

Study of Extreme Magnetopause Distortions under Varying Solar Wind Conditions

Niklas Grimmich¹, Ferdinand Plaschke¹, Martin Archer², Daniel Heyner¹,
Johannes Z. D. Mieth¹, Rumi Nakamura³, David G. Sibeck⁴

¹Institut für Geophysik und Extraterrestrische Physik, Technische Universität Braunschweig,
Braunschweig, Germany

²Department of Physics, Imperial College London, London, UK

³Space Research Institute, Austrian Academy of Sciences, Graz, Austria

⁴NASA Goddard Space Flight Center, Greenbelt, Maryland, US

Key Points:

- More than 160.000 magnetopause crossings (MPCs) identified in THEMIS data between 2007 and 2022 using a Random Forest Classifier
- MPCs that extremely deviate in location from the Shue et al. (1998) model are quite common
- Important solar wind parameters associated with deviations include the IMF cone angle, solar wind velocity and Alfvén Mach number

Corresponding author: Niklas Grimmich, n.grimmich@tu-braunschweig.de

Abstract

To first order, the magnetopause (MP) is defined by a pressure balance between the solar wind and the magnetosphere. The boundary moves under the influence of varying solar wind conditions and transient foreshock phenomena, reaching unusually large and small distances from the Earth. We investigate under which solar wind conditions such extreme MP distortions occur. Therefore, we construct a database of magnetopause crossings (MPCs) observed by the THEMIS spacecraft in the years 2007 to mid-2022 using a simple Random Forest Classifier. Roughly 7% of the found crossing events deviate beyond reported errors in the stand-off distance from the Shue et al. (1998) MP model and are thus termed extreme distortions. We find the occurrence of these extreme events in terms of expansion or compression of the MP to be linked to different solar wind parameters, most notably to the IMF magnitude, cone angle, velocity, Alfvén Mach number and temperature. Foreshock transients like hot-flow anomalies and foreshock bubbles could be responsible for extreme magnetospheric expansions. The results should be incorporated into future magnetopause models and may be helpful for the reconstruction of the MP locations out of soft x-ray images, relevant for the upcoming SMILE mission.

1 Introduction

Earth’s magnetopause is the boundary layer between the solar wind and the terrestrial magnetosphere. It is an obstacle for the incoming super-magnetosonic solar wind. A bow shock (BS) upstream of the MP decelerates the solar wind and then deflects the plasma around the magnetosphere. The region between the magnetopause and the bow shock is called magnetosheath (e.g., Baumjohann & Treumann, 1997). Depending on the angle between the interplanetary magnetic field (IMF) vector and the bow shock normal, the respective bow shock region (and the magnetosheath) may be denoted as quasi-parallel (angle $< 45^\circ$) or quasi-perpendicular (angle $> 45^\circ$). Upstream of the quasi-parallel bow shock, an extended foreshock region can form, permeated by waves which are excited due to the interaction of the solar wind with particles reflected at and back streaming from the BS (e.g., Eastwood et al., 2005).

Dynamical changes in the solar wind and subsequently in its interaction with the BS influence the magnetosheath flow and impact the MP location and shape. In the absence of reconnection, when the MP can be described as a rotational discontinuity, the MP is well-characterized as a tangential discontinuity at which pressure balance should hold. On the magnetospheric side, the magnetic pressure is the most important contributor to that balance, while on the magnetosheath side both plasma (thermal) and magnetic pressures (from the draped IMF) contribute significantly (e.g., Shue & Chao, 2013). Thus, variations of the total pressure in the solar wind and in the magnetosheath lead to inward and outward motion of the MP. Additionally, strong southward IMF conditions lead to magnetic flux erosion from the dayside MP via magnetic reconnection and therefore inward motion of the dayside MP (Aubry et al., 1970; Sibeck et al., 1991; Shue et al., 1997, 1998). Solar wind dynamic pressure, IMF strength and orientation can be identified as the parameters influencing the MP location. Consequently, many empirical MP models use the solar wind dynamic pressure p_{dyn} and the IMF B_z -component as input parameters (e.g., Fairfield, 1971; Sibeck et al., 1991; Shue et al., 1997; Chao et al., 2002; Lin et al., 2010; Nguyen et al., 2022c, and many others). In these models, the MP stand-off distance R_0 serves as an indicator for the overall location of the boundary layer, which is directly influenced by one or both of the two parameters p_{dyn} and B_z .

Newer models like the one from Lin et al. (2010) or Nguyen et al. (2022c) use additional parameters like the solar wind magnetic pressure and the dipole tilt to take asymmetries and cusp indentation into account, enhancing the forecasting accuracy of the model, e.g., shown by Case and Wild (2013) for the Lin et al. (2010) model. Nevertheless, most models fail to predict magnetopause locations under extreme pressure conditions (e.g.,

68 Tátrallyay et al., 2012; Suvorova & Dmitriev, 2015). In these cases, other parameters
 69 can become more significant. One of those parameters, which to our knowledge is not
 70 included in the models and also describes the IMF orientation, is the IMF cone angle
 71 ϑ_{cone} between the Earth-Sun-line and the IMF vector. Magnetospheric expansions be-
 72 yond the magnetopause model predictions are often found when the IMF is quasi-radial
 73 ($\vartheta_{\text{cone}} < 30^\circ$) (Fairfield et al., 1990; Suvorova et al., 2010; Dušík et al., 2010; Samsonov
 74 et al., 2012; Grygorov et al., 2017).

75 In addition to changes in the dynamic pressure and/or IMF orientation, other phe-
 76 nomena have been discussed as origins of MP disturbances, which can lead to extreme
 77 R_0 values. Phenomena originating near the magnetopause include magnetic reconnec-
 78 tion and associated flux transfer events (FTE, e.g., Elphic, 1995) or the Kelvin-Helmholtz
 79 instability (KHI, e.g., Johnson et al., 2014). In the magnetosheath, so called magnetosheath
 80 or high-speed jets (HSJs) can travel from their point of origin at the bow shock down
 81 to the magnetopause and cause an indentation and excitation of surface waves (Shue et
 82 al., 2009; Plaschke et al., 2018; Archer et al., 2019). Finally, kinetic transients in the fore-
 83 shock region like hot-flow anomalies (HFAs) or foreshock bubbles (FBs) and ULF-wave
 84 generated phenomena like foreshock cavitons, short large-amplitude magnetic structures
 85 (SLAMS) or shocklets can impact the MP in different ways as well (Sibeck et al., 1999;
 86 Jacobsen et al., 2009; Turner et al., 2011; Archer et al., 2015; H. Zhang et al., 2022). Some
 87 of these phenomena only result in localized distortions (e.g. HFAs, Sibeck et al., 1999;
 88 Turner et al., 2011), others could have global impacts (e.g. FBs, Archer et al., 2015).

89 These phenomena and the solar wind magnetosphere interactions have been stud-
 90 ied for two decades using data from several multi-spacecraft missions. Cluster (Escoubet
 91 et al., 2001) contributed significantly to the exploration of different plasma regions of
 92 the magnetosphere, advancing our understanding of reconnection and the movement of
 93 the magnetopause (see Haaland et al., 2021, for a comprehensive overview). The Time
 94 History of Events and Macro-scale Interactions during Substorms (THEMIS) mission (Angelopoulos,
 95 2008) enabled observations of solar wind phenomena and direct responses in the mag-
 96 netosphere due to the special orbit configuration of the multiple spacecraft. The aim of
 97 the most recent mission MMS is to study in detail magnetic reconnection at the small-
 98 est scales (Burch et al., 2016).

99 Typically, all these spacecraft can only observe the MP at the position and time
 100 they cross this boundary or when the MP is in motion and moves over the spacecraft.
 101 So far, global observations of the MP have not been possible. The upcoming Solar Wind
 102 Magnetosphere Ionosphere Link Explorer (SMILE) mission will provide the first oppor-
 103 tunity to observe the location, shape and motion of the dayside MP at any given time
 104 (Raab et al., 2016; Branduardi-Raymont et al., 2018), based on measurements of soft x-
 105 rays. Soft x-rays are emitted during solar wind charge exchange with neutrals from the
 106 Earth's exosphere (e.g., see review by Kuntz, 2019). Studies of this phenomenon in the
 107 near-Earth regions showed the possibility to image the magnetospheric boundary lay-
 108 ers in soft x-ray wavelengths and reconstruct the magnetopause surface from the images
 109 (e.g., Sibeck et al., 2018; C. Wang & Sun, 2022). SMILE will take advantage of this to
 110 study the whole dayside magnetosphere from a polar orbit and image the soft x-rays with
 111 a Soft X-ray Imager (SXI) to track the magnetopause motion on global scales. Additional
 112 instrumentation of SMILE will include a Magnetometer (MAG), a Light Ion Analyser
 113 (LIA) and an Ultra-Violet Imager (UVI) which will monitor the plasma environment,
 114 in particular the solar wind conditions, and the auroral oval in UV wavelengths, respec-
 115 tively. Thereby, the motion of the magnetopause can be linked to the upstream plasma
 116 conditions and the ionospheric response.

117 The SMILE mission is expected to launch in late 2024 or early 2025. In prepara-
 118 tion, much effort is put in the development of MP reconstruction techniques based on
 119 simulated SXI images, for which fundamental knowledge about the magnetopause shape
 120 and behaviour is needed (see C. Wang & Sun, 2022, and references there in). The in-

fluence of IMF parameters on it has been subject to several statistical studies (e.g., Plaschke, Glassmeier, Auster, Angelopoulos, et al., 2009; Dušík et al., 2010; Staples et al., 2020; Nguyen et al., 2022b). In this study, we focus on extreme MP locations, which cannot be explained with a common MP model like the improved Shue et al. (1998) model, theoretically capable of predicting the MP location under extreme solar wind conditions. The reason for this is most likely due to the fact that such models are designed to be optimal around the typical conditions. Therefore, extreme and unusual conditions are given less weight in fitting the models, resulting in model predictions deviating under such conditions.

In particular, we are interested under which solar wind conditions these events occur. In the following, we give a brief introduction to the used spacecraft data (section 2). We describe the construction of our database of magnetopause crossings observed by the THEMIS spacecraft (section 3) and show the results (section 4). We then compare the solar wind conditions for which extreme events occur with the standard solar wind distributions (section 5). Eventually, we discuss and summarize our findings (section 6).

2 Spacecraft Data

Since 2007 the spacecraft of the THEMIS mission have been orbiting Earth near the equatorial plane to investigate the plasma environment in the near-Earth region (Angelopoulos, 2008). For the identification of MPCs in the timespan of 2007 to 2022, we use the magnetic field data from the Fluxgate Magnetometer (FGM, Auster et al., 2008), and particle data and moments from the Electrostatic Analyzer (ESA, McFadden et al., 2008). Data from the entire 15 years interval are used from probes THA, THD, and THE, while THB and THC only contribute data until the end of 2009, as they were then sent into lunar orbits, becoming the ARTEMIS mission (Angelopoulos, 2011).

FGM and ESA data are used in the spin-resolution (FGM) and reduced mode (ESA) with cadences of about 3 to 4 s. Low resolution FGM data and full mode ESA data are used to bridge bigger data gaps (> 15 min). This occurs almost exclusively in the velocity data of ESA, leading to some uncertainties in this data which can be compensated in our detection method. The FGM and ESA data are synchronized and resampled to common 3 s time stamps. Finally, we average the data in a moving 60 s window for each time step, to smooth out turbulent fluctuations which could be misidentified as MPCs. The data is processed in 1-hour intervals with an overlap of 2 minutes into the next interval. Intervals were omitted for data gaps that could not be bridged, i.e. if less than 15 mins of data were available in both the high and low resolution. This is necessary, as large data gaps lead to jumps in the data which could be misinterpreted as MPC. Results are combined to a bigger dataset afterwards.

All vector quantities in the dataset are transformed into the AGSE (aberrated geocentric solar ecliptic) coordinate system with an average aberration angle of $\varphi \sim 4.3^\circ$ resulting from the Earth's orbital velocity of 30 km/s around the Sun and an average solar wind speed of 400 km/s. Taking this aberration effect on the MP into account, improves the prediction of MP models (e.g., Safránková et al., 2002).

We limit our investigation to the dayside magnetosphere (position in AGSE x-direction larger than 0) outside the nominal plasmasphere (minimal radial distance from the Earth's center larger than $4 R_E$). These conditions have to be fulfilled throughout the 1-hour intervals. This can result in fewer observations near the terminator.

In addition to the observational data, we construct a dataset containing the orbital data of the THEMIS probes in the cartesian (x, y, z) and the spherical AGSE coordinates (r, θ, λ) , see Fig. 1) in 1-minute resolution for the all selected time intervals. This dataset also comprises the equivalent stand-off distance $R_{0,sc}$ calculated with the Shue et al. (1997,

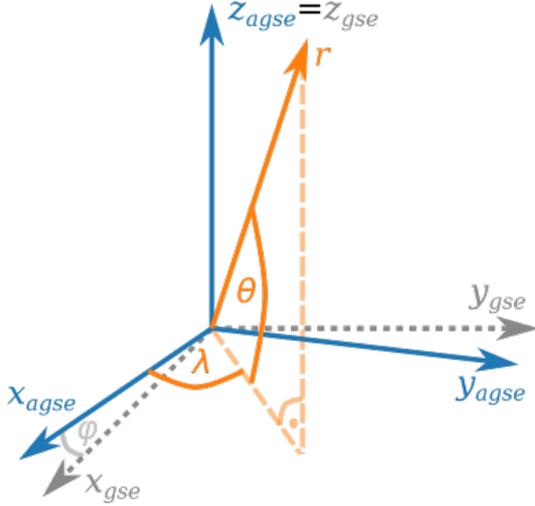


Figure 1. Orientations and relations of the two main coordinate systems. The grey axes depict the standard GSE (geocentric solar ecliptic) system. The cartesian and spherical AGSE (aberrated geocentric solar ecliptic) coordinate axes are depicted in blue and orange, respectively.

