ÖAW — AUSTRIAN ACADEMY OF SCIENCES

Predicting the magnetic flux rope fields at the Sun-Earth L1 point

U.V. Amerstorfer[1], C. Möstl[1,2], R.L. Bailey[1], A.J. Weiss[1,2], M.A. Reiss[1,3], T. Amerstorfer[1], J. Hinterreiter[1], M. Bauer[1]

[1]Space Research Institute, Austrian Academy of Sciences, Graz, Austria; [2] Graz University of Technology, Graz, Austria; [3]NASA (Heliophysics Science Division), NASA Goddard Space Flight Center, Greenbelt, USA

Forecasting of coronal mass ejection magnetic flux rope fields at L1 is a long-standing challenge and one of the major problems in space weather forecasting. We use machine learning algorithms (e.g., linear regression, lars lasso, RANSAC, or random forest) to predict the scalar mean magnetic field of the flux rope. For our study, we take events observed at the Wind, Stereo-A and Stereo-B satellites from the ICME list created within the EU-project HELCATS. We analyse different scores (STD, RMSE, or the skill of the model) of the presented methods and show the importance of different used features.

INTRODUCTION

The work presented here is embedded in a larger project called PREDSTORM. The goal of this project is to improve the accuracy and lead time for predicting the occasionally destructive effects of coronal mass ejections (CMEs) at Earth with currently available spacecraft, by combining different approaches, from pattern recognition on large solar wind in situ data sets to machine learning (ML) algorithms for prediction of the magnetic field in flux ropes and a CME magnetic flux rope model (3DCORE; Möstl et al. 2018). As a first approach in trying to predict the magnetic flux rope (MFR) field within ICMEs, we utilize ML algorithms and investigate the use of different features.

DATA

We take the ICME list created within the EU-project HELCATS. Out of this list we take the ICME events observed at the Wind, Stereo-A and Stereo-B satellites. Of all of these events we only take those ICMEs, which have a clearly defined sheath, because we want to investigate the use of sheath values for forecasting MFR values. In total, we have 258 events.

FEATURE SELECTION

We will try to predict the mean total magnetic field of the flux rope, $< B_{\rm tot}^{\rm MFR} >$ (also called $< B >$ label), when we know the sheath values and/or the first few hours of the MFR. We investigate 6 different feature cases:

1. only sheath values (case **sheath**)
2. sheath values and the first hour of the MFR (case **sheath + mfr (1h)**)
3. sheath values and the first 3 hours of the MFR (case **sheath + mfr (3h)**)
4. the first hour of the MFR (case **mfr (1h)**)
5. the first 3 hours of the MFR (case **mfr (3h)**)
6. the first 5 hours of the MFR (case **mfr (5h)**)

We have a lot of different features, which we can use, and we try to select the most important ones. For this purpose, we have a look at the correlation matrix. An example of the correlation matrix for the case **mfr (5h)** is shown in Figure 1. The correlation coefficients regarding $< B_{\rm tot}^{\rm MFR} >$ are shown in Table 1 for all of the feature cases.

