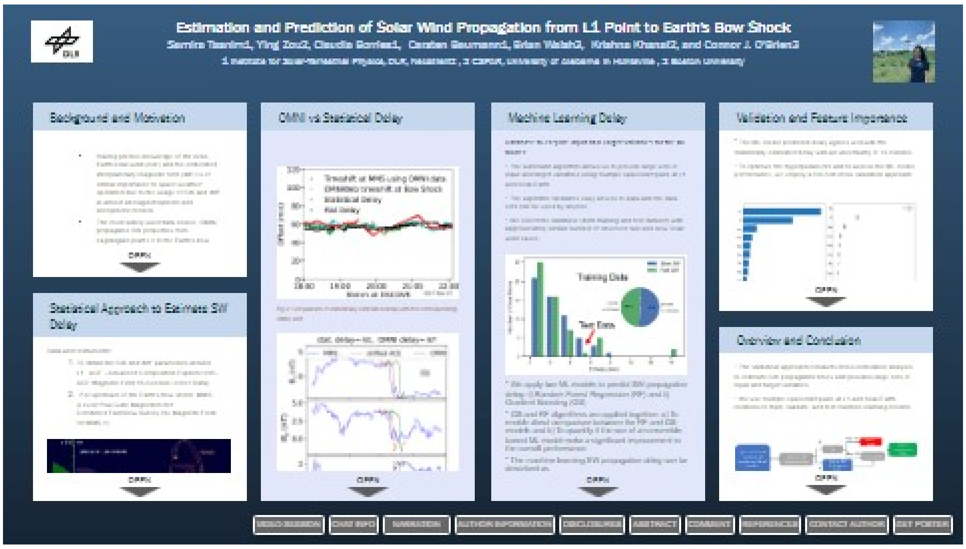
Estimation and Prediction of Solar Wind Propagation from L1 Point to Earth's Bow Shock



Samira Tasnim1, Ying Zou2, Claudia Borries1, Carsten Baumann1, Brian Walsh3, Krishna Khanal2, and Connor J. O'Brien3

1 Institute for Solar-Terrestrial Physics, DLR, Neustrelitz , 2 CSPAR, University of Alabama in Huntsville , 3 Boston University

PRESENTED AT:

BACKGROUND AND MOTIVATION

- Having precise knowledge of the near-Earth solar wind (SW) and the embedded interplanetary magnetic field (IMF) is of critical importance to space weather operation due to the usage of SW and IMF in almost all magnetospheric and ionospheric models.
- The most widely used data source, OMNI, propagates SW properties from Lagrangian point L1 to the Earth's bow shock by estimating the propagation time of the SW. However, the uncertainty of this time can reach ~30 min
- The overarching goal of the project is to deliver machine learning models to specify and forecast near-Earth SW conditions based on spacecraft measurements around L1 by marrying the long history of multi-point SW measurements with the gradient boosting and random forest prediction models in the form of ensemble of decision trees.
- We train the ML model to specify and/or predict the propagation time from L1 monitors to a given location upstream or at the bow shock

STATISTICAL APPROACH TO ESTIMATE SW DELAY

Data and Instruments:

1. To obtain the SW and IMF parameters around L1: ACE - Advanced Composition Explorer (H0 - ACE Magnetic Field 16-Second Level 2 Data)

2. For upstream of the Earth's bow shock: MMS (Level2 Flux Gate Magnetometer Combined Fast/Slow Survey DC Magnetic Field for MMS 1)

