

# Identifying MESSENGER Magnetospheric Boundary Crossings Using a Random Forest Region Classifier

Daragh M. Hollman<sup>1,2</sup>, Caitríona M. Jackman<sup>1</sup>, Katarina Domijan<sup>2</sup>, Charles F. Bowers<sup>1,3</sup>, Simon J. Walker<sup>1</sup>, Matthew J. Rutala<sup>1</sup>, Alexandra R. Fogg<sup>1</sup>

<sup>1</sup>Astronomy & Astrophysics Section, School of Cosmic Physics, Dublin Institute for Advanced Studies, DIAS Dunsink Observatory, Dublin D15 XR2R, Ireland.

<sup>2</sup>Department of Mathematics and Statistics, Maynooth University, Maynooth, Co. Kildare, Ireland.

<sup>3</sup>Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA

## Key Points:

- A random forest was used to classify magnetospheric regions based on MESSENGER magnetometer and ephemeris data
- The model had a high accuracy on unseen data, with a score of  $0.9868 \pm 0.0002$
- 20,143 bow shock crossings and 11,180 magnetopause crossings were determined for the MESSENGER mission

---

Corresponding author: Daragh M. Hollman, [daragh.hollman@dias.ie](mailto:daragh.hollman@dias.ie)

## Abstract

We present a new list of bow shock and magnetopause crossings based on automated region classification for the MESSENGER (Mercury Surface, Space Environment, Geochemistry and Ranging) mission. We fit a random forest model to magnetometer and ephemeris data to classify the solar wind, magnetosheath, and magnetosphere regions surrounding Mercury. The random forest was highly accurate when predicting the testing dataset, with an accuracy of  $0.9868 \pm 0.0002$ . We apply this model to the orbital phase (March 2011 to April 2015) of the MESSENGER mission, and determine crossings automatically where changes in region classification occur. We apply some post-processing steps to this list of crossings and determine 20,143 bow shock crossings and 11,180 magnetopause crossings for the MESSENGER mission. We find good visual comparison with the spatial distribution of these new crossings and existing datasets of boundary crossing intervals.

## Plain Language Summary

Surrounding the planet Mercury is a large and dynamic magnetic bubble (known as a magnetosphere), which changes in size and shape due to interactions between Mercury’s magnetic field and the solar wind (a stream of charged particles from the Sun). Due to changing external conditions, the location of the boundaries of this bubble can vary rapidly and dramatically. NASA’s MESSENGER spacecraft spent 4 years orbiting Mercury between 2011 and 2015, and sampled from different regions of space with respect to this bubble which have distinct physical differences. We use a simple, yet powerful, machine learning technique to build a model to identify these regions, to ultimately determine a list of times when the spacecraft crossed the boundaries between these regions.

## 1 Introduction

Mercury has the weakest intrinsic magnetic field of all the planets in the Solar System. This, coupled with its close proximity to the Sun, produces extreme and highly variable magnetospheric interactions (Milillo et al., 2020). Among other effects, variations in the upstream solar wind conditions cause highly dynamic fluctuations in the morphology of its magnetospheric boundaries, the bow shock and the magnetopause. The bow shock marks the discontinuity where the solar wind is rapidly decelerated, compressed and deflected as Mercury’s magnetosphere acts as an obstacle in its path (Milillo et al., 2020). The magnetopause lies at the equilibrium between the internal magnetic pressure of the magnetosphere and the external solar wind pressure (Shue & Song, 2002). Relative to other magnetised planets in the solar system, Mercury has the smallest magnetosphere, with an average magnetopause standoff distance of only 1.45 Mercury radii ( $R_M = 2,438$  km) (Winslow et al., 2013),  $\approx 7$  times closer than Earth’s (Shue et al., 1997; Shue & Song, 2002) and  $\approx 48$  times closer than Jupiter’s (Joy et al., 2002; Rutala et al., 2025) relative to their radii. Measuring the locations of these boundaries is important for understanding the timescales and magnitudes of magnetospheric compressions (Slavin et al., 2019), separating plasma populations to enable large scale studies (Bowers et al., 2024; Karlsson et al., 2024), and informing models (Winslow et al., 2013; Zhong et al., 2015; Philpott et al., 2020; He et al., 2022).

Three spacecraft have previously sampled Mercury’s boundaries. Mariner 10 was the first to encounter Mercury’s bow shock and magnetopause across three flybys between 1974 and 1975 (Ness et al., 1974). MESSENGER (Mercury Surface, Space Environment, Geochemistry and Ranging) (Leary et al., 2007), launched in 2004, made three Mercury flybys (two in 2008 and one in 2009 (Slavin et al., 2008; Baker et al., 2011)) before entering continual orbit in 2011. This was followed by almost 4 years of exploration, with orbits taking the spacecraft within the vicinity of the boundaries thousands of times. Bepi-Colombo is a two spacecraft mission on the way to Mercury which has undergone 6 fly-

bys to date (though much of that data has yet to be made available), its expected arrival is the end of 2026 (Benkhoff et al., 2021).

Crossings of the bow shock and magnetopause are visually identifiable in data from spacecraft magnetometers as well as other instruments such as plasma, energetic particle and wave sensors. Bow shock crossings are typically identified by large changes in the magnitude or variability of the magnetic field (Winslow et al., 2013; Jackman et al., 2019; Kruparova et al., 2019). Bow shock crossings can be split into two broad categories: quasi-perpendicular and quasi-parallel, based on the angle between interplanetary magnetic field (IMF) vector and the bow shock surface normal. Quasi-perpendicular crossings are typically identified through rapid changes in the magnitude of the magnetic field. They also often include a shock foot and overshoot (Dimmock et al., 2019). Quasi-parallel crossings are instead typically identified through changes in variability (Nevskii et al., 2023). Back-streaming particles create turbulent conditions around the shocked front which result in increased variability in the magnetic field (Jarvinen et al., 2020; Milillo et al., 2020). For these reasons, quasi-parallel shocks are often more difficult to visually determine.