171 1998), hereafter SH98, model equation, as done in previous studies (Plaschke, Glassmeier,
172 Auster, Constantinescu, et al., 2009; Staples et al., 2020):

$$173 \quad R_{0,sc} = r \left(\frac{2}{1 + \cos \zeta} \right)^{-\alpha}. \quad (1)$$

174 Here r is the radial distance from the Earth's center to the spacecraft and ζ is the zenith
175 angle between the x -axis and the Earth-spacecraft-line (denoted by θ in Shue et al., 1997,
176 1998). The flaring parameter α is calculated with the formula given by Shue et al. (1998),
177 using the appropriate dynamic solar wind pressure p_{dyn} and IMF component $B_{z,\text{IMF}}$ for
178 all orbital points:

$$179 \quad \alpha = \left(0.58 - 0.007 \frac{B_{z,\text{IMF}}}{\text{nT}} \right) \left[1 + 0.024 \ln \left(\frac{p_{\text{dyn}}}{\text{nPa}} \right) \right]. \quad (2)$$

180 We take into account that our approximation of a static solar wind speed for the
181 aberration effect results in mean errors of $0.034 R_E$ for $R_{0,sc}$ and 0.823° for the λ -position,
182 which have no drastic influence on our study.

183 The appropriate solar wind parameters are obtained from the high resolution 1-
184 min OMNI dataset, which mainly combines the WIND (Lepping et al., 1995; Ogilvie et
185 al., 1995) and ACE (Stone et al., 1998; Smith et al., 1998; McComas et al., 1998) space-
186 craft data, time-shifted to the bow shock nose (King & Papitashvili, 2005). Smaller data
187 gaps up to 5 minutes in the OMNI dataset are bridged by linear interpolation.

188 3 Magnetopause Crossing Identification Method

189 Our identification process utilizes a combination of supervised machine learning
190 methods and a threshold-based classification, to infer crossing events from automatically
191 labelled data.

192 Recent studies have already shown the efficiency of classifying the near-Earth re-
193 gions from spacecraft data with machine learning methods (e.g., Breuillard et al., 2020;

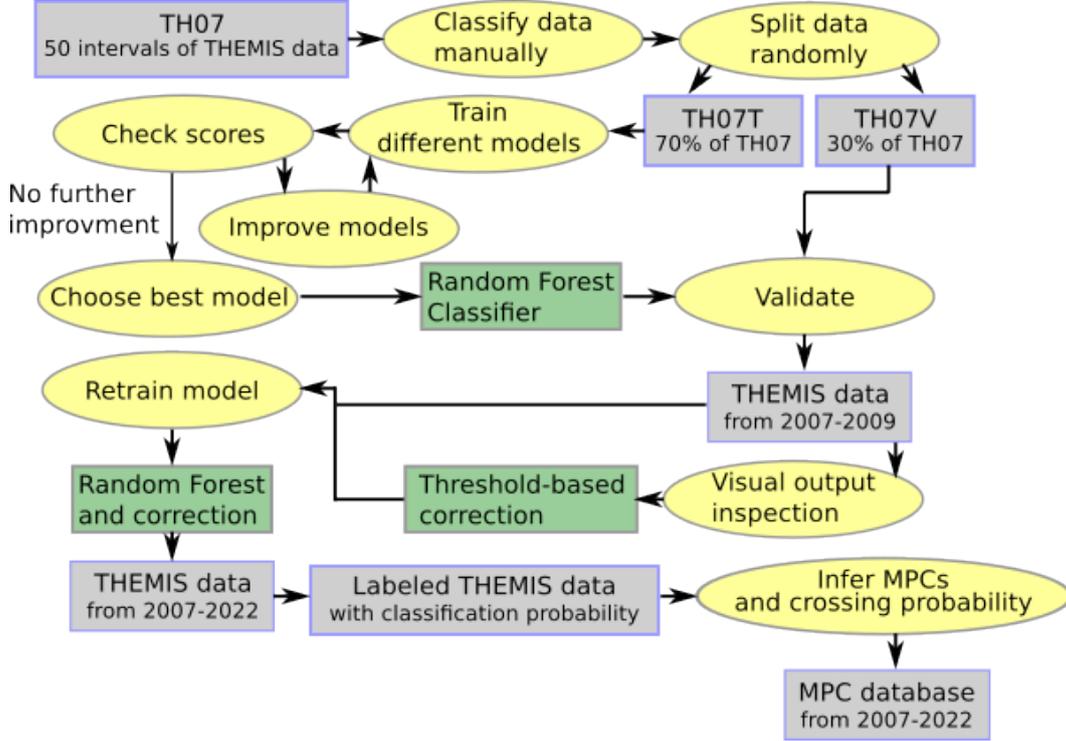


Figure 2. Flow diagram outlining our identification process.

194 Olshevsky et al., 2021; Nguyen et al., 2022a). In particular Nguyen et al. (2022a) showed
 195 that even a simple machine learning algorithm like the Gradient Boosting Classifier can
 196 outperform manually set threshold based detection methods of the three typical near-
 197 Earth regions (solar wind, magnetosheath and magnetosphere), reaching more than sat-
 198 isfying accuracies.

199 Unfortunately, Nguyen et al. (2022a) only inferred if one MPC is found in a 1-hour
 200 interval, finding only a limited amount of MPCs with an uncertain location. This is not
 201 suitable for our study, as we can not be certain to infer the right model deviations from
 202 their catalogue. We aim to construct a database in which extreme MPCs are clearly iden-
 203 tified on smaller timescales and with a clear spacecraft location, which can be used in
 204 future studies on extreme MP distortions. Nevertheless, we can use the same approach
 205 as Nguyen et al. (2022a) in giving every data point a label according to the near-earth
 206 region it most likely pertains to, and then infer the boundary crossings from the labels.
 207 For our study, we only need to distinguish between data points that are in the magne-
 208 tosphere labelled 1 and data points that are not in the magnetosphere labelled 0, facil-
 209 itating the identification of magnetopause crossings.

210 In Fig. 2 we present a flow diagram summarizing our identification process. De-
 211 tailed description can be found in the following sections.

212 3.1 Machine Learning Algorithms

213 For our study, we only need to distinguish between data points that are in the mag-
 214 netosphere labelled 1 and data points that are not in the magnetosphere labelled 0, fa-
 215 cilitating the identification of magnetopause crossings.

216 Nguyen et al. (2022a) trained their algorithm with data resampled to 1-minute res-
 217 olution consisting of the magnetic field components (B_x , B_y , B_z), the ion velocity (v_x , v_y , v_z),
 218 the ion density n_{ion} and the ion temperature T_{ion} . We include the magnitude of mag-
 219 netic field and velocity as well as a flux index $F_{\text{idx}}(t)$ which describes the omnidirectional
 220 energy flux of ions with energies between 10^2 eV and 10^4 eV, where the solar wind and
 221 magnetosheath regions are easily identified:

$$222 \quad F_{\text{idx}}^{1e2,1e4}(t) = \log_{10} \left(\int_{10^2 \text{ eV}}^{10^4 \text{ eV}} \frac{E_{\text{Flux}}(t)}{\frac{\text{eV}}{\text{cm}^2 \cdot \text{s} \cdot \text{sr}}} dE \right). \quad (3)$$

223 The index reaches high values of 10 and above if the magnetosheath ion population is
 224 observed, otherwise $F_{\text{idx}} < 10$ holds (compare panels (5) and (6) of Fig. 3). This $F_{\text{idx}}(t)$
 225 can be better handled by the model than the total energy flux distribution for each time
 226 step.

227 To train and compare different machine learning algorithms, we built a dataset of
 228 50 randomly selected time intervals with different lengths of the initial phase of the THEMIS
 229 mission in 2007 (TH07), that represent well outer-magnetospheric dayside observations
 230 (see Fig. 4). Each interval contains at least one magnetopause crossing. All data points
 231 are then labelled manually by visually inspecting n_{ion} and B_z changes, as well as ion en-
 232 ergy flux density measurements, yielding roughly 30,000 labelled data points from in-
 233 side (Label 1) and outside (Label 0) the magnetosphere with ~ 1300 MPCs for train-
 234 ing. Data points in a smeared out MPC or boundary layer are attempted to be separated
 235 in the middle of the crossing. Fig. 3 displays one of the intervals from TH07 with all in-
 236 put parameters for the algorithms; it also shows labels given manually and by the trained
 237 Random Forest machine learning classifier.

238 We randomly divide our dataset TH07 into a training set (70% of the data points,
 239 TH07T) and a validation set (30% of data points, TH07V). With TH07T we train, test
 240 and compare different models to decide which model to utilize for the identification. TH07V
 241 is later used to verify the training scores of the best model, assuring the model has not
 242 overfitted the trainings data. The nature of our problem, inside (class/label 1) or out-
 243 side (class/label 0) the magnetosphere, is a binary classification problem which can be
 244 tackled with a number of different algorithms (e.g., described in Géron, 2019).

245 One of the simplest binary classifiers is the Logistic Regression (LR, e.g., Cox &
 246 Snell, 1970), predicting the probability of a data point belonging to the positive class (la-
 247 bel 1) by calculating a logistic (sigmoid) function of a linear fit of the input data. This
 248 algorithm assumes that the data points are linearly distributed in parameter space. Ad-
 249 ditionally, the data has to be normalized for the algorithm to work probably.

250 Another often used method is the Decision Tree (DT, e.g., Breiman et al., 1984).
 251 This algorithm can directly (with only little preprocessing) predict a class from differ-
 252 ent input data using simple if-then-else decision rules inferred from data features/input
 253 parameters. A common problem with DTs if not restricted correctly is overfitting, i.e.,
 254 adapting too tightly to the training data, reducing the adaptability of the model to new
 255 data.

256 More advanced algorithms like the Random Forest (RF, e.g., Breiman, 2001) or Gra-
 257 dient Boosting (GB, e.g., Friedman, 2001) use ensemble methods for their prediction:
 258 multiple simple models are trained on the data and the final prediction are then derived
 259 from the predictions of all contributing simple models. Both RF and GB algorithms use
 260 DTs as basis. The RF algorithm trains a group of DTs on random training data sub-
 261 sets and use the most common prediction in the group as final prediction, therefore re-
 262 ducing the problem of overfitting of the individual DTs. The GB on the other hand se-
 263 quentially fit DTs on the residual errors of the previously trained DT until the ensem-
 264 ble convergences on the smallest errors, and predicts the class via the sum of the ensem-
 265 ble predictions. These ensemble methods are widely used in many machine learning ap-

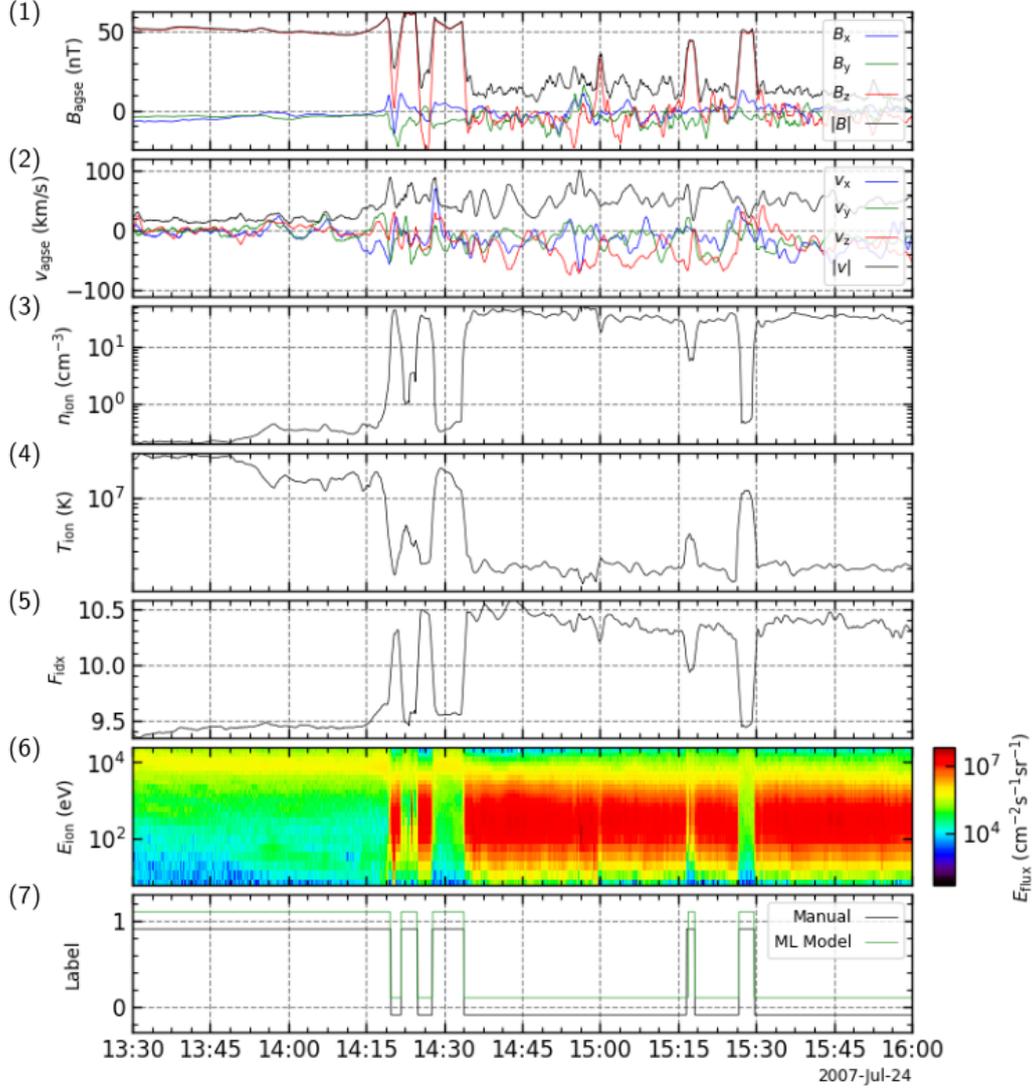


Figure 3. Time series plot of THEMIS data (THE) on the 24 July 2007. From top to bottom the panels display the averaged magnetic field data, the ion velocity, the ion density, the ion temperature, the energy flux density, the flux index and the data label given manually and by a Random Forest Classifier. The label values were shifted slightly for better visual comparison.

