Massive multi-mission statistical study and analytical modeling of the Earth's magnetopause: 1 - A gradient boosting based automatic detection of near-Earth regions

Gautier Nguyen, Nicolas Aunai, Bayane Michotte de Welle, Alexis Jeandet, Benoit Lavraud, Dominique Fontaine

Massive multi-mission statistical study and analytical modeling of the Earth's magnetopause: 1 - A gradient boosting based automatic detection of near-Earth regions

G. Nguyen [1], N.Aunai [1], B.Michotte de Welle [1], A.Jeandet [1], B.Lavraud [2,3]and D.Fontaine [1]

[1]CNRS, Ecole polytechnique, Sorbonne Université, Univ Paris Sud, Observatoire de Paris, Institut Polytechnique de Paris, Université Paris-Saclay, PSL Research Univsersity, Laboratoire de Physique des Plasmas, Palaiseau, France
[2]Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[3]Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse, CNRS, CNES, Toulouse, France

Key Points:

- A gradient boosting algorithm is used to classify the magnetosphere, magnetosheath and solar wind in-situ data from multiple missions
- The method outperforms the detection methods based on manually-set threshold and is trained faster than existing machine-learning based methods
- The method is used to identify 15 062 magnetopause crossings and 17 227 bow shock crossings

Corresponding author: Gautier Nguyen, gautier-mahe.nguyen@intradef.gouv.fr

Abstract

We present an automatic classification method of the three near-Earth regions, the magnetosphere, the magnetosheath and the solar wind from their in-situ data measurement by multiple spacecraft. Based on gradient boosting classifier, this very simple and very fast method outperforms the detection routines based on manually-set thresholds. The method is used to identify 15 062 magnetopause crossings and 17 227 bow shock crossings in the data of 11 different spacecraft of the THEMIS, ARTEMIS, Cluster, MMS and Double Star missions and for a total of 83 cumulated years. These multi-mission catalogs are easily reproducible, can be automatically enlarged with additional data and their elaboration paves the way for future massive statistical analysis of near-Earth boundaries.

1 Introduction

The magnetopause is the boundary where magnetospheric and magnetosheath pressures balance, and where magnetic fields of terrestrial and solar origin interact. It acts like an obstacle for the upcoming supersonic solar wind and is thus located downstream a collisionless bow shock (Burgess, 1995) across which the solar wind becomes subsonic. The magnetopause and the bow shock are the boundaries of the three main near-Earth regions: the magnetosphere, the magnetosheath and the solar wind. By definition, the shape, location and properties of these boundaries depend on the upstream solar wind conditions (Fairfield, 1971). The ever-growing quantity of near-Earth in-situ data allowed the realisation of statistical studies dedicated to the physical properties of the different near-Earth regions and to the position, shape and dynamics of both the magnetopause (Paschmann et al. (2018); Němeček et al. (2020); Hasegawa (2012) and references therein) and the bow shock (Kruparova et al. (2019) and references therein). Such studies also led to the development of numerous magnetopause (Shue et al. (1997); Lin et al. (2010); Wang et al. (2013); Liu et al. (2015) and references therein) and bow shock (Jeřáb et al. (2005); Farris and Russell (1994) and references therein) surface models.

The first step of both empirical modelling and statistical studies is always the same: establishing a consistent catalog of boundary crossings from the streaming in-situ data provided by missions of interest. This, in addition to being time-consuming, is an ambiguous task, strongly linked to the interpretation of an external observer and thus poorly reproducible. As a result, catalogs of events are difficult to make and their size represents, with time, an ever decreasing proportion of the total and massive amount of public multi-mission data that has been accumulating for decades. This severely hampers the development of a statistically relevant and global vision of our near space environment and plasma processes therein. Consequently, the elaboration of automatic event detection methods in streaming in-situ time series data provided by spacecraft appears as an interesting option to accelerate the collection of boundary crossings and improve the reproducibility and robustness of statistical studies. Figure 1 shows typical observations from a spacecraft travelling outbound through the three regions. From top to bottom are represented the proton density, the magnetic field components, the ion velocity components and the omnidirectional energy flux of ions measured by THEMIS B. The last panel will be explained in the following sections. The three regions are easily distinguishable by eye and the first method we could think about in this classification task would be to use manually set thresholds on wisely chosen physical quantities. Using the data provided by the five THEMIS spacecraft coupled with the solar wind conditions provided by WIND, Jelínek et al. (2012) established a method based on thresholds on the magnetic field amplitude B and the proton density N_p normalized by the interplanetary magnetic field (IMF) amplitude and proton density. They used this method to identify the three near-Earth regions and eventually build lists of crossings from this classification. The principle of the method consists in manually setting the two straight lines that best separate the three regions in the (N_p, B) plane in a way that is similar to what is shown in Figure 10. Nevertheless, this still requires the manual setting of thresholds on a reduced number of parameters. There is no guarantee on how well they will

do on an unknown set of data and the separability of the two features presented here is not guaranteed on the whole magnetopause, especially in the case of nightside, flanks or high latitude boundary crossings [1]. The method could thus be improved with additional features such as the amplitude of the ion bulk velocity or the ion temperature but this would lead to the establishment of manual thresholds in a N-dimensional space, which is a tricky task if done manually.

A way to go beyond this solution stands in using supervised machine learning algorithms that usually have the advantage of rapidly finding the intrinsic differences between different labeled points in complex multi-dimensional datasets. The use of these algorithms to classify time series into several categories is not new in the field of space physics. They have especially proved their effectiveness in classifying the solar wind into several categories (Camporeale et al., 2017) or to determine if an interval of data contains a Flux Transfer Event (FTE) (Karimabadi et al., 2009). Recently, Olshevsky et al. (2019), Breuillard et al. (2020) and Argall et al. (2020) all proposed neural network based methods to classify the different near-Earth regions from the MMS data. The high performances reached by the three methods confirms the potential and the efficiency of statistical learning algorithms for such a classification task. Although providing interesting results, the high level of flexibility offered by such deep learning methods is generally balanced by the large amount of labeled data samples needed and the very long time they require to converge in their training. It thus appears important to establish reliable methods that require shorter convergence time.

In this paper, we establish such a method by training a gradient boosting algorithm to automatically classify the three near-Earth regions from the magnetic field and plasma moments of the THEMIS mission. After presenting the data and the associated labels, we present the algorithm we use and explain this choice. We then evaluate its performances and investigate its adaptability to the various missions that explore the different parts of the magnetosphere boundaries: Double Star, MMS, Cluster and ARTEMIS. The outcome is then compared to the one obtained by setting thresholds manually for the different missions. The gradient boosting prediction is then used to automatically elaborate multi-mission catalogs of boundary crossings[2].

In particular, the massive magnetopause crossing catalog obtained in this study is the preamble to companion studies, focusing on the statistical analysis of the shape and location of the surface and its dependency on solar wind and seasonal parameters (hereafter (Nguyen et al., 2020a), the subsequent building of a new analytical and dynamical model of the magnetopause surface as a function of relevant upstream and seasonal control parameters (hereafter (Nguyen et al., 2020b)), and that re-visits the question of the indentation of the magnetopause surface in the near-cusp regions (hereafter (Nguyen et al., 2020c)).

2 THEMIS dataset

We used plasma moments and magnetic field data from the five THEMIS spacecraft, between April 2007 and January 2010 for THEMIS B and C and until June 2019 for the three remaining spacecraft. In particular, we considered the data measured during the dayside, dawn and dusk operation phases. The magnetic field data were provided by the Fluxgate Magnetometer (FGM, Auster et al. (2008)) with a temporal resolution of 3s. Concerning the plasma moments, we used the Fast-Survey mode of the data provided by the electrostatic analyzer (ESA, McFadden et al. (2008)) for which the distribution functions are composed of 24 energy channels and 50 solid-angle distributions with a temporal resolution of 4s. We use the onboard moments to fill in the data gaps in the Slow-Survey mode. The remaining holes in the plasma moments are filled with the data measured in the full mode and linearly

[1] Additional, less obvious observational examples of this case are shown in the Appendix C.

[2] Catalogs are available online at : `https://github.com/gautiernguyen/in-situ_Events_lists`

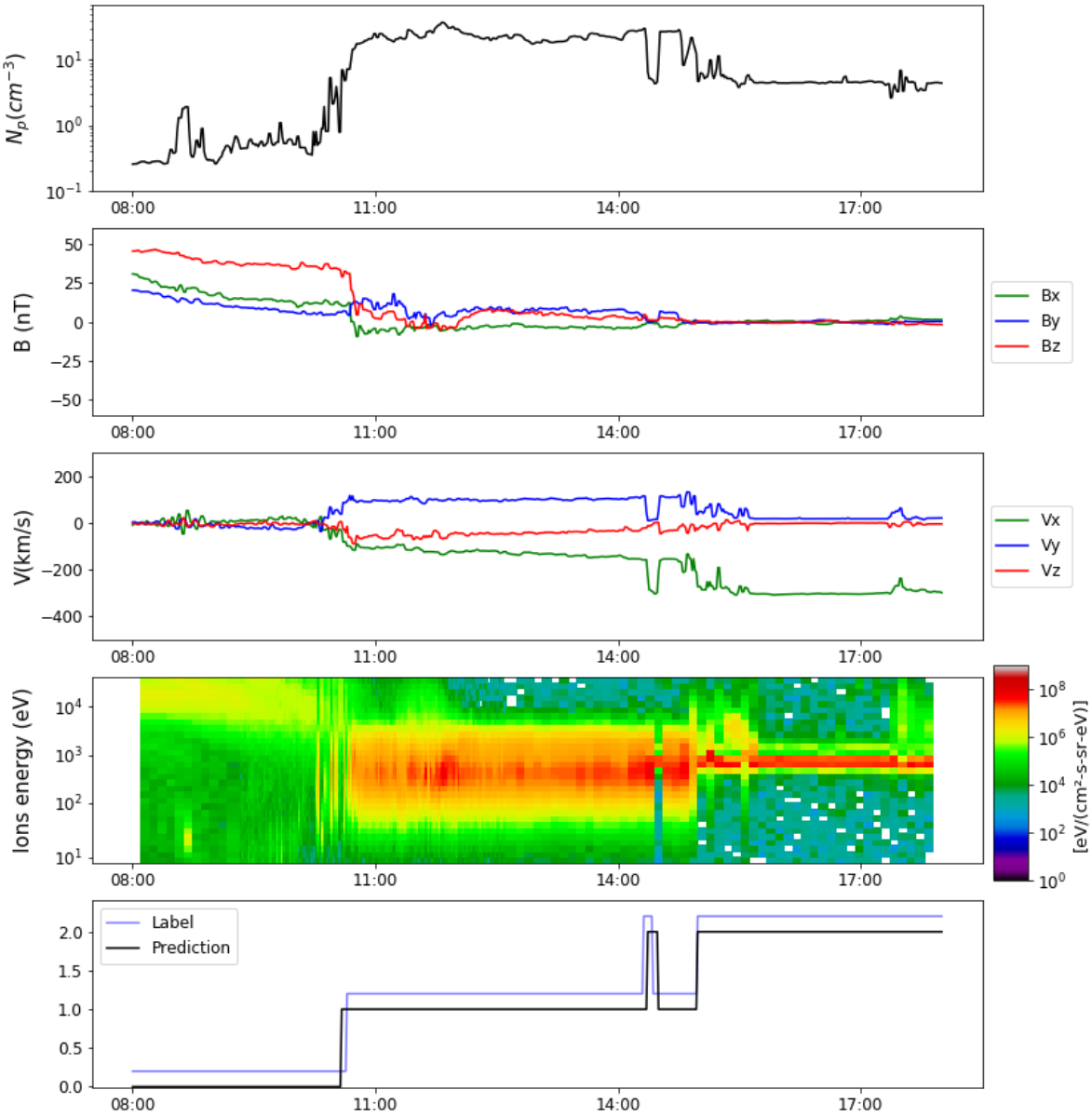

Figure 1. In-situ measurement provided by THEMIS B spacecraft on the 12^{th} of May 2008. From the top to the bottom are represented: the ion density, the magnetic field components, the velocity components the omnidirectional differential energy fluxes of ions. The last bottom panel represents the evolution of the label (blue) , intentionally shifted for visual inspection and the prediction made by our algorithm (black).

time interpolated in order to obtain streaming time series of the ion density, velocity and temperature with a uniform resolution of 4s. The ESA and FGM measurements are then synchronized to obtain a unique dataset with a common resolution of 1 minute in order to erase the noise due to very punctual partial crossings that will be particularly hard to label and detect.

Due to the important differences existing between the different missions in the specificities of the distribution functions and particle energy or pitch angle spectrograms, we chose to focus on the plasma moments and magnetic field only. The ion omnidirectional differential energy fluxes shown in Figure 1 are then only be used for visual inspection of data and to provide visual guidance in our labeling process.

For each spacecraft, the associated dataset then consists in 8 distinct input features: the ion bulk velocity components, V_x, V_y, V_z, the magnetic field and its components, B_x, B_y, B_z, the ion density N_p and the temperature T.

3 Label

We start our work by considering the THEMIS B dataset, the remaining sets will be used when we will perform the massive detection of boundary crossings in section 7.

Each datapoint is associated to a given label that indicates the region in which the spacecraft is at the measurement time:

- Points in tenuous regions with almost no ion bulk flow and important magnetic field amplitude are identified as magnetosphere points.
- Points in comparatively dense regions with a fast ion bulk flow and low temperature are identified as solar wind points.
- Points that are not identified as solar wind or magnetosphere are identified as magnetosheath. Those points correspond to the denser regions with an intermediate plasma velocity with a wide range energy flux. With this definition, any region downstream of the bow shock that is not the magnetosphere is considered as the magnetosheath. This will thus concern pristine magnetosheath points but also the regions composed of mixed plasmas such as the reconnection outflows, the cusp dense and hot plasma or the different magnetosphere and magnetosheath boundary layers.

While only considering the above three classes is enough for the purpose of the statistical analysis performed in the companion papers of this study (Nguyen et al., 2020a, 2020b, 2020c), the classification could be extended to additional classes. Having a single model classifying many different regions (e.g. solar wind, foreshock, cusps, boundary layers, lobes, plasma sheet etc.) may appear appealing at first glance. However it is worth mentioning that statistical studies rarely need that many classes, and moreover that multiplying classes often needlessly complicates the classification. Indeed, not only this may require more evolved algorithms thereby increasing the training time, it also introduces errors that basic knowledge of the Earth environment would prevent. For instance classifying automatically the ion foreshock from pristine solar wind can much more easily be done on a dataset where all but solar wind data (in a sense of our classification above) has been removed by a first pass of our classification than on a whole-orbit dataset. By construction, this prevents any non-physical errors an observer would never make in confusing unrelated regions, but also provides a better chance to fine tune the algorithm to a well defined task, and reduce training time.

We make those labels by inspecting the data visually and deciding, by selecting intervals, to which class their points belong to.