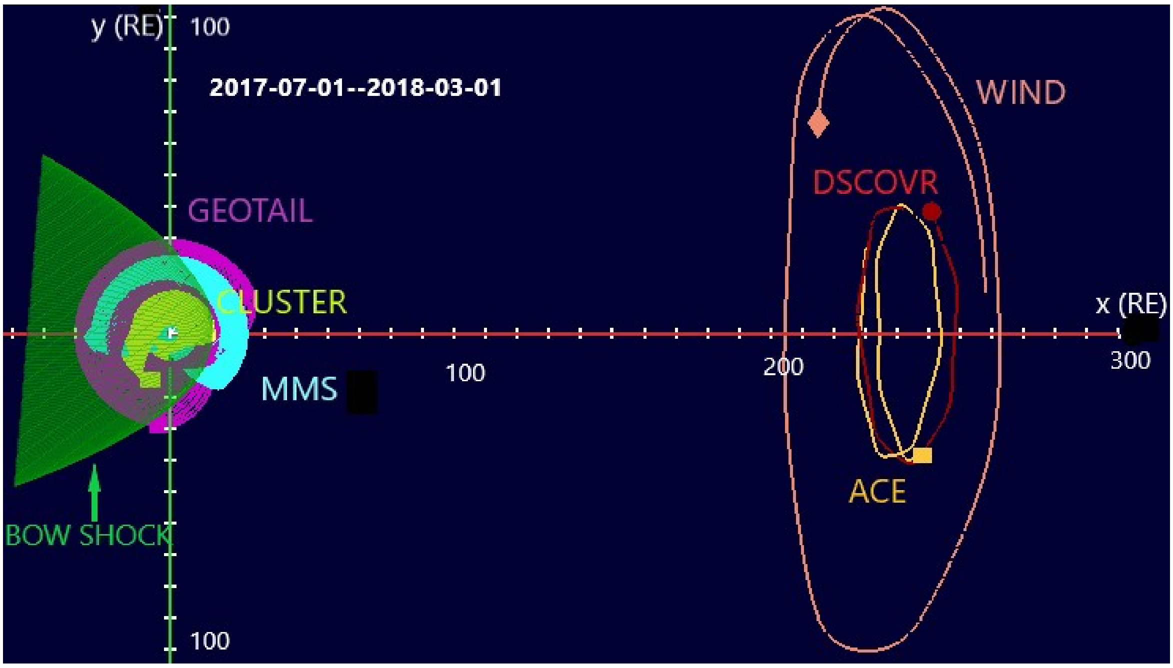


Fig 1: Orbits of ACE, WIND, DSCOVR, GEOTAIL, CLUSTER, and MMS 1

Criterion I: near-Earth monitors (GEOTAIL, CLUSTER, ARTIMES, and MMS) are located at X>15 RE and |Y|< 15 RE

Criterion II: have an ion temperature <1 keV [Mailyan et al., 2008; Case and Wild, 2012]

Criterion III: assume a constraint on highly fluctuating magnetic filed to avoid foreshock conditions

Methodology:

* To trace SW propagation, we perform the analysis on IMF clock angle:

$$\theta = \tan^{-1}\frac{B_y}{B_z}$$

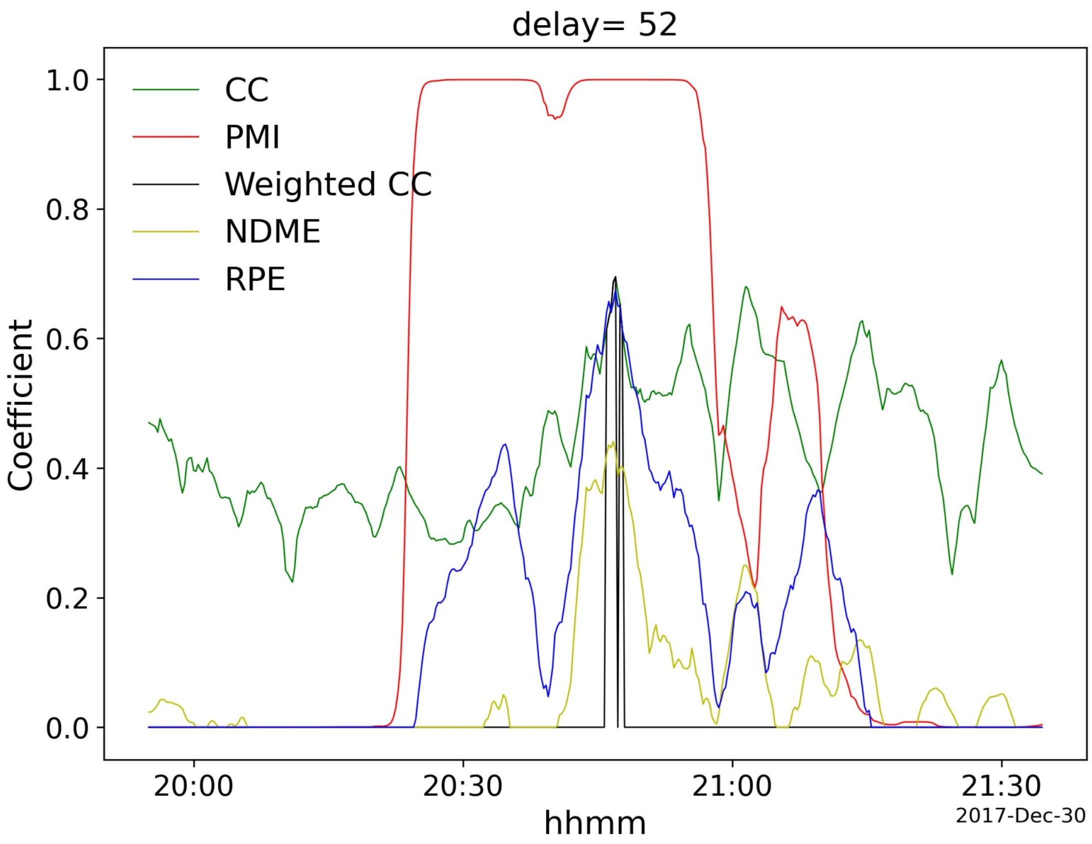


Fig 2: Cross-correlation profile with weighted CC, PMI, and NDME indices

* We segment ACE in 20 minutes window and find the MMS data that best match the ACE data by sliding along 2 hours of data incrementing 1 minute at a time