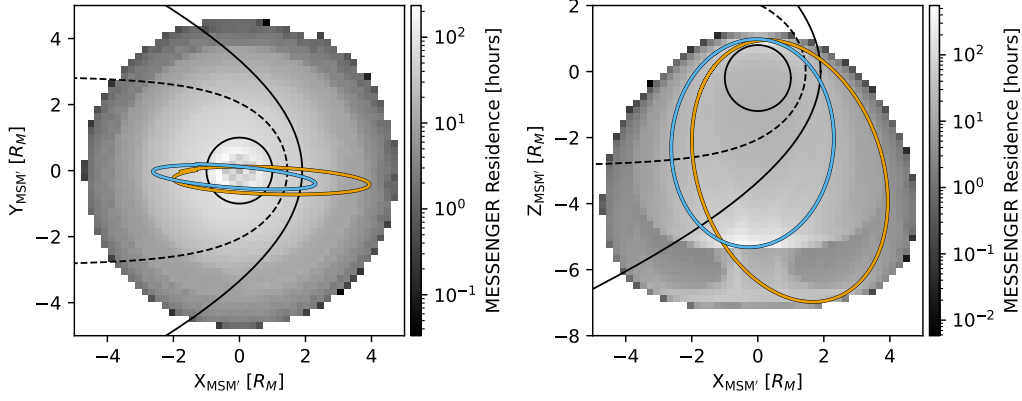
Magnetopause crossings are typically identified through changes in magnetic field variability or through rotations of the magnetic field. The magnetosheath is often turbulent and compressed, producing a highly variable magnetic field in magnitude while the magnetosphere is much less variable (Haaland et al., 2014). Furthermore, the magnetic field orientation within the magnetosheath depends on the IMF orientation, while the magnetic field orientation within the magnetosphere is relatively steady as the direction of the field is determined by the planetary dipole. Magnetopause crossings are hence most easily identified when the magnetosheath magnetic field vector and magnetosphere magnetic field vector are separated by a large angle, or when there is a large change in variability of the magnetic field (Winslow et al., 2013). Due to the extreme variability of Mercury’s magnetospheric boundaries (Milillo et al., 2020), as a spacecraft passes across a boundary, we often observe multiple crossings as the boundary moves back and forth over the spacecraft. Because of this, crossings at Mercury have previously been considered instead as intervals of time in which multiple crossings occur, rather than a series of individually identified boundary crossings. Intervals of crossings of Mercury’s bow shock and magnetopause have been visually identified with the above methods by Winslow et al. (2013); Philpott et al. (2020) and Sun et al. (2020); Sun (2023) for the duration of the mission. Further detail on this labelling can be found in section 2.1.

Automated boundary detection methods, including machine learning techniques, were explored previously at Mercury and other planets, discussed further in section 3.1. Automated crossing detection was previously achieved through both the direct detection of changes within data [e.g. Daigavane et al. (2022); Nevskii et al. (2023)], but also through indirect methods, where crossings were determined based on changes in region classification [e.g. Bakrania et al. (2020); Olshevsky et al. (2021); Wang et al. (2025)]. Yet, no automated method currently exists to determine individual crossings of Mercury’s boundaries. The magnetospheric region in which a spacecraft is at a given time is often trivial to determine visually for single event case studies by those with domain knowledge. However, over longer timescales automated identification offers a quicker approach. Furthermore, visual-based selection may also contain inherent biases potentially producing an inconsistent result. In this work, we present a region classifier model based on numerical parameters of the data, resulting in an unambiguous and reproducible identification of boundary crossings. This method could easily be rolled out onto future unseen data without significant additional effort.

In this manuscript, section 2 outlines the MESSENGER mission, its available data and their applicability to boundary identification. Section 3 describes the methods used to model region classification, along with the application of that model. Section 4 details the results of model testing and the resulting list of boundary crossings along with example applications of the model. Section 5 further explores the new crossing list and ex-

amines spatial distributions and trends in heliocentric distance. This section also outlines avenues of future research.

## 2 Data

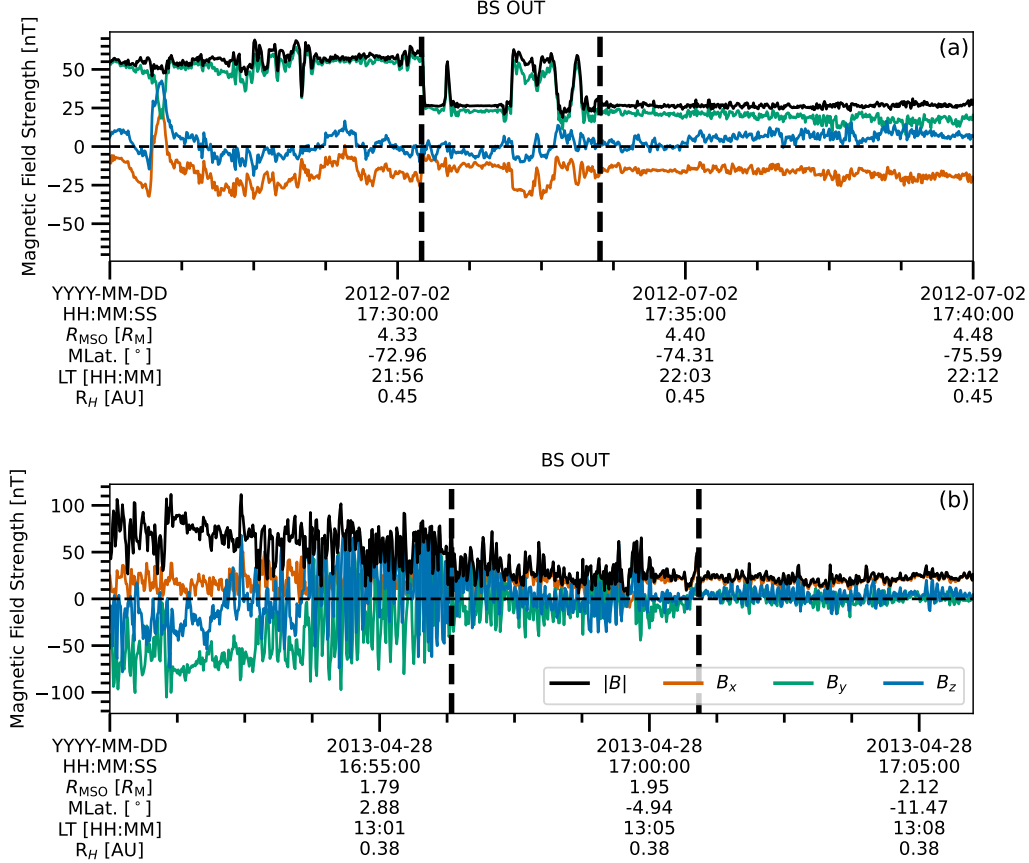


**Figure 1.** Two example orbits of the MESSENGER mission plotted in the  $XY$  and  $XZ$  planes of the aberrated Mercury-Solar-Magnetic (MSM') coordinate system. The background histogram denotes the residence time of MESSENGER as projected to a given spatial bin in each plane. A pale blue background denotes regions where MESSENGER did not sample. A circle is drawn to represent the size and location of Mercury. The average boundary shape for the bow shock (solid curve) and magnetopause (dashed curve) as determined by Winslow et al. (2013) are included for context. The orange line plots the orbit of MESSENGER starting 25th December 2011 02:00, the blue line plots the orbit of MESSENGER starting 5th June 2013 01:00.