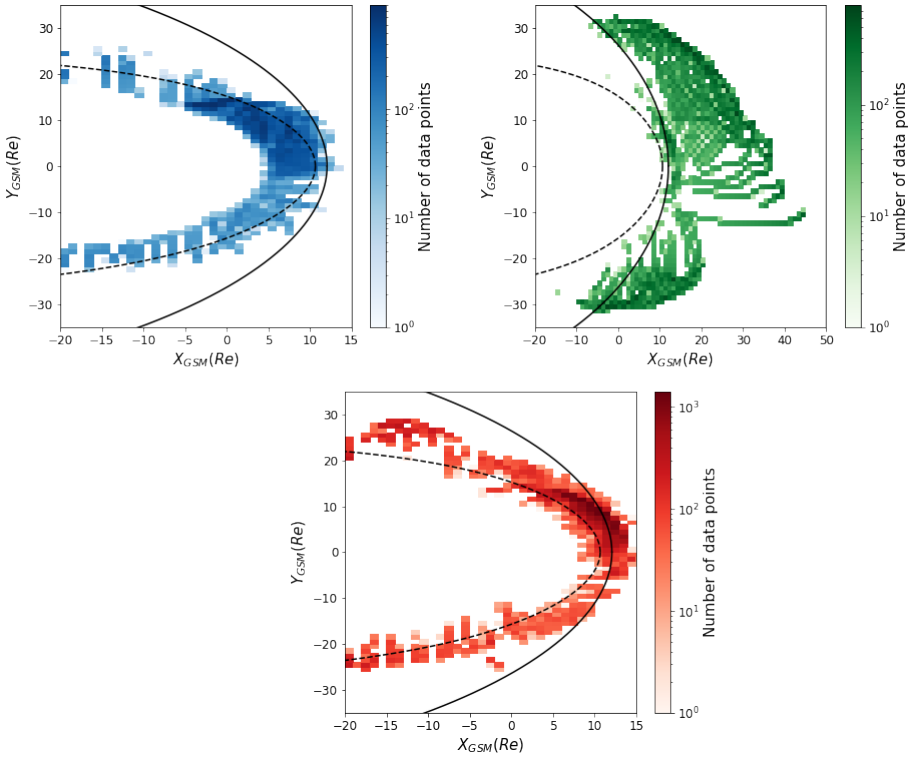

Figure 2. Spatial coverage of our labeled THEMIS dataset projected in the (X-Y) GSM plane, the solid black line represent a stand-off position the bow shock following (Jeřáb et al., 2005) model while the dotted black line represent the magnetopause model of (Lin et al., 2010). Labels are spatially represented in a log-scale 2D histogram. Magnetosphere bins in blue vary between 1 and 901, Magnetosheath bins in red vary between 1 and 1 421, solar wind bins in green vary between 1 and 788

The typical labeling of the three regions for a 1 minute resampled data interval is shown on the last panel of Figure 1 where the theoretical label, shown in blue has been slightly shifted vertically for visualization purposes. With modern visualization and data science tools (Wes McKinney, 2010; Génot et al., 2010), the interval selection and labeling of all the points enclosed is an easy and fast task, in particular because the regions are easily identified visually and without much ambiguity. Following this process, our dataset is made of 59 798 magnetosphere points, 48 056 magnetosheath points and 150 415 solar wind points.

We selected data measured during the dawn, dayside and dusk operation phases of THEMIS and thus expect a good Magnetic Local Time (MLT) coverage of both magnetopause and shock surfaces. This is confirmed by the spatial coverage of our labeled dataset shown in Figure 2. With such coverage, we expect the method to be robust enough regarding the variability of the data through the three different THEMIS operation phases.

In the following, we will designate the subset that has been used to fit our algorithm by *training set*. We will designate by *test set* the remaining subset of data that is used to evaluate the performance of our model. For each of the configurations we have been testing our algorithm with, the *training set* represents 70% of the dataset while the *test set* represents the remaining 30% of the dataset.To ensure there is no bias in our evaluation of

the performance, the train and test sets are chosen in distinct time intervals o the THEMIS B mission.

4 Algorithm

The recently developed automatic near-Earth classification routines were all based on the application of a neural network algorithms (Breuillard et al., 2020; Argall et al., 2020; Olshevsky et al., 2019). These deep learning methods typically offer a great flexibility leading to good results on complex problems, at the cost of requiring lots of training data points and long convergence times. In this study, we choose to train a Gradient Boosting algorithm(Friedman, 2001). Such a method may not offer as much flexibility as deep learning methods, but has been recognized to perform well on complex, eventually imbalanced classification problems (Brown & Mues, 2012) while typically needing much less labeled data and being much lighter to train. The Gradient Boosting is based on the iterative fit of the residuals obtained by the successive training and predictions made by weak learner algorithms, here a decision tree. The final prediction results from the convergence of the ensemble of decision trees on the smallness of the residuals or when the maximum number of trees is reached, and corresponds to the class with the highest probability.

Since our massive prediction relies on this probabilistic output, and since the Gradient Boosting is known to result in possibly not-well calibrated probabilities (Niculescu-Mizil & Caruana, 2005), we calibrate it before performing the massive prediction. We show in Appendix B that our model is well-calibrated and that its probabilistic output can then be used as is.

We computed the method using its Python implementation provided by Scikit-Learn (Pedregosa et al., 2011) with the default hyperparameters[3]. With the actual size of our dataset, it took two minutes for our algorithms to train on an AMD ryzen ^{TM}threadripper ^{TM}2990wx processor.

5 Results

5.1 Performances

After fitting our Gradient Boosting model to our *training set*, we evaluate its performance by comparing its prediction on the *test set* with the corresponding labels. The value of the prediction for a given time interval can be shown in the last subplot of Figure 1. From then on, a prediction made by our model can be split into four categories for each class:

- A true positive (TP) is a point of a class that has been predicted correctly as such,
- A true negative (TN) is a point not belonging to the concerned class that has been predicted correctly (e.g a magnetosheath point that has not been predicted as a magnetosphere point when considering the magnetospheric case)
- A false negative (FN) is a point of a class that has not been correctly predicted as such (e.g a Magnetosphere point that has been predicted as a Magnetosheath point)
- A false positive (FP) is a point not belonging to the concerned class that has been predicted as belonging to the class (e.g a Magnetosheath point that has been predicted as a Magnetosphere point when considering the magnetospheric case)

[3] described here: `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html`

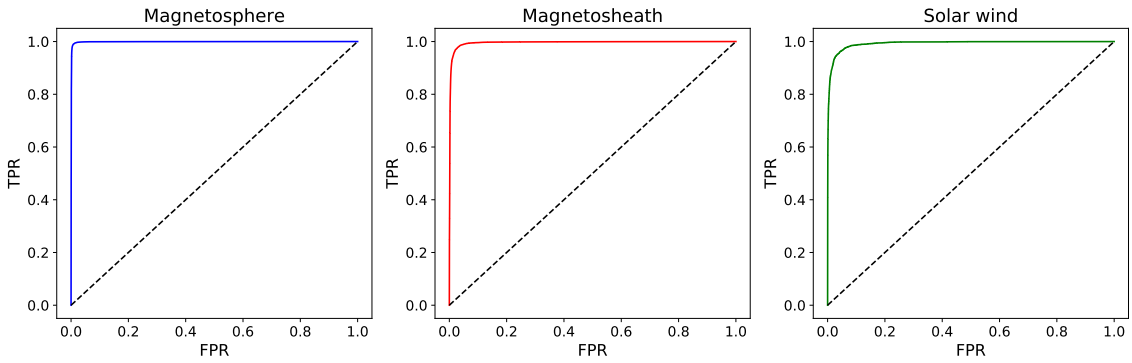

Figure 3. ROC curve of our model trained and predicting on THEMIS B data in the case of a temporal split between the *training set* and the *test set*. From left to right are represented the ROC curve concerning the magnetosphere, the magnetosheath and the solar wind.

With these four categories, we can define the *true positive rate* TPR as the ratio between the number of TPs over the total number of expected positives points:

$$TPR = \frac{N_{TPs}}{N_{TPs} + N_{FNs}} \tag{1}$$

The *false positive rate* FPR is defined as the ratio between the number of FNs over the total number of expected negative points:

$$FPR = \frac{N_{FPs}}{N_{FPs} + N_{TNs}} \tag{2}$$

An ideal model would be a model without any FN or FP. In this case, we would then expect the TPR to be equal to 1 and the FPR to be equal to 0 for the three classes. These two values are obtained for a given decision threshold based on the predicted probability as explained in the previous subsection. Logically, low decision thresholds would imply more points predicted as belonging to a certain class and then raise both FPR and TPR. By contrast, higher decision thresholds would decrease the number of positive points and thus decrease the FPR and the TPR.

The evolution of the TPR as a function of the FPR for continuously varying decision threshold can be represented as the Receiving Operator Curve (ROC) shown in the Figure 3 for the three classes. As expected, we notice an increasing TPR with an increasing FPR. The main interest in this curve stands in the inflexion point that correspond to the best compromise we can find between low FPR and high TPR. We want this point to be as close to the top left of each curve as possible as this would imply a FPR close to 0 and a TPR close to 1. A random classifier would, for each decision threshold, increase the TPR and FPR by the same amount and the associated ROC curve would then be a straight line of slope 1 as shown with the black dashed lines of Figure 3.

The quality of the ROC can be quantified by computing the area under curve (AUC) of each of the ROCs and we then expect this AUC to be as close to 1 as possible. To ensure the independence of the result from the split we made between *training* and *test* set, we trained and predicted our model 10 times for 10 different splits and computed the AUC. The average AUC we obtained for each class is shown on the first row of Table 1. At first, the high AUC obtained for the three classes indicate how well our model perform in classifying the three regions. Moreover, the standard deviation we obtain is lower than $1e-3$. This shows that our method is independent from the split we make between our two sets.

Mission	AUCMagnetosphere	AUC Magnetosheath	AUC Solar Wind
THEMIS	0.999	0.997	0.999
Cluster 1 (without retraining)	0.988	0.983	0.996
Cluster 1 (with retraining)	0.999	0.998	0.999
Double Star TC1 (without retraining)	0.996	0.992	0.996
Double Star TC1 (with retraining)	0.999	0.998	0.999
MMS (without retraining)	0.997	0.994	0.995
ARTEMIS	0.999	0.999	0.999

Table 1. Comparison of the AUC of the ROC of our detection algorithms for different missions.

Compared with the short required training time, this legitimates our initial choice of fitting a gradient boosting classifier.

In addition to this metrics, we can define the *Heidke Skill Score* HSS that compares the performance of our algorithm to what would come from a random classifier:

$$HSS = \frac{\frac{N_{TPs}+N_{TNs}}{N} - \frac{(N_{TPs}+N_{FNs})*(N_{TPs}+N_{FPs})+(N_{FN}+N_{TNs})*(N_{FP}+N_{TN})}{N^2}}{1 - \frac{(N_{TPs}+N_{FNs})*(N_{TPs}+N_{FPs})+(N_{FN}+N_{TNs})*(N_{FP}+N_{TN})}{N^2}} \quad (3)$$

where N denotes the total number of points on which the prediction was lead. A negative HSS indicates randomness performs better than the classifier while a perfect forecast would be associated to a HSS of 1. The Evolution of the HSS for varying probabilistic decision threshold is shown in Figure 4. The high value reached by the HSS for each class for a wide range of decision thresholds confirms the efficiency of our model. Finding decreasing HSS for high decision threshold is not surprising as the number of FN will slightly increase in this case. The main interest in this curve then stands in the value of the HSS we do find for the decision threshold we set for our prediction, that is to say 0.5. The value of the HSS we have in this case is shown in Table 3 and finding it pretty close to 1 indicates how well our model performs in the classification of the three regions.

5.2 Influence of the manual labeling

The manual labeling process can be an important source of prediction errors. Thus, the label can eventually contain errors that could affect the quality of our prediction and high AUC would then not indicate the classification ability of our model but its ability to learn from an erroneous label. To figure this out, we perform trainings and evaluations of the algorithm by voluntarily mislabeling an ever-growing percentage of the dataset. If our model completely follows the indicated label in the training set, we expect a high AUC whatever this percentage might be. The mislabeling process is done as follows:

- We select a fraction of random points in the dataset
- The magnetosphere and the solar wind points are labeled as magnetosheath points
- Magnetosheath points are randomly mislabeled between the two other classes

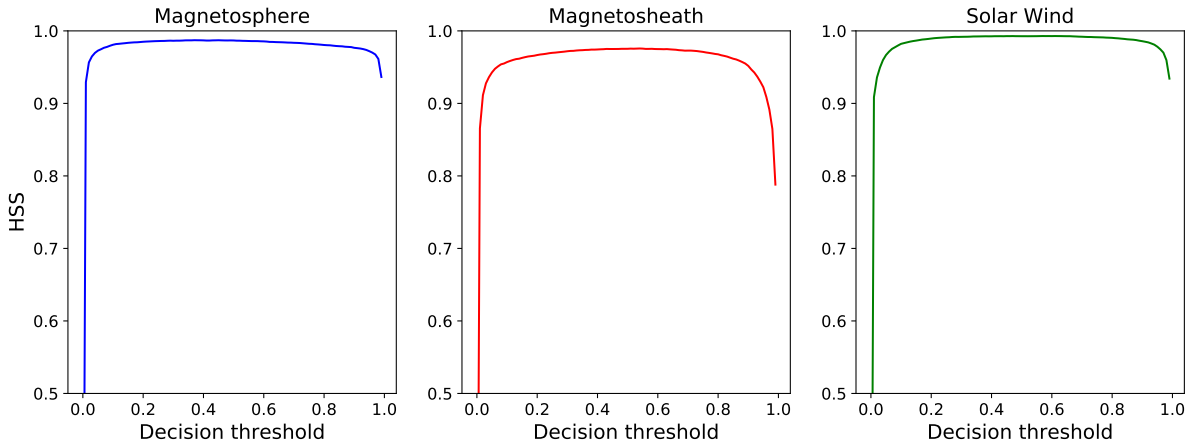

Figure 4. Evolution of the Heidke Skill Score as a function of the decision threshold for our model trained on THEMIS data for a temporal split. . From left to right are represented the HSS curve concerning the magnetosphere, the magnetosheath and the solar wind.

The main reason that justifies this process stands in the fact that a human observer will never confuse magnetosphere and solar wind and and the fact there is of course some ambiguity in the labeling for classes concerned with a physical interface, where data points do not strictly belong to either one or the other but rather represent the finite transition region, omitted in our model. We repeat the process for an ever growing percentage of the dataset until the proportion of the mislabeled points reaches 50% of the dataset. The random mislabeling and associated training and AUC computation are repeated 10 times at each step. The evolution of the AUC with the mislabeling proportion is shown in Figure 5 for the three classes of the THEMIS dataset. The grey dashed lines represent the standard deviation we have between the different iterations of a given percentage of mislabeling.

Having a more significant drop in the performances for the magnetosheath is not surprising as this is the class that will be most affected by our mislabeling process. Noticing that drop for the three different classes proves the model does not simply follow the indications provided from the labels but tries to find an intrinsic difference in the physical parameters of the three classes.

This shows the real capacity of our algorithm to classify the three near-Earth regions as well as the reliability of our label.

5.3 Comparison with other algorithms

As we just saw in the previous subsections, gradient boosting performs well after a very short required training time. This indicates that more complex algorithms such as neural networks (Argall et al., 2020; Breuillard et al., 2020; Olshevsky et al., 2019) are not necessary for this classification task. However, one legitimate question could be to ask whether even lighter machine learning algorithm would perform as well. Therefore, in addition to the gradient boosting classifier, we train and evaluate the performances of two simpler classification algorithms: a Logistic Regression (Berkson, 1944) and a Classification Tree (Breiman et al., 1984). From the AUC and HSS averaged over three different temporal splits that are shown for each class and each algorithm in Table 2, Gradient Boosting appears as being the algorithm that performs best on differentiating the three regions. However, it should be noted here that even the simplest algorithms result in fair performances already after a training phase as long as the one we have for Gradient Boosting.