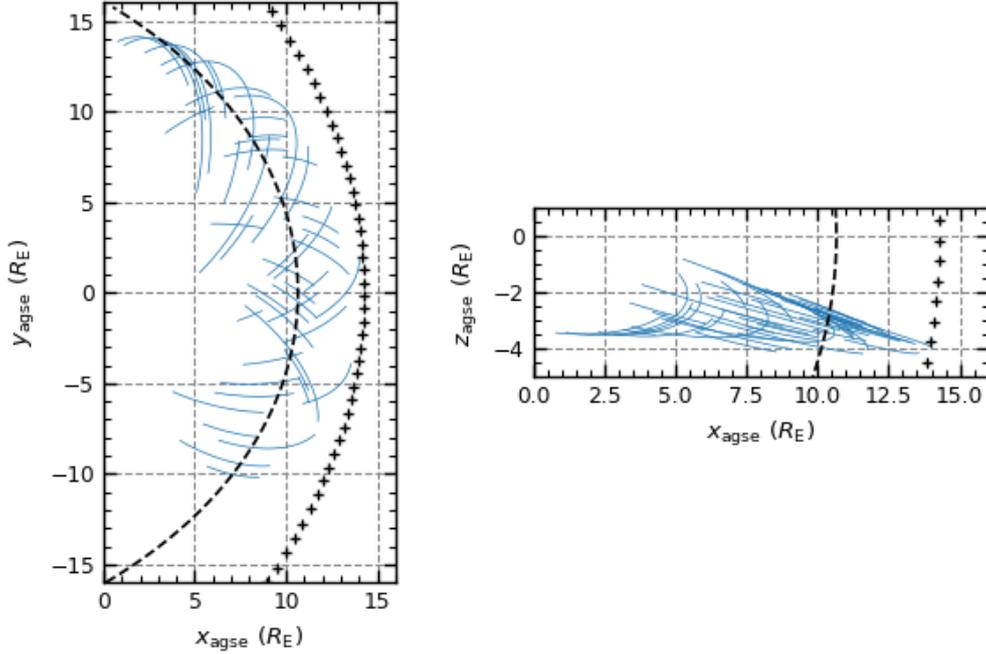


Figure 4. Spatial distribution of 50 training intervals in the AGSE x-y-plane (top panel) and x-z-plane (bottom panel), respectively. The dashed line represents the Shue et al. (1998) model magnetopause and the black crosses represent the Chao et al. (2002) model bow shock for $B_{z,IMF} = -1$ nT and $p_{dyn} = 1.5$ nPa.

266 plications, reaching high accuracies (Géron, 2019). Nguyen et al. (2022a) used the GB
 267 algorithm in their work for the identification of the near-Earth regions in spacecraft data.

268 All the presented algorithms except the RF were also compared by Nguyen et al.
 269 (2022a). We start the training with more input parameters, hence, we repeat the model
 270 comparison here to ensure using the optimal model. For the comparison we have to split
 271 our training data TH07T again into training subset (TH07TC) and into a validation sub-
 272 set (TH07TV) with a data ratio of 70/30.

273 For the first testing round, we utilize the default implementation of the algorithms
 274 from Python’s Scikit-learn library (Pedregosa et al., 2011) and evaluate the models via
 275 the cross validation (CV) scores. Cross validation means that the training data (TH07TC)
 276 is split into n equally sized subsets. The model is then trained and evaluated n times
 277 with all possible combinations of these subsets as training ($n-1$ subsets) and validation
 278 data (1 subset). Thus, the CV scores give us a mean accuracy (fraction of correct pre-
 279 dictions) and standard deviation over all n subsets, working as an indicator for the in-
 280 dependence of the data split into training and validation data. Here we utilize a 10-folded
 281 CV, i.e., we split the TH07TC into $n = 10$ subsets. Based on this first CV, we can al-
 282 ready conclude that the two ensemble classifiers perform better. Nevertheless, as sug-
 283 gested by Géron (2019), we aim at improving all the models by adjusting some impor-
 284 tant hyperparameters (specific boundaries for the algorithms) using a grid search method:
 285 We train and evaluate the models via CV with different parameter combinations in search
 286 for the best scores.

287 In the case of the LR the default hyperparameters yield the best results, while for
 288 the other algorithms the grid search shows that setting hyperparameters like the max-

Table 1. Final validation scores of different ML algorithms.

Score	Logistic Regression	Decision Tree	Random Forest	Gradient Boosting
CV	0.9633 ± 0.0012	0.9877 ± 0.0007	0.9939 ± 0.0005	0.9937 ± 0.0005
Precision	0.9606	0.9889	0.9939	0.9937
Recall	0.9722	0.9882	0.9938	0.9938
AUC	0.9944	0.9881	0.9998	0.9998

289 imal tree depth and the number of estimators (here: DTs) in the ensemble resulted in
 290 better scores. The maximal tree depth limits the number of if-then-else decisions in the
 291 DTs, reducing the risk of overfitting the models. The best results are obtained by set-
 292 ting the parameters as follows: for the simple DT the maximal depth is set to 20, for the
 293 RF it is set to 40 and for the GB it is set to 15. The number of estimators is set to 600
 294 and 400 for the RF and GB classifier, respectively. Additionally, the learning rate in the
 295 GB classifier is changed from 0.1 to 0.5, i.e., the fitting of the base estimators is accel-
 296 erated slightly, without risking overfitting, by setting a higher number of estimators.

297 In addition to the CV score, we look at other scores that are often used for vali-
 298 dating (binary) classifiers (Géron, 2019): the precision is the ratio of correct predictions
 299 out of all *inside magnetosphere* algorithm predictions; the recall or sensitivity is the ra-
 300 tio of correct predictions out of all true *inside* labels. For example, a recall of 0.95 for
 301 a classifier means that 95% of the data points inside the magnetosphere are predicted
 302 correctly.

303 To ascertain which model can distinguish best between the two classes, we also uti-
 304 lize the AUC (area under the curve) score. This score is derived as the integral from the
 305 receiver operating characteristic (ROC) curve, which describes the true positive rate (which
 306 is identical to the recall) as a function of the false positive rate (ratio of false predictions
 307 out of all true *outside* labels). For a purely random classifier, the AUC score would be
 308 0.5, while a value of 1 would indicate a classifier perfectly distinguishing between the two
 309 classes.

310 The final validation scores for the algorithms (after grid search and setting of the
 311 hyperparameters) are obtained by application of the trained models on the validation
 312 set TH07TV; They are depicted in Table 1. As can be seen, all algorithms share scores
 313 with values over 0.96 in all categories, and therefore could distinguish between the two
 314 classes and predict many magnetospheric data points correctly. By looking at the dif-
 315 ferent scores in detail, it's clear that the DT performs better than the LR in regard to
 316 CV score, precision and recall; only in the AUC score LR shows higher values. Overall,
 317 the ensemble methods (RF and GB) perform even better than the simpler models, yield-
 318 ing nearly identical scores. The CV scores show the lowest standard deviation of $5 \cdot 10^{-4}$.
 319 Higher accuracies of 0.994 indicate a slightly better independence from the chosen train-
 320 ing data. RF and GB also have precisions and recalls over 0.994 and AUC scores of 0.999.
 321 Thus, the ensemble methods are slightly better suited for the classification: they can dis-
 322 tinguish very well between the two classes while also correctly predicting the labels in
 323 over 99% of the cases, matching the model comparison results of Nguyen et al. (2022a).

324 Finally, we compared the feature/input parameter importance of the RF and GB
 325 classifiers for the prediction of data points. The feature importance is a calculation of
 326 the relative contribution of each feature to the final decision, showing easily the influ-
 327 ence of parameters to the model results. While the GB classifier mainly utilizes the ion
 328 density for its prediction, the RF uses many of the input parameters in its decision. This
 329 leads to the RF classifier being not as much affected by spurious density changes as the

330 GB classifier; the latter tends to label density peaks erroneously as magnetosheath data,
 331 even if other observations suggested a different classification.

332 Thus, we decide to utilize the RF classifier to label the THEMIS data. We train
 333 it on our complete training dataset TH07T. The previous obtained scores are again ver-
 334 ified by validating the RF classifier on the validation set TH07V.

335 3.2 Additional Threshold-based Corrections

336 Visual inspection of ~ 100 randomly selected intervals from 2007 to 2009 that were
 337 labelled with the Random Forest Classifier revealed some identification mistakes asso-
 338 ciated to foreshock phenomena or BS crossings. In addition, some mistakes were also found
 339 related to cold plasma observations deep inside the magnetosphere. To correct these mis-
 340 takes, we use the following threshold-based label correction:

- 341 1. Southward IMF ($B_z \leq 0$ nT) and large ion velocities in AGSE x-direction ($v_x \leq$
 342 -250 km/s) should only be observed outside the dayside magnetosphere. If either
 343 (or both) of these criteria is fulfilled and if, in addition, ion densities above $n_{\text{ion}} >$
 344 0.5 cm $^{-3}$ are observed, then the associated points are relabelled as outside the mag-
 345 netosphere.
- 346 2. High magnetic field magnitudes ($B > 150$ nT) and small deviations between the
 347 flux index and a high energy flux index ($F_{\text{idx}}^{1e2,1e4} - F_{\text{idx}}^{6.5e3,1e4} \leq 0.5$) should only
 348 be observed inside the magnetosphere. If either (or both) of these criteria is ful-
 349 filled and if, in addition, ion densities below $n_{\text{ion}} < 0.75$ cm $^{-3}$ are observed, then
 350 the associated points are relabelled as inside the magnetosphere.

351 Roughly one percent of the labels have been corrected. The classification proba-
 352 bility of these corrected labels is manually set to 0.85, indicating the correction.

353 We retrain our model on the gathered dataset of THEMIS data between 2007 and
 354 2009 with corrected labels, trying to improve the classifier with these new labels. From
 355 here on, since directly adjacent points often share the same label, we choose a new data
 356 sampling rate of 12 s. Hence, we could accelerate the classification process without loos-
 357 ing the accuracy of our model. Then we utilize the retrained Random Forest Classifier
 358 to label the remaining data up to 2022, while also applying the threshold-based label cor-
 359 rection for 1% of the data.

360 3.3 Identification of Magnetopause Crossings

361 We search for MPCs by automatically identifying the times where labels change
 362 from one region to the other. We only count a label change as a MPC if at least two points
 363 before and after the change belong to the same region. That means the spacecraft have
 364 to be at least 24 s in a different region for a crossing to count.

365 The identification process results in an average of 13,164 MPCs per year. In to-
 366 tal, 184,292 MPCs have been observed by the THEMIS spacecraft over the 15 years stud-
 367 ied. These MPCs are collected into the dataset TH-MPC (Grimmich et al., 2023)

368 We calculate the deviation from the theoretical model stand-off distance ΔR_0 given
 369 by the SH98 model for each identified crossing

$$370 \quad \Delta R_0 = R_{0,\text{sc}} - R_{0,\text{Shue}}, \quad (4)$$

$$371 \quad R_{0,\text{Shue}} = \left[10.22 + 1.29 \tanh \left(0.184 \left(\frac{B_{z,\text{IMF}}}{\text{nT}} + 8.14 \right) \right) \right] \left(\frac{p_{\text{dyn}}}{\text{nPa}} \right)^{-\frac{1}{6.6}}, \quad (5)$$

372 where equation (5) corresponds to equation (10) in SH98 and $B_{z,\text{IMF}}$ and p_{dyn} are taken
 373 as the mean values in an event-preceding 8-minute interval from the solar wind OMNI

374 dataset, taking the time delay from the bow shock to the magnetopause and the termi-
375 nator into account. With definition (4), a negative ΔR_0 corresponds to a compression
376 and a positive ΔR_0 to an expansion of the magnetopause to the spacecraft location.

377 In some cases ($\sim 11\%$) the stand-off distance and the deviation could not be cal-
378 culated due to a lack of OMNI data for an entire interval, we have excluded the corre-
379 sponding MPC entries from our database.

380 For each found MPC, we infer a crossing probability from the prediction probabil-
381 ity $p_{\text{RF}}(t)$ given by the RF classifier. The calculation is a weighted average of the prob-
382 ability of the 2 points before and after the jump in the labels:

$$383 \quad p_{\text{MPC}}(t_0) = \frac{1}{3} [p_{\text{RF}}(t_0 - 12 \text{ s}) + 0.5p_{\text{RF}}(t_0) + 0.5p_{\text{RF}}(t_0 + 12 \text{ s}) + p_{\text{RF}}(t_0 + 24 \text{ s})]. \quad (6)$$

384 The points are weighted with increasing time distance from the jump with 0.5 or 1 (see
385 (6)), as the RF classifier predicts the labels with higher precision further away from the
386 jump. The two points directly adjacent to the label change have the biggest prediction
387 uncertainty and should contribute less to the probability calculation.

388 MPCs with low crossing probability are more likely misidentified or ambiguous. Thus,
389 it's reasonable from here on to only use the roughly 75% of the database with high ($>$
390 0.75) crossing probabilities (121,770 MPCs of TH-MPC). Additionally, as can be seen
391 in Fig. 5, the MPC distributions with and without low crossing probability deviate es-
392 sentially in count of events.

393 We point out that some misidentified crossings may still be left in the database,
394 particularly in the high longitude region near the terminator, where a clean identifica-
395 tion of crossings can be difficult, due to KHI-induced plasma mixing. Other misidenti-
396 fied crossings which are still included in the database are multiple crossings associated
397 to a single extended magnetopause adjacent Low Latitude Boundary Layer (LLBL). This
398 layer contains a mixture of magnetospheric and magnetosheath plasmas (e.g., Hasegawa,
399 2012), making a clear separation of the regions difficult and false multiple crossing de-
400 tections more likely.

401 4 Magnetopause Crossing Statistics

402 Fig. 5 displays the distribution of all identified MPCs in the database over the stand-
403 off distance, the deviation from the SH98 model in that distance and the latitude and
404 longitude angles of the crossing positions. Separate distributions are shown for higher
405 (> 0.75) and lower (< 0.75) crossing probabilities. In the top panel (a), the stand-off
406 distance distribution is shown. We see a clear asymmetry around the maximum which
407 lies roughly between 10.5 and $11 R_E$: At $11.5 R_E$ a sharp decrease is seen, while for the
408 smaller R_0 we see a smooth slope. The ΔR_0 distribution (panel (b)) indicates a tendency
409 of the SH98 model to predict the MP a little nearer to Earth, as the maximum is at about
410 $0.25 R_E$. This may result from the fact that Shue et al. (1997, 1998) only used the in-
411 nermost MPCs for fitting their model, while we do not restrict the database. Most of
412 the MPCs are found between -1 and $1 R_E$ ($\sim 80\%$) which is consistent with reported SH98
413 model accuracies of $\sim \pm 1 R_E$ (Case & Wild, 2013; Staples et al., 2020). As can be seen
414 in the bottom two panels (c) and (d), the THEMIS orbits lead to MPC observations (1)
415 being widely distributed in longitude ($|\lambda| < 90$) over the dayside and (2) being restricted
416 in latitude to the near-equatorial region ($|\theta| < 30$).