MESSENGER orbited Mercury from March 2011 to April 2015, after which it was de-orbited (Milillo et al., 2020). Its instrumentation included a magnetometer (MAG) which sampled the magnetic field at a cadence of up to 20 Hz (Leary et al., 2007; Anderson et al., 2007). MESSENGER initially orbited Mercury with a period of 12 hours, however, after April 2012 it was shortened to 8 hours. Figure 1 shows examples of MESSENGER's orbit in the aberrated Mercury Solar Magnetic (MSM') coordinate system, where the origin is centred at Mercury's magnetic dipole (offset  $\approx 479$  km northward of the planet's centre (Anderson et al., 2011; Johnson et al., 2012)),  $X_{\text{MSM}'}$  points into the flow of solar wind (aberrated from sunwards due to Mercury's orbital velocity - which is significant with respect to the solar wind velocity).  $Z_{\text{MSM}'}$  points northward along the rotation axis, and  $Y_{\text{MSM}'}$  completes the right-handed orthogonal set. In our work, we follow methods employed in previous works and calculate this aberration angle daily, assuming an average solar wind velocity of  $400 \text{ km s}^{-1}$  (Bowers et al., 2024). The aberration angle varies between  $5.6^\circ$  at aphelion and  $8.4^\circ$  at perihelion. MESSENGER's periaapsis stays close to the surface north of the planet, while its apoapsis extends approximately  $7R_{\text{M}}$  south of the planet before shortening to  $5R_{\text{M}}$ . This highly elliptical orbit enables the spacecraft to sample from the magnetosphere, magnetosheath, and solar wind, all in one orbit. In Figure 1, the orange line shows an orbit with a period of 12 hours. The orbit is aligned to the  $XZ$  plane, and tilted sunward (positive  $X$ ), meaning it spends a large portion of the orbit within the solar wind. In contrast, the blue line shows an or-

bit tilted anti-sunward (towards negative  $X$ ), and spends a smaller portion of the orbit within the solar wind. This is furthered by this orbit having a lower apoapsis, and hence a shorter period. The background of Figure 1 shows a 2D histogram of MESSENGER's residence as projected to each plane. The edges of MESSENGER's extent are darker and therefore less sampled, while bins closer to the planet were sampled more frequently.

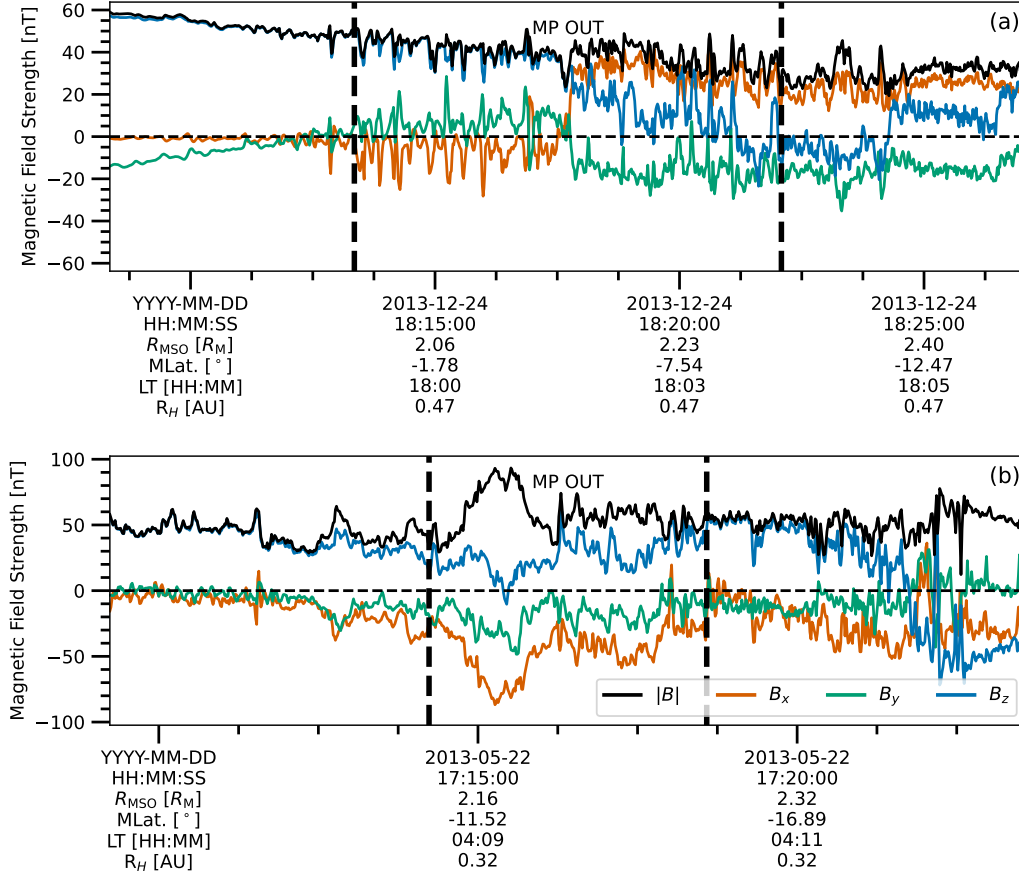
## 2.1 Instrumentation



**Figure 2.** Contrasting examples of magnetometer data from MESSENGER for two out-bound bow shock crossing intervals as labelled by Philpott et al. (2020). Vertical dashed lines mark these intervals, along with the label BS OUT. Panels (a) and (b) show examples of quasi-perpendicular and quasi-parallel cases respectively.