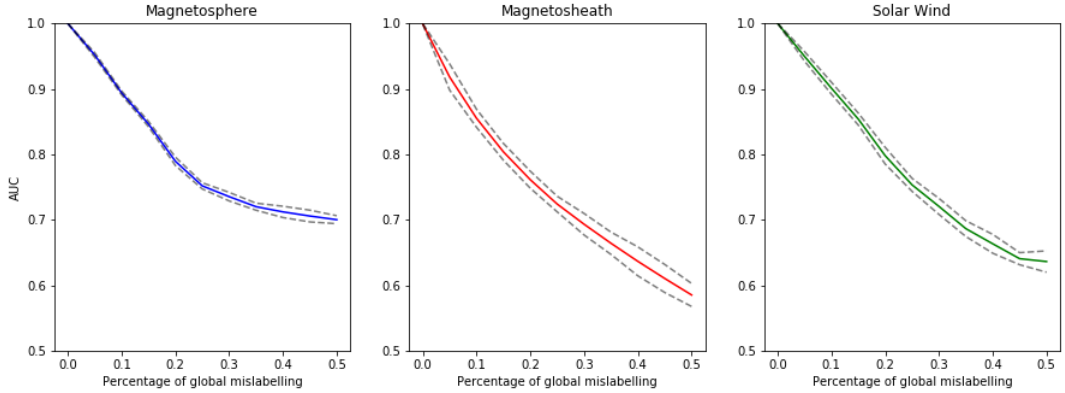

Figure 5. Evolution of the AUC as a function of the mislabeling ratio for the three different classes: magnetosphere (blue), magnetosheath (red) and solar wind (green). The gray dashed line represent the standard deviation we have between the different AUC scores of a same mislabeling percentage.

	Logistic Regression	**Decision Tree**	**Gradient Boosting**
AUC magnetosphere	0.998	0.976	**0.999**
AUC magnetosheath	0.954	0.937	**0.997**
AUC solar wind	0.937	0.881	**0.999**
HSS magnetosphere	0.974	0.953	**0.987**
HSS magnetosheath	0.846	0.878	**0.975**
HSS solar wind	0.560	0.701	**0.992**

Table 2. AUC and HSS obtained for different algorithms for several train-test split.

6 Adaptability

Having trained an algorithm to detect the three near-Earth regions with high reliability, we should not have difficulties to adapt it to the data provided by additional spacecraft that go through those regions. Even if a similar work can be adapted on the numerous past missions that went through the three near-Earth regions, we focus here on the most recent missions that offer the advantage of providing the data with the best quality, which removes an additional complexity that would appear with older missions.

To do so, we label data points of each of the missions we are working on and compare this label to the predictions of our model trained with THEMIS data.

6.1 Double Star

We use the data of the TC1 spacecraft on the whole mission period (between the 1st of January 2004 and the 1st of january 2008) The magnetic field data are provided by the Fluxgate Magnetometer (Carr et al., 2005) with a temporal resolution of 2s. The plasma moments are provided by the CIS-HIA instrument (Fazakerley et al., 2005) with a temporal resolution of 4s. Just like for the THEMIS dataset, we resampled the data to a 1 minute resolution.

A typical representation of the data is shown in Figure 6.

Here the part of the data we labeled is made of 20671 magnetosphere points, 23091 magnetosheath points and 4944 solar wind points taken at the beginning of the year 2005. The main reason explaining the noticed imbalance in the data stands in the orbit of TC1 itself that is not supposed to cross the bow shock.

The spatial distribution of our labeled data is also shown in Appendix A.

Since Double Star also has an equatorial orbit, we expect the model trained on THEMIS to perform well even without having to be retrained and this is the main reason why our label does not have to provide an entire coverage of the (X-Z) plane. And this is confirmed by the high AUC and HSS we have in Tables 1 and the comparison of the HSS obtained for the different missions shown in the Table 3.

Refitting the model would then allow a finer detection that would be specific to the quality of the data provided by Double Star in comparison to the THEMIS data but can be skipped as it does not bring a significant gain in AUC according to Table 1.

6.2 MMS

We used the data of the MMS 1 spacecraft between September 2015 and July 2019. The magnetic field data were provided by the Fluxgate Magnetometer (Russell et al., 2016) with a temporal resolution of 4.5s. The plasma moments were provided by the Fast-survey mode of the Fast Plasma Investigation instrument (FPI, Pollock et al. (2016)) with a temporal resolution of 4.5s. Just like THEMIS, the data were resampled to a 1 minute resolution.

A typical representation of the data is shown in Figure 7.

Since MMS also has an equatorial orbit, we once again expect the model trained on THEMIS to provide a very good classification of the three regions on MMS data as for the case of what has been shown for Double Star.

To figure it out, we label 7 612 magnetosphere points, 19 272 magnetosheath points and 3 651 solar wind points during the first year of MMS and these labels the associated prediction of the classifier. The spatial coverage of these labeled points is shown in Figure A2

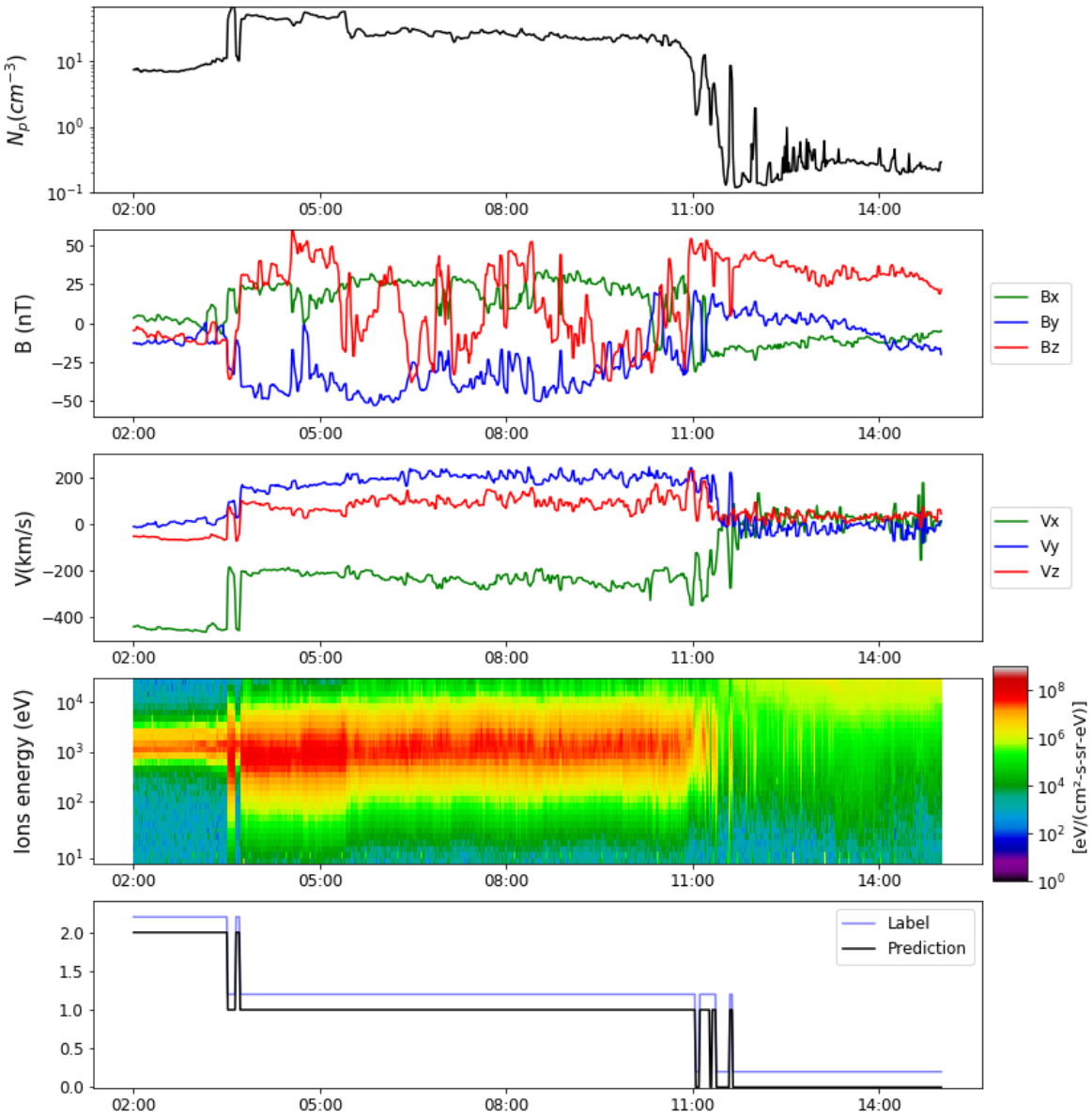

Figure 6. In-situ measurement provided by Double Star TC1 spacecraft on the 1^{st} of January 2005. The legend is the same than in 1

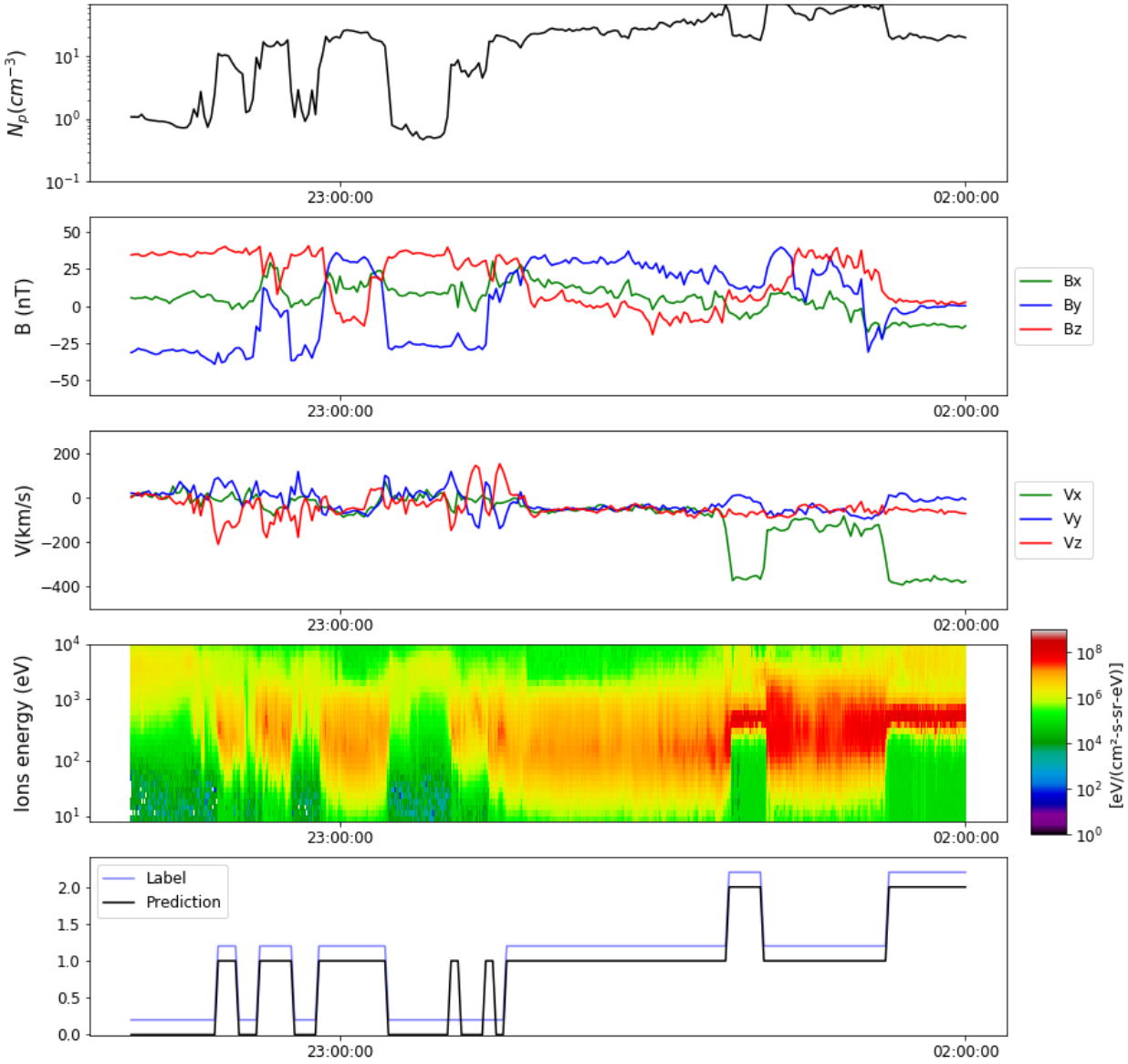

Figure 7. In-situ measurement provided by MMS spacecraft on the 31^{st} of December 2015. The legend is the same than in 1

The high AUC and HSS shown in the Tables 1 and 3 confirms the adaptability of our classifier to equatorial missions without further additional fitting.

6.3 Cluster

We use the available data from Cluster 1 spacecraft between the 1st of January 2001 and the 1st of January 2013 and from Cluster 3 spacecraft between the 1st of January 2001 and the 1st of December 2009. The magnetic field data are provided by the Fluxgate Magnetometer with a temporal resolution of 4s (Balogh et al., 2001). The plasma moments are provided by the Hot Ion Analyzer instrument (Rème et al., 2001) when the instrument is working under the magnetosphere or the magnetosheath mode. Here again, the data is resampled to a 1 minute resolution.

In comparison with Double Star and MMS, this case might be more challenging because of the orbit, here polar, and the regions visited that have different physical properties than the one visited by equatorial missions. The data provided THEMIS and Cluster can therefore be substantially different and there is no real clue on how an algorithm trained on equatorial orbit data would perform on predicting on polar orbit data.

One minute sampled Cluster data are shown in Figure 8 and we here label 50 277 points of magnetosphere, 76 468 points of magnetosheath and 22 017 of solar wind between the years 2005 and 2006 which spatial distribution is shown in Figure A3. Those two years will constitute the time period on which we will test the adaptability of the region classifier. One third of these labeled points are used to evaluate the performances of the models while we kept the remaining two thirds in the case refitting the algorithm is needed. Applying our THEMIS-trained model, we notice a lower AUC for each of the three classes. This indicates the difficulties the classifier has to adapt to polar orbit data.

We then adapt our classifier to the polar case by refitting the model trained on THEMIS with the Cluster labels. The increasing AUC we obtained, shown in Table 1 and the associated high HSS also shown in 3 proves the necessity we had to adapt our algorithm to the specificity of the Cluster data. It also shows our model can be easily adapted to the data of another mission, exploring regions with significant statistical deviations of the features, after a small labeling and refitting phase.

6.4 ARTEMIS

The mission ARTEMIS actually corresponds to the THEMIS B and C spacecraft when they were moved from a terrestrial to a lunar orbit at the end of 2009. The data we used in this case are then the one provided by the same THEMIS B instruments than in section 2 between the 1st of January 2010 and the 1st of June 2019.

The orbit of the ARTEMIS spacecraft is different from the orbit of the mission we have been investigating so far. This difference comes with a lot of change in the nature of the data measured by the spacecraft.

First of all, the spacecraft orbit the moon and are much farther (around 60 Re) from the Earth than the spacecraft of the other missions we have used. This implies the spacecraft does not explore the dayside regions and crosses the magnetopause and the bow shock in the nightside. At these distances, the magnetosheath plasma becomes almost as fast and as tenuous as the solar wind and small magnetosheath fluctuation scould easily be confused with either a magnetopause or a bow shock crossing.

Second, the spacecraft spend most of their time in the solar wind, which may make statistical properties of their measurements more sensitive to the data variability induced by the solar cycle that we neglected for the previous missions.