* To correlate IMF clock angle at L1 and near-Earth and obtaining propagation times, the algorithm computes

1) Cross-correlation (CC) coefficient

2) Plateau-shaped Magnitude Index (PMI)

3) Dimensionless Measures of Average Error (NDME)

* Our analysis uses, Weighted CC = CC*PMI

* when max(CC) > 0.5 and NDME > 0.4

OMNI VS STATISTICAL DELAY

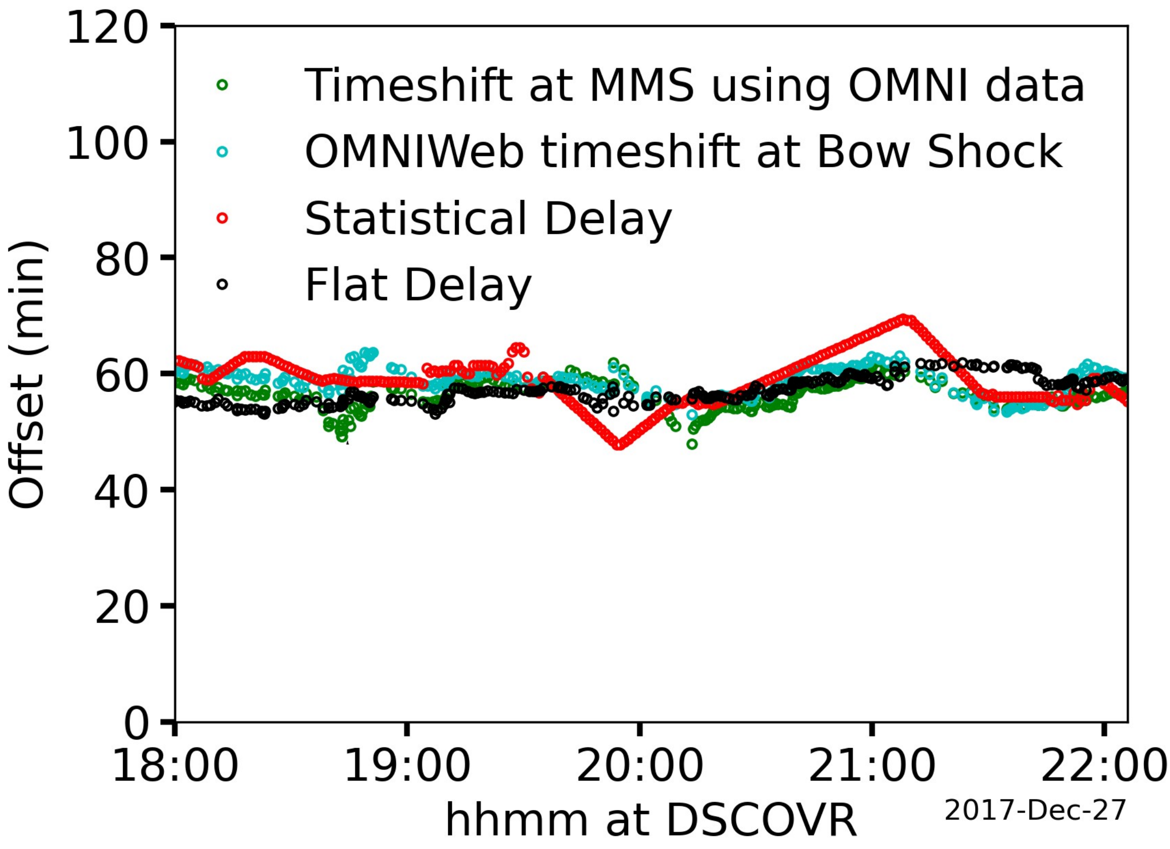


Fig 3: Comparison of statistically estimated delay with the corresponding OMNI shift