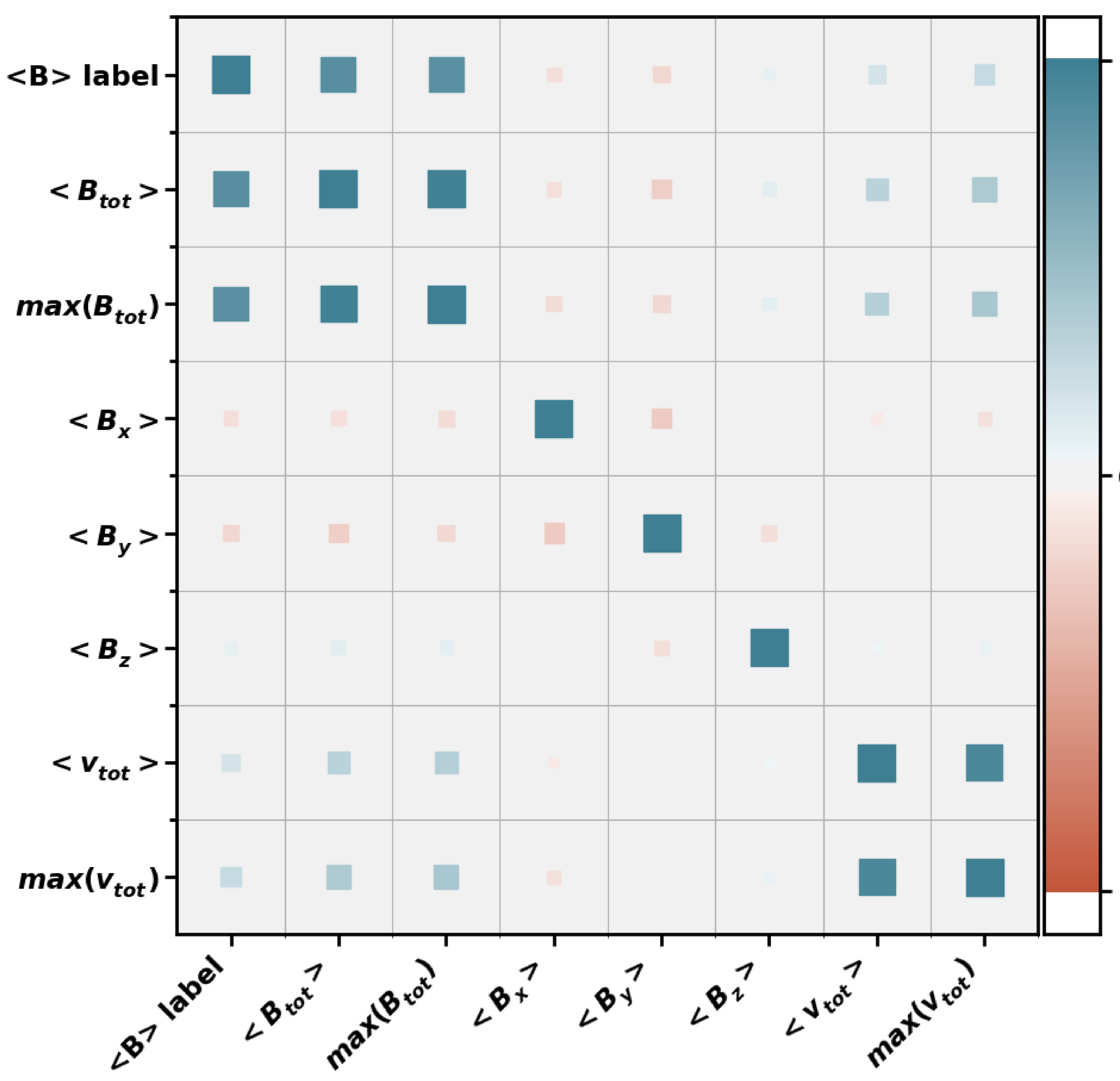

Figure 1: Example correlation map for the feature case **mfr (5h)**.

It is clearly visible that, apart from $< B_{\rm tot} >$ and $\max(B_{\rm tot})$, there are no significant correlations with $< B_{\rm tot}^{\rm MFR} >$. Another fact that can be seen from the correlation matrix is that the two features $< B_{\rm tot} >$ and $\max(B_{\rm tot})$ are also strongly correlated with each other. This means we do not need to take both features for training the ML algorithm, but in order to avoid redundance and over-fitting, it is sufficient to take only one. We decided to take $< B_{\rm tot} >$, since it has the slightly higher correlation coefficient.

feature	sheath	sheath + mfr (1h)	sheath + mfr (3h)	mfr (1h)	mfr (3h)	mfr (5h)
$< B_{\rm tot} >$	0.59	0.64	0.71	0.85	0.86	0.84
$\max(B_{\rm tot})$	0.57	0.64	0.64	0.86	0.86	0.82
$< B_z >$	0.19	0.15	0.11	0.01	0.05	-0.01
$< v_{\rm tot} >$	0.14	0.14	0.13	0.18	0.16	0.18
$\max(v_{\rm tot})$	0.09	0.16	0.14	0.27	0.20	0.18
$< B_y >$	-0.16	-0.14	-0.19	-0.14	-0.22	-0.15
$< B_z >$	-0.16	-0.23	-0.29	-0.09	-0.16	-0.16

Table 1: Correlation coefficients of different features regarding the label $< B_{\rm tot}^{\rm MFR} >$.

MACHINE LEARNING ALGORITHMS

As a first step in trying to predict the scalar mean magnetic field of MFRs within ICMEs we investigate the possibility of using traditional ML algorithms. We implement 11 models (*linear regression*, *lasso*, *ridge*, *elastic-net*, *huber*, *lars*, *lars lasso*, *passive-aggressive*, *RANSAC*, *SGD*, *random forest*). Our algorithm determines cross-validation scores in the training data set and then decides upon those scores, which model is the best final model to make the prediction with the testing data set.

RESULTS

Different scores of the best final model on the testing data set are shown in Table 2. The skill is defined as $\mathrm{skill} = 1 - \mathrm{mse}/\mathrm{mse}_{\rm ref}$, where mse and $\mathrm{mse}_{\rm ref}$ are the mean squared error of the proposed and of a reference (mean) model. A skill of 0 means the proposed model does not improve the prediction over a mean model, whereas a skill of 1 means a perfect prediction.

feature cases	RMSE	STD	SKILL	best algorithm
sheath	3.6	1.2	0.31	*huber*
sheath + mfr (1h)	3.2	0.85	0.1	*SGD*
sheath + mfr (3h)	2.75	0.73	0.03	*lars*
mfr (1h)	2.32	0.8	0.56	*huber*
mfr (3h)	2.17	0.61	0.51	*ridge*
mfr (5h)	2.05	0.62	0.66	*ridge*

Table 2: Cross-validation scores (root mean squared error RMSE and standard deviation STD) and skill scores for the best ML model for the different feature cases. We also name the best algorithm determined for the specific feature case.