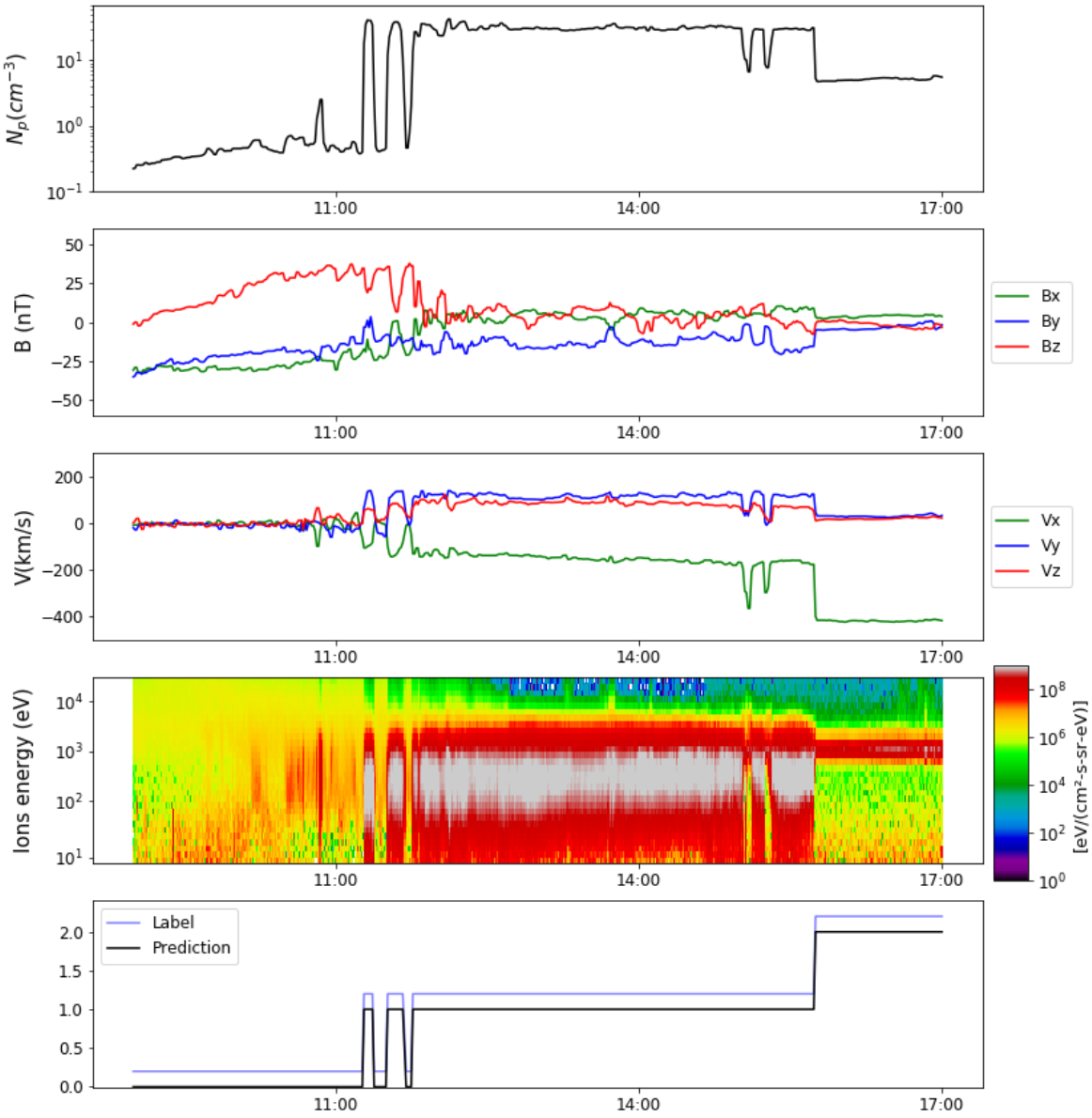

Figure 8. In-situ measurement provided by Cluster 1 spacecraft on the 6^{th} of February 2005. The legend is the same than in 1

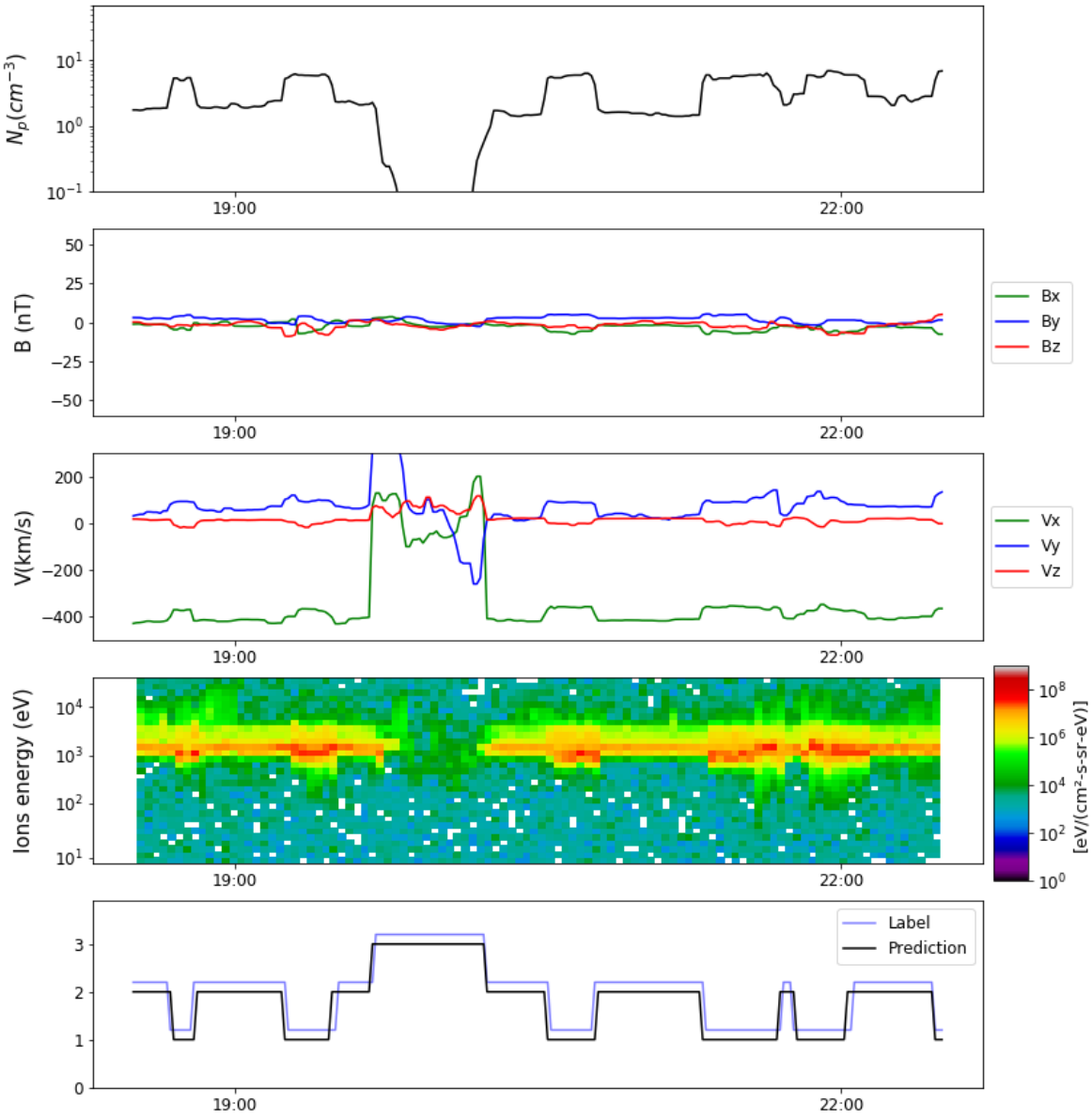

Figure 9. In-situ measurement provided by the ARTEMIS B spacecraft on the 13^{rd} of August 2016. The legend is the same than in 1

Finally, this specific type of orbit also introduces time intervals during which the data does not take values statistically close to any of our regions of interest. Indeed, once per orbit, ARTEMIS explores the lunar wake, characterized by an extremely low density and fluctuating velocity in many directions. These intervals, for which a typical representation of the data is shown in Figure 9, cannot be considered to belong to any of our existing region classes.

For this three reasons, the method we presented in the previous sections and successfully adapted to Double Star, MMS and Cluster cannot be used as is and the entire process from the labeling to the choice of the feature has to be designed from scratch.

Mission	HSS Magnetosphere	HSS Magnetosheath	HSS Solar Wind
THEMIS B	0.987	0.975	0.993
Cluster 1	0.976	0.972	0.981
Double Star TC1	0.980	0.974	0.983
MMS	0.982	0.973	0.987
Artemis	0.976	0.962	0.974

Table 3. Comparison of the HSS of our detection algorithms for different missions.

To cope with the variability induced by the solar cycle we label a month per year. We furthermore add the lunar wake as a fourth explored region (with an associated value of 3). The final labeled dataset is made of 26 560 magnetosphere points, 131 656 magnetosheath points, 429 283 solar wind points and 15 070 points of lunar wake which spatial distribution is shown in Figure A4.

We cope with the increasing difficulty to distinguish magnetosheath and solar wind by adding the spacecraft GSM coordinates as a feature of the dataset which will then consist in 11 input variables.

Having a different dataset and a different number of classes, we here cannot use the model trained in the previous section and we will then focus on the specific model we trained for this mission. The resulting high AUC shown in Table 1 shows the gradient boosting also performs well in a significantly different region and with more classes. This especially confirmed with the AUC and the HSS we found for the lunar wake region, that we respectively found equal to 0.97 and 0.947.

7 Comparison with manually-set thresholds

Having shown the efficiency of gradient boosting on different missions [4], we want to compare it to the state of the art non-learning methods such as the one elaborated by Jelínek et al. (2012) that we described in the introduction.

Figure 10 represents the 2D histogram of B and N_p for THEMIS B, Double Star and Cluster 1 on the periods on which we labeled the different associated datasets. We divided these parameters by the corresponding solar wind density and the IMF amplitude that we obtained from the OMNI data shifted from the actual measurement time using the two-step propagation algorithm described in Šafránková et al. (2002). At first sight, one can easily distinguish three main regions that are separated with the solid red lines for the three missions. Nevertheless, these linear boundaries have been set manually and we cannot ensure these could be the best choice for the three missions. To evaluate the quality of the classification, we compute the TPR and the FPR for the three missions and for varying boundary lines. We then use these values to compute the AUCs that are shown in the Table 4.

Once again, we notice a lower AUC in the case of Cluster which is consistent with the difference we have between equatorial and polar orbits as explained in the previous section. Additionally, even if the boundaries plotted in Figure 10 seem to provide a decent separation between the three regions, the AUC is lower than the one we obtained with the

[4] Additional prediction examples can be found in the appendix B.

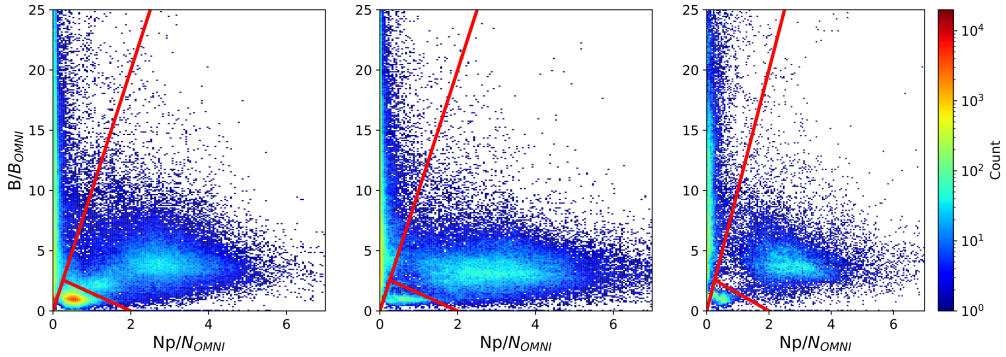

Figure 10. 2d histogram of B and N_p divided by the corresponding OMNI data for the three missions: THEMIS B (left), Cluster 1 (middle) and Double Star TC1 (right). The solid red lines indicate a possible set of linear boundaries we could define to separate the three regions

Mission	AUC Magnetosphere	AUC Magnetosheath	AUC Solar Wind
THEMIS B	0.915	0.908	0.859
Cluster 1	0.897	0.852	0.828
Double Star TC1	0.913	0.894	0.843

Table 4. AUC for the threshold-based method

gradient boosting. This indicates our model performs better in classifying the three regions by setting more flexible boundaries on supplementary features while requiring less fitting time than the one required to manually set the thresholds used in the Figure 10.

The same kind of histogram gets messier with a much less obvious transition from the magnetosheath to the solar wind and the addition of the moon's wake as shown with the ARTEMIS data in Figure 11. This shows the difficulty manually set thresholds would have for far night side oriented missions and the interest of using machine learning in this case.

8 Massive detection of boundary crossings

In the previous sections, we have shown the efficiency, the reliability and the adaptability of our classifiers on data from several missions and spacecraft. From now on, these classifiers can be used to elaborate our magnetopause and bow shock crossings catalogs by classifying the in-situ data provided by any near-Earth spacecraft and by selecting time intervals enclosing two predicted regions. To do so, we train our 3 different models, THEMIS, Cluster and ARTEMIS on their whole labeled datasets [5]. In the following, these models will be respectively named the equatorial, high-latitude and lunar models.

In addition to the performances on the labeled dataset of the different missions in use in this paper, shown in the previous sections, we make sure the massive prediction of the 3 models are consistent with the labeled data by comparing the physical characteristics of the classified and the labeled data points of each regions. Figure 12 compares the average prediction of the three different models to their associated average training label for each bin of the (N_p, B) plane.

[5] Those trained models can be found at `https://github.com/gautiernguyen/in-situ_Events_lists`

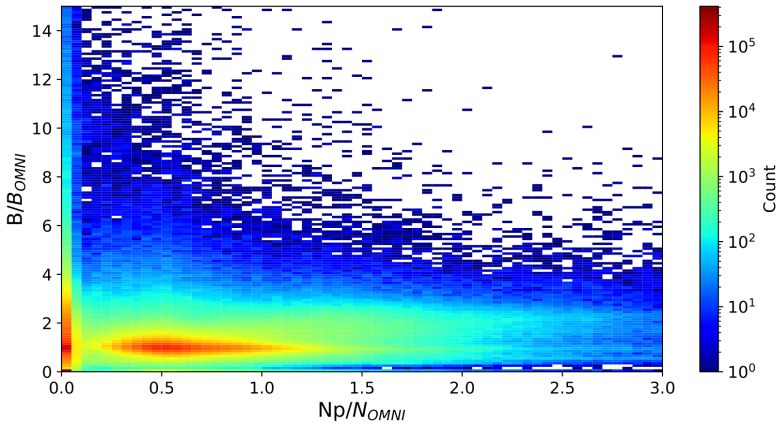

Figure 11. 2d histogram of B and N_p divided by the corresponding OMNI data for ARTEMIS B

At first, the binned average of the labels, shown in the right column exhibits three main zones where there is almost no doubt on the region visited by the spacecraft. The transition from a zone to another appears as a zone where the label is transient consistently with what is expected for the crossing of one of the near-Earth boundaries. It is worth noting that the transitions between the different colored regions are far from being linear and this is even more true when we look at the distributions for the polar and lunar models. Following the discussion of section 7, this confirms the limits of an automatic detection method based on the manual setting of thresholds on the densities and magnetic field amplitudes and give a further support in favor of the application of machine learning for such classification task. Despite an expected increased noise, the pattern in the left column is similar to the one in the right column. This suggests that the massive prediction obtained from the equatorial, the polar and the lunar model is consistent with our labeled data. For both the equatorial and the polar model, we notice bins at low densities for which the average prediction is rather equal to magnetosheath or solar wind than magnetosphere. This indicate the presence of low density datapoints that have either been classified as magnetosheath or magnetosphere. However, these bins all contain less than 100 datapoints and are actually within the margin of erro r of our model.