417 If we compare the R_0 -distribution with the distribution of the five THEMIS space-
418 craft dwell times at specific locations (Fig. 6), we see that the probes spent much more
419 time in regions with $R_0 < 11 R_E$. Thus, the asymmetry in the MPC distribution re-
420 sults from this orbit bias which naturally leads to more MPCs at smaller stand-off dis-

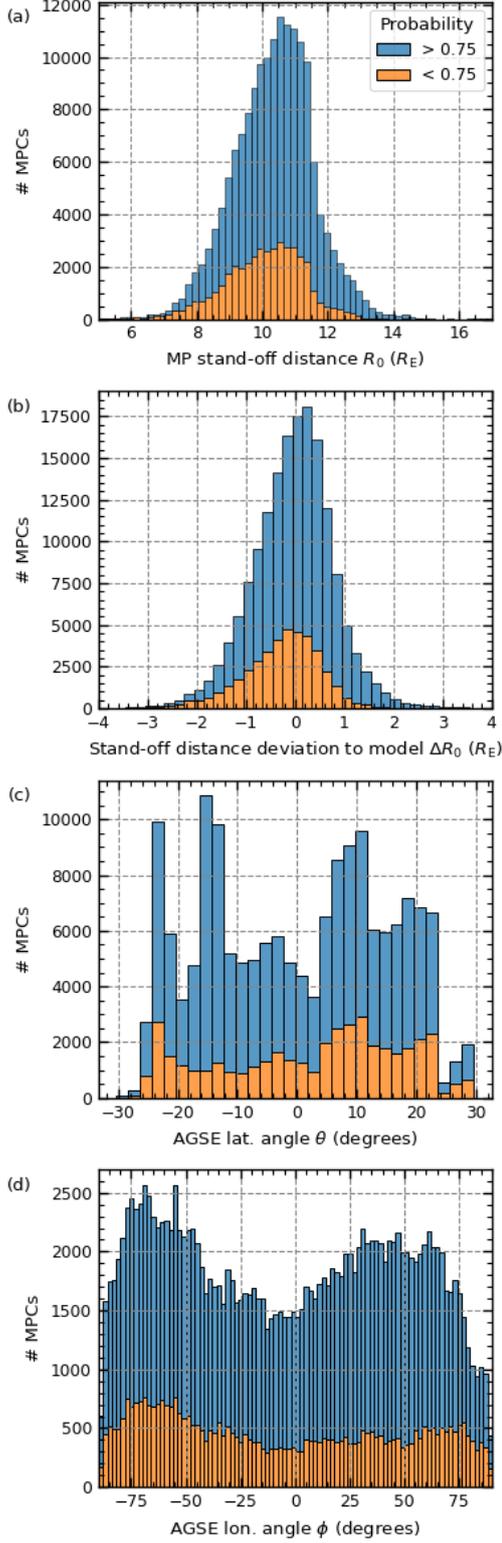


Figure 5. Distribution of detected MPCs, with detection probability ≤ 0.75 in blue and > 0.75 in orange. The panels show from top to bottom the stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, the latitude angle and the longitude angle of the respective MPCs in AGSE coordinates.

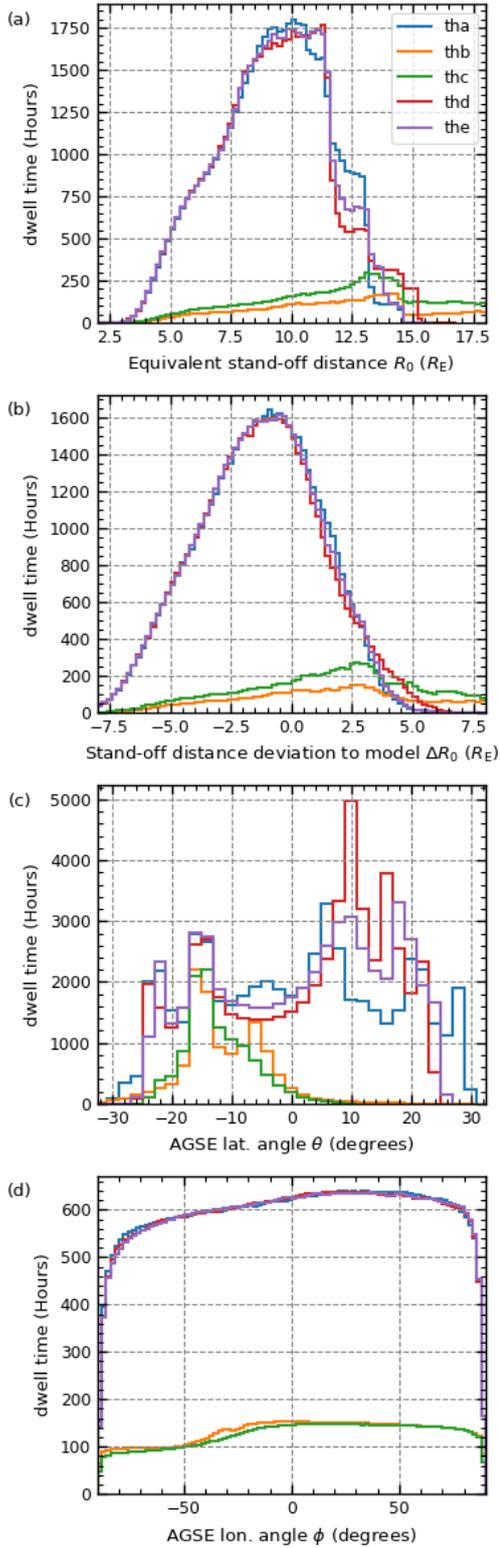


Figure 6. Dwell time distributions of the five THEMIS spacecraft with respect to the stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, the latitude angle and the longitude angle in AGSE coordinates (top to bottom).

421 tances. To compensate this orbit bias, we normalize the distributions, dividing the MPC
422 count by the corresponding cumulative dwell time of all THEMIS spacecraft in each bin.

423 The normalization results can be seen in Fig. 7 showing the probability distribu-
424 tion of MPCs per hour of spacecraft observation time and also a comparison between
425 subsolar ($|\lambda| \leq 30^\circ$) and flank ($|\lambda| > 30^\circ$) MPCs. The orbital bias in the stand-off dis-
426 tance (top panel) is no longer visible and the distribution is quasi symmetrical around
427 $10.7 R_E$ indicated by the very similar mean and median values of the distribution. In-
428 terestingly, the subsolar MPCs occur slightly less frequently (0.86 MPCs/h) and the cor-
429 responding distribution is quite narrow in comparison to the broader flank MPCs dis-
430 tribution, which is centred around $10.5 R_E$.

431 The SH98 model MP is dependent on the flaring parameter α and the stand-off dis-
432 tance R_0 . On the day side, the flaring parameter has little influence on the MP position.
433 Thus, adapting the SH98 MP to the MPC observations is achieved by changing the stand-
434 off distance. At the flanks, motion of the MP results in variability of α . Since we fix the
435 value of α with the prevalent solar wind conditions, all MP motion is attributed to changes
436 in R_0 , potentially leading to a broader distribution in this parameter (see Fig. 7a).

437 In the distribution of the deviations to the model (Fig. 7, panel (b)), the tendency
438 to observe MPCs further away from Earth in comparison to model predictions is visi-
439 ble. Significant positive deviations from the SH98 model ($\Delta R_0 > 1$) result from expan-
440 sions of the MP in the subsolar and flank regions while the significant negative devia-
441 tions ($\Delta R_0 < -1$) result almost only from MP compressions in the flank regions.

442 Looking at the angular distributions of the MPCs, we find a notable asymmetry
443 between the dawn and dusk sectors in the longitude distribution (bottom panel). The
444 mean occurrence rate between -90° and -30° (dawn) is 0.79 MPCs/h while the rate be-
445 tween 30° and 90° (dusk) is 0.63 MPCs/h. In the subsolar sector the occurrence is in
446 general lower than at the flank sectors (0.59 MPCs/h).

447 The MPCs are more or less equally distributed in latitude (panel (c)).

448 5 Solar Wind Statistics

449 5.1 Data Selection

450 The SH98 model magnetopause's location and shape are solely influenced by the
451 solar wind dynamic pressure p_{dyn} and the IMF B_z -component. The model is nominally
452 suitable to make predictions under extreme solar wind conditions which can lead to large
453 deformations of the magnetopause (Shue et al., 1998). However, as shown in panel (b)
454 of Fig. 7, we find numerous MPCs ($\sim 20\%$) outside the model uncertainties of $\pm 1 R_E$ oc-
455 ccurring with rates ≤ 1.0 MPCs per hour.

456 About 7% of the MPCs in the database are classified as extreme deviations from
457 the model stand-off distances, surpassing $\pm 1.5 R_E$. Positive deviations correspond to mag-
458 netospheric expansions and negative deviations to magnetospheric compressions, in the
459 following called expanded MPCs and compressed MPCs, respectively. From Fig. 7 we
460 can infer that extreme expansions occur with rates ≤ 0.57 MPCs per hour and extreme
461 compressions with rates ≤ 0.38 MPCs per hour.

462 These considerably deviating MPCs may be influenced by solar wind parameters
463 that are not considered in the SH98 model. For this study, we associate each MPC from
464 the high probability TH-MPC database with one set of solar wind parameters, comprised
465 of the medians of the IMF magnitude B_{IMF} , the cone angle ϑ_{cone} between the Earth-Sun-
466 line and the IMF vector, the clock angle ϑ_{clock} between the IMF B_y - and B_z -components,
467 the solar wind velocity u_{sw} , the ion density n_{ion} , the ion temperature T_{ion} , the dynamic

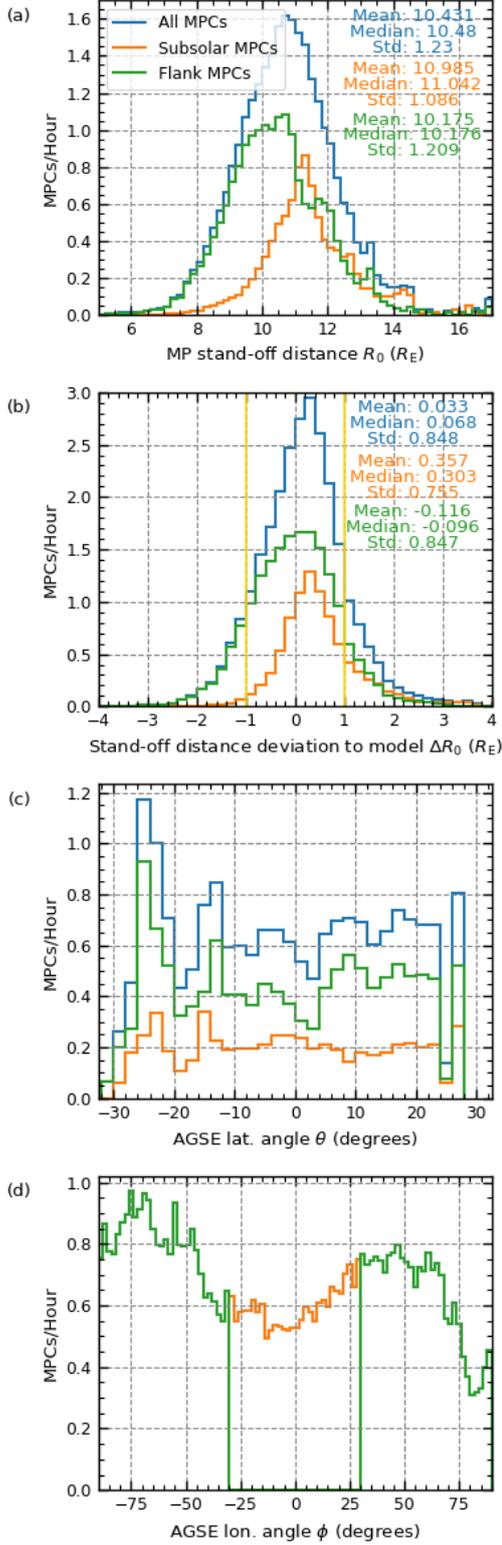


Figure 7. Histograms of the normalized distributions, showing crossing events per hour for each bin. The panels show the same variables as Fig. 3. The blue histogram depicts the hole dataset, while the orange and green histograms depicts the subsolar ($|\lambda| < 30$) and the flank ($|\lambda| > 30$) magnetopause subsets, respectively. The yellow line in panel 2 represents the uncertainty of the SH98 model.

468 pressure p_{dyn} , the plasma β and the Alfvénic Mach number M_A , based on OMNI mea-
 469 surements form 8-minute intervals preceding each MPC.

470 5.2 Parameter Influence

471 To quantify the contribution of different solar wind parameters to the magnetopause
 472 distortions, we compare the whole distribution of the solar wind parameters from our
 473 OMNI dataset with the solar wind parameters associated with the TH-MPC database
 474 and the two extreme MPC subsets of expanded MPCs and compressed MPCs. We nor-
 475 malize each distribution individually by the total number of contributing data points.

476 The distributions with respect to B_{IMF} , ϑ_{cone} , ϑ_{clock} , u_{sw} , n_{ion} , T_{ion} , p_{dyn} , plasma
 477 β and M_A are shown in Figure 8. The OMNI data are shown in black and serves as ref-
 478 erence. The solar wind data during the MPCs are shown in blue, while the orange and
 479 green lines display the distributions associated with extreme MPCs. The maxima and
 480 medians of the datasets are displayed as well, equally colour coded.