A few example plots of the prediction of $< B_{\rm tot}^{\rm MFR} >$ for the feature case **mfr (5h)** is shown in Figure 2.

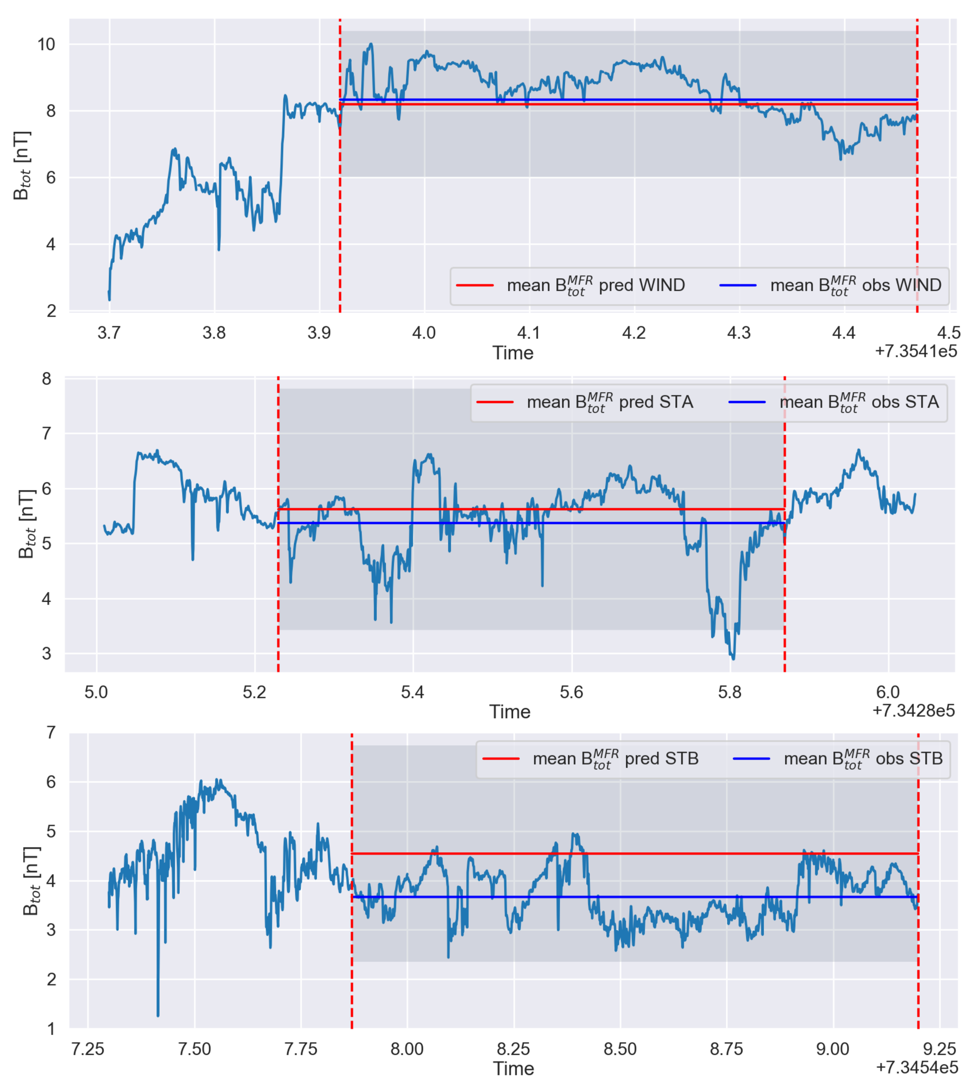

Figure 2: Example results of the prediction of MFR values with traditional ML algorithms for the feature case **mfr (5h)**. The shaded area gives the mean RMSE of the data set when using cross-validation. The vertical red dashed lines denote the beginning and end of the MFR, the red line shows the prediction of $< B_{\rm tot}^{\rm MFR} >$, and the blue line depicts $< B_{\rm tot}^{\rm MFR} >$ calculated from the observed values.

SUMMARY & OUTLOOK

Within the project PREDSTORM we attempt to forecast the magnetic fields in magnetic flux ropes with machine learning algorithms. We investigate the use of different features from different regions of the interplanetary coronal mass ejection to improve the forecast. Including features of the sheath region leads to worse scores and predictions than taking only features from the flux rope. As next steps we will (1) derive synthetic magnetic field and plasma data from empirical flux rope modeling with our own 3DCORE simulation and (2) combine the results with past analogues ensemble algorithms.

References

Möstl, C., Amerstorfer, T., Palmerio, E., et al. 2018, Space Weather, 16, 216

Acknowledgements. This work is supported by the Austrian Science Fund (FWF): P31659, P31521, P31265, J4160. The HELCATS catalogue is available from https://www.helcats-fp7.eu.