8.1 Magnetopause catalog

We then elaborate a complete magnetopause crossing catalog by running the THEMIS classifier on the data provided by THEMIS A, B, C, D and E spacecraft. To gain time in the construction of the crossings and because we do not expect any magnetopause crossing in the nightside operation phase, we restrict ourselves to the dayside, dawn and dusk operation phase. As no crossing is expected far away in the solar wind or close to the Earth dipole, we also only the parts of the spacecraft orbit that were less than 5 Re away from the magnetopause distance predicted by Lin et al. (2010) for a dynamic pressure of 2 nPa and a null IMF B_z. The raw predictions of the classifier are then smoothed by applying a median filter with a window size of 3 minutes. We then define a magnetopause crossing as a 1 hour interval that contains as many magnetosheath points as magnetosphere points making sure every detected events were disjoints.

The same model is used on the in-situ data provided by Double Star between 2004 and 2007 and MMS between 2015 and 2020.

We finally apply the same process on the in-situ data provided by Cluster 1 on the 2001-2013 period, by Cluster 3 on the 2001-2009 period and on ARTEMIS between 2010 and

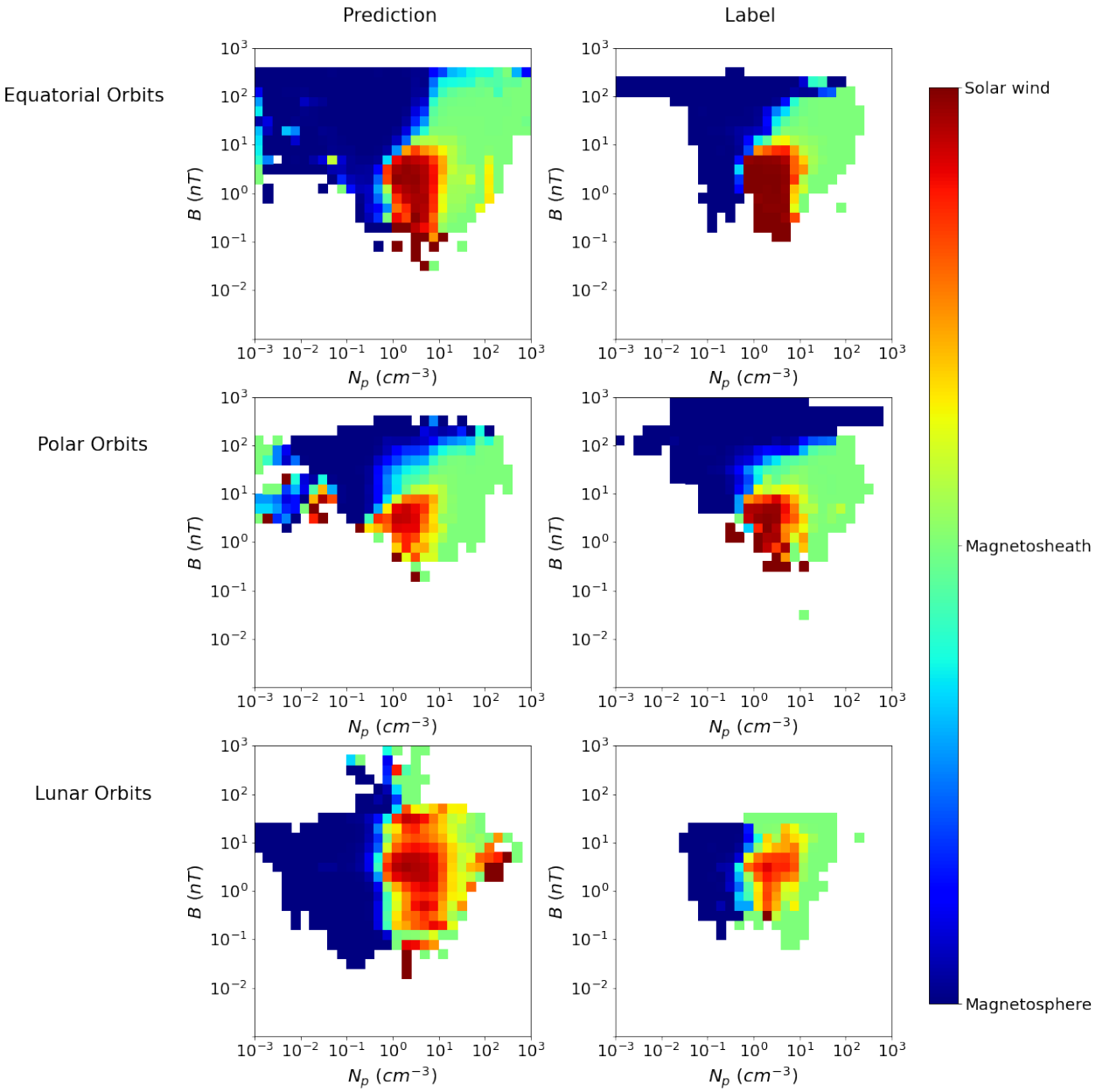

Figure 12. Binned average of the massive prediction (*left column*) and the training label (*right column*) of the equatorial (*first row*), the polar (*second row*) and the lunar (*third row)* models projected in the (N_p, B) plane

Mission	Magnetopause crossings	Bow shock crossings
THEMIS A	2 822	1 590
THEMIS B	376	1 030
THEMIS C	670	1 238
THEMIS D	2 705	1 520
THEMIS E	2 738	1 511
Cluster 1	1 782	3 225
Cluster 3	1 571	2 004
Double Star TC1	891	846
MMS 1	805	1 035
ARTEMIS B	215	1 602
ARTEMIS C	487	1 626
Total	15 062	17 227

Table 5. Number of magnetopause and bow shock crossings we have for different missions

2019 by using the corresponding trained model. The total number of crossings we obtained are summarized in the Table 5.[6]

Our detection method has been evaluated on regions where the spacecraft is not expected to cross a boundary. In these regions, the algorithm is less likely to hesitate on its prediction. On the other hand, it is more probable it hesitates on the predictions made close to the boundaries. Consequently, we have to ensure the classification is still of decent quality there.

Figure 13 represents the ROC we have on the classification between magnetosphere and magnetosheath points for THEMIS B, Cluster 1 and Double Star for the subset of our test set that lies in the proximity of a magnetopause or shock crossing. These predictions have been obtained with a model that has been trained with the complement part of the dataset, i.e. the subset that excludes the proximity of the crossings. The obtained AUC is here lower than the one presented in the previous sections. This can be explained by the fact, the manual labels made in this region is more ambiguous than the one made in the parts of orbit that are far from one of the boundaries, resulting in a more hesitant classifier. The AUC value is still high enough to ensure the good quality of the classification when a spacecraft arrives close to the magnetopause and thus our capacity of building crossings from the prediction made by our model.

We then computed the mean probability of each crossing by averaging the probabilities of belonging to the predicted class of each point present in the crossing.

Events with high probability would correspond to undoubtful crossings while the events with the lowest probability would be less likely to be actual crossings. The probability distribution of our 15 062 events is shown in Figure 14. Having a high probability for the greatest part of our events then ensures the consistency of our magnetopause list.

[6] All of the magnetopause lists can be found at the same address.

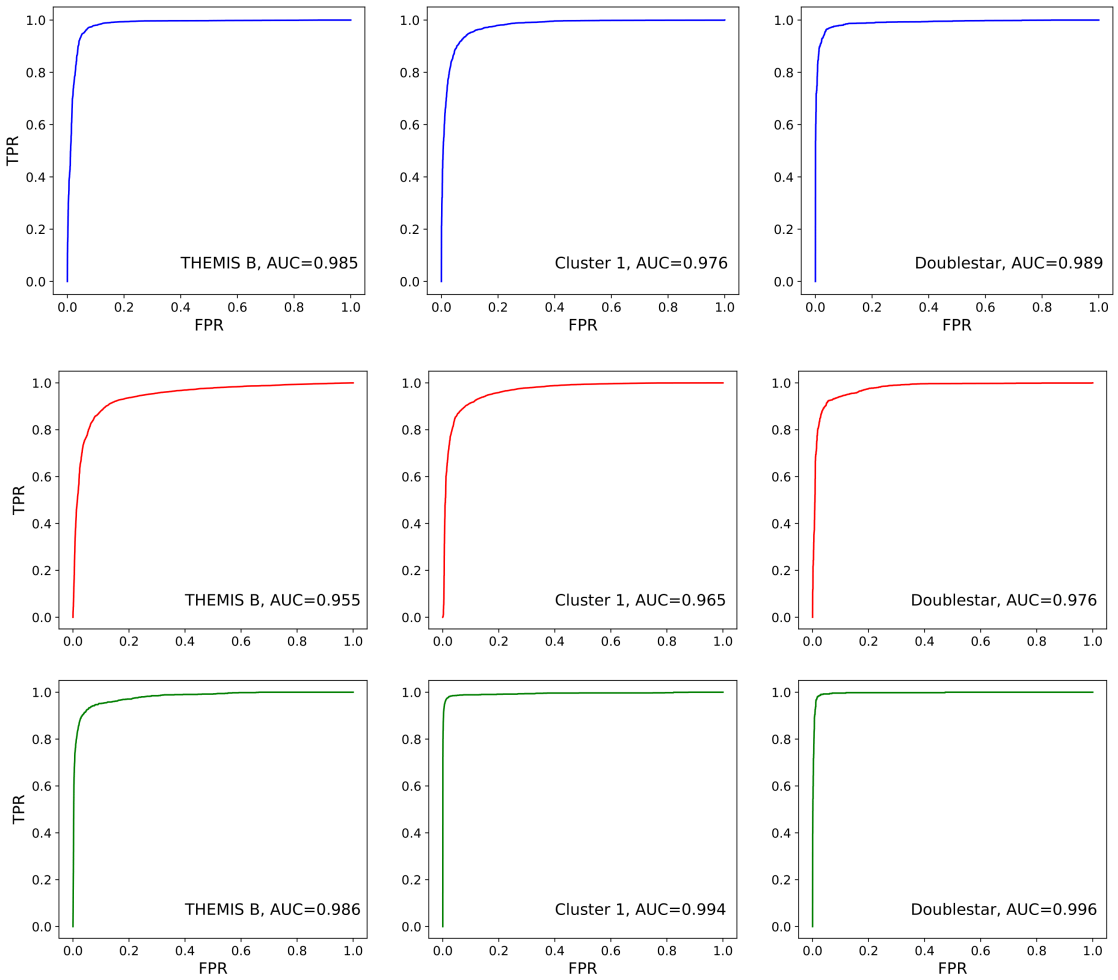

Figure 13. ROC curves evaluated on the labeled crossings for the three missions THEMIS B (left), Cluster 1 (middle) and Double Star (right) for the three classes: magnetosphere(top), magnetosheath(middle) and solar wind (bottom)

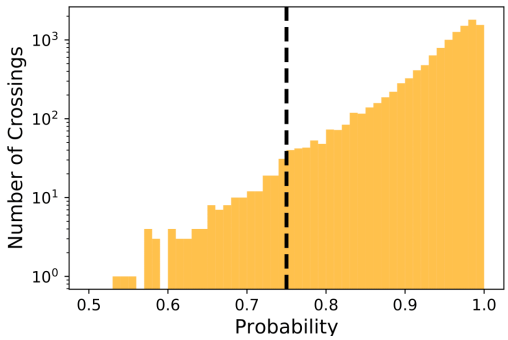

Figure 14. Distribution of the probability of the 15 062 magnetopause crossings we built and summarized in Table 5. The solid dashed line represent the probability threshold we chose for the Figure 15

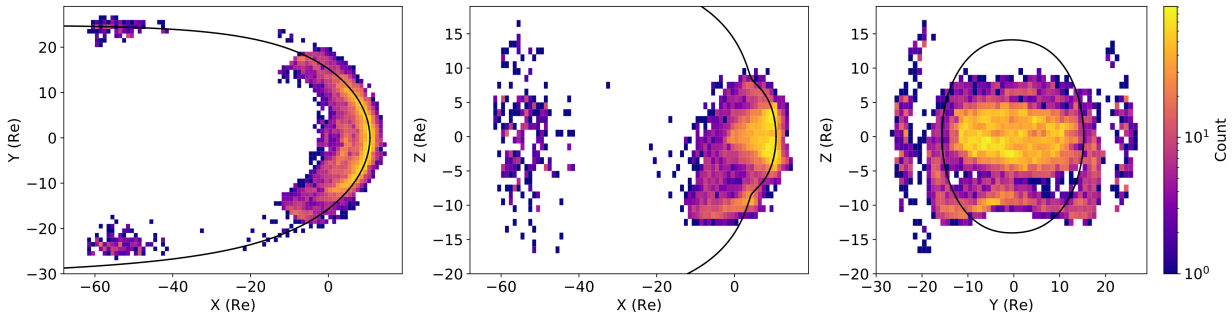

Figure 15. Spatial distribution of the crossings above the threshold in Figure 14 in the XY (left), XZ (middle), YZ (right) GSM planes. The solid black line indicate the Lin et al. (2010) magnetopause model with a dynamic pressure of 2 nPa and a null B_z.

Finally, the spatial distributions of the crossings that have a probability higher than 75% in the GSM XZ, XY and YZ planes is shown in Figure 15. These crossings represent 98.5% of the crossings built with our models and are then expected to be the most likely to be actual magnetopause crossings. The solid black lines represent the position of the magnetopause model established by (Lin et al., 2010) and computed for a dynamic pressure of 2 nPa, a null B_z and assuming no dipole-tilt. The proximity between this distance and our actual crossings ends up proving the capacity our method has to elaborate a magnetopause crossings catalog with a decent coverage of the surface at different latitudes and longitudes.

8.2 Bow shock catalog

We define a bow shock crossing event as 10 minutes interval that contains as much magnetosheath points as solar wind points. We then run the models we trained for the different missions detailed in Section 3 on the same dataset we used to elaborate the magnetopause crossing catalog. The total number of obtained crossings is once again summarized in Table 5 [7].

The spatial distribution of the crossings with a probability higher than 75% in the GSM XZ, XY and YZ planes is shown in the Figure 17. The solid black line here represents the location of the Jeřáb et al. (2005) bow shock model computed for a dynamic pressure of 2 nPa, a null B_z and an Alfven Mach of 8.

9 Conclusion

Using a Gradient Boosting Classifier, we established an automatic detection method of the different near-Earth regions as observed by the THEMIS spacecraft during the dawn, dusk and dayside mission phases. This method was successfully adapted on other equatorial dayside missions (Double Star and MMS) and, after a small retraining phase necessary to consider the orbital differences between different missions, was also successful on non-equatorial dayside missions such as Cluster. The adaptability of the method has even been tested on missions with a substantially different orbit such as ARTEMIS for which we provided a successful region classification after a small redesign of the observed features, implying the addition of the spacecraft GSM position, and the way the label was made, by considering an additional region with the lunar wake. Having proved this adaptability, we

[7] And the bow shock lists can once again be found at `https://github.com/gautiernguyen/in-situ_Events_lists`.