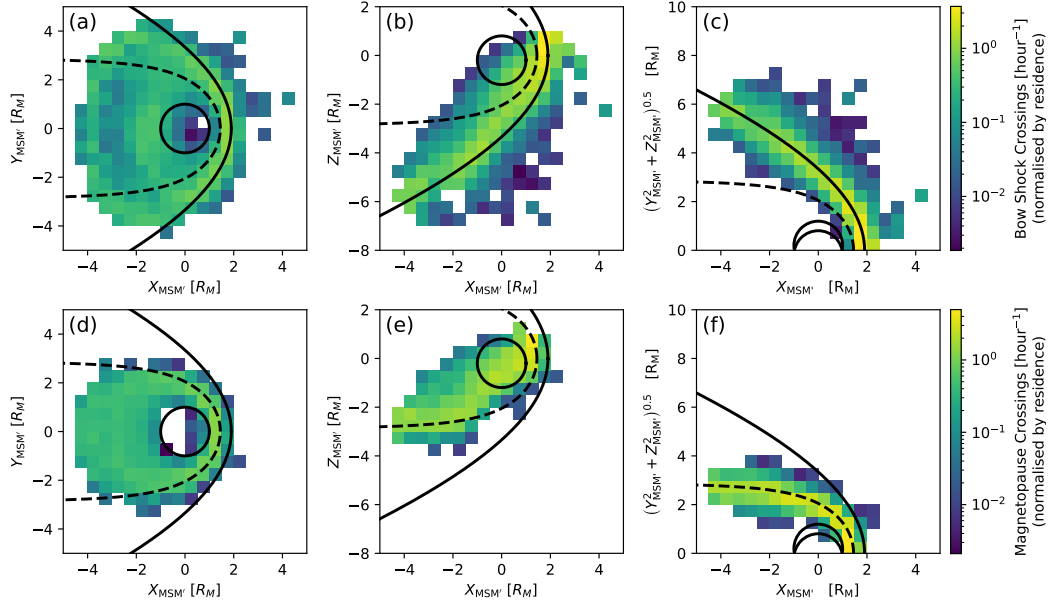
We utilise the MESSENGER MAG Calibrated Data Bundle (Anderson et al., 2007; Korth & Anderson, 2016) during the orbit phase of the MESSENGER mission. This dataset consists of the time, vector magnetic field, and spacecraft position for the duration of the mission. Measurements were recorded with a maximum sample rate of 20 Hz (down-sampled averages were also made available at 1, 5, 10, and 60 second intervals).

Onboard MESSENGER was also the Fast Imaging Plasma Spectrometer (FIPS) instrument which measured the energy, angular distributions, and compositional distributions of charged particles between  $46 \text{ eV } q^{-1}$  and  $13 \text{ keV } q^{-1}$  (Andrews et al., 2007). Generally, information about the plasma in a magnetospheric environment is extremely use-



**Figure 3.** Contrasting examples of magnetometer data from MESSENGER for two outbound magnetopause crossing intervals as labelled by Philpott et al. (2020). Vertical dashed lines mark these intervals, along with the label MP OUT. Panel (a) shows an example containing a clear and singular rotation of the field orientation, while panel (b) shows a less clear example.

ful in categorising spacecraft location with respect to that environment. Many previous efforts to classify spacecraft location with respect to magnetospheric region at Mercury and other planets in the solar system have utilised plasma spectra and derived products in their methods: Bakrania et al. (2020); James et al. (2020); Daigavane et al. (2022); Grison et al. (2025). However, FIPS data have several constraints which make them less useful for this work in particular. The nominal measurement cadence of FIPS of 10 seconds (burst mode, near planet), and 65 seconds (normal mode) (Andrews et al., 2007), limits the scale at which classifications based on these data can be made. Through visual inspection of MAG data we find individual crossings can occur just seconds apart. To classify changes in region on this short of a timescale, we require a measurement cadence on a similar or shorter timescale. FIPS can also run in a faster mode, with a 2 second measurement cadence (Andrews et al., 2007), however this was not consistently in use. Secondly, part of FIPS’s field of view was occluded by MESSENGER’s sunshade (Andrews et al., 2007), restricting the distribution of plasma measured as sources from the sunward direction were obstructed, making the measurements less reliable. For these two reasons, we choose to not use FIPS plasma spectra or derived products in this work.



**Figure 4.** The spatial spread of bow shock (a-c) and magnetopause (d-f) crossing intervals from Philpott et al. (2020) are shown as projected to the  $XY$  and  $XZ$  planes of the aberrated Mercury-Solar-Magnetospheric (MSM') coordinate system. The density of crossing intervals is normalised by the time spent in a given spatial bin, resulting in a crossing rate in units of crossings per hour. A circle is drawn to represent the size and location of Mercury. The average boundary shape for the bow shock (solid curve) and magnetopause (dashed curve) by Winslow et al. (2013) (based on these data) are included for context.

## 2.2 Previous Labelling of Boundary Crossing Intervals

We based this study on previous labelling of the MESSENGER MAG dataset by Philpott et al. (2020) which further expanded labelling by Winslow et al. (2013) to the entire mission. They identified intervals of time grouping sequences of crossings of the same magnetospheric boundary together. These intervals were labelled by boundary (BS: bow shock, MP: magnetopause) and direction (IN: inbound - towards the magnetosphere, OUT: outbound, away from the magnetosphere), e.g. BS\_IN, MP\_OUT. These intervals were visually determined based on manual inspection of the MAG data alone, and marked sections of time which span the first and last crossing in a given pass through the boundary. Though multiple lists of boundary crossing intervals are available [e.g. Philpott et al. (2020); Sun (2023)], for this study we focus on the list by Philpott et al. (2020), which was validated by members of the MESSENGER MAG science team. This list labels 16,266 intervals: 8,108 for the bow shock, 8,131 for the magnetopause, and 27 data gaps.

Examples of these intervals within the context of the data are shown in Figures 2 and 3 for bow shock and magnetopause intervals respectively. Figure 2a shows a typical case for a quasi-perpendicular crossing of Mercury's bow shock. We observe two distinct regions in the MAG data: high amplitude magnetosheath before the crossing interval and low magnitude solar wind after the crossing interval. Figure 2b instead shows a typical case for a quasi-parallel crossing of Mercury's bow shock, characterised by significantly



higher variability than the above case. In Figure 3a, the magnetopause crossing itself was visually distinct as there was a rotation in the field resulting in a large change in  $B_x$  across the boundary. Figure 3b instead shows an example where the field vector in the magnetosphere and magnetosheath are closely aligned. There was very little rotation across the boundary, and the exact point of crossing is difficult to discern. We show the spatial distribution of these crossing intervals in Figure 4 as projected to the  $XY_{\text{MSM}'}$  (a, d),  $XZ_{\text{MSM}'}$  (b, e), and cylindrical (c, f) planes. Panels (c) and (f) show the spatial spread matching well to the average boundary location by Winslow et al. (2013). This was to be expected, as this boundary shape was determined using a subset of these intervals. Due to the geometric constraints of Mercury’s orbit (as described in Figure 1) crossings were predominantly observed in negative  $Z_{\text{MSM}'}$  and never northward of  $Z_{\text{MSM}'} = 1.5$ . Crossings were observed at all local times - with a skew towards noon.

While useful for visualising the system, it is somewhat arbitrary to describe an interval of time with a single position. As shown in Figure 2a, bow shock and magnetopause crossing intervals often contain more than one individual crossing. By treating multiple crossings as one entity, we lose relevant statistical information about the boundaries, such as the exact locations/times and the temporal density of each individual crossing (and hence the dynamic temporal variability of Mercury’s magnetosphere). The aim of this work was to build on this list of boundary crossing intervals by utilising them to first isolate regions of pure solar wind, magnetosheath, and magnetosphere within MESSENGER MAG and ephemeris data. Based on this region classification, we constructed a dataset upon which to train a machine learning model, to automatically identify every individual boundary crossing within each crossing interval.