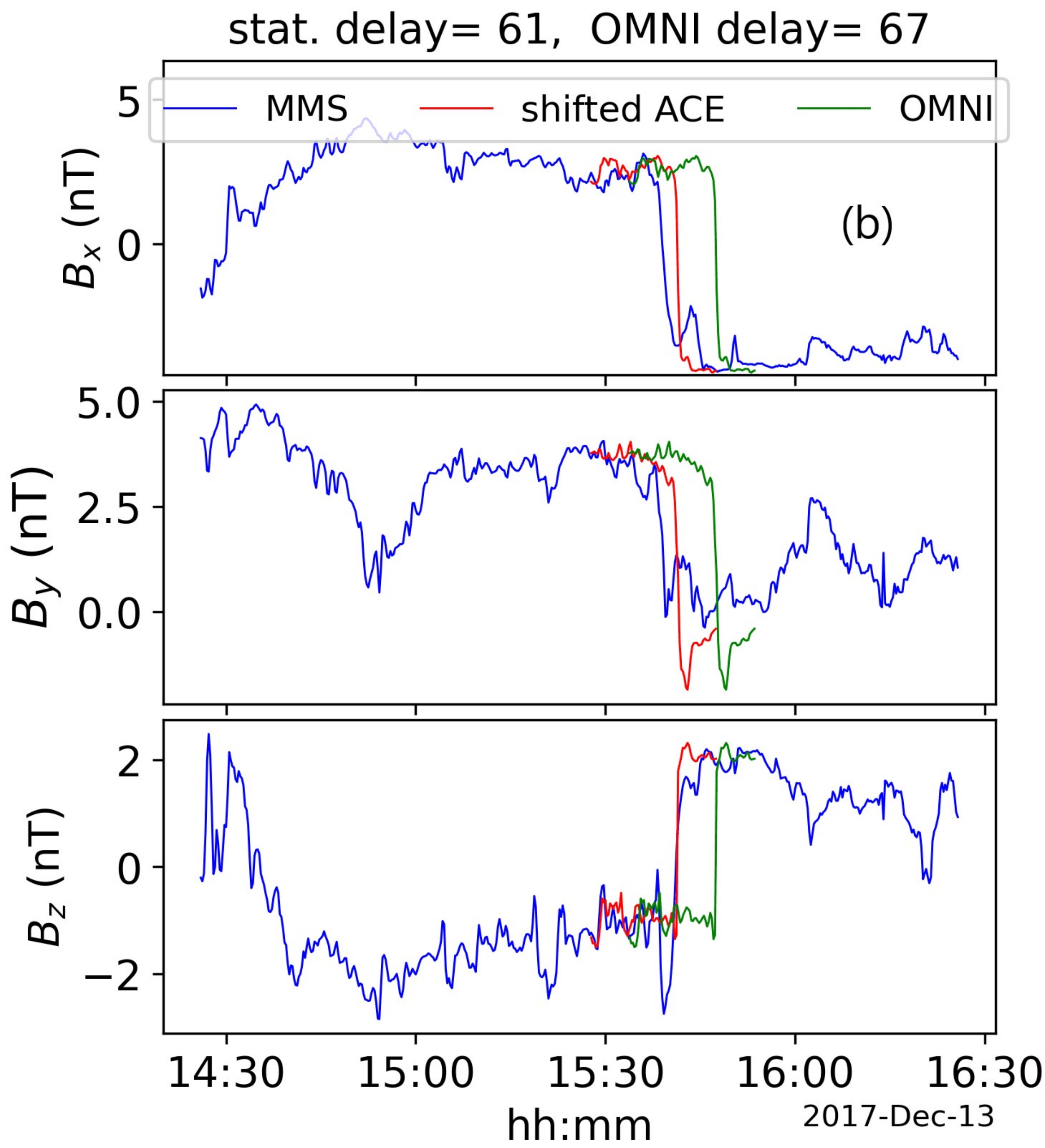


Fig 4: The magnetic field observed by MMS (blue), at ACE lagged by the calculated cross-correlation delay in red and by the OMNI delay in green.