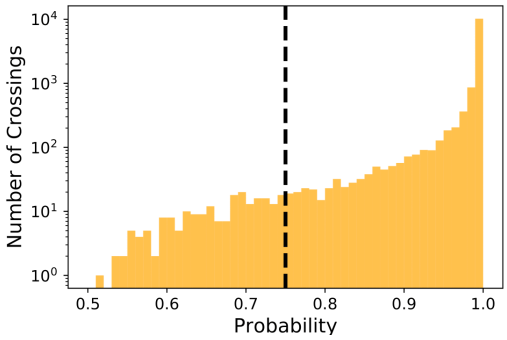

Figure 16. Distribution of the probability of the 17227 bow shock crossings we built and summarized in Table 5. The solid dashed line represent the probability threshold we chose for the Figure 17

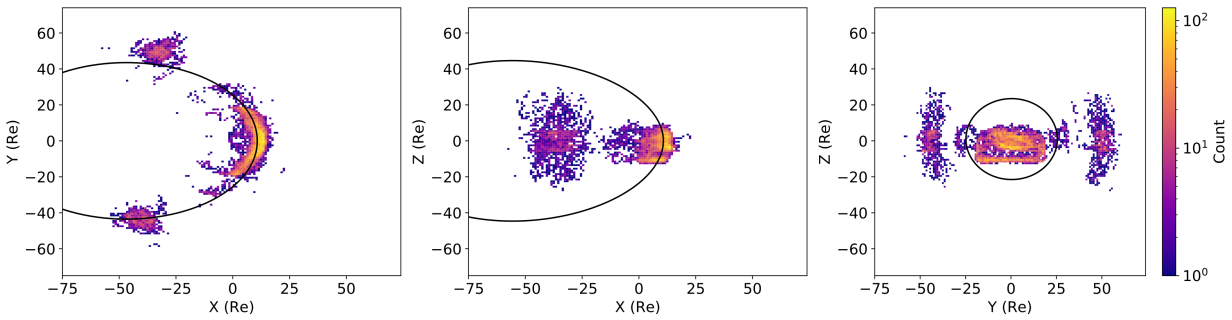

Figure 17. Spatial distribution of the crossings above the threshold in Figure 16 in the XY (left), XZ (middle), YZ (right) GSM planes. The solid black line indicate the Jeřáb et al. (2005) bow shock model with a dynamic pressure of 2 nPa, a null B_z and an Alfven Mach of 8.

may also think of using the data of additional near-Earth missions, such as WIND, Geotail, Hawkeye, Polar or Interball, provided enough information about the plasma moments with a sufficient resolution is provided.

For every mission we considered, our method outperformed the quality of the detection provided by manually-set thresholds and reached similar AUC values as the one achieved by neural networks based methods (Breuillard et al., 2020; Olshevsky et al., 2019; Argall et al., 2020), with the advantage of being much faster to train. Using the plasma moments paved the way to an easy adaptability from a specific type of mission to another and the production of light-weight algorithms that could eventually be taken onboard of upcoming missions to automatically select the data of interest and thus automatically decide the data that should be stored (and at which resolution) for further analysis. This wouldallow a significant gain in time regarding the data selection process that are either threshold triggered or human monitored like the Scientist In The Loop process in charge of the manual selection of MMS data. Moreover, the method does not use the specificity of being in the near-Earth environment and could also be adapted to other planetary missions in the solar system.

For simplification, we only considered 3 classes and defined as magnetosheath any region where plasma differed from pristine solar wind and magnetospheric ones. The classification could be enhanced by the consideration of additional regions depending on the scientific objectives.

We used this method to elaborate one of the most exhaustive public magnetopause and bow shock crossing catalogs to our knowledge. A bonus to our method is that these catalogs can be readily and automatically grown as new data is made available. Having a large list of events also gives the opportunity to study these two near-Earth boundaries and physical processes occurring in their vicinity, from a statistical point of view . One could especially think of exploiting the magnetopause crossings to provide an automatic detection method of the typical in-situ signature of accelerated plasma flow induced by magnetic reconnection, which will be the topic of a forthcoming study.

Last but not least, the fast and reproducible of one of the most exhaustive existing boundary crossings catalogs is the preamble for the massive statistical analysis of the position of the position of both the magnetopause and the bow shock for various solar wind and seasonal conditions. In the specific case of the magnetopause, this is the purpose to the three companion papers of this study (Nguyen et al., 2020a, 2020b, 2020c).

Appendix A Spatial distribution of the different labeled datasets

In this section, we represent the spatial distribution of the labeled dataset of Double Star (Figure A1), MMS (Figure A2), Cluster (Figure A3) and ARTEMIS (Figure A4).

Appendix B Probability calibration of the model

A well calibrated classifier is a classifier for which the probabilistic output gives a correct representation of the data seen by the algorithm. For instance, we expect 50% of the points predicted with a probability of 50% to be actual positives (either TP or FN). This verification is necessary as soon as the probabilistic output of a model is at stake. Nevertheless, boosted algorithms such as gradient boosting are known to have calibration issues (Niculescu-Mizil & Caruana, 2005). We then have to ensure our model is well-calibrated before using its probabilistic output in the way we did it in Section 5. To do so, we plot the Calibration curve shown in Figure B1that represents the evolution of the fraction of actual positive in our test set for each probability bin. For a perfectly-calibrated classifier, the calibration curves should be linear and stick to the black dashed line for the three classes. Having a linear calibration curve close enough to the perfect calibration curve for the three regions, we consider the probabilistic output of our model to be decently well-calibrated.

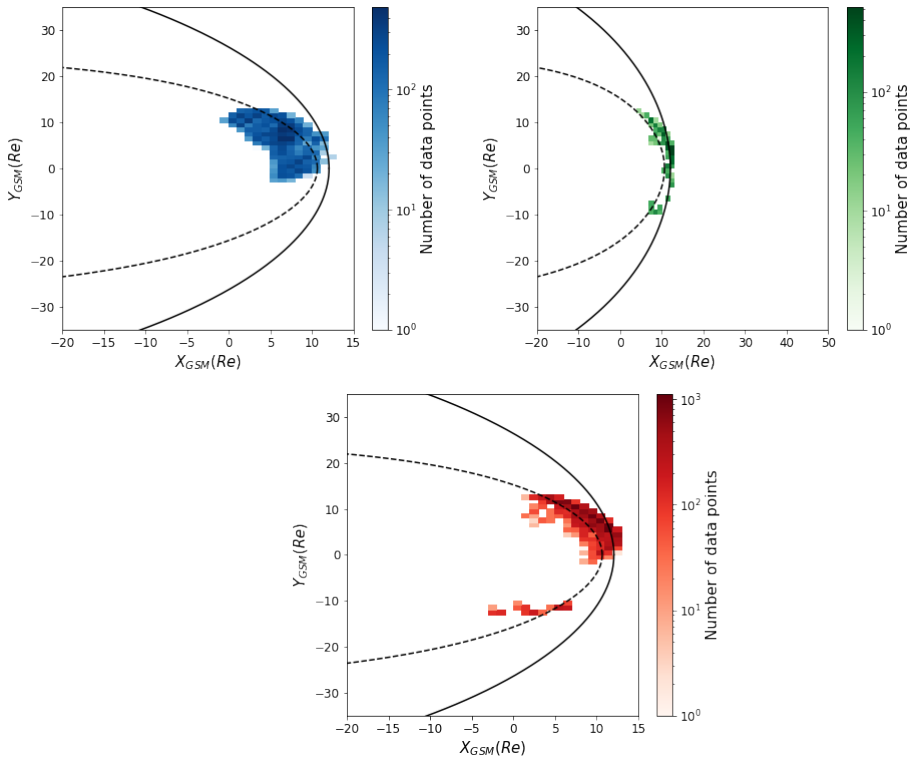

Figure A1. Spatial coverage of the Double Star labeled dataset. The legend is the same than in Figure 2

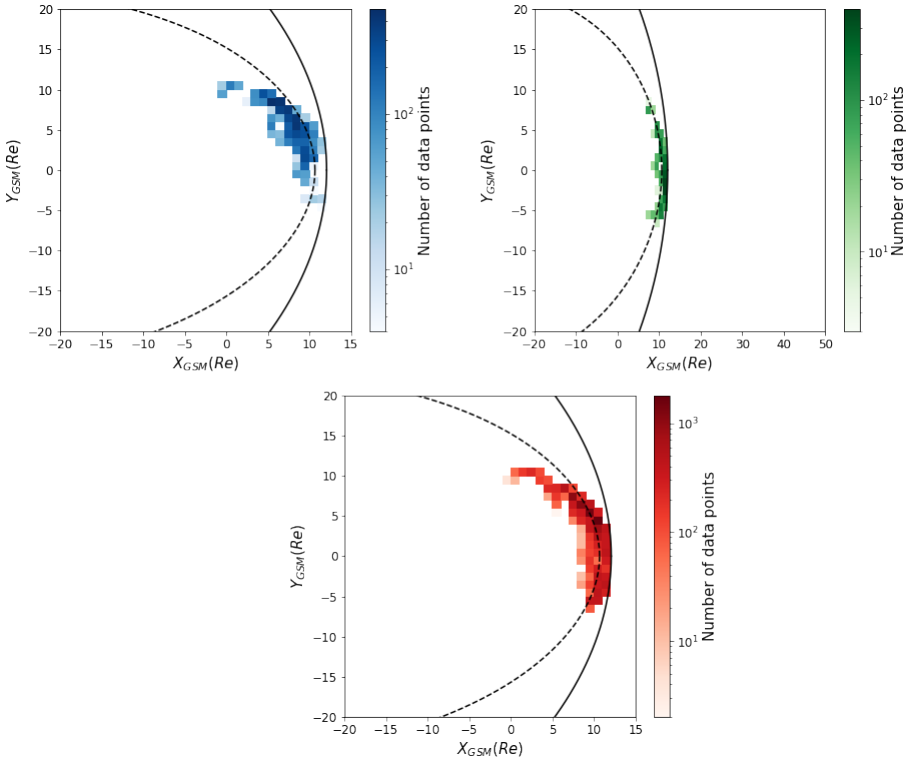

Figure A2. Spatial coverage of the MMS labeled dataset. The legend is the same than in Figure 2

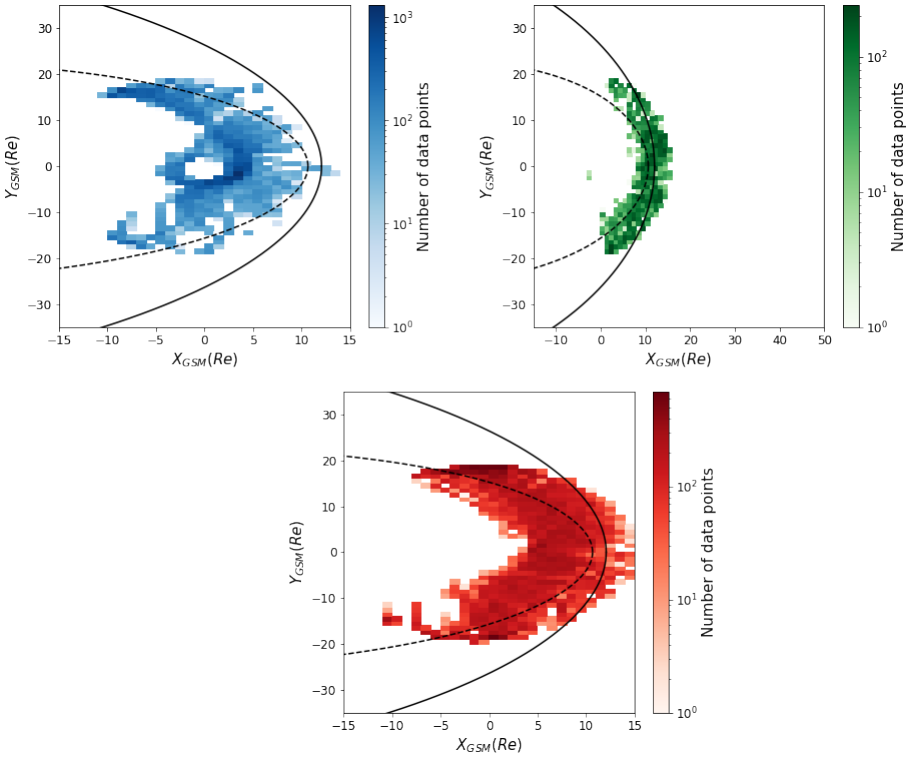

Figure A3. Spatial coverage of the Cluster labeled dataset. The legend is the same than in Figure 2

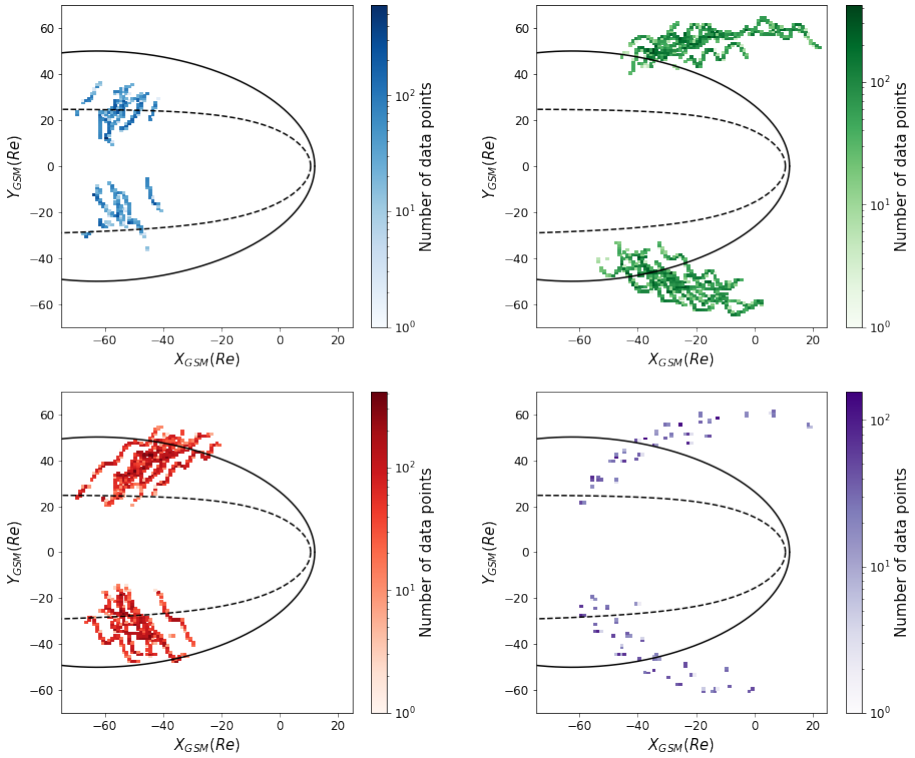

Figure A4. Spatial coverage of the ARTMIS labeled dataset. The legend is the same than in Figure 2 with the addition of the Moon's wake bins in purple which vary between 1 and 157

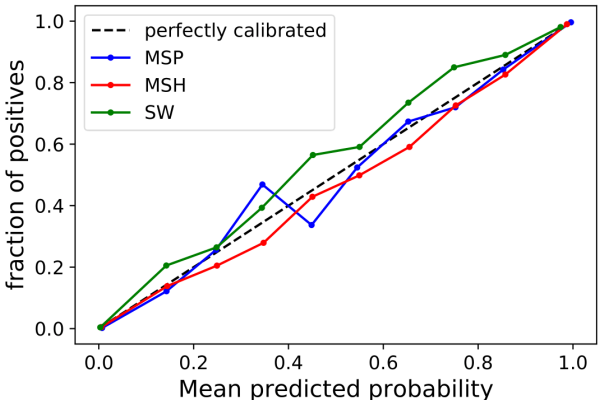

Figure B1. Calibration curve of our model trained on THEMIS data for the three regions. The black dashed line represent the calibration a perfectly-calibrated classifier would have.