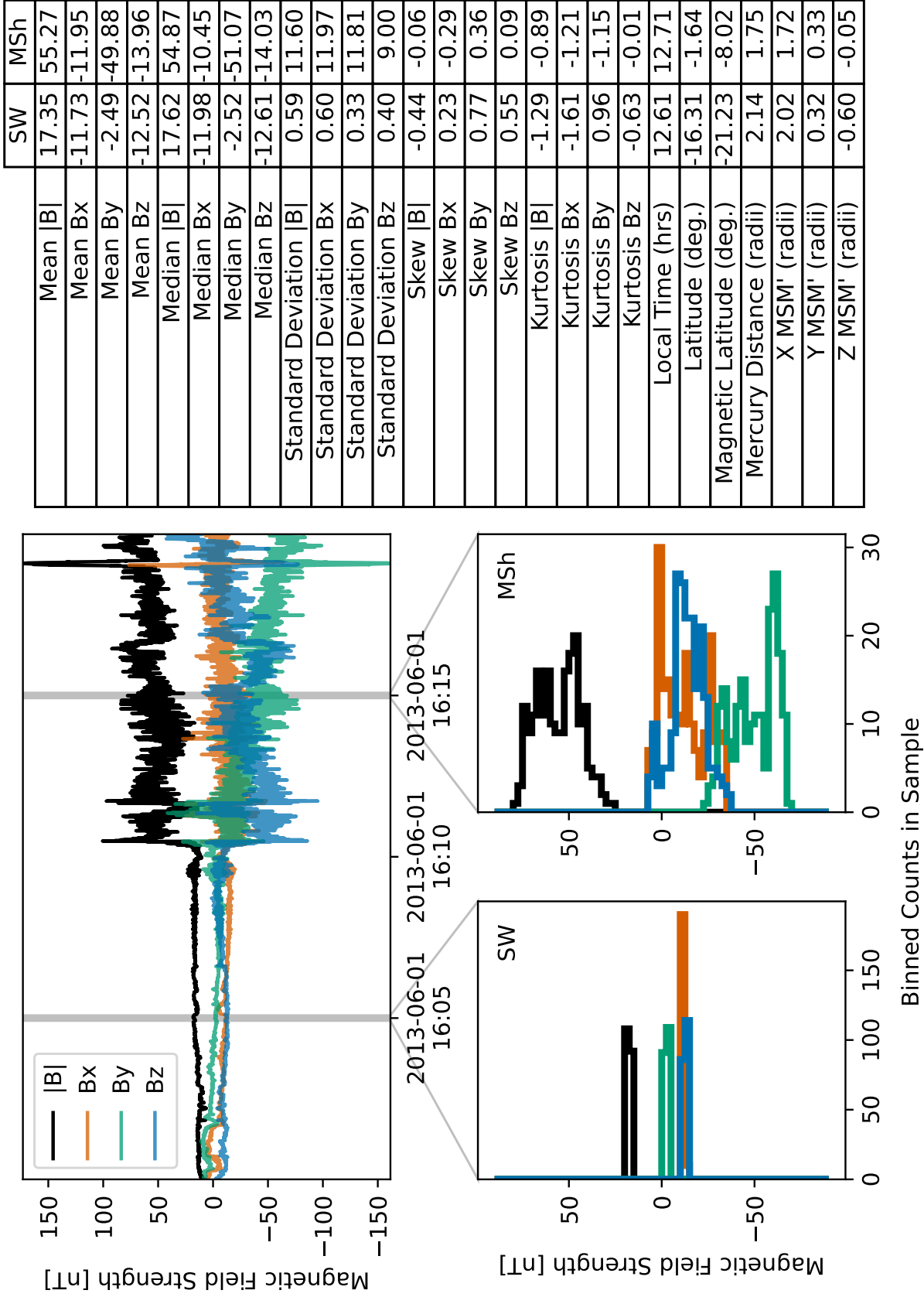
### 2.3 Data Reduction

We opted to use a supervised machine learning technique to first classify magnetospheric region (i.e. solar wind, magnetosheath, magnetosphere) at Mercury, and only then determine spacecraft crossings of the boundaries based on changes in that region classification. To create a labelled dataset of magnetospheric regions, we treated each MESSENGER observation in time as an independent random sample, but incorporated some time-dependency information by deriving features from a 10 second window around each second in time. The length of this window was chosen as a compromise between minimising temporal variation within the window, and maximising the length to ensure the data were representative of the overall region distribution. We then constructed features from the distribution of data within that 10 second window. The MESSENGER MAG dataset records 6 covariates from each time step: the vector magnetic field,  $(B_x, B_y, B_z)$ , and the vector position  $(X_{\text{MSO}}, Y_{\text{MSO}}, Z_{\text{MSO}})$ , both in MSO coordinates - which we converted to MSM’ by rotating by the aberration angle, and offsetting the position by the dipole offset described at the start Section 2. We derived statistical features from the distribution of data within each 10-second window of data. For ephemeris features (any feature derived from the position of the spacecraft), we only recorded the position at the mid-point of the window as a feature for that window. In total, we derived 28 features from these data: the mean, median, standard deviation, skew, and kurtosis of the magnitude and components of the magnetic field, the magnitude and components of MESSENGER’s position in MSM’ coordinates and further metrics of MESSENGER’s position relative to Mercury: latitude, magnetic latitude, local time, and heliocentric distance. Figure 5 shows an example of two such reduced samples, their distribution of magnetic field data, and the derived features.

We constructed a labelled dataset of 10-second windows with the following method:

1. Iterating through the list of crossing intervals by Philpott et al. (2020), we loaded data within 10 minutes before and after each interval. These ‘interval extensions’ delineated the regions within which we selected training data. Figure 6 shows an





**Figure 5.** An example of how features are drawn from a given sample of data. Panel (a) shows an example of MESSENGER MAG data before and after an in-bound bow shock crossing. Two 10 second long samples are arbitrarily selected from before and after this crossing - shaded in grey. Below, in panels (b, solar wind) and (c, magnetosheath) we show the distributions of magnetic field data in each of these samples. Panel (d) shows the numerical value of each feature derived from these samples.

example of these extensions surrounding an outbound bow shock crossing interval. In panel (a), the orange and yellow regions denote the two interval extensions corresponding to magnetosheath and solar wind training data respectively. Panels (b) and (c) highlight how the distributions of magnetic field data are distinct between the two regions. We limited these selection regions to 10 minutes to avoid large overlap between successive interval extensions.

