Deutsches Zentrum für Luft- und Raumfahrt e.V. in der Helmholtz-Gemeinschaft

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

Samira Tasnim[1], Ying Zou[2], Claudia Borries[1], Carsten Baumann[1], Brian Walsh[3], Krishna Khanal[2], and Connor J. O'Brien[3]

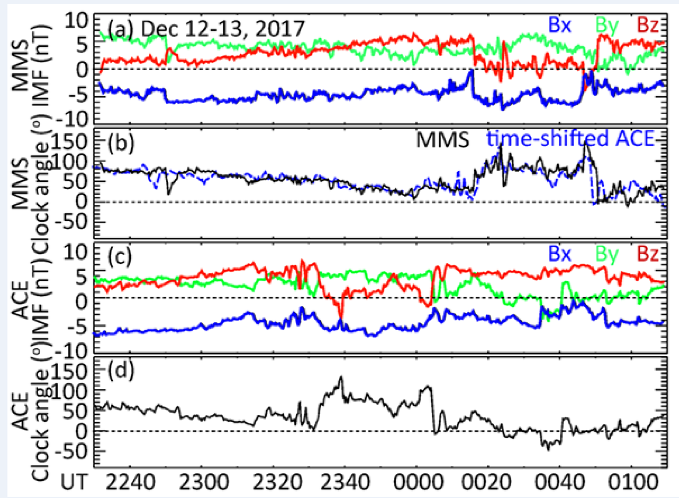
1 Institute for Solar-Terrestrial Physics, DLR, Neustrelitz , 2 CSPAR, University of Alabama in Huntsville , 3 Department of Electrical and Computer Engineering, Boston University

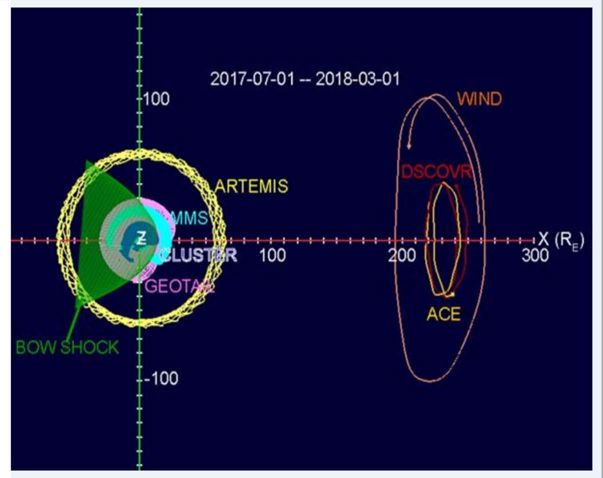
Background and Motivation

- 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
- Specifically, we train the model to specify and/or predict the propagation time from L1 monitors to a given location upstream or at the bow shock

Data and Instruments

- To obtain the SW and IMF parameters around L1: ACE - Advanced Composition Explorer (H0 - ACE Magnetic Field 16-Second Level 2 Data)
- For upstream of the Earth’s bow shock: MMS (Level2 Flux Gate Magnetometer Combine Fast/Slow Survey DC Magnetic Field for 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

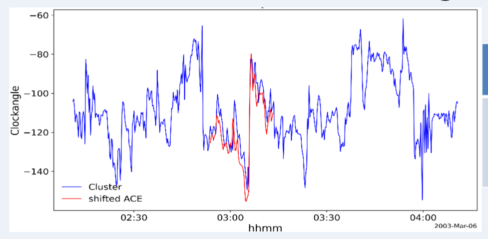
Database to Prepare Input and Target Variables for the ML Model