Appendix C Additional detection examples

In this section, we show additional detection examples of the region classifier on the data of THEMIS B (Figure C1), Double Star (Figure C2), MMS (Figure C3), Cluster (Figure C4) and ARTEMIS (Figure C5).

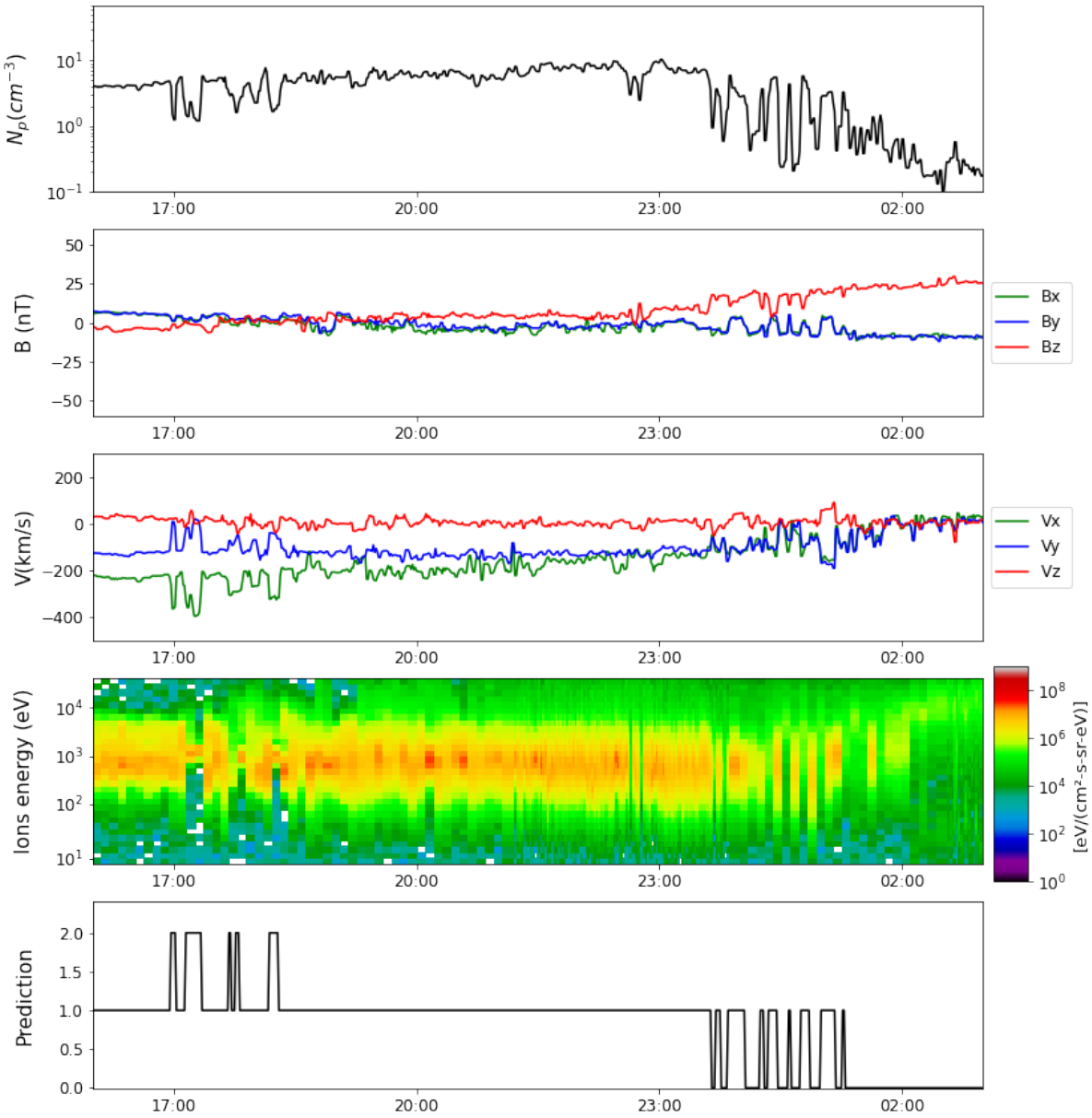

Figure C1. In-situ measurement provided by THEMIS B spacecraft on the 10^{th} of November 2008. The legend is the same than in 1.

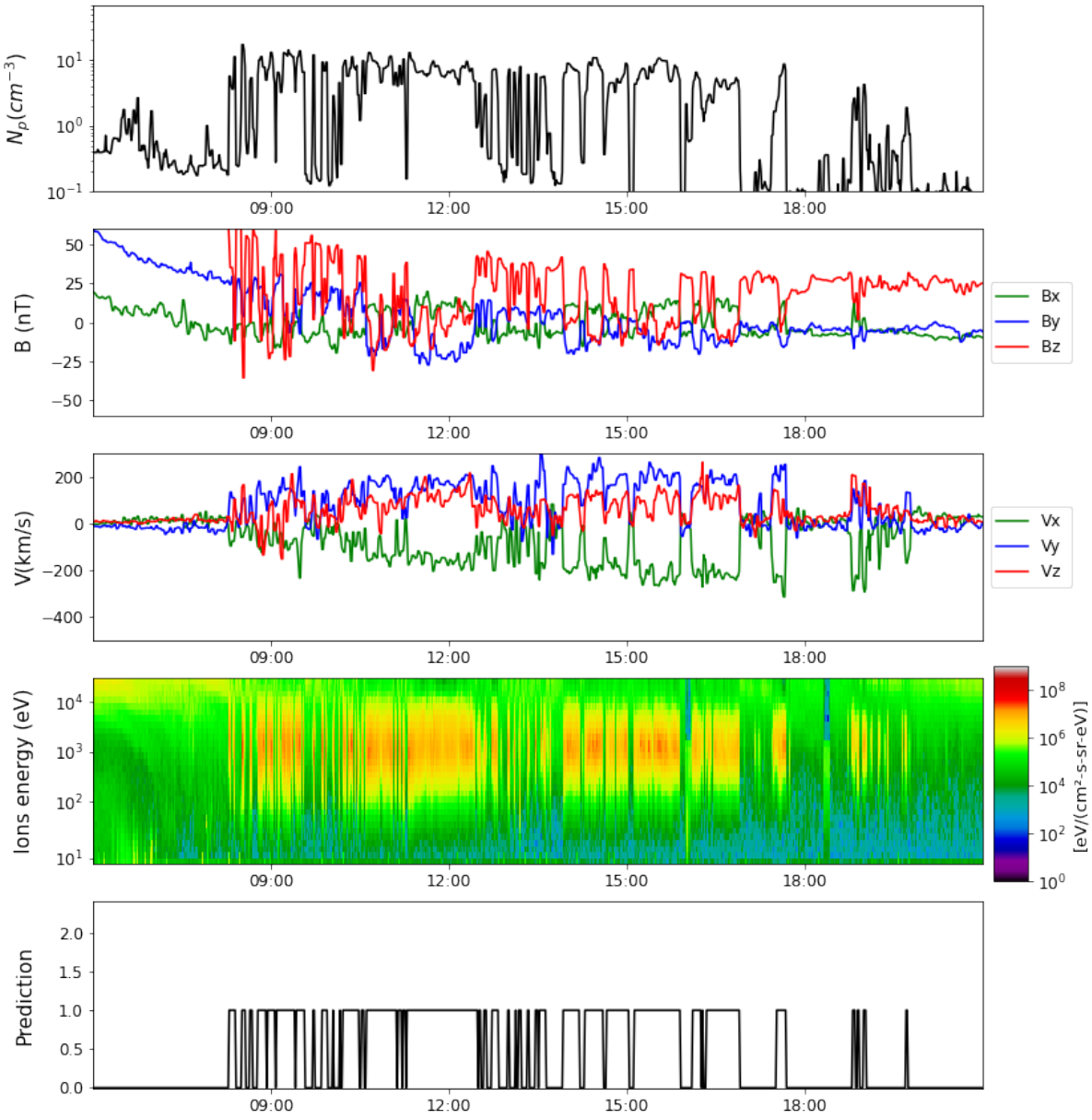

Figure C2. In-situ measurement provided by Double Star TC 1 spacecraft on the 15^{th} of January 2005. The legend is the same than in 1.

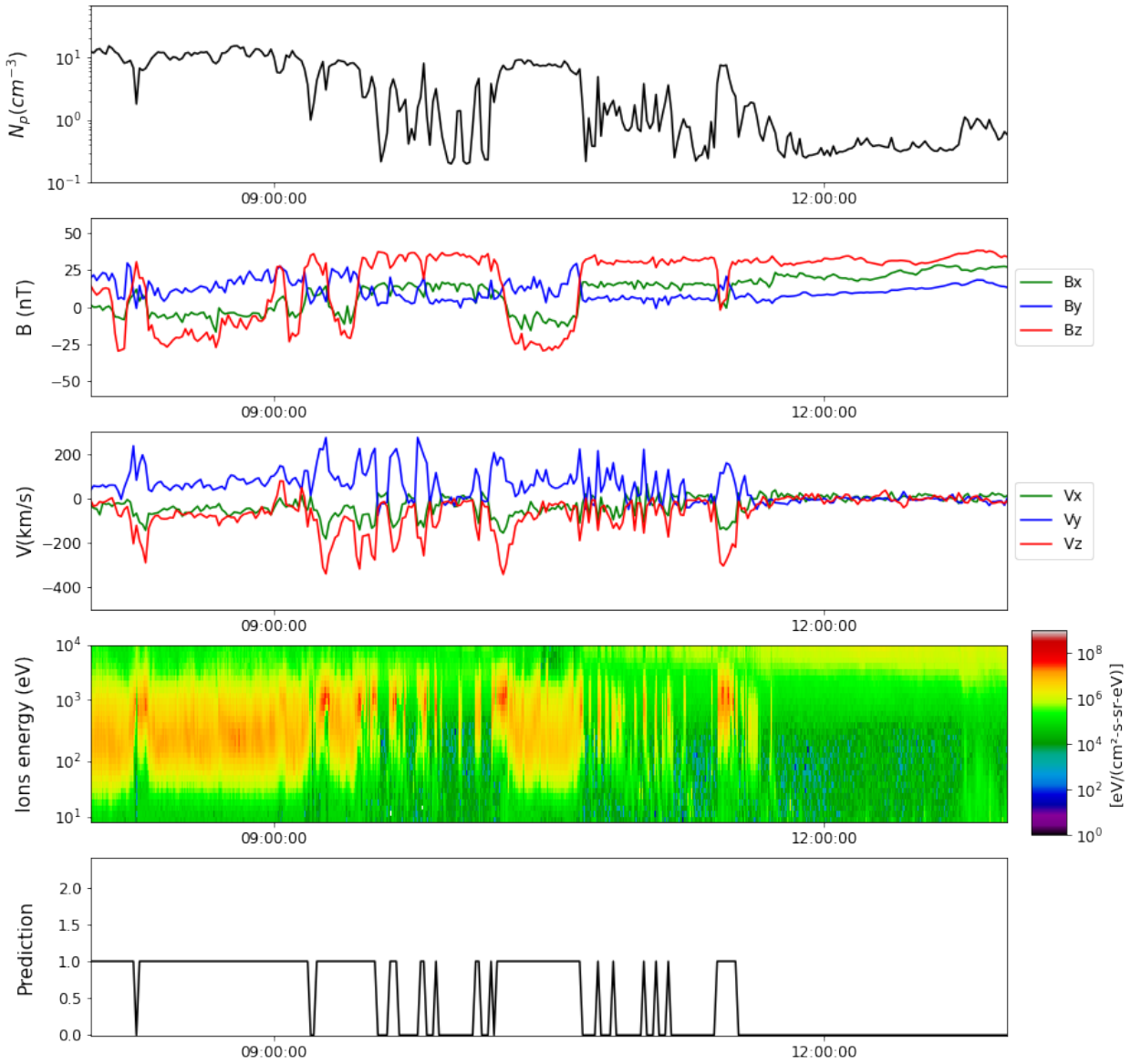

Figure C3. In-situ measurement provided by MMS 1 spacecraft on the 2^{nd} of December 2015. The legend is the same than in 1.

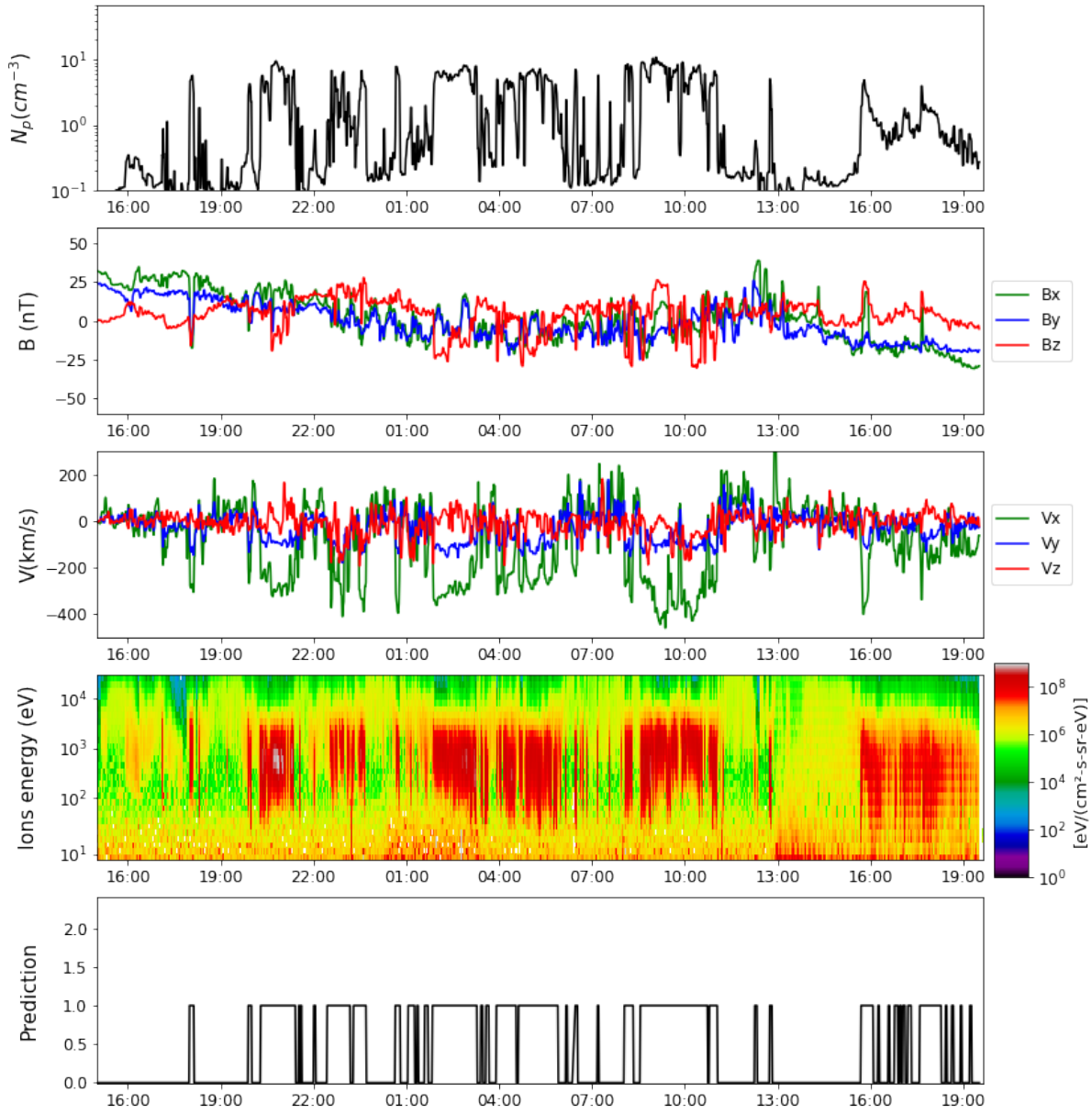

Figure C4. In-situ measurement provided by Cluster 3 spacecraft on the 23^{rd} of June 2003. The legend is the same than in 1.