481 The solar wind data distributions (in black) agree nicely with results from previ-
 482 ous studies (e.g., Plaschke et al., 2013; L. Q. Zhang et al., 2019; Larrodera & Cid, 2020;
 483 Ma et al., 2020). Furthermore, we find for all parameters an expected similarity in shape
 484 and maximum values between the blue and black distributions, as MPCs should be ob-
 485 served under all possible solar wind conditions over the long time range considered in
 486 this study. However, some of the distributions associated with extreme MPCs notably
 487 differ from the reference distributions, particularly with respect to ϑ_{cone} , u_{sw} , T_{ion} and
 488 M_A , indicating an influence of these parameters on the occurrence of extreme MP dis-
 489 tortions. We compute the quotient of the distributions corresponding to the extreme MPCs
 490 with the reference solar wind distributions to indicate favourable occurrence conditions
 491 in the solar wind parameters. These favourable conditions are visible in quotient max-
 492 ima above 1 and unfavourable conditions in minima under 1. In Fig. 9 these deviations
 493 from the reference distributions are displayed. The errors are computed using the mean
 494 detection rate of 15 MPCs per 1-hour interval as typical count error. In the following,
 495 we discuss the solar wind parameter distributions in the order of ascending influence on
 496 the extreme MPCs.

497 All clock angle distributions (Fig. 8C) show a double peak structure representing
 498 the known feature of the Parker spiral (e.g., L. Q. Zhang et al., 2019). In addition, we
 499 see small deviations in shape with respect to the reference solar wind distribution over
 500 all angles. Some clock angle orientations appear to be slightly more beneficial for the oc-
 501 currence of extreme MPCs (see Fig. 9C). For example, the compressed MPCs show a
 502 tendency to occur under southward IMF conditions ($|\vartheta_{\text{clock}}| \geq 100^\circ$) and the distribu-
 503 tion for the expanded MPCs deviates noticeable around 0° , corresponding to occurrences
 504 during northward IMF. However, the deviations from 1 seen in Fig. 9C are rather small.

505 Although, the influence of the dynamic pressure on the magnetopause location should
 506 be captured by the SH98 model, we still see some subtle deviations in the distributions
 507 (panel G in Fig. 8) hinting at a further influence. We ignore the high peak at 0.3 nPa
 508 for the compressed MPCs in Fig. 9G as this large deviation results from only very few
 509 MPCs in this bin. Favourable conditions for extreme compressed MPCs are slightly higher
 510 pressures between 1.8 nPa and 3.5 nPa. The extreme expanded MPCs occur preferably
 511 under weaker pressures around 1.1 nPa. In both cases, however, the deviations in the
 512 distribution quotients found are less than 1.5. We conclude that the effect of dynamic
 513 pressure on extreme MPCs is already well captured by the SH98 model, as expected.

514 The distributions of plasma β (Fig. 8H) for extreme MPCs are slightly shifted with
 515 respect to the reference distribution. This shift is clearly visible in the maximum and
 516 the median values of the distributions. From Fig. 9H we infer that MP expansions oc-
 517 cur more frequently for slightly higher β values between 2 and 5, and compressions are

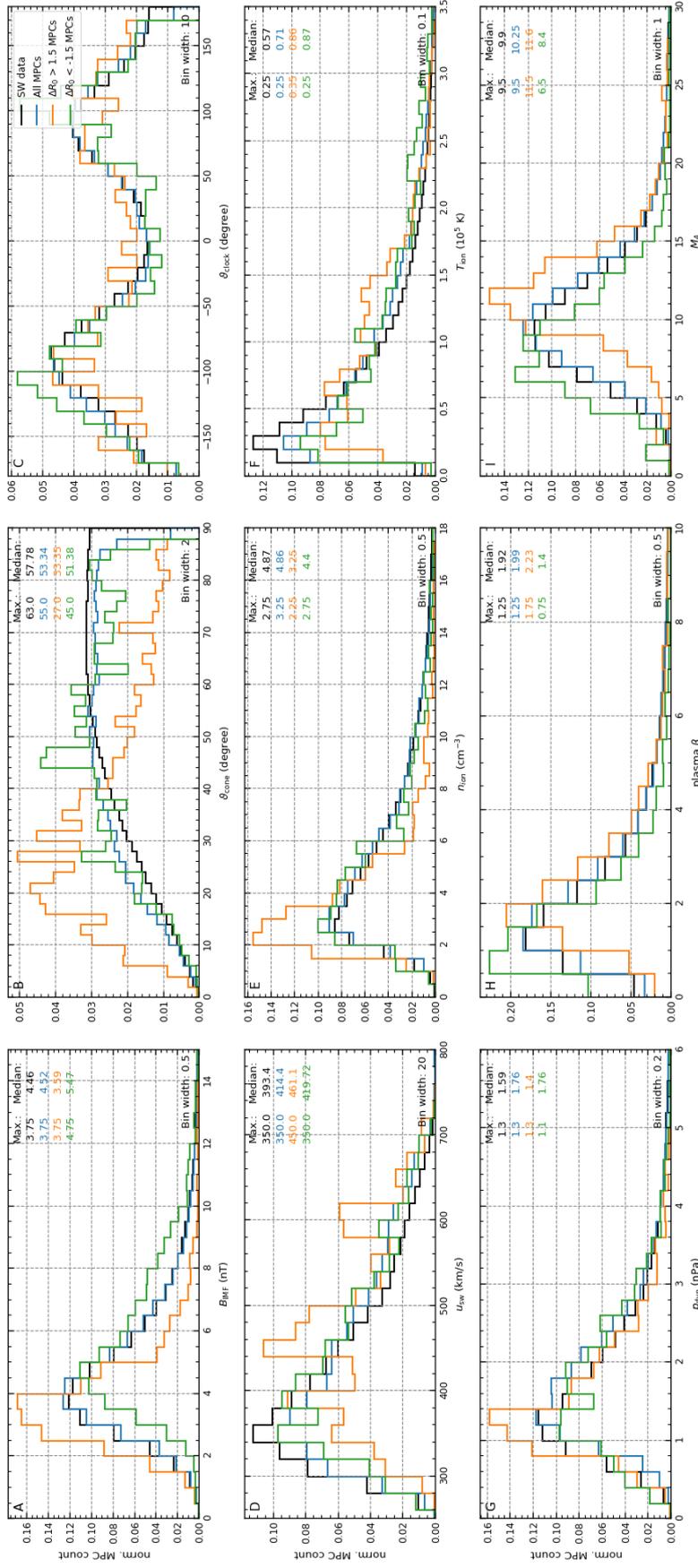


Figure 8. Probability distributions of different solar wind parameters. In each panel the whole solar wind distribution of a specific parameter (black lines) is compared to the parameter distribution pertaining to all MPCs (blue lines) and the extreme MPCs (orange and green lines, respectively). The maximum and the median values are indicated in each panel. All distributions are normalized by the total number of contributing data points.

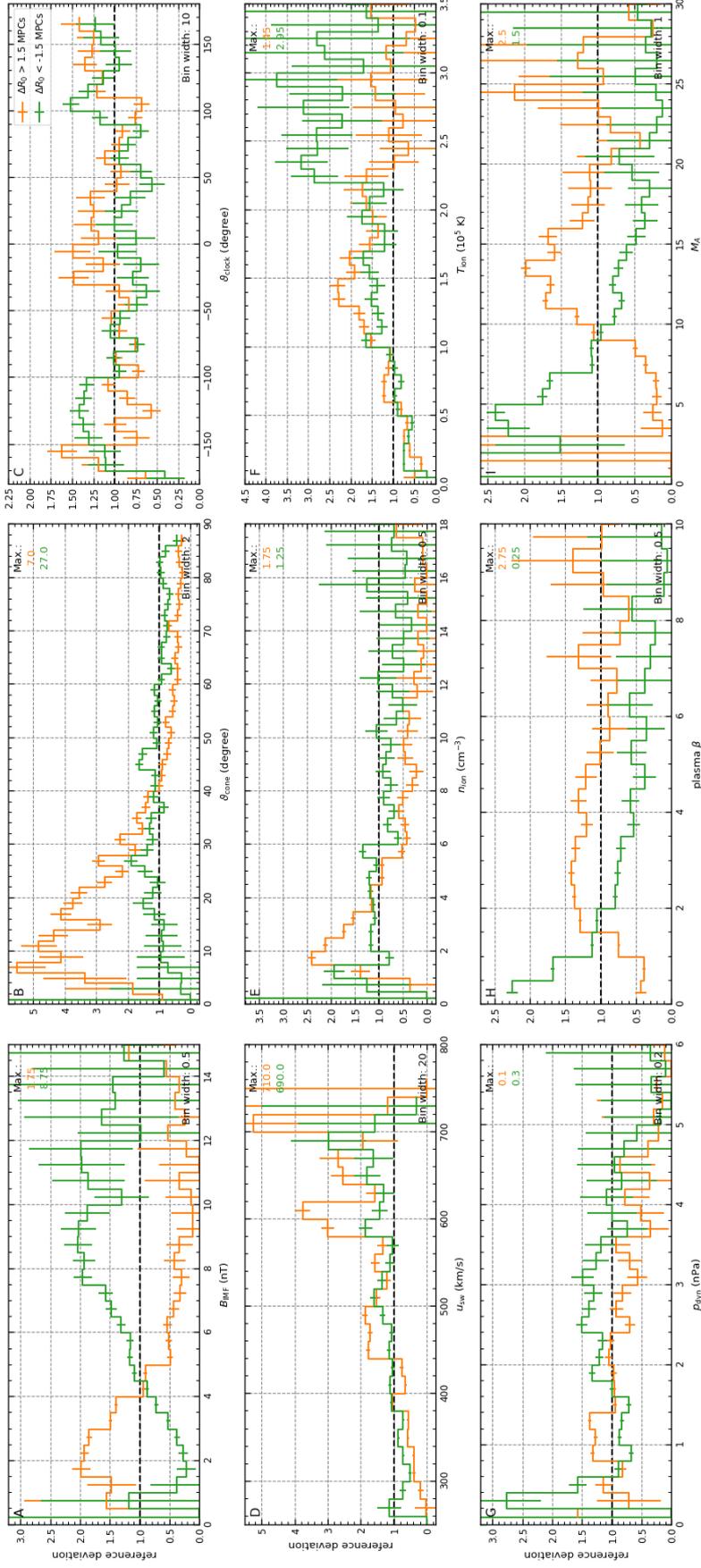


Figure 9. Deviations from the reference distributions (see Fig. 8) corresponding to extreme MPCs, calculated by dividing the distribution by the reference. The orange lines correspond to expanded MPCs and the green lines correspond to compressed MPCs.

518 more frequent for lower values below 1. Thus, higher/lower values lead to more frequent
519 expansions/compressions.

520 In the ion density distributions (Fig. 8E), we find quite different deviations of the
521 distributions for expanded and compressed MPCs. For the expanded MPCs, we can infer
522 from Fig. 9E a clear tendency of higher occurrence rates between density values of
523 1.5 cm^{-3} and 3.5 cm^{-3} . For the compressed MPCs, we find one peak at $n_{\text{ion}} = 1.25$
524 cm^{-3} which might be not reliable, as the bin contains only few MPCs. The other positive
525 deviation for density values between 2 cm^{-3} and 6 cm^{-3} in the distribution quotient
526 is very small.

527 Interestingly, all temperature distributions (Fig. 8F) share a common maximum
528 around $3 \cdot 10^4 \text{ K}$, but differ quite a lot in the median values. We find that the distribu-
529 tions for extreme MPCs are shifted to higher T_{ion} . Both compressed and expanded MPCs
530 seem to occur more frequently in the temperature range between $1.0 \cdot 10^5 \text{ K}$ and $2.1 \cdot 10^5$
531 K (see Fig. 9F). Higher T_{ion} are favourable only for the compressed MPCs. However,
532 in this temperature range we only observed very few MPCs.

533 In panel A of Fig. 8 and 9 we can see that extreme expanded MPCs occur more
534 frequently for smaller IMF magnitudes, with B_{IMF} between 1.5 nT and 4 nT. In con-
535 trast, the distribution of the extreme compressed MPCs is shifted to higher IMF mag-
536 nitudes, indicating favourable conditions above 6 nT.

537 Fig. 8I, depicting the Alfvén Mach number distributions, shows obvious deviations
538 between the reference and the extreme MPC distributions. The maxima and medians
539 for the compressed and expanded MPCs deviate substantially from the reference, and
540 we can clearly infer favourable conditions from Fig. 9I: For the expanded MPCs, we see
541 the maximal occurrence rate at $M_A = 11.5$ and favourable conditions of M_A between
542 11 and 16. For the compressed MPCs, we find the maximum at $M_A = 4.5$ and favourable
543 conditions of M_A between 3 and 7.

544 Both expanded and compressed MPCs seem to occur more frequently under high
545 u_{sw} conditions (above 440 km/s). This trend is more clearly visible for the expanded MPCs
546 (see Fig. 8D and 9D).

547 Lastly, we find a significant influence of ϑ_{cone} on extreme expanded MPCs. Quasi-
548 radial IMF conditions ($\vartheta_{\text{cone}} < 35^\circ$) clearly favour expanded MPCs (see panel B in Fig.
549 8 and 9). No similar feature can be seen with respect to the occurrence of compressed
550 MPCs as a function of ϑ_{cone} . However, $\vartheta_{\text{cone}} 25^\circ$ and 30° could be a favourable condi-
551 tion for the compressed MPCs.

552 6 Discussion

553 In Fig. 7 (a), we find a quite symmetrical distribution of stand-off distances around
554 $10.7 R_E$, which can be regarded as typical (e.g. Baumjohann & Treumann, 1997). In com-
555 parison with stand-off distance predictions by the Shue et al. (1998) model (panel (b)),
556 we find a slight tendency of the model to underestimate the stand-off distance, which
557 probably results from the fact that Shue et al. (1998) only used the innermost crossings
558 of MP encounters for fitting the model parameters.