- 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

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

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

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



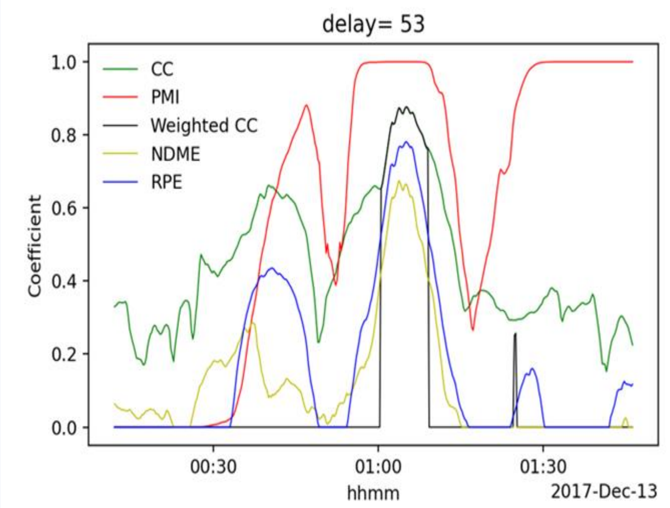
Input Feature Variables					Target Variable
IMF vectors (Bx, By, Bz)	SW velocity (Vx)	Ion temperature (Ti)	Position of L1 monitor (x, y, z)	Position of near-Earth monitor (x, y, z)	SW propagation Delay

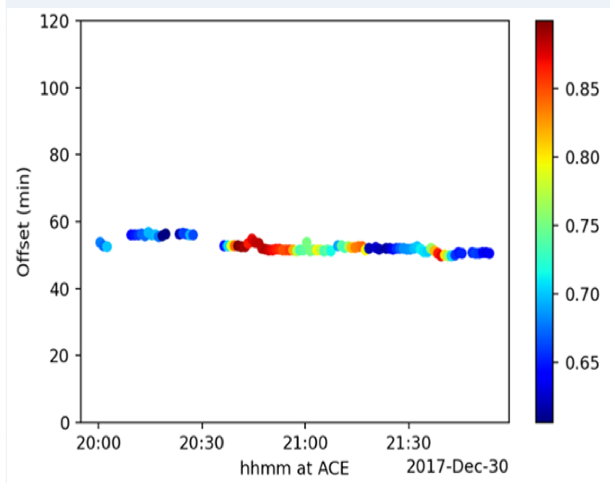
Methodology

- 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





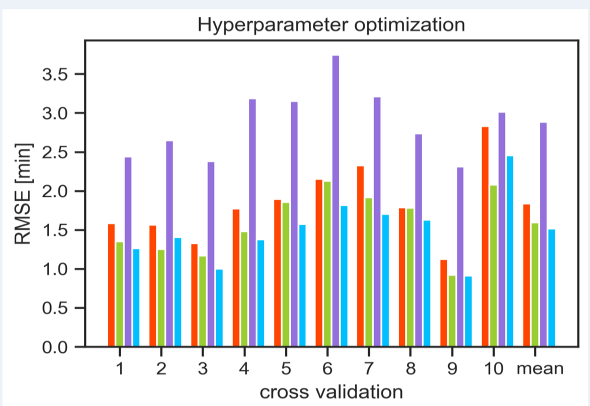
ML Models and Predictions

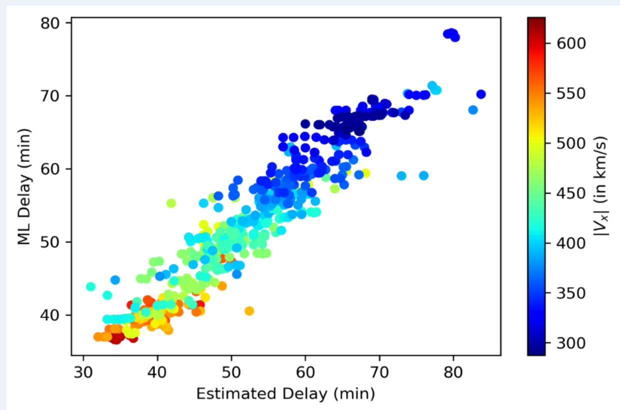
- We apply two ML models to predict SW propagation delay: i) Random Forest Regression (RF) and ii) Gradient Boosting (GB)
- GB and RF algorithms are applied together: a) To enable direct comparison between the RF and GB models and b) To quantify if the use of an ensemble-based ML model make a significant improvement to the overall performance
- The machine learning SW propagation delay can be described as