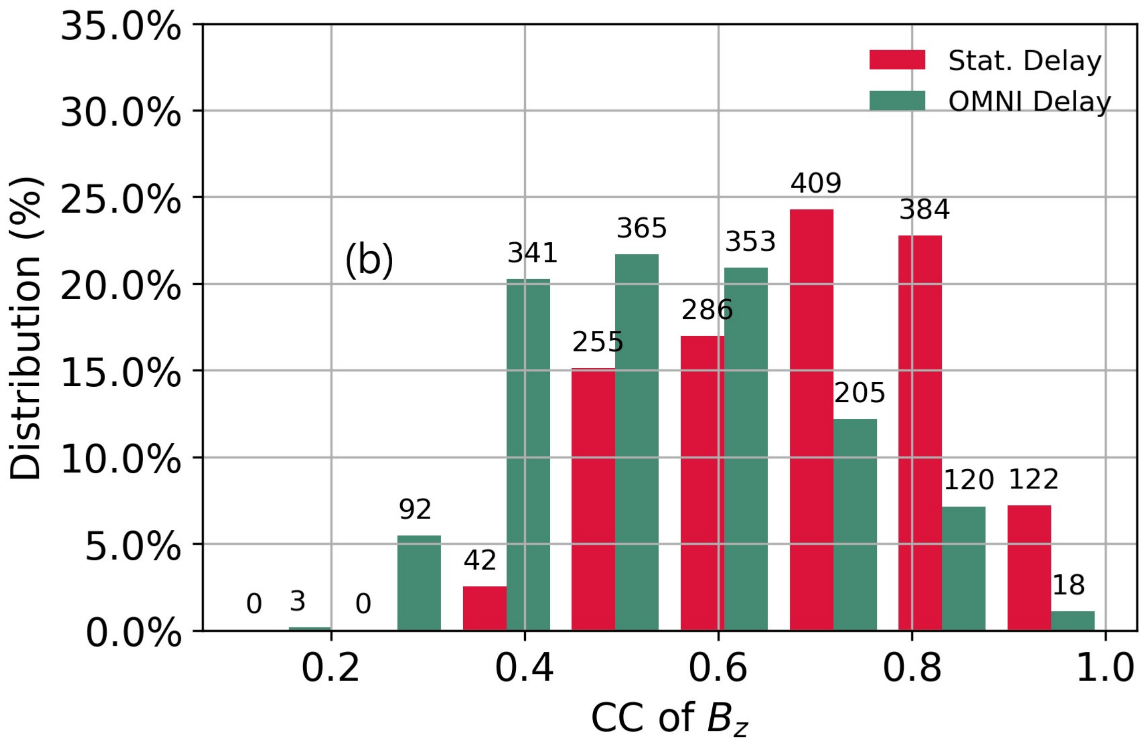


Fig 5: Histograms of the cross correlation coefficients (CC) between DSCOVR and MMS data of B_z

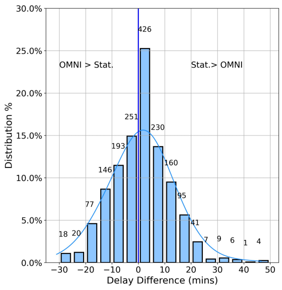


Fig 6: Distribution of the differences between statistical delay and OMNI delay using DSCOVR and MMS data

* Among 1058 case (data points) of B_z, about 87% cases result in large $CC \geq 0.5$ using statistical delay, whereas 60% cases of OMNI provides similar high CC values

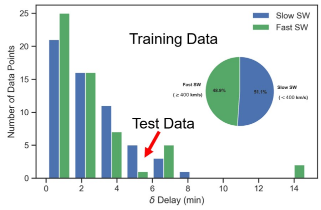
MACHINE LEARNING DELAY

Database to Prepare Input and Target Variables for the ML Model:

* The automatic algorithm allows us to provide large sets of input and target variables using multiple spacecraft pairs at L1 and near-Earth

* The algorithm facilitates easy access to data and the data sets can be used by anyone

* We select the database (both training and test dataset) with approximately similar number of observed fast and slow solar wind cases



* We apply two ML models to predict SW propagation delay: i) Random Forest Regression (RF) and ii) Gradient Boosting (GB)

* GB and RF algorithms are applied together: a) To enable direct comparison between the RF and GB models and b) To quantify if the use of an ensemble-based ML model make a significant improvement to the overall performance

* The machine learning SW propagation delay can be described as

$$\Delta t\_ML=f\_D\ (x)$$

Here f_D describes the ML algorithm trained on the data set $\mathcal{D}$ and x contains the feature

vectors

* We follow Baumann and McCloskey [2021]'s method,where we use Bayesian optimization based on the Gaussian process

Algorithm Default/Optimized	Trees	Max features/Tree	Min samples split	Min sample leaf	Learning rate
Random Forest	150/800	9/3	2/20	1/20	NA
Gradient Boost	150/800	9/3	2/20	1/20	0.01