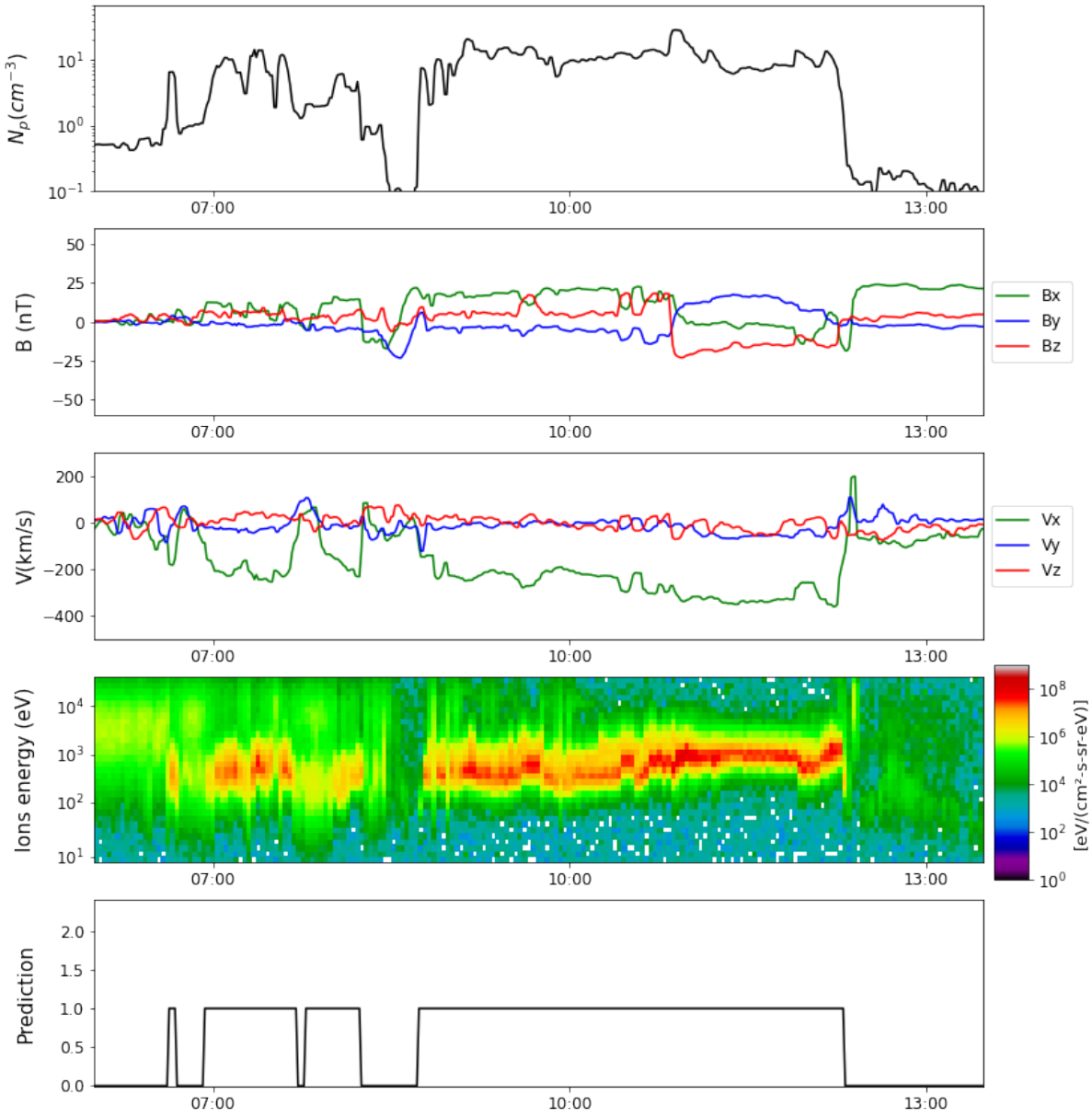

Figure C5. In-situ measurement provided by ARTEMIS B spacecraft on the 24^{th} of April 2013. The legend is the same than in 1.

598 THEMIS data are accessible via the NASA Coordinated Data Analysis web (`https://`
599 `cdaweb.sci.gsfc.nasa.gov/index.html/`). Cluster and Double Star data are accessible
600 via the Cluster and Double Star Science archive (`http://csa.esac.esa.int/`). All of
601 our trained algorithms can be found here `https://github.com/gautiernguyen/in-situ`
602 `_Events_lists`.

603 References

604 Argall, M. R., Small, C. R., Piatt, S., Breen, L., Petrik, M., Kokkonen, K., ... Burch, J. L.
605 (2020). Mms sitl ground loop: Automating the burst data selection process. *Frontiers*
606 *in Astronomy and Space Sciences*, *7*, 54. Retrieved from `https://www.frontiersin`
607 `.org/article/10.3389/fspas.2020.00054` doi: 10.3389/fspas.2020.00054

608 Auster, H. U., Glassmeier, K. H., Magnes, W., Aydogar, O., Baumjohann, W., Constanti-
609 nescu, D., ... Wiedemann, M. (2008, Dec). The THEMIS Fluxgate Magnetometer.
610 *Scientific Studies of Reading*, *141*(1-4), 235-264. doi: 10.1007/s11214-008-9365-9

611 Balogh, A., Carr, C., Acuña, M., Dunlop, M., Beek, T., Brown, P., ... Schwingenschuh, K.
612 (2001, 10). The cluster magnetic field investigation: Overview of in-flight performance
613 and initial results. *Annales Geophysicae*, *19*. doi: 10.5194/angeo-19-1207-2001

614 Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American*
615 *Statistical Association*, *39*(227), 357–365. Retrieved from `http://www.jstor.org/`
616 `stable/2280041`

617 Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and*
618 *regression trees*. Monterey, CA: Wadsworth and Brooks.

619 Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020).
620 Automatic classification of plasma regions in near-earth space with supervised machine
621 learning: Application to magnetospheric multi scale 2016–2019 observations. *Frontiers*
622 *in Astronomy and Space Sciences*, *7*, 55. Retrieved from `https://www.frontiersin`
623 `.org/article/10.3389/fspas.2020.00055` doi: 10.3389/fspas.2020.00055

624 Brown, I., & Mues, C. (2012, 02). An experimental comparison of classification algorithms
625 for imbalanced credit scoring data sets. *Expert Syst. Appl.*, *39*, 3446-3453. doi: 10
626 .1016/j.eswa.2011.09.033

627 Burgess, D. (1995, Jan). Foreshock-shock interaction at collisionless quasi-parallel shocks.
628 *Advances in Space Research*, *15*(8-9), 159-169. doi: 10.1016/0273-1177(94)00098-L

629 Camporeale, E., Carè, A., & Borovsky, J. E. (2017, Nov). Classification of Solar Wind
630 With Machine Learning. *Journal of Geophysical Research (Space Physics)*, *122*(11),
631 10,910-10,920. doi: 10.1002/2017JA024383

632 Carr, C., Brown, P., Zhang, T. L., Gloag, J., Horbury, T., Lucek, E., ... Richter, I. (2005,
633 Nov). The Double Star magnetic field investigation: instrument design, performance
634 and highlights of the first year's observations. *Annales Geophysicae*, *23*(8), 2713-2732.
635 doi: 10.5194/angeo-23-2713-2005

636 Fairfield, D. H. (1971, Jan). Average and unusual locations of the Earth's magne-
637 topause and bow shock. *Journal of Geophysical Research*, *76*(28), 6700. doi:
638 10.1029/JA076i028p06700

639 Farris, M. H., & Russell, C. T. (1994, Sep). Determining the standoff distance of the bow
640 shock: Mach number dependence and use of models. *Journal of Geophysical Research*,
641 *99*(A9), 17681-17690. doi: 10.1029/94JA01020

642 Fazakerley, A. N., Carter, P. J., Watson, G., Spencer, A., Sun, Y. Q., Coker, J., ...
643 Schwartz, S. J. (2005, Nov). The Double Star Plasma Electron and Current Ex-
644 periment. *Annales Geophysicae*, *23*(8), 2733-2756. doi: 10.5194/angeo-23-2733-2005

645 Friedman, J. H. (2001, 10). Greedy function approximation: A gradient boosting machine.
646 *Ann. Statist.*, *29*(5), 1189–1232. Retrieved from `https://doi.org/10.1214/aos/`
647 `1013203451` doi: 10.1214/aos/1013203451

648 Génot, V., Jacquey, C., Bouchemit, M., Gangloff, M., Fedorov, A., Lavraud, B., ... Pinçon,
649 J. L. (2010, May). Space Weather applications with CDPP/AMDA. *Advances in Space*
650 *Research*, *45*(9), 1145-1155. doi: 10.1016/j.asr.2009.11.010

Hasegawa, H. (2012). Structure and Dynamics of the Magnetopause. , *1*(2), 71–119. doi: 10.5047/meep.2012.00102.0071

Jelínek, K., Němeček, Z., & Šafránková, J. (2012, May). A new approach to magnetopause and bow shock modeling based on automated region identification. *Journal of Geophysical Research (Space Physics)*, *117*(A5), A05208. doi: 10.1029/2011JA017252

Jeřáb, M., Němeček, Z., Šafránková, J., Jelínek, K., & Měrka, J. (2005, Jan). Improved bow shock model with dependence on the IMF strength. *Planetary and Space Science*, *53*(1-3), 85-93. doi: 10.1016/j.pss.2004.09.032

Karimabadi, H., Sipes, T. B., Wang, Y., Lavraud, B., & Roberts, A. (2009, Jun). A new multivariate time series data analysis technique: Automated detection of flux transfer events using Cluster data. *Journal of Geophysical Research (Space Physics)*, *114*(A6), A06216. doi: 10.1029/2009JA014202

Kruparova, O., Krupar, V., Å afránková, J., Němeček, Z., Maksimovic, M., Santolik, O., ... Merka, J. (2019, Mar). Statistical Survey of the Terrestrial Bow Shock Observed by the Cluster Spacecraft. *Journal of Geophysical Research (Space Physics)*, *124*(3), 1539-1547. doi: 10.1029/2018JA026272

Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010, Apr). A three-dimensional asymmetric magnetopause model. *Journal of Geophysical Research (Space Physics)*, *115*(A4), A04207. doi: 10.1029/2009JA014235

Liu, Z., Lu, J. Y., Wang, C., Kabin, K., Zhao, J. S., Wang, M., ... Zhao, M. X. (2015). Journal of Geophysical Research : Space Physics A three-dimensional high Mach number asymmetric magnetopause model from global MHD simulation. *Journal of Geophysical Research*, 5645–5666. doi: 10.1002/2014JA020961.Received

McFadden, J. P., Carlson, C. W., Larson, D., Ludlam, M., Abiad, R., Elliott, B., ... Angelopoulos, V. (2008, Dec). The THEMIS ESA Plasma Instrument and In-flight Calibration. *Scientific Studies of Reading*, *141*(1-4), 277-302. doi: 10.1007/s11214-008-9440-2

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2020a). *Massive multi-missions statistical study and analytical modeling of the Earth magnetopause: 2 - Shape and location.* (Submitted)

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2020b). *Massive multi-missions statistical study and analytical modeling of the Earth magnetopause: 3 - An asymmetric magnetopause analytical model.* (Submitted)

Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine, D. (2020c). *Massive multi-missions statistical study and analytical modeling of the Earth magnetopause: 4- On the near-cusp magnetopause indentation.* (Submitted)

Niculescu-Mizil, A., & Caruana, R. (2005). Obtaining calibrated probabilities from boosting. In *Proceedings of the twenty-first conference on uncertainty in artificial intelligence* (p. 413–420). Arlington, Virginia, USA: AUAI Press.

Němeček, Z., Šafránková, J., & Šimůnek, J. (2020). An examination of the magnetopause position and shape based upon new observations. In *Dayside magnetosphere interactions* (p. 135-151). American Geophysical Union (AGU). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/9781119509592.ch8` doi: 10.1002/9781119509592.ch8

Olshevsky, V., Khotyaintsev, Y. V., Divin, A., Delzanno, G. L., Anderzen, S., Herman, P., ... Markidis, S. (2019, Aug). Automated classification of plasma regions using 3D particle energy distribution. *arXiv e-prints*, arXiv:1908.05715.

Paschmann, G., Haaland, S. E., Phan, T. D., Sonnerup, B. U. Ö., Burch, J. L., Torbert, R. B., ... Fuselier, S. A. (2018, Mar). Large-Scale Survey of the Structure of the Dayside Magnetopause by MMS. *Journal of Geophysical Research (Space Physics)*, *123*(3), 2018-2033. doi: 10.1002/2017JA025121

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., ... Zeuch, M. (2016,

Mar). Fast Plasma Investigation for Magnetospheric Multiscale. *Scientific Studies of Reading*, *199*(1-4), 331-406. doi: 10.1007/s11214-016-0245-4

Rème, H., Aoustin, C., Bosqued, J. M., Dand ouras, I., Lavraud, B., Sauvaud, J. A., ... Sonnerup, B. (2001, Oct). First multispacecraft ion measurements in and near the Earth's magnetosphere with the identical Cluster ion spectrometry (CIS) experiment. *Annales Geophysicae*, *19*, 1303-1354. doi: 10.5194/angeo-19-1303-2001

Russell, C. T., Anderson, B. J., Baumjohann, W., Bromund, K. R., Dearborn, D., Fischer, D., ... Richter, I. (2016, Mar). The Magnetospheric Multiscale Magnetometers. *Scientific Studies of Reading*, *199*(1-4), 189-256. doi: 10.1007/s11214-014-0057-3

Shue, J. H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., & Singer, H. J. (1997, May). A new functional form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical Research*, *102*(A5), 9497-9512. doi: 10.1029/97JA00196

Šafránková, J., Němeček, Z., Dušík, v., Přech, L., Sibeck, D. G., & Borodkova, N. N. (2002). The magnetopause shape and location: a comparison of the interball and geotail observations with models. *Annales Geophysicae*, *20*(3), 301–309. Retrieved from `https://www.ann-geophys.net/20/301/2002/` doi: 10.5194/angeo-20-301-2002

Wang, Y., Sibeck, D. G., Merka, J., Boardsen, S. A., Karimabadi, H., Sipes, T. B., ... Lin, R. (2013, May). A new three-dimensional magnetopause model with a support vector regression machine and a large database of multiple spacecraft observations. *Journal of Geophysical Research (Space Physics)*, *118*(5), 2173-2184. doi: 10.1002/jgra.50226

Wes McKinney. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (p. 56 - 61). doi: 10.25080/Majora-92bf1922-00a