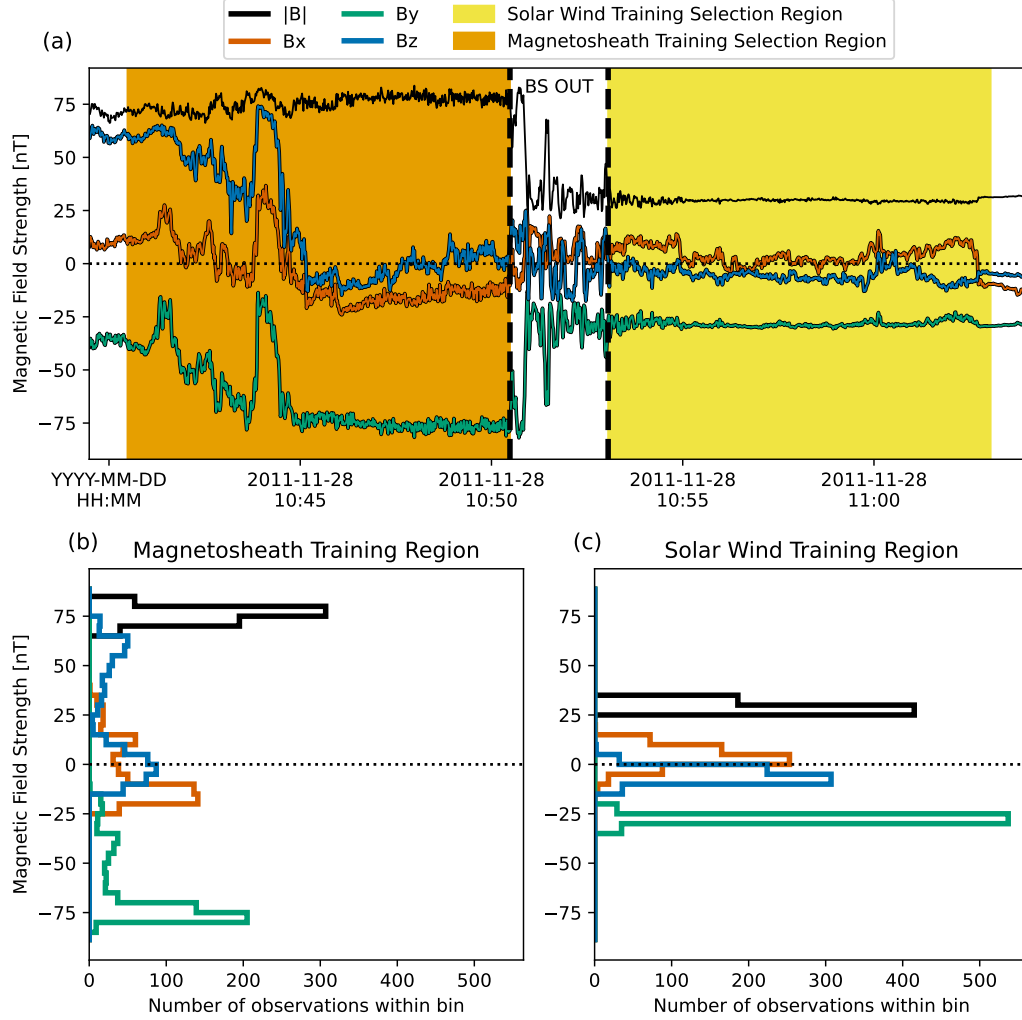
2. We then checked for overlap between interval extensions and the neighbouring intervals. For some crossing intervals, primarily those on the dayside, the length of time between successive bow shock and magnetopause crossing intervals can be less than 10 minutes. As a result, the 10 minute interval extension following one crossing interval could overlap with the start of the next, inadvertently including and mislabelling data from an interval in the training data. These cases occur for approximately 13% of intervals. To prevent this, we limited the length of the extensions to avoid any overlap with crossing intervals such that all extensions were  $\leq 10$  minutes.
3. Within each interval extension, we randomly selected 10 of these 10-second windows to add to the training dataset. We selected only 10 samples from each extension to reduce computational time, however, selecting multiple samples from each interval was critical to avoid potential anomalies in the data, such as the brief rotation of the magnetic field on the 28th November 2011 10:44 in Figure 6.
4. For each sample, we calculated the features as described above and labelled the sample's region based on the label of its respective crossing interval. E.g. if the sample was taken after a BS\_IN interval, it was labelled as magnetosheath.

As a result of this method of construction, this dataset contained a class imbalance. There were only two interval types which yielded samples of solar wind (BS\_IN and BS\_OUT) and magnetosphere (MP\_IN and MP\_OUT), however, all four interval types yielded magnetosheath samples. This resulted in twice as many magnetosheath samples as solar wind or magnetosphere. In addition to this, we identified extreme (non-physical) values from the training dataset by stipulating that mean and standard deviation of  $|B|$  were less than 5,000 nT. We chose this value, based on visual inspection of the training dataset, to exclude obvious outliers. Extreme samples were first removed, before randomly re-sampling all classes until all classes were the same size as the minority class. This left us with a dataset containing 243,234 samples, 81,078 in each class. The distribution of features in this dataset can be found in supplementary information S1. We randomly split these data into training and testing datasets, with a ratio of 9:1. When later applying the model to the unseen data (i.e. within the crossing intervals), the same data reduction technique is applied.

## 3 Methodology

### 3.1 Machine Learning Approach

Multiple approaches to the automatic identification of individual crossings have been explored in previous works for many planets in the Solar System. Individual crossings have previously been identified directly through the use of change point detection (Daigavane et al., 2022; Nevskii et al., 2023), or based on the thresholding of data (Case & Wild, 2013). Indirect methods are also viable, where magnetospheric regions (e.g. solar wind, magnetosheath, magnetosphere) are first classified, and individual crossings are then placed where these regions change. Region classifiers have also been explored previously, including decision tree methods (Nguyen et al., 2022; Wang et al., 2025), unsupervised methods (Bakrania et al., 2020; Toy-Edens et al., 2024; Edmond et al., 2024), and neural networks (Argall et al., 2020; Breuillard et al., 2020; Olshevsky et al., 2021; Kit Cheng & Achilleos, 2022; Julka et al., 2023). Each of these studies was tailored to the nature of the datasets used and spacecraft available.



**Figure 6.** Example MAG data from MESSENGER during and surrounding an outbound bow shock crossing. Panel (a) shows the three-component magnetic field vector in MSM' coordinates, along with the magnitude (black). Shaded regions mark sections of 'pure' magnetosheath (left, orange) and solar wind (right, yellow) within which we sampled training data. Unshaded regions are not considered for training purposes. Panels (b) and (c) show the distributions of these data (binned every 5 nT) within each shaded region for the magnetosheath and solar wind respectively.