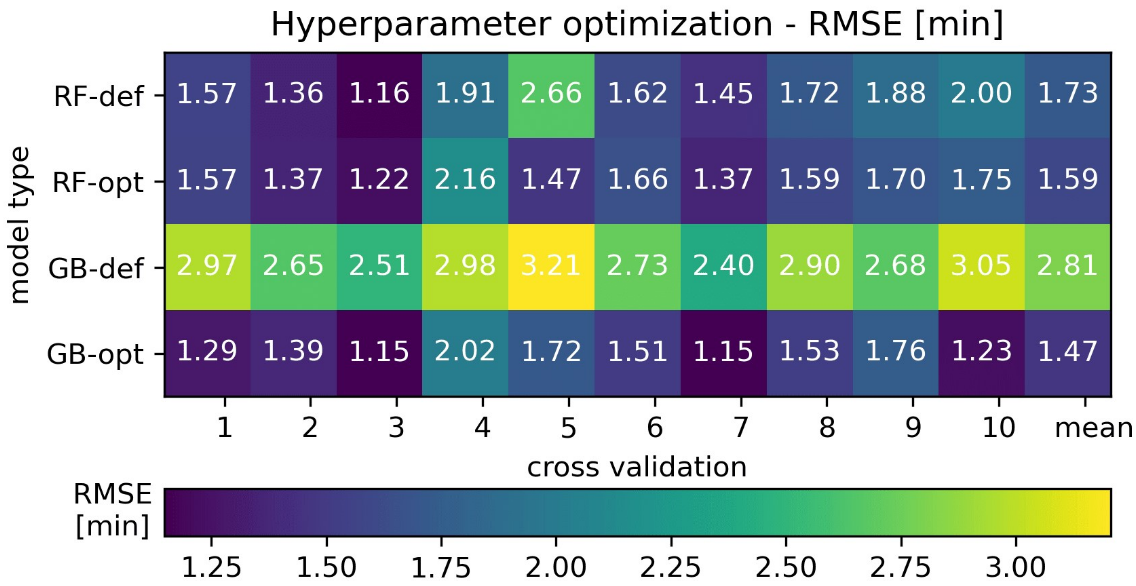


Fig 7: Comparisons of RMSE for the ten-fold cross-validation of default (RF-df) and optimized (RF-opt) random forest with the default (GB-df) and optimized (GB-opt) gradient boosting algorithm

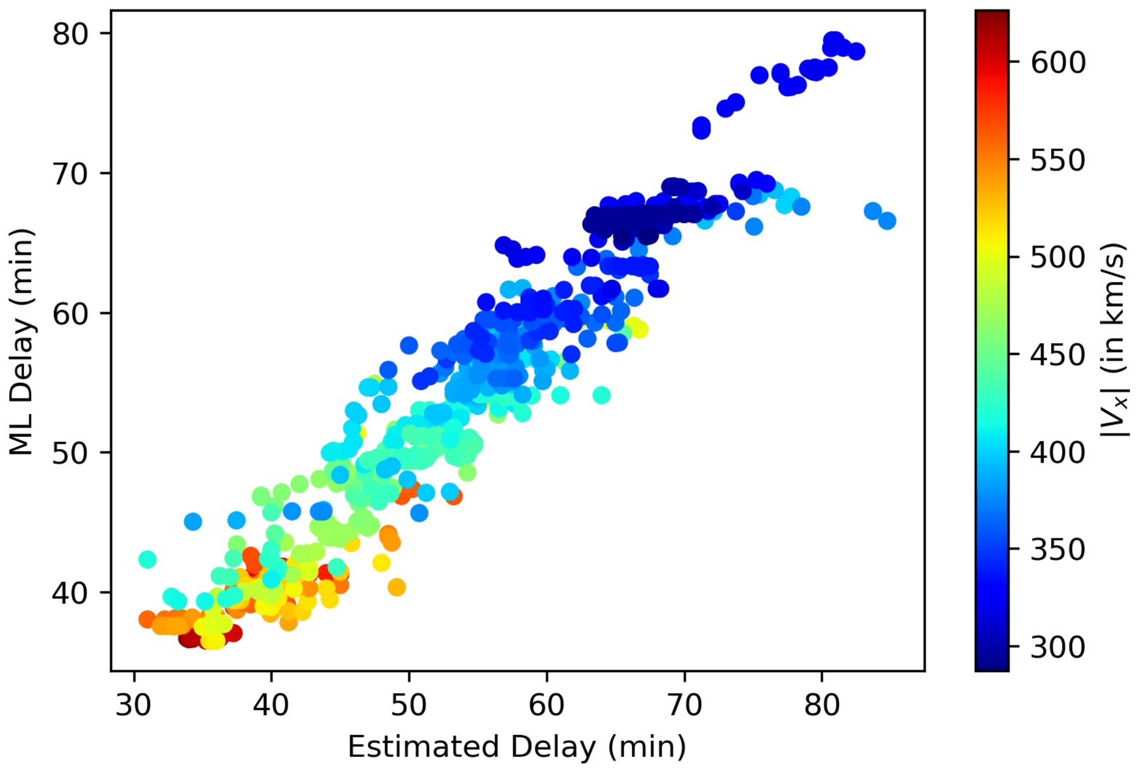


Fig 8: Machine learning model (Gradient Boosting) predicted delay versus statistically estimated delays

