Ann. Geophys. Discuss.,
https://doi.org/10.5194/angeo-2019-149-AC3, 2020

Annales Geophysicae Discussions
Open Access
EGU

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

Gautier Nguyen et al.

gautier.nguyen@lpp.polytechnique.fr

Received and published: 29 January 2020

We thank the referee for the numerous comments regarding our methodology and the various suggestions proposed to better our work. Please find below the detailed answers to the referee's questions and 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.

The labels were made by inspecting the data visually and deciding, by selecting inter-

Printer-friendly version

Discussion paper

vals, to which class their points belonged to. This requires to zoom in and out many intervals and is thus a long and fastidious process. To make it faster, in particular to zoom in regions of interest, we decided to guide our eyes with the preliminary predictions of a GB classifier trained on a dataset iteratively widened by our labels, plotted over the data.

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.

The final labeled dataset we used for THEMIS data does cover the 2007-2009 period. Past this period, THEMIS B and C became Artemis and this specific case was mentioned in the paper. The concern of the variability due to the solar cycle then concerns the quality of the massive prediction lead on the data provided by THEMIS A, D and E.

Even if the solar cycle induces variability in the physical parameters of the three regions, we will find the same differences between the three different classes on the dayside part of the near-Earth environment:

- The magnetosphere will still be characterized by a low density, a high temperature, a high magnetic field and a plasma almost at rest
- The magnetosheath will still be characterized by a high density, a subsonic ion flow, and a lower magnetic field amplitude than in the magnetosphere
- The solar wind will still be characterized by a supersonic ion flow, a high density being lower to the one in the magnetosheath and a low magnetic field amplitude.

At first order, this physical differences we have between each classes will prevail on the variability induced by the solar cycle. The latter can then be neglected in the specific case of these missions.

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.

Gradient boosting algorithms have proven their capability to rapidly deal with complex, eventually imbalanced (Brown et al. (2012)) classification problems. This is the reason for which we chose this class of algorithms. We computed the method using its python implementation provided by Scikit-learn (Pedregosa et al. (2011)). The method has been computed with the standard hyperparameters provided by Scikit-learn that is to say:

- 100 decision tree
- The multinominal deviance loss function which is a standard loss function used for multiclass classification

These precisions will be added in the revised manuscript

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

We agree with the referee and attach to our answer the evolution of the HSS as a function of the probabilistic decision threshold. For the decision threshold we chose for our massive detection (e.g 0.5), the high value of the HSS we obtain for the three classes confirms the efficiency of our model. This figure will be added to the revised version of the paper where we will also introduce this metric.

Table 1: Since the scores for all three classes are almost perfect, I would expect the mislabeled AUC to be 75C2 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)?

The two only reasons of a data mislabel by a human observer are the confusion of magnetosheath points with either magnetosphere or solar wind points. Consequently, we mislabeled our dataset following this process:

- we selected a fraction of random points of the dataset
- The magnetosphere and the solar wind points were mislabelled as magnetosheath points
- magnetosheath points were randomly mislabelled between the two different classes

This operation has been done for various percentages of mislabelling 10 times each. We attach to our answer the evolution of the AUC for each class with the mislabelling.

Even if the classifier is almost perfect, we do not expect a particular evolution in the AUC for each class apart from a drop in the model performance. From then on, having the same performances for the solar wind and the magnetosphere for this percentages is a coincidence that is seen by the evolution of the AUC for these two classes.

This mislabelling process and the evolution of the AUC with the mislabelling will be added in the annex of the revised paper.

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.

Cluster has a polar orbit while THEMIS and Double Star stay in the equatorial plane. This physical difference in latitude of the different regions traversed by the spacecraft is

the main reason that explains the variability we notice in the data from a mission to another and prevails on the instrumental specificity of each spacecraft. This is especially shown with the score we obtain on Double Star and Cluster without retraining.

A global model that takes into account this orbital variability could be a nice improvement of our models in a future work.