559 In the longitude distribution of the MPCs (panel (d) of Fig. 7) we see a tendency
560 to observe more MPCs at the magnetospheric flanks and a clear asymmetry between the
561 occurrence rates in the dawn and dusk sectors. At the flanks, occurrences of KHI waves
562 are likely (Taylor et al., 2012; Johnson et al., 2014) which should lead to frequent move-
563 ment of the MP and more observations of MPCs. Additionally, as already mentioned,
564 the Random Forest machine learning algorithm has some difficulties to clearly distinguish
565 the magnetosphere and magnetosheath regions in case of thicker boundary layers lead-

566 ing to multiple crossing detections. We try to mitigate this problem by only studying
 567 MPCs with high crossing probabilities. Some remnants of this multiple MPCs might still
 568 be in the database, resulting in a tendency to observe more flank MPCs.

569 Interestingly, KHIs are reported to occur more frequently at the dusk flank (e.g.
 570 Taylor et al., 2012). Thus, if the KHI is responsible for the higher occurrence rates, we
 571 should see more MPCs at the dusk than at the dawn flank. However, we observe a slightly
 572 higher occurrence rate for MPCs at the dawn flank of the magnetosphere, which has a
 573 thicker magnetopause boundary layer (Walsh et al., 2014). This result may give more
 574 weight to a possible explanation in terms of multiple MPC detections by the Random
 575 Forest Classifier. Another explanation for the asymmetry could be that the magnetopause
 576 moves more frequently in the dawn sector due to the thinner and more turbulent mag-
 577 netosheath (Walsh et al., 2014). The foreshock will more often be located in this sector
 578 and excite more frequent MP movement, resulting in more frequent MPCs and there-
 579 fore in higher occurrence rates.

580 By comparing our database to previous magnetopause studies, we can find out which
 581 explanation might be more reasonable. For example, Staples et al. (2020) used a threshold-
 582 based detection algorithm to study MPCs observed by THEMIS. Overall, they looked
 583 at THEMIS data from 2007 to 2016 and only kept the innermost crossings of multiple
 584 MPCs in a 10-min interval. Their MPC distributions (see Fig. A1) and ours are very
 585 similar, giving us confidence in our detection method using the Random Forest Classi-
 586 fier. By removing of multiple crossings, the above-mentioned higher occurrence rates due
 587 to KHIs or the detection method should not be visible in the distributions from Staples
 588 et al. (2020). However, their database also is subject to the dawn-dusk asymmetry in
 589 MPC occurrences. Hence, the reason for this asymmetry is more likely the more frequent
 590 occurrence of MP movement in the dawn sector possible linked to the foreshock or the
 591 more turbulent magnetosheath downstream of the quasi-parallel shock. Nevertheless, fur-
 592 ther investigations are necessary to fully understand this dawn-dusk asymmetry in MPC
 593 occurrences.

594 Let us now have a look at the roughly 7% of the identified MPCs that deviate dras-
 595 tically from the model predictions, that may not be immediately explained by changes
 596 in the solar wind dynamic pressure or the B_2 -component of the IMF. From the compar-
 597 ison of the solar wind parameters during these extreme MPCs with the standard solar
 598 wind parameter distributions, we are able to infer some significant solar wind param-
 599 eter influences on magnetopause location:

600 The most obvious influence pertains to the IMF cone angle, which controls the ex-
 601 pansion of the magnetosphere as reported before (e.g., Slavin et al., 1996; Merka et al.,
 602 2003; Suvorova et al., 2010; Park et al., 2016; M. Wang et al., 2020). Under radial or quasi-
 603 radial IMF conditions, the dayside bow shock location is closer to Earth than on aver-
 604 age, the magnetosheath thickness decreases, and the dayside magnetopause moves sun-
 605 ward. This happens in parts due to the establishment of a quasi-parallel foreshock in the
 606 subsolar region, which redistributes the dynamic pressure of the solar wind plasma and
 607 yields a lower magnetic pressure, affecting the magnetosphere. Additionally, the total
 608 plasma pressure is strongly modified in the bow shock crossing and distributed due to
 609 the flow diversion in the magnetosheath across the dayside magnetopause surface (Suvorova
 610 et al., 2010; Samsonov et al., 2012) leading to an expanding magnetopause to re-establish
 611 the pressure balance.

612 Extreme compressions might also occur under quasi-radial IMF conditions ($\vartheta_{\text{cone}} \approx$
 613 30°). As Archer and Horbury (2013), Plaschke et al. (2013) and LaMoury et al. (2021)
 614 point out, HSJs occur more often under these conditions. Shue et al. (2009) and Archer
 615 et al. (2019) observed significant indentations of the magnetopause caused by a HSJ un-
 616 der radial IMF. Thus, the higher occurrence rate for compressions may be linked with
 617 such HSJs.

618 Substantial influences on extreme MP distortions stemming from the magnitude
 619 of the IMF, the plasma β and the Alfvén Mach number might in fact result from the same
 620 source: Extreme expansions of the magnetopause occur more frequently for small IMF
 621 magnitudes, i.e., values like the Alfvén velocity or the magnetic pressure are small as well.
 622 Naturally, small magnetic pressures and Alfvén velocities lead to higher plasma β and
 623 Alfvén Mach numbers, respectively, which are favourable conditions for extreme mag-
 624 netospheric expansions as well. The expansions are possibly related to the higher par-
 625 ticle reflection rates, leading to stronger kinetic energy dissipation at the bow shock un-
 626 der these condition (Winterhalter & Kivelson, 1988; Treumann, 2009), strengthening the
 627 foreshock region and thereby reducing the pressure on the dayside magnetosphere. On
 628 the contrary, the compressed MPCs occur more frequently for low Mach numbers and
 629 low plasma β , which result from higher IMF magnitudes. This might be connected to
 630 the fact, that the solar wind Mach number controls the magnetosheath plasma β . For
 631 low magnetosheath plasma β resulting from low solar wind Mach numbers, reconnection
 632 is more likely, leading to flux erosion and compression of the magnetosphere. Further-
 633 more, we can infer that for magnetospheric compressions the magnetic pressure is dom-
 634 inant in the solar wind ($\beta < 1$), and for the magnetospheric expansions thermal pres-
 635 sure is dominant ($\beta > 1$).

636 Both in the velocity and in the temperature distributions (Fig. 8 D and F), we iden-
 637 tified a shift to higher values for extreme MPCs, especially noticeable for magnetospheric
 638 expansions. These higher values in the velocity probably lead to an increase in the ve-
 639 locity shear across the magnetopause triggering stronger KHIs and subsequently ex-
 640 citing extreme oscillatory MP motion (Kavosi & Raeder, 2015). Additionally, Chu et al.
 641 (2017) and Vu et al. (2022) found that many different foreshock structures like FBs and
 642 HFAs were observed during radial IMF with solar wind velocities around 600 km/s. Gen-
 643 erally, the favourable conditions of fast solar wind with large Alfvén Mach numbers and
 644 low cone angles for the extreme expansions coincide with favourable conditions for the
 645 occurrence of these foreshock transients (H. Zhang et al., 2022). These phenomena are
 646 characterized by hot tenuous plasma regions in the foreshock region, in which flow de-
 647 flection and pressure reduction occur (Turner et al., 2013). On impact on the MP this
 648 pressure "hole" lead to an expansion followed by a compression of the magnetosphere
 649 (e.g., Sibeck et al., 1999; Turner et al., 2011; Archer et al., 2014, 2015). Thus, these fore-
 650 shock transients might play an additional role in the extreme expansions and compres-
 651 sions of the magnetopause.

652 Weak but noticeable deviations in the clock angle distributions may stem from resid-
 653 ual effects of the IMF B_z component, which is captured in the SH98 model. The pos-
 654 sible favourable condition around 0° for expanded MPCs coincide with fact that the mag-
 655 netopause stand-off distance is larger for northward IMF, when no reconnection and flux
 656 erosion is happening at the dayside. For southward IMF reconnection and flux erosion
 657 occur, driving the MP inwards and favouring observations of compressed MPCs under
 658 this condition.

659 The influence of the dynamic pressure is already prominently captured in the SH98
 660 model. Therefore, the pressure effect would not appear as substantial deviation in our
 661 plots. All influences on the magnetopause presented here are additional effects.

662 The deviation in the density distributions (Fig. 9E) for the compressed MPCs are
 663 also negligible, showing no significant favourable condition. In the distribution associ-
 664 ated with the expanded MPCs we can see a minor preference for more tenuous solar wind
 665 plasma. Tenuous plasma causes a decrease in the thermal pressure, therefore reducing
 666 the total pressure impacting the magnetosphere.

667 Overall, we find more significant deviations from the reference distributions asso-
 668 ciated with the expanded MPCs, i.e., magnetospheric expansions are less well covered

669 by the SH98 model. Therefore, we identify more favourable conditions for a extreme mag-
 670 netospheric expansions.

671 7 Conclusion

672 In this study, the last 15 years of THEMIS observations have been used to build
 673 a very large MPC database. This database allows us to examine extreme MP distortions
 674 in detail in special case or statistical studies.

675 Our statistical study shows that parameters such as the Alfvén Mach number, the
 676 IMF cone angle and the ion velocity are responsible for quite frequent occurrences of ex-
 677 treme magnetopause distortions. Quasi-radial IMF conditions with a plasma $\beta > 1$, higher
 678 Alfvén Mach numbers and ion velocities above 450 km/s are favourable for magnetospheric
 679 expansions beyond the SH98 model predictions, while magnetospheric compressions are
 680 associated with more southward IMF conditions with plasma $\beta < 1$, lower Alfvén Mach
 681 numbers and IMF strengths above 5 nT.

682 The expansions of the magnetopause under high Mach number and velocity con-
 683 ditions are possible linked to KHIs and also foreshock transients, while other phenom-
 684 ena like magnetosheath jets might be responsible for some compressions. This could be
 685 studied further by comparing the observation times of such phenomena with our database.
 686 In sorting the extreme MPCs by possible origin mechanisms, we also hope to learn more
 687 about main drivers behind the extreme events.

688 Here we only study MPCs in low latitudes, observed on the dayside. With the uti-
 689 lization of CLUSTER data, we plan to expand our database to high latitudes (e.g., Panov
 690 et al., 2008), allowing for a comparison between the equatorial and more polar regions.

691 With the upcoming SMILE mission, the shape and location of the MP will be di-
 692 rectly inferred and linked to in-situ measurements of solar wind conditions. This will al-
 693 low for an immediate comparison with the results of this study and open the door for
 694 further investigations of extreme MP distortions.

695 Appendix A Threshold based MPC database

696 Staples et al. (2020) used the following criteria for the identification of MPCs on
 697 the dayside magnetosphere:

- 698 1. During a THEMIS crossing from the magnetosphere to the magnetosheath $\Delta B_{z,\text{gsm}} <$
 699 $-0.6 \frac{\text{nT}}{\text{s}}$ and $\Delta n_{\text{ion}} > 0.08 \frac{1}{\text{cm}^3 \text{s}}$ should hold over the crossing. These criteria are
 700 reversed for crossings from the magnetosheath to the magnetosphere.
- 701 2. In average, $B_{z,\text{gsm}} > 5 \text{ nT}$ and $n_{\text{ion}} < 7 \text{ cm}^{-3}$ should hold for a 48-s interval
 702 within the magnetosphere before/after a possible event.
- 703 3. These two criteria must be satisfied in a 60-s-interval.

704 They completed their database with additional crossings identified by Plaschke, Glass-
 705 meier, Auster, Angelopoulos, et al. (2009), then removed duplicate crossings and reduced
 706 the database to the innermost crossings. Fig. A1 displays the distributions of their database
 707 identically normed as our database distributions.

708 Open Research

709 The magnetopause crossing event database Grimmich et al. (2023) used for this
 710 study is publicly available under <https://osf.io/b6kux/>, hosted by the Open Science
 711 Framework (OSF). To collect and plot data, we used the open source Python Space Physics
 712 Environment Data Analysis Software (pySPEDAS) which can be found here: <https://>

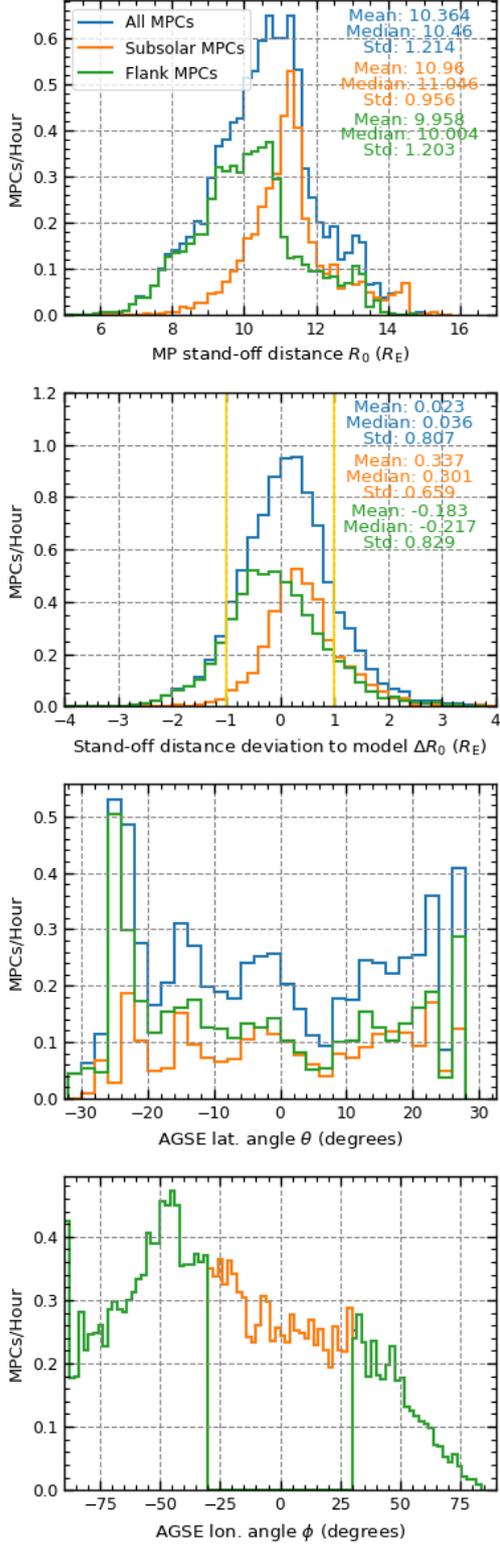


Figure A1. Histogram of normalized MPC distribution based on the database by Staples et al. (2020), showing crossing events per hour for each bin. The panels show from top to bottom: The stand-off distance of the MP, the deviation of this distance from the SH98 model stand-off distance, and the latitude and the longitude angles in AGSE coordinates. The blue histogram depicts the hole dataset, while the orange and green histograms depict the subsolar ($|\lambda| < 30$) and the flank ($|\lambda| > 30$) magnetopause subsets, respectively.