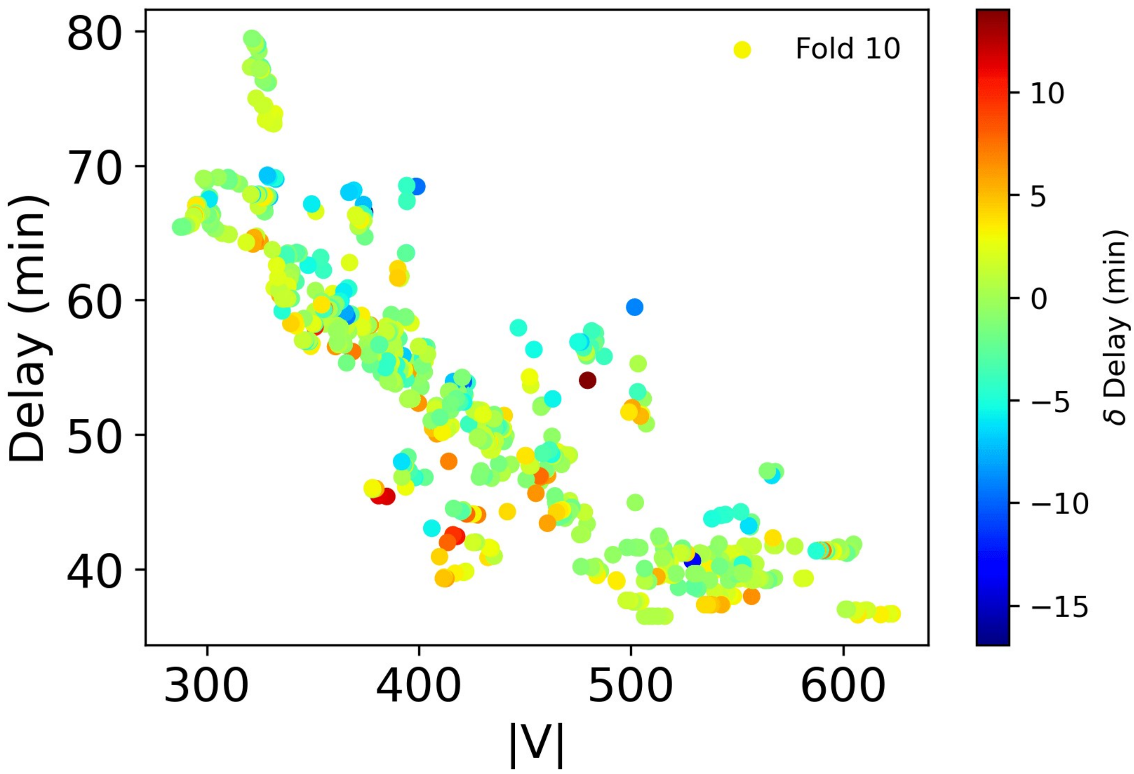


Fig 9: ML delay variation with the magnitude of V_x

VALIDATION AND FEATURE IMPORTANCE

* The ML model predicted delay agrees well with the statistically estimated delay with an uncertainty of ±5 minutes

* To optimize the hyperparameter and to assess the ML model performance, we employ a ten-fold cross validation approach

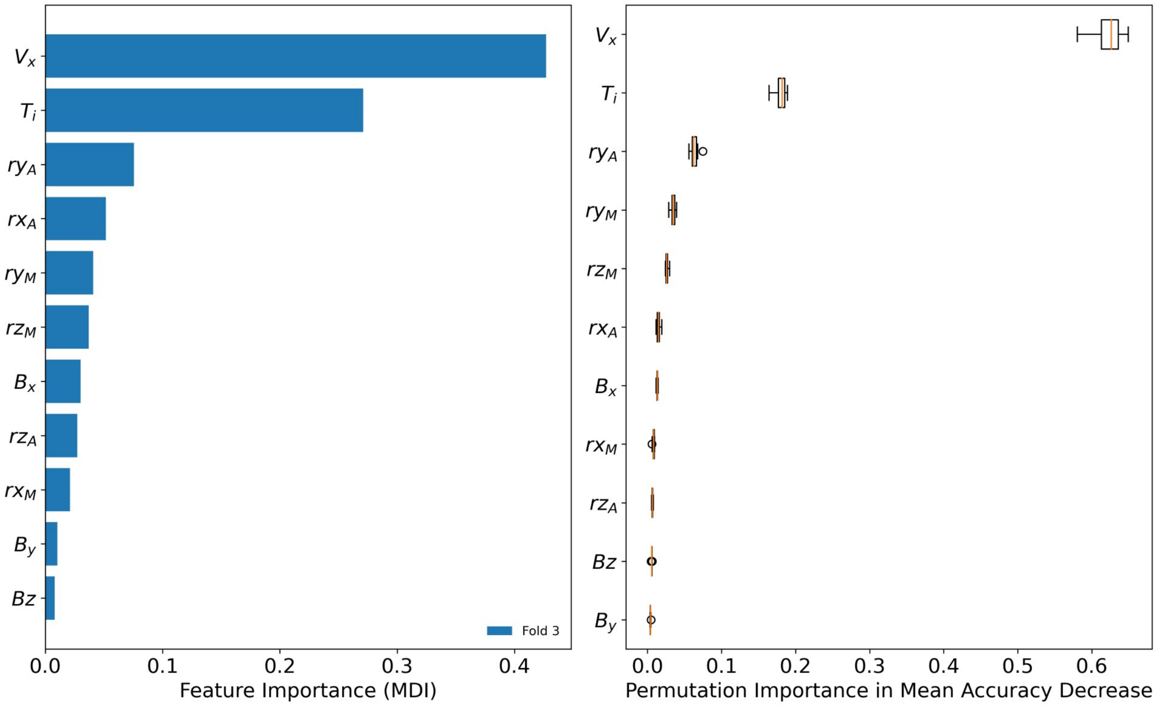
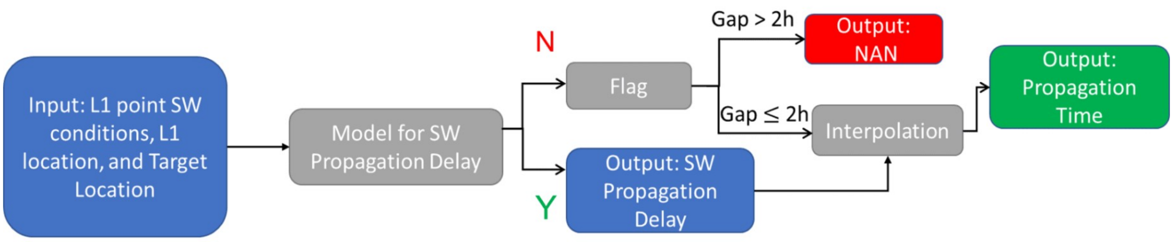