**Table 1.** Evaluation metrics for 6 different feature-sets. The mean and standard deviation training accuracy and out-of-bag score are shown for 10 model runs, each with different but fixed random seed.

Feature Set	Training Accuracy		OOB Score	
	Mean	SD	Mean	SD
All Features	$\approx 1$	$1 \times 10^{-6}$	0.9843	0.0002
Reduced Features	$\approx 1$	$1 \times 10^{-6}$	0.9866	0.0001
No Ephemeris	$\approx 1$	$3 \times 10^{-6}$	0.9313	0.0003
Only Mean	0.999982	$7 \times 10^{-6}$	0.8306	0.0003
Only Median	0.999984	$7 \times 10^{-6}$	0.826	0.0003
Only Standard Deviation	0.999986	$6 \times 10^{-6}$	0.6721	0.0004

We choose to use a random forest model (Breiman, 2001) in our work for their computational efficiency, flexibility of fit, robustness to outliers and noise, and insensitivity to overfitting (Breiman, 2001; Belgiu & Drăguț, 2016). Specifically, we use the RANDOMFORESTCLASSIFIER implementation from Scikit-learn (version 1.6.1), a Python machine learning toolkit (Pedregosa et al., 2011). Random forests are an unconventional tool for working with time series, as the time-dependent structure is not modelled. Instead, we treated each observation in time as an independent random sample (as described in Section 2.3). We evaluated 5 different feature-sets to use in modelling. We trained 10 random forest models with default hyperparameters on all of the training data and recorded training accuracy and out-of-bag (OOB) score (each with different but fixed random seed). Mean and standard deviation of the training accuracy, and the OOB Score are included in Table 1. ‘All Features’ is a feature-set containing all of the features described above. ‘Reduced Features’ removes skew and kurtosis features. ‘No Ephemeris’ removes all positional features of the spacecraft, such as X, Y, and Z MSM’, local time, latitude, etc. ‘Only Mean’, ‘Only Median’, and ‘Only Standard Deviation’ contain only 4 features each, for the mean, median, and standard deviation of the field magnitude and components respectively.

We used the Python package OPTUNA to vary a selection of model hyperparameters during model training (Akiba et al., 2019). Using the optimal feature-set from Section 3.1, we trained 100 random forest models on the entire reduced dataset, allowing hyperparameters to vary while optimising for minimum OOB error. Each model had a different random seed (each fixed to ensure reproducibility). We recorded the hyperparameters which were allowed to vary, and their optimum values in Table 2. To measure the accuracy and variance of this final model architecture, we trained 30 further models (again with different but fixed random seed) with the selected features and these optimal hyperparameters, evaluating their performance on the testing dataset. We saved the model with the largest accuracy score for further application.

### 3.2 Model Application

Due to the regularity of MESSENGER’s orbit and the pre-labelled boundary intervals, and to save on computational time we restricted our application to the crossing intervals themselves along with the surrounding 10 minutes on either side. However, we have included an example classification to a full orbit of MESSENGER in supplementary information S2. While we treated the crossing intervals by Philpott et al. (2020) as ground truth for the purposes of training, we do not want to rule out the possibility of mislabelling in the intervals dataset. In application of the model, we grouped crossing intervals into pairs where applicable to ensure a consistent classification between in-

**Table 2.** A subset of hyperparameters used by the random forest model which were optimised during training. Further detail, and a full list of parameters, can be found in the Scikit-Learn documentation for `RANDOMFORESTCLASSIFIER`.

Hyperparameter	Value	Sampling Bounds	Description
<code>n_estimators</code>	478	50, 500	The number of trees in the forest
<code>max_depth</code>	58	10, 100	The maximum depth of the trees.
<code>max_features</code>	$\log_2(N_{\text{features}})$	Choice of $\log_2 N$ , $\sqrt{N}$ , or No Limit	The number of features to consider for a given decision tree
<code>min_samples_split</code>	2	2, 10	The minimum number of samples required to split a node

tervals. This consisted of pairing consecutive bow shock and magnetopause intervals (i.e. BS\_IN followed by MP\_IN, or MP\_OUT followed by BS\_OUT), the average time between intervals being  $\approx 42$  minutes. Not all boundary crossing intervals exist in pairs however, due to the existence of data gaps and missing intervals in the dataset.

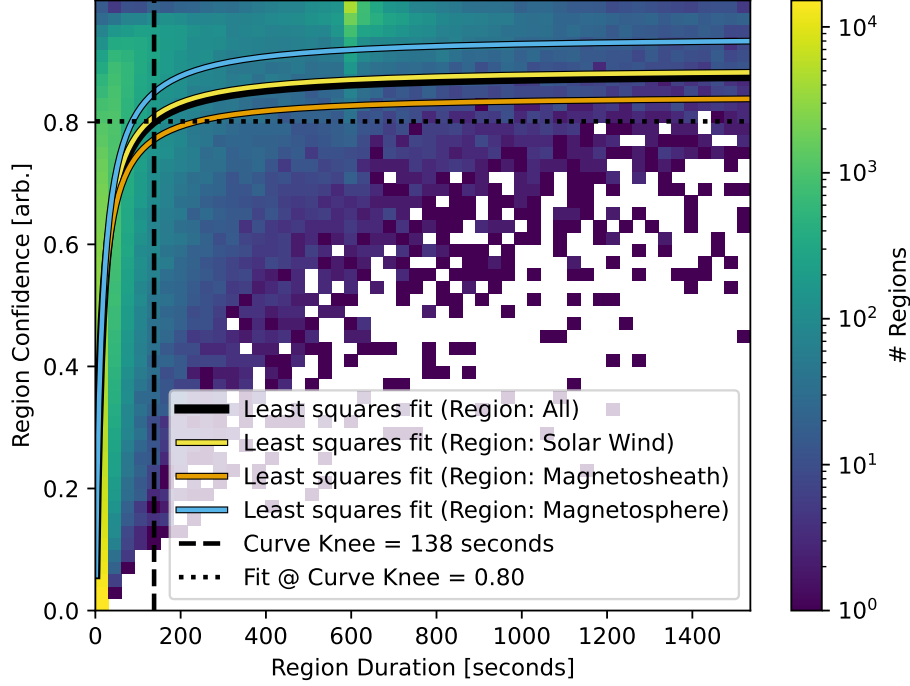
Iterating through these groups of crossing intervals (either paired or unpaired), we took the following steps to apply the model:

1. First, the full cadence MESSENGER MAG time series was loaded from 10 minutes before the first crossing interval of the pair, to 10 minutes after the second crossing interval. When unpaired, data was loaded from 10 minutes before, to 10 minutes after the single interval.
2. For each second of data loaded, we calculated features from the surrounding 10 seconds of data as described in Section 2.3.
3. These features were passed to the model to determine both a class prediction, along with class probability for the three classes (with the class probability of a single tree being the fraction of samples of the same class in a leaf, and the class probability from the model being the average probability of all trees).