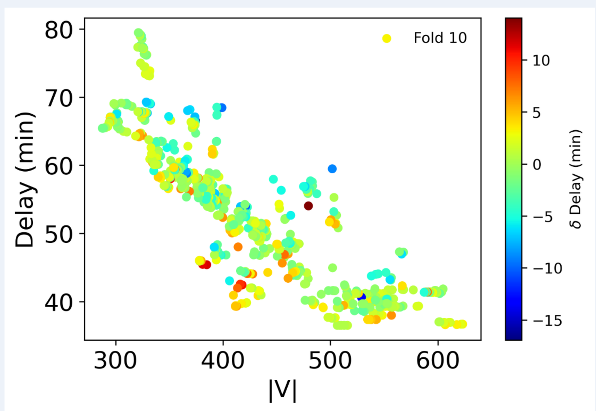
$$\Delta t_{ML} = f_{\mathcal{D}}(x)$$

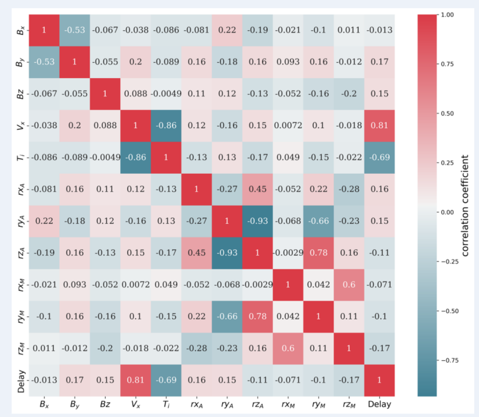
Here $f_{\mathcal{D}}$ describes the ML algorithm trained on the data set $\mathcal{D}$ and x contains the feature vectors