Fig 10: Feature Importance Bar Chart using the first set (left) of input vectors for a single data fold. (Right) Feature importance of the GB model using 10-fold cross validation

OVERVIEW AND CONCLUSION

* The statistical approach conducts cross-correlation analysis to estimate SW propagation times and provides large sets of input and target variables

* We use multiple spacecraft pairs at L1 and near-Earth locations to train, validate, and test machine learning models



* The ML algorithm using these data sets helps to specify and predict (1) the propagation time from L1 monitors to a given location upstream or at the bow shock and (2) to forecast near-Earth SW conditions

* The obtained propagation times are then compared to OMNI. Factors that limit the OMNI accuracy are also examined

DISCLOSURES

We would like to thank CDAWeb and NASA for ACE, WIND, DSCOVR, GEOTAIL, CLUSTER, and MMS data. We acknowledge the support of DLR for this research.

AUTHOR INFORMATION

Contact: Samira Tasnim

DLR, Institute for Solar-Terrestrial Physics

Kalkhorstweg 53, 17235 Neustrelitz

Tel.: +49 3981 480-316, Email: samira.tasnim@dlr.de

ABSTRACT

Having precise knowledge of the near-Earth solar wind (SW) and the embedded interplanetary magnetic field (IMF) is of critical importance to space weather operation due to the usage of SW and IMF in almost all magnetospheric and ionospheric models. The most widely used data source, OMNI, propagates SW properties from Lagrangian point L1 to the Earth's bow shock by estimating the propagation time of the SW. However, the time difference between OMNI timeshifted IMF and the best match-up of IMF can reach ˜15 min. Firstly, we aim to develop an improved statistical algorithm to contribute to the SW propagation delay problem of space weather prediction. The algorithm focuses on matching SW features around the L1 point and upstream of the bow shock by computing the variance, cross-correlation coefficient, the plateau-shaped magnitude index, and the non-dimensional measure of average error index between the measurements at the two locations. The obtained propagation times are then compared to OMNI. Factors that limit the OMNI accuracy are also examined. Secondly, the automatic algorithm allows us to generate large sets of input and target variables using multiple spacecraft pairs at L1 and near-Earth locations to train, validate, and test machine learning models to specify and forecast near-Earth SW conditions. Finally, we offer a machine learning (ML) approach to specify and predict the propagation time from L1 monitors to a given location upstream or at the bow shock and forecast near-Earth SW conditions with the gradient boosting and random forest prediction models in the form of an ensemble of decision trees.

REFERENCES

Baumann, C., and McCloskey, A, Timing of the solar wind propagation delay between L1 and Earth based on machine learning, J. Space Weather Space Climate, 11, 2021

Mailyan, B., Munteanu, C., and Haaland, S., What is the best method to calculate the solar wind propagation delay?, Annales Geophysicae, 26, 2008

Case, N. A., and Wild, J. A., A statistical comparison of solar wind propagation delays derived from multispacecraft techniques, J. Geophys. Res., 117:A02101, 2012