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?

Gradient boosting has proven its efficiency to deal with imbalanced dataset (Brown et al. 2012) and this is one of the reason for which we chose this algorithm, this will be precised in the paper.

Additionally, we show in the paper that the model trained on THEMIS data already gets on well with every categories of data measured by Double Star. As the retrain already takes into account the differences between the three classes represented in the THEMIS dataset , the imbalance will have a tiny influence on the model performances that are already good.

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.

We use the plasma moments to take into account all of the possible complexity of the underlying nature of the object we are trying to detect, without simplifying a priories coming from modelization

This approach is preferable as long as the different type of signatures we are trying to detect have strong intrinsic properties. Which is not the case for the magnetosheath

and the solar wind in the distant night side. This is why we help the algorithm by giving the position of the spacecraft in this specific case.

We agree we didn't mention the AUC score for the fourth class that we found equal to 0.997. We will mention this score in the revised version of the manuscript.

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".

We agree with the referee on the omission we made on this sentence. This will be corrected.

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.

We agree such study would be interesting and this is why we mention it in the conclusion of our paper. The construction of a data-driven bow shock and magnetopause model is one of the objectives we can have by performing massive crossing detection from in-situ data and this was studied by an intern from our group. The preliminary plots we obtained showed consistency with what we know on the position of the bow shock for varying solar wind dynamic pressure. Nevertheless, this work is very preliminary and would need a specific focus that goes beyond the scope of this paper.

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

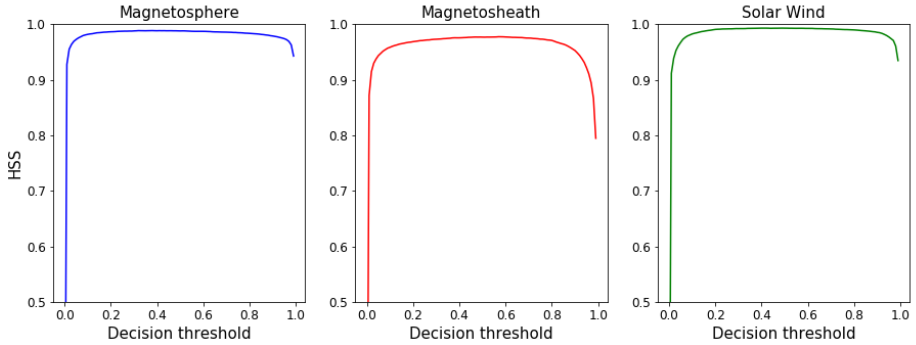

Fig. 1. Heidke Skill Score of our model trained on THEMIS data for varying decision threshold for the three classes

Printer-friendly version

Discussion paper

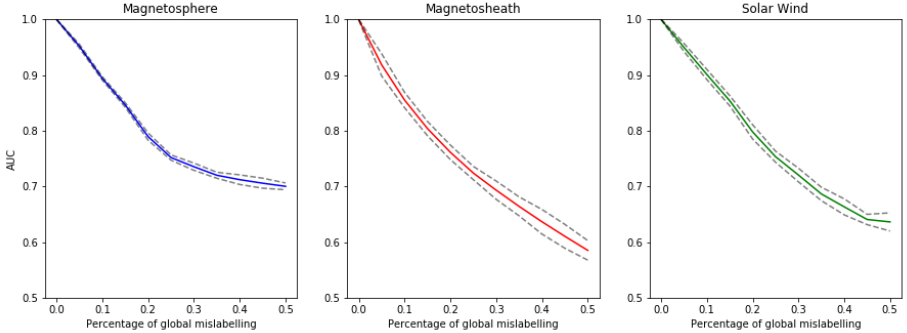

Fig. 2. Evolution of the AUC as a function of the percentage of mislabelled points in the dataset. The grey dashed lines indicate the error bars obtained for each 10 repetitions of mislabelled training

Printer-friendly version

Discussion paper