713 github.com/spedas/pyspedas. THEMIS data can be retrieved from <http://themis>
 714 [.ssl.berkeley.edu/data/themis/](http://ssl.berkeley.edu/data/themis/) and OMNI data from the GSFC/SPDF OMNIWeb
 715 interface at <https://omniweb.gsfc.nasa.gov>. The machine learning task were performed
 716 with the scikit-learn Python library, from which we utilized the implementations of the
 717 different algorithms. The documentation can be found here: [https://scikit-learn.org/
 718 stable/supervised_learning.html#supervised-learning](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning).

719 Acknowledgments

720 We acknowledge NASA contract NAS5-02099 and V. Angelopoulos for use of data from
 721 the THEMIS Mission. Specifically: C. W. Carlson and J. P. McFadden for use of ESA
 722 data and K. H. Glassmeier, U. Auster and W. Baumjohann for the use of FGM data pro-
 723 vided under the lead of the Technical University of Braunschweig and with financial sup-
 724 port through the German Ministry for Economy and Technology and the German Cen-
 725 ter for Aviation and Space (DLR) under contract 50OC 0302. We thank J. King and N.
 726 Papitashvili of the National Space Science Data Center (NSSDC) in the NASA/GSFC
 727 for the use of the OMNI 2 database. FP and JZDM are supported by the Deutsches Zen-
 728 trum für Luft- und Raumfahrt under contract 50OC2201. DH was supported by the Ger-
 729 man Ministerium für Wirtschaft und Klimaschutz and the German Zentrum für Luft-
 730 und Raumfahrt under contracts 50QW1501 and 50QW2202. The authors want to thank
 731 Nick Hatzigeorgiu and Eric Grimes for the ongoing development of PySPEDAS.

732 References

- 733 Angelopoulos, V. (2008, apr). The THEMIS Mission. *Space Sci. Rev.*, *141*(1-4), 5-
 734 34. doi: 10.1007/s11214-008-9336-1
- 735 Angelopoulos, V. (2011, dec). The ARTEMIS Mission. *Space Sci. Rev.*, *165*(1-4), 3-
 736 25. doi: 10.1007/s11214-010-9687-2
- 737 Archer, M. O., Hietala, H., Hartinger, M. D., Plaschke, F., & Angelopoulos, V.
 738 (2019, feb). Direct observations of a surface eigenmode of the dayside magne-
 739 topause. *Nature Communications*, *10*, 615. doi: 10.1038/s41467-018-08134-5
- 740 Archer, M. O., & Horbury, T. S. (2013, feb). Magnetosheath dynamic pressure en-
 741 hancements: occurrence and typical properties. *Annales Geophysicae*, *31*(2),
 742 319-331. doi: 10.5194/angeo-31-319-2013
- 743 Archer, M. O., Turner, D. L., Eastwood, J. P., Horbury, T. S., & Schwartz, S. J.
 744 (2014, oct). The role of pressure gradients in driving sunward magnetosheath
 745 flows and magnetopause motion. *Journal of Geophysical Research (Space*
 746 *Physics)*, *119*(10), 8117-8125. doi: 10.1002/2014JA020342
- 747 Archer, M. O., Turner, D. L., Eastwood, J. P., Schwartz, S. J., & Horbury, T. S.
 748 (2015, feb). Global impacts of a Foreshock Bubble: Magnetosheath, magne-
 749 topause and ground-based observations. *Planetary and Space Science*, *106*,
 750 56-66. doi: 10.1016/j.pss.2014.11.026
- 751 Aubry, M. P., Russell, C. T., & Kivelson, M. G. (1970, jan). Inward motion of
 752 the magnetopause before a substorm. *Journal of Geophysical Research*, *75*(34),
 753 7018. doi: 10.1029/JA075i034p07018
- 754 Auster, H. U., Glassmeier, K. H., Magnes, W., Aydogar, O., Baumjohann, W.,
 755 Constantinescu, D., ... Wiedemann, M. (2008, dec). The THEMIS Flux-
 756 gate Magnetometer. *Space Sci. Rev.*, *141*(1-4), 235-264. doi: 10.1007/
 757 s11214-008-9365-9
- 758 Baumjohann, W., & Treumann, R. (1997). *Basic Space Plasma Physics*. Imperial
 759 College Press.
- 760 Branduardi-Raymont, G., Wang, C., Escoubet, C., Adamovic, M., Agnolon, D.,
 761 Berthomier, M., ... Zhu, Z. (2018). SMILE definition study report. *European*
 762 *Space Agency, ESA/SCI*, *1*.
- 763 Breiman, L. (2001, jan). Random Forests. *Machine Learning*, *45*, 5-32. doi: 10

- 764 .1023/A:1010933404324
765 Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. S. (1984). *Classification*
766 *and Regression Trees (1st Edition)*. Chapman & Hall/CRC. doi: 10.1201/
767 9781315139470
768 Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G.
769 (2020, sep). Automatic classification of plasma regions in near-Earth space
770 with supervised machine learning: application to Magnetospheric Multi Scale
771 2016-2019 observation. *Frontiers in Astronomy and Space Sciences*, 7, 55. doi:
772 10.3389/fspas.2020.00055
773 Burch, J. L., Moore, T. E., Torbert, R. B., & Giles, B. L. (2016, mar). Magneto-
774 spheric Multiscale Overview and Science Objectives. *Space Sci. Rev.*, 199(1-4),
775 5-21. doi: 10.1007/s11214-015-0164-9
776 Case, N. A., & Wild, J. A. (2013, oct). The location of the Earth's magne-
777 topause: A comparison of modeled position and in situ Cluster data. *Jour-*
778 *nal of Geophysical Research (Space Physics)*, 118(10), 6127-6135. doi:
779 10.1002/jgra.50572
780 Chao, J. K., Wu, D. J., Lin, C. H., Yang, Y. H., Wang, X. Y., Kessel, M., ... Lep-
781 ping, R. P. (2002, jan). Models for the Size and Shape of the Earth's Mag-
782 netopause and Bow Shock. In L.-H. Lyu (Ed.), *Space weather study using*
783 *multipoint techniques* (p. 127).
784 Chu, C., Zhang, H., Sibeck, D., Otto, A., Zong, Q., Omid, N., ... Angelopou-
785 los, V. (2017, mar). THEMIS satellite observations of hot flow anom-
786 alies at Earth's bow shock. *Annales Geophysicae*, 35(3), 443-451. doi:
787 10.5194/angeo-35-443-2017
788 Cox, D. R., & Snell, E. J. (1970). *Analysis of Binary Data (2nd Edition)*. Chapman
789 & Hall/CRC. doi: 10.1201/9781315137391
790 Dušík, Š., Granko, G., Šafránková, J., Němeček, Z., & Jelínek, K. (2010, oct). IMF
791 cone angle control of the magnetopause location: Statistical study. *Geophysical*
792 *Research Letters*, 37(19), L19103. doi: 10.1029/2010GL044965
793 Eastwood, J. P., Lucek, E. A., Mazelle, C., Meziane, K., Narita, Y., Pickett, J., &
794 Treumann, R. A. (2005, jun). The Foreshock. *Space Sci. Rev.*, 118(1-4), 41-94.
795 doi: 10.1007/s11214-005-3824-3
796 Elphic, R. C. (1995, jan). Observations of Flux Transfer Events: A Review. *Geo-*
797 *physical Monograph Series*, 90, 225. doi: 10.1029/GM090p0225
798 Escoubet, C. P., Fehringer, M., & Goldstein, M. (2001, oct). Introduction The Clus-
799 ter mission. *Annales Geophysicae*, 19, 1197-1200. doi: 10.5194/angeo-19-1197
800 -2001
801 Fairfield, D. H. (1971, jan). Average and unusual locations of the Earth's magne-
802 topause and bow shock. *Journal of Geophysical Research*, 76(28), 6700. doi:
803 10.1029/JA076i028p06700
804 Fairfield, D. H., Baumjohann, W., Paschmann, G., Luehr, H., & Sibeck, D. G.
805 (1990, apr). Upstream pressure variations associated with the bow shock
806 and their effects on the magnetosphere. *Journal of Geophysical Research*,
807 95(A4), 3773-3786. doi: 10.1029/JA095iA04p03773
808 Friedman, J. H. (2001, oct). Greedy function approximation: A gradient boost-
809 ing machine. *The Annals of Statistics*, 29(5), 1189 – 1232. Retrieved from
810 <https://doi.org/10.1214/aos/1013203451> doi: 10.1214/aos/1013203451
811 Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and Tensor-*
812 *Flow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Me-
813 dia, Inc.
814 Grimmich, N., Plaschke, F., Archer, M., Heyner, D., Mieth, J., Nakamura, R., &
815 Sibeck, D. (2023). *Database: THEMIS magnetopause crossings between 2007*
816 *and mid-2022* [dataset]. Open Science Framework (OSF). Retrieved from
817 <https://osf.io/b6kux/> doi: 10.17605/OSF.IO/B6KUX
818 Grygorov, K., Šafránková, J., Němeček, Z., Pi, G., Přech, L., & Urbář, J. (2017,

- 819 nov). Shape of the equatorial magnetopause affected by the radial inter-
 820 planetary magnetic field. *Planetary and Space Science*, 148, 28-34. doi:
 821 10.1016/j.pss.2017.09.011
- 822 Haaland, S., Hasegawa, H., Paschmann, G., Sonnerup, B., & Dunlop, M. (2021,
 823 aug). 20 Years of Cluster Observations: The Magnetopause. *Journal*
 824 *of Geophysical Research (Space Physics)*, 126(8), e29362. doi: 10.1029/
 825 2021JA029362
- 826 Hasegawa, H. (2012, aug). Structure and Dynamics of the Magnetopause and Its
 827 Boundary Layers. *Monographs on Environment, Earth and Planets*, 1(2), 71-
 828 119. doi: 10.5047/meep.2012.00102.0071
- 829 Jacobsen, K. S., Phan, T. D., Eastwood, J. P., Sibeck, D. G., Moen, J. I., An-
 830 gelopoulos, V., ... Fornaçon, K. H. (2009, aug). THEMIS observations
 831 of extreme magnetopause motion caused by a hot flow anomaly. *Jour-*
 832 *nal of Geophysical Research (Space Physics)*, 114(A8), A08210. doi:
 833 10.1029/2008JA013873
- 834 Johnson, J. R., Wing, S., & Delamere, P. A. (2014, nov). Kelvin Helmholtz Insta-
 835 bility in Planetary Magnetospheres. *Space Sci. Rev.*, 184(1-4), 1-31. doi: 10
 836 .1007/s11214-014-0085-z
- 837 Kavosi, S., & Raeder, J. (2015, may). Ubiquity of Kelvin-Helmholtz waves at
 838 Earth's magnetopause. *Nature Communications*, 6, 7019. doi: 10.1038/
 839 ncomms8019
- 840 King, J. H., & Papitashvili, N. E. (2005, feb). Solar wind spatial scales in and
 841 comparisons of hourly Wind and ACE plasma and magnetic field data.
 842 *Journal of Geophysical Research (Space Physics)*, 110(A2), A02104. doi:
 843 10.1029/2004JA010649
- 844 Kuntz, K. D. (2019, jan). Solar wind charge exchange: an astrophysical nuisance.
 845 *The Astronomy and Astrophysics Review*, 27(1), 1. doi: 10.1007/s00159-018
 846 -0114-0
- 847 LaMoury, A. T., Hietala, H., Plaschke, F., Vuorinen, L., & Eastwood, J. P. (2021,
 848 sep). Solar Wind Control of Magnetosheath Jet Formation and Propagation to
 849 the Magnetopause. *Journal of Geophysical Research (Space Physics)*, 126(9),
 850 e29592. doi: 10.1029/2021JA029592
- 851 Larrodera, C., & Cid, C. (2020, mar). Bimodal distribution of the solar wind at
 852 1 AU. *Astronomy and Astrophysics*, 635, A44. doi: 10.1051/0004-6361/
 853 201937307
- 854 Lepping, R. P., Acuña, M. H., Burlaga, L. F., Farrell, W. M., Slavin, J. A., Schat-
 855 ten, K. H., ... Worley, E. M. (1995, feb). The Wind Magnetic Field Investiga-
 856 tion. *Space Sci. Rev.*, 71(1-4), 207-229. doi: 10.1007/BF00751330
- 857 Lin, R. L., Zhang, X. X., Liu, S. Q., Wang, Y. L., & Gong, J. C. (2010, apr). A
 858 three-dimensional asymmetric magnetopause model. *Journal of Geophysical*
 859 *Research (Space Physics)*, 115(A4), A04207. doi: 10.1029/2009JA014235
- 860 Ma, X., Nykyri, K., Dimmock, A., & Chu, C. (2020, oct). Statistical Study of Solar
 861 Wind, Magnetosheath, and Magnetotail Plasma and Field Properties: 12+
 862 Years of THEMIS Observations and MHD Simulations. *Journal of Geophysical*
 863 *Research (Space Physics)*, 125(10), e28209. doi: 10.1029/2020JA028209
- 864 McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley,
 865 P., & Griffee, J. W. (1998, jul). Solar Wind Electron Proton Alpha Monitor
 866 (SWEPAM) for the Advanced Composition Explorer. *Space Sci. Rev.*, 86,
 867 563-612. doi: 10.1023/A:1005040232597
- 868 McFadden, J. P., Carlson, C. W., Larson, D., Ludlam, M., Abiad, R., Elliott,
 869 B., ... Angelopoulos, V. (2008, dec). The THEMIS ESA Plasma Instru-
 870 ment and In-flight Calibration. *Space Sci. Rev.*, 141(1-4), 277-302. doi:
 871 10.1007/s11214-008-9440-2
- 872 Merka, J., Szabo, A., Šafránková, J., & Němeček, Z. (2003, jul). Earth's bow shock
 873 and magnetopause in the case of a field-aligned upstream flow: Observation