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

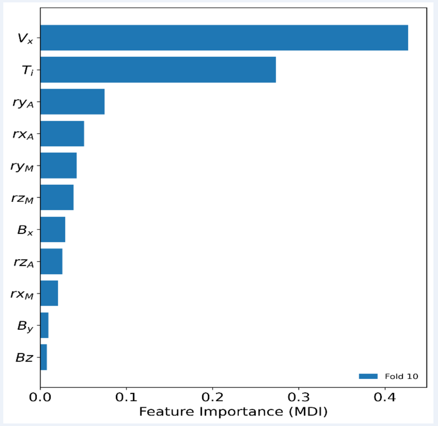


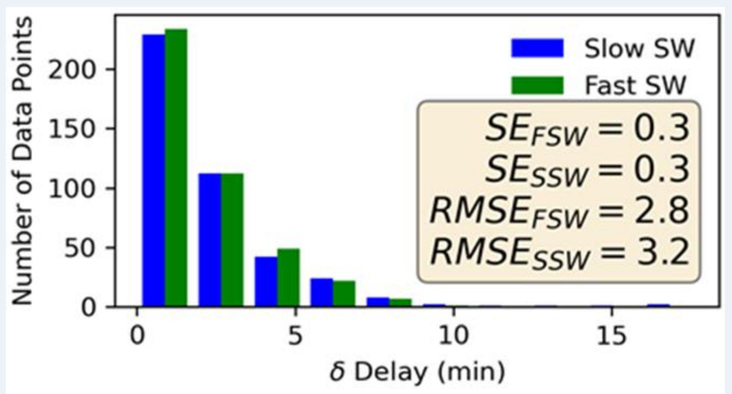


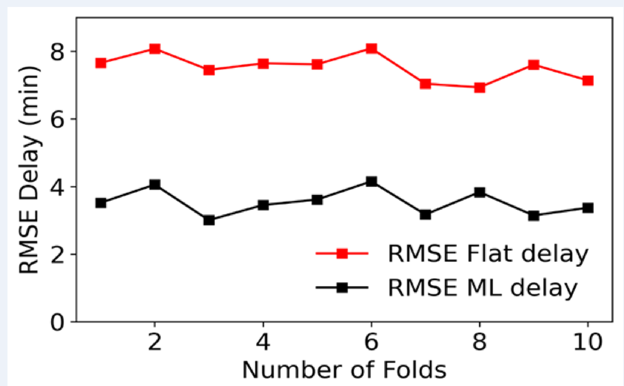




- The ML model predicted delay agrees well with the statistically estimated delay with an uncertainty of ±5 minutes
- To optimize the hyperparameter and to assess the ML model performance, we employ a ten-fold cross validation approach
- Feature importance of the ML model and Correlations between feature vectors and the target SW delay are investigated
- Validation: ML model predicted delays are compared with the results of physical models: 1) Flat Delay and 2) OMNI shifted delay using Phase Front Normal





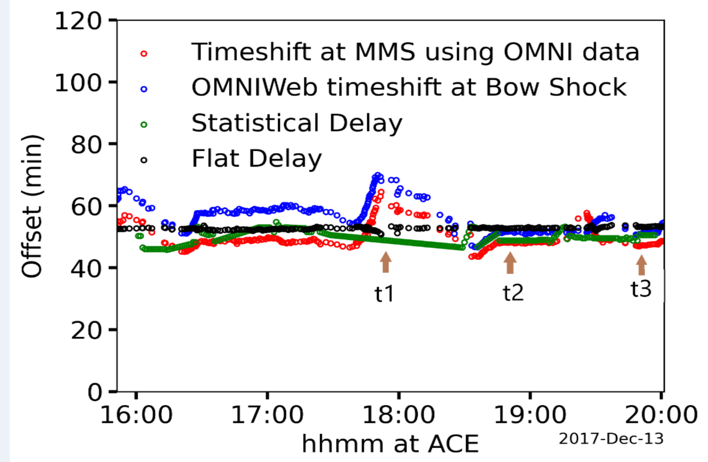


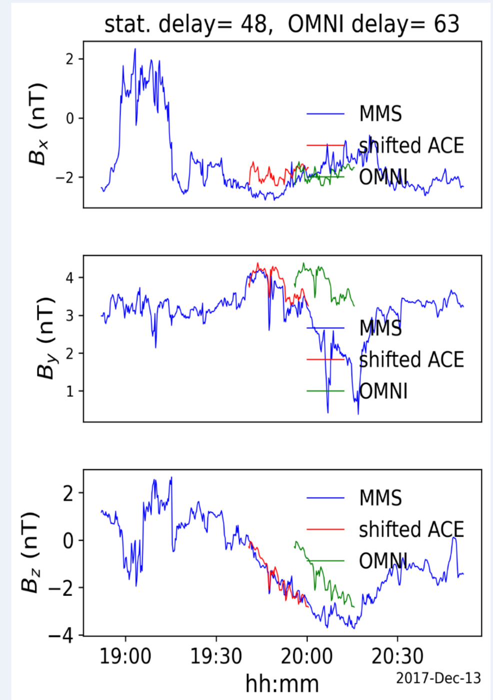
Model	RMSE	mean absolute error	r2 score
Gradient Boosting Regressor	9.12	2.18	0.92
Random Forest Regressor	1.26	3.34	0.96
Neural Network	1.73	1.197	0.97

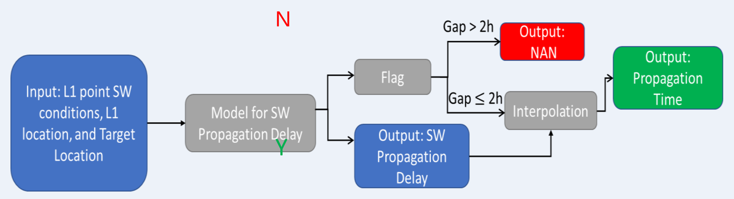
Model	RMSE	mean absolute error	r2 score
RNN	12.50	8.84	0.013

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

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

Contact

Samira Tasnim
German Aerospace Center (DLR)
Institute for Solar-Terrestrial Physics
Kalkhorstweg 53, 17235 Neustrelitz
Tel.: +49 3981 480-316, Email: samira.tasnim@dlr.de