Ann. Geophys. Discuss.,
https://doi.org/10.5194/angeo-2019-149-RC3, 2020

Annales Geophysicae Discussions

Interactive comment on "Automatic detection of the Earth Bow Shock and Magnetopause from in-situ data with machine learning" *by* Gautier Nguyen et al.

Anonymous Referee #2

Received and published: 10 January 2020

This work presents a Decision Tree based classification scheme for the detection of the various parts of the Magnetoshpere, i.e. the inner magnetosphere the magnetosheath and the solar wind, based on in-situ magnetic field and plasma measurements. The method is built on data from the THEMIS mission and then used (with and without retraining) for other missions as well, such as the Cluster and Double Star. In addition they claim that the same method could potentially be used to separate other regions of the interplanetary space, as is the "Lunar Wake" area demonstrated by using data from the THEMIS/Artemis satellites. The method outperforms simpler classification schemes, such as a typically used threshold based set of criteria and can be utilized to run in a large set of data, thus providing results for the accurate detection of the

Magnetopause and the Bow Shock for a large time period.

The method is highly promising and specifically the comparison of the Gradient Boosted Decision Trees against a typical (and similar) threshold-based approach can be a powerful demonstration of the capabilities of Machine Learning techniques. Unfortunately, the authors do not provide enough details on the method use, while some elements of the construction of the test and training datasets seem somewhat unclear. An updated version of this paper with an expanded Section 2.2 and with a more thorough analysis of the labeling procedure could be fit for publication.

Additional Comments:

Lines 52-55: I would require more detail in the labeling process. What exactly is meant by "successive, eventually corrected, predictions"? The term "training set" is used here, even though the proper training set is defined later, on line 61. Does the final dataset cover the full time range of the 2007-2009 period? Are the authors concerned that their dataset might not be representative of a variety of solar/magnetospheric conditions, since 2007-2009 was near the solar minimum and was a rather quiet time period with regards to geospace activity.

Section 2.2 Algorithm: I would advise the authors to extend this section with more details on the specific way with which they have implemented the method (number of decision trees, cost function, how exactly does the final probability score emerge etc). Providing an entire book as a reference is not particularly helpful and I believe that many readers would be interested in the technical details, since ML is being used in an increasing number of applications these days.

Section 2.3: An additional metric would be welcome here, e.g. the Heidke SS, especially since the AUC scores are pretty close to one another.

Table 1: Since the scores for all three classes are almost perfect, I would expect the mislabeled AUC to be 75% everywhere, but its 70.7% for the Magnetosheath class.

Do you have any suspicions or thoughts on why that might be? Did you also perform 3 different misslabelings to verify this result (as you did with the training-test dataset selection)?

Section 3: Since different satellites carry different instruments with varying sensitivities it would be interesting to see if using some sort of normalization scheme in the data can help the method to yield high scores without re-training.

Also, it is generally advised to use an as-equal-as-possible sample size for all the classes in a dataset. Especially in the Double Star case the Solar Wind category seems significantly under-represented compare to the other two. Have the authors tried to replicate their results with a more balanced dataset?

Section 3.3: Wouldn't a set of Lunar coordinates (selenocentric) be more useful in properly identifying the fourth class? Or alternatively and additional parameter that captures the Moon's Local Time position? Also, I do not see the AUC scores for the fourth category in Table 1.

Lines 227-228: "Events with high probability would then correspond to undoubtful crossings while the events with the lowest probability would be the most likely to be actual crossings". Is this correct or was it meant to be "while the events with the lowest probability would be LESS likely to be actual crossings".

Section 5: It would be very interesting to see the difference in the position of the Magnetopause and the Bow Shock for quiet vs disturbed conditions (e.g. low vs high solar pressure) as predicted by this method and a comparison against an analytical model.

Interactive comment on Ann. Geophys. Discuss., https://doi.org/10.5194/angeo-2019-149, 2019.