- 874 and model comparison. *Journal of Geophysical Research (Space Physics)*,
875 *108*(A7), 1269. doi: 10.1029/2002JA009697
- 876 Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine,
877 D. (2022a, jan). Massive Multi-Mission Statistical Study and Analytical Mod-
878 eling of the Earth’s Magnetopause: 1. A Gradient Boosting Based Automatic
879 Detection of Near-Earth Regions. *Journal of Geophysical Research (Space*
880 *Physics)*, *127*(1), e29773. doi: 10.1029/2021JA029773
- 881 Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., &
882 Fontaine, D. (2022b, jan). Massive Multi-Mission Statistical Study and
883 Analytical Modeling of the Earth’s Magnetopause: 2. Shape and Location.
884 *Journal of Geophysical Research (Space Physics)*, *127*(1), e29774. doi:
885 10.1029/2021JA029774
- 886 Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., & Fontaine,
887 D. (2022c, jan). Massive Multi-Mission Statistical Study and Analytical
888 Modeling of the Earth’s Magnetopause: 3. An Asymmetric Non Indented Mag-
889 netopause Analytical Model. *Journal of Geophysical Research (Space Physics)*,
890 *127*(1), e30112. doi: 10.1029/2021JA030112
- 891 Ogilvie, K. W., Chornay, D. J., Fritzenreiter, R. J., Hunsaker, F., Keller, J., Lo-
892 bell, J., ... Gergin, E. (1995, feb). SWE, A Comprehensive Plasma In-
893 strument for the Wind Spacecraft. *Space Sci. Rev.*, *71*(1-4), 55-77. doi:
894 10.1007/BF00751326
- 895 Olshevsky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzén,
896 S., ... Markidis, S. (2021, oct). Automated Classification of Plasma Regions
897 Using 3D Particle Energy Distributions. *Journal of Geophysical Research*
898 *(Space Physics)*, *126*(10), e29620. doi: 10.1029/2021JA029620
- 899 Panov, E. V., Büchner, J., FränZ, M., Korth, A., Savin, S. P., Rème, H., &
900 Fornaçon, K. H. (2008, jan). High-latitude Earth’s magnetopause outside the
901 cusp: Cluster observations. *Journal of Geophysical Research (Space Physics)*,
902 *113*(A1), A01220. doi: 10.1029/2006JA012123
- 903 Park, J.-S., Shue, J.-H., Kim, K.-H., Pi, G., Němeček, Z., & Áfránková, J. (2016,
904 jul). Global expansion of the dayside magnetopause for long-duration radial
905 IMF events: Statistical study on GOES observations. *Journal of Geophysical*
906 *Research (Space Physics)*, *121*(7), 6480-6492. doi: 10.1002/2016JA022772
- 907 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ...
908 Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of*
909 *Machine Learning Research*, *12*, 2825–2830.
- 910 Plaschke, F., Glassmeier, K. H., Auster, H. U., Angelopoulos, V., Constantinescu,
911 O. D., Fornaçon, K. H., ... Nakamura, R. (2009, jan). Statistical study of the
912 magnetopause motion: First results from THEMIS. *Journal of Geophysical*
913 *Research (Space Physics)*, *114*, A00C10. doi: 10.1029/2008JA013423
- 914 Plaschke, F., Glassmeier, K. H., Auster, H. U., Constantinescu, O. D., Magnes, W.,
915 Angelopoulos, V., ... McFadden, J. P. (2009, jan). Standing Alfvén waves
916 at the magnetopause. *Geophysical Research Letters*, *36*(2), L02104. doi:
917 10.1029/2008GL036411
- 918 Plaschke, F., Hietala, H., & Angelopoulos, V. (2013, oct). Anti-sunward high-speed
919 jets in the subsolar magnetosheath. *Annales Geophysicae*, *31*(10), 1877-1889.
920 doi: 10.5194/angeo-31-1877-2013
- 921 Plaschke, F., Hietala, H., Archer, M., Blanco-Cano, X., Kajdič, P., Karlsson, T., ...
922 Sibeck, D. (2018, aug). Jets Downstream of Collisionless Shocks. *Space Sci.*
923 *Rev.*, *214*(5), 81. doi: 10.1007/s11214-018-0516-3
- 924 Raab, W., Branduardi-Raymont, G., Wang, C., Dai, L., Donovan, E., Enno, G.,
925 ... Zheng, J. (2016, jul). SMILE: a joint ESA/CAS mission to investigate
926 the interaction between the solar wind and Earth’s magnetosphere. In J.-
927 W. A. den Herder, T. Takahashi, & M. Bautz (Eds.), *Space telescopes and*
928 *instrumentation 2016: Ultraviolet to gamma ray* (Vol. 9905, p. 990502). doi:

- 10.1117/12.2231984
- 929 Safránková, J., Nemeček, Z., Dusík, S., Prech, L., Sibeck, D. G., & Borodkova, N. N.
930 (2002, mar). The magnetopause shape and location: a comparison of the In-
931 terball and Geotail observations with models. *Annales Geophysicae*, *20*(3),
932 301-309. doi: 10.5194/angeo-20-301-2002
- 933 Samsonov, A. A., Nemeček, Z., Safránková, J., & Jelínek, K. (2012, may). Why does
934 the subsolar magnetopause move sunward for radial interplanetary magnetic
935 field? *Journal of Geophysical Research (Space Physics)*, *117*(A5), A05221. doi:
936 10.1029/2011JA017429
- 937 Shue, J. H., & Chao, J. K. (2013, jun). The role of enhanced thermal pressure in
938 the earthward motion of the Earth's magnetopause. *Journal of Geophysical
939 Research (Space Physics)*, *118*(6), 3017-3026. doi: 10.1002/jgra.50290
- 940 Shue, J. H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., &
941 Singer, H. J. (1997, may). A new functional form to study the solar wind
942 control of the magnetopause size and shape. *Journal of Geophysical Research*,
943 *102*(A5), 9497-9512. doi: 10.1029/97JA00196
- 944 Shue, J. H., Chao, J. K., Song, P., McFadden, J. P., Suvorova, A., Angelopoulos, V.,
945 ... Plaschke, F. (2009, sep). Anomalous magnetosheath flows and distorted
946 subsolar magnetopause for radial interplanetary magnetic fields. *Geophysical
947 Research Letters*, *36*(18), L18112. doi: 10.1029/2009GL039842
- 948 Shue, J. H., Song, P., Russell, C. T., Steinberg, J. T., Chao, J. K., Zastenker, G.,
949 ... Kawano, H. (1998, aug). Magnetopause location under extreme solar
950 wind conditions. *Journal of Geophysical Research*, *103*(A8), 17691-17700. doi:
951 10.1029/98JA01103
- 952 Sibeck, D. G., Allen, R., Aryan, H., Bodewits, D., Brandt, P., Branduardi-Raymont,
953 G., ... Wing, S. (2018, jun). Imaging Plasma Density Structures in the Soft
954 X-Rays Generated by Solar Wind Charge Exchange with Neutrals. *Space Sci.
955 Rev.*, *214*(4), 79. doi: 10.1007/s11214-018-0504-7
- 956 Sibeck, D. G., Borodkova, N. L., Schwartz, S. J., Owen, C. J., Kessel, R., Kokubun,
957 S., ... Zastenker, G. N. (1999, mar). Comprehensive study of the magne-
958 tospheric response to a hot flow anomaly. *Journal of Geophysical Research*,
959 *104*(A3), 4577-4594. doi: 10.1029/1998JA900021
- 960 Sibeck, D. G., Lopez, R. E., & Roelof, E. C. (1991, apr). Solar wind control of the
961 magnetopause shape, location, and motion. *Journal of Geophysical Research*,
962 *96*(A4), 5489-5495. doi: 10.1029/90JA02464
- 963 Slavín, J. A., Szabo, A., Peredo, M., Lepping, R. P., Fitzenreiter, R. J., Ogilvie,
964 K. W., ... Steinberg, J. T. (1996, may). Near-simultaneous bow shock cross-
965 ings by WIND and IMP 8 on December 1, 1994. *Geophysical Research Letters*,
966 *23*(10), 1207-1210. doi: 10.1029/96GL01351
- 967 Smith, C. W., L'Heureux, J., Ness, N. F., Acuña, M. H., Burlaga, L. F., & Scheifele,
968 J. (1998, jul). The ACE Magnetic Fields Experiment. *Space Sci. Rev.*, *86*,
969 613-632. doi: 10.1023/A:1005092216668
- 970 Staples, F. A., Rae, I. J., Forsyth, C., Smith, A. R. A., Murphy, K. R., Raymer,
971 K. M., ... Imber, S. M. (2020, apr). Do Statistical Models Capture the
972 Dynamics of the Magnetopause During Sudden Magnetospheric Compres-
973 sions? *Journal of Geophysical Research (Space Physics)*, *125*(4), e27289. doi:
974 10.1029/2019JA027289
- 975 Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D.,
976 Ormes, J. F., & Snow, F. (1998, jul). The Advanced Composition Explorer.
977 *Space Sci. Rev.*, *86*, 1-22. doi: 10.1023/A:1005082526237
- 978 Suvorova, A. V., & Dmitriev, A. V. (2015, apr). Magnetopause inflation under ra-
979 dial IMF: Comparison of models. *Earth and Space Science*, *2*(4), 107-114. doi:
980 10.1002/2014EA000084
- 981 Suvorova, A. V., Shue, J. H., Dmitriev, A. V., Sibeck, D. G., McFadden, J. P.,
982 Hasegawa, H., ... Nemeček, Z. (2010, oct). Magnetopause expansions for
983

- 984 quasi-radial interplanetary magnetic field: THEMIS and Geotail observations.
 985 *Journal of Geophysical Research (Space Physics)*, 115(A10), A10216. doi:
 986 10.1029/2010JA015404
- 987 Tatrallyay, M., Erdos, G., Nemeth, Z., Verigin, M. I., & Vennerstrom, S. (2012,
 988 dec). Multispacecraft observations of the terrestrial bow shock and magne-
 989 topause during extreme solar wind disturbances. *Annales Geophysicae*, 30(12),
 990 1675-1692. doi: 10.5194/angeo-30-1675-2012
- 991 Taylor, M. G. G. T., Hasegawa, H., Lavraud, B., Phan, T., Escoubet, C. P., Dun-
 992 lop, M. W., . . . Wild, J. A. (2012, jun). Spatial distribution of rolled up
 993 Kelvin-Helmholtz vortices at Earth’s dayside and flank magnetopause. *Annales*
 994 *Geophysicae*, 30(6), 1025-1035. doi: 10.5194/angeo-30-1025-2012
- 995 Treumann, R. A. (2009, dec). Fundamentals of collisionless shocks for astrophysical
 996 application, 1. Non-relativistic shocks. *The Astronomy and Astrophysics Re-*
 997 *view*, 17(4), 409-535. doi: 10.1007/s00159-009-0024-2
- 998 Turner, D. L., Eriksson, S., Phan, T. D., Angelopoulos, V., Tu, W., Liu, W., . . .
 999 Glassmeier, K. H. (2011, apr). Multispacecraft observations of a foreshock-
 1000 induced magnetopause disturbance exhibiting distinct plasma flows and an
 1001 intense density compression. *Journal of Geophysical Research (Space Physics)*,
 1002 116(A4), A04230. doi: 10.1029/2010JA015668
- 1003 Turner, D. L., Omid, N., Sibeck, D. G., & Angelopoulos, V. (2013, apr). First ob-
 1004 servations of foreshock bubbles upstream of Earth’s bow shock: Characteristics
 1005 and comparisons to HFAs. *Journal of Geophysical Research (Space Physics)*,
 1006 118(4), 1552-1570. doi: 10.1002/jgra.50198
- 1007 Vu, A., Liu, T. Z., Zhang, H., & Pollock, C. (2022, feb). Statistical Study of Fore-
 1008 shock Bubbles, Hot Flow Anomalies, and Spontaneous Hot Flow Anomalies
 1009 and Their Substructures Observed by MMS. *Journal of Geophysical Research*
 1010 *(Space Physics)*, 127(2), e2021JA030029. doi: 10.1029/2021JA030029
- 1011 Walsh, A. P., Haaland, S., Forsyth, C., Keese, A. M., Kissinger, J., Li, K., . . . Tay-
 1012 lor, M. G. G. T. (2014, jul). Dawn-dusk asymmetries in the coupled solar
 1013 wind-magnetosphere-ionosphere system: a review. *Annales Geophysicae*, 32(7),
 1014 705-737. doi: 10.5194/angeo-32-705-2014
- 1015 Wang, C., & Sun, T. (2022, dec). Methods to derive the magnetopause from soft X-
 1016 ray images by the SMILE mission. *Geoscience Letters*, 9(1), 30. doi: 10.1186/
 1017 s40562-022-00240-z
- 1018 Wang, M., Lu, J. Y., Kabin, K., Yuan, H. Z., Zhou, Y., & Guan, H. Y. (2020,
 1019 may). Influence of the Interplanetary Magnetic Field Cone Angle on the
 1020 Geometry of Bow Shocks. *The Astronomical Journal*, 159(5), 227. doi:
 1021 10.3847/1538-3881/ab86a7
- 1022 Winterhalter, D., & Kivelson, M. G. (1988, sep). Observations of the Earth’s bow
 1023 shock under high Mach number/high plasma beta solar wind conditions. *Geo-*
 1024 *physical Research Letters*, 15(10), 1161-1164. doi: 10.1029/GL015i010p01161
- 1025 Zhang, H., Zong, Q., Connor, H., Delamere, P., Facsko, G., Han, D., . . . Yao,
 1026 S. (2022, aug). Dayside Transient Phenomena and Their Impact on
 1027 the Magnetosphere and Ionosphere. *Space Sci. Rev.*, 218(5), 40. doi:
 1028 10.1007/s11214-021-00865-0
- 1029 Zhang, L. Q., Wang, C., Wang, J. Y., & Lui, A. T. Y. (2019, may). Statistical
 1030 properties of the IMF clock angle in the solar wind with northward and south-
 1031 ward interplanetary magnetic field based on ACE observation from 1998 to
 1032 2009: Dependence on the temporal scale of the solar wind. *Advances in Space*
 1033 *Research*, 63(10), 3077-3087. doi: 10.1016/j.asr.2019.01.023