After iterating through each group of crossing intervals and applying the model with the above steps, the classification times, and class probabilities were saved together - referred herein as the ‘raw model output’.

### 3.3 Determining Boundary Crossings

To place crossings based on the three class probabilities, we first split the raw model output back up into the groups of paired crossings. For each of these groups, we determined the classified region by the largest class probability from the model. We then placed crossings where this classification changed, halfway between the two surrounding classifications. When classifications were uncertain - meaning there was no distinct classification, i.e. the ratio of the second highest class probability over the highest class probability was close to 1 - variation in each consecutive classification resulted in many crossings in short succession. This was unwanted behaviour as we lacked confidence in the region classifications which dictate those crossings, and hence could not be certain that these were physical crossings. To handle this, we first combine successive region classification of the same region into a contiguous region describing a longer time. For each of these contiguous regions, we determine an associated ‘region confidence’ as the complement of the mean of the aforementioned probability ratio within the region. For the first and last region of any application, we assume a region confidence of 1. To mark uncertain regions as unknown, we introduce the following criterion. Figure 7 shows a 2D



**Figure 7.** A 2D histogram of region confidence and duration for all regions, determined based on the raw model outputs. A solid black line shows a least squares regression with the functional form  $f(x) = \frac{Ax}{B+x}$  ( $A = 0.883 \pm 0.001$ ,  $B = 13.99 \pm 0.04$ ). Horizontal and vertical black dashed lines intersect the curve at the knee point - determined with the *kneedle* method (Satopaa et al., 2011). Yellow, orange, and blue lines show this function as fit to only regions classified as solar wind, magnetosheath, or magnetosphere respectively (with parameters:  $A_{\text{SW}} = 0.89 \pm 0.001$ ,  $B_{\text{SW}} = 13.7 \pm 0.07$ ,  $A_{\text{MSh}} = 0.846 \pm 0.001$ ,  $B_{\text{MSh}} = 13.13 \pm 0.05$ ,  $A_{\text{MSp}} = 0.942 \pm 0.002$ ,  $B_{\text{MSp}} = 15.21 \pm 0.09$ ).

histogram of region confidence against region duration for all regions classified (with outliers in duration,  $> 3\sigma \approx 1.7\%$ , removed.). We applied a non-linear least squares regression from the SCIPY Python package (Virtanen et al., 2020) of functional form  $f(x) = \frac{Ax}{B+x}$  to parametrise a curve describing the average relationship between region confidence and duration. This function was chosen as its form matched our visual observations of the data: a sharp decrease in average confidence at short durations, and converging to a constant average confidence at large durations. Based on the knee-point of this curve (calculated using the *kneedle* method (Satopaa et al., 2011) - determined by the maximum vertical distance of the curve from a straight line with the average slope of the curve), we imposed that any region shorter than the curve duration at the knee (138 seconds) would be re-classified as unknown, unless it was more confident than the curve confidence at the knee (0.8), i.e. removing the lower left quadrant in Figure 7. We choose this metric (rather than rejecting all regions under the curve) as we want to keep long-duration regions - even those with lower than average confidence. While the average confidence across the region is low, the continuous classification is indicative of a correct region classification. We can think of region duration as another metric of confidence. We additionally applied curves to this regions dataset, subset by region type, however, we use the fit to all regions for determining unknown regions. An increase in columns near 600 seconds (10 minutes) corresponded to applications in which the first crossing was placed at the start of the crossing interval, yielding a region which spans from the start of the application to the start of the crossing interval, which by definition is 10 minutes. This could similarly occur from the end of the crossing interval to the end of the application.

The result was a list of contiguous region classifications of the following 4 classes: solar wind, magnetosheath, magnetosphere, and unknown. We hence defined 12 region transitions under 3 broad categories:

1. Direct crossings: Bow shock inbound and outbound (BS\_IN, BS\_OUT), and magnetopause inbound and outbound (MP\_IN, MP\_OUT).
2. Unknown crossing types for unknown transitions between the three regions (SW  $\rightarrow$  UKN, UKN  $\rightarrow$  SW, MSh  $\rightarrow$  UKN, UKN  $\rightarrow$  MSh, MSp  $\rightarrow$  UKN, UKN  $\rightarrow$  MSp).
3. Unphysical crossing types (SW  $\rightarrow$  MSp, or MSp  $\rightarrow$  SW) where the intermediate region is skipped. While this violates the nominal magnetospheric geometry, crossings classed as unphysical are not necessarily incorrect in all cases, as magnetospheric geometry can change under extreme conditions [e.g. Bowers et al. (2025)]. Table 3 records the number of crossings grouped by these types.

Examples of applications of this method, including the determination of unknown regions and labelling of crossings are shown in Figures 8 and 9. Figure 8 shows an example classification with resulting crossing placement. In this example, we find three possible types of region transition. On the 3rd April 2011 at 16:02 the model describes a transition from magnetosphere to magnetosheath through a short unknown region. This results in two unknown crossings labelled as UKN (MSp  $\rightarrow$  UKN) for the magnetosphere to unknown region transition and UKN (UKN  $\rightarrow$  MSh) for unknown to magnetosheath. Between the MP\_OUT and BS\_OUT intervals, this method described several unknown regions placed wholly within the same region (magnetosheath before and after). Unlike the the first unknown region in this example, these don't necessitate a physical crossing as no confident change in region classification has occurred (and to eye there is no real crossing for these). Unknown crossings were marked here as UKN (MSh  $\rightarrow$  UKN) and UKN (UKN  $\rightarrow$  MSh). Another example of this crossing type occurred within the BS\_OUT interval. Lastly, the model placed several direct bow shock crossings, where transitions between magnetosheath and solar wind occurred directly, without an unknown region. There was one of these crossings at 17:30, marking brief motion of the boundary over and back across the spacecraft, before finally crossing into the solar wind at 17:43, with no more crossings occurring for this outbound passage. This example application highlights a relatively simple case for clarity of explanation, however in practice, more complex cases are common -



an example of which is shown in Figure 9. This figure shows a quasi-parallel bow shock crossing interval, where the model application was uncertain (difference in class probabilities in Figure 9c was small), yielding frequent changes in prediction. This example highlights how the introduction of unknown regions aids in removing false crossings which are due to fluctuations in uncertain predictions. Simpler cases also exist, with single crossing intervals also being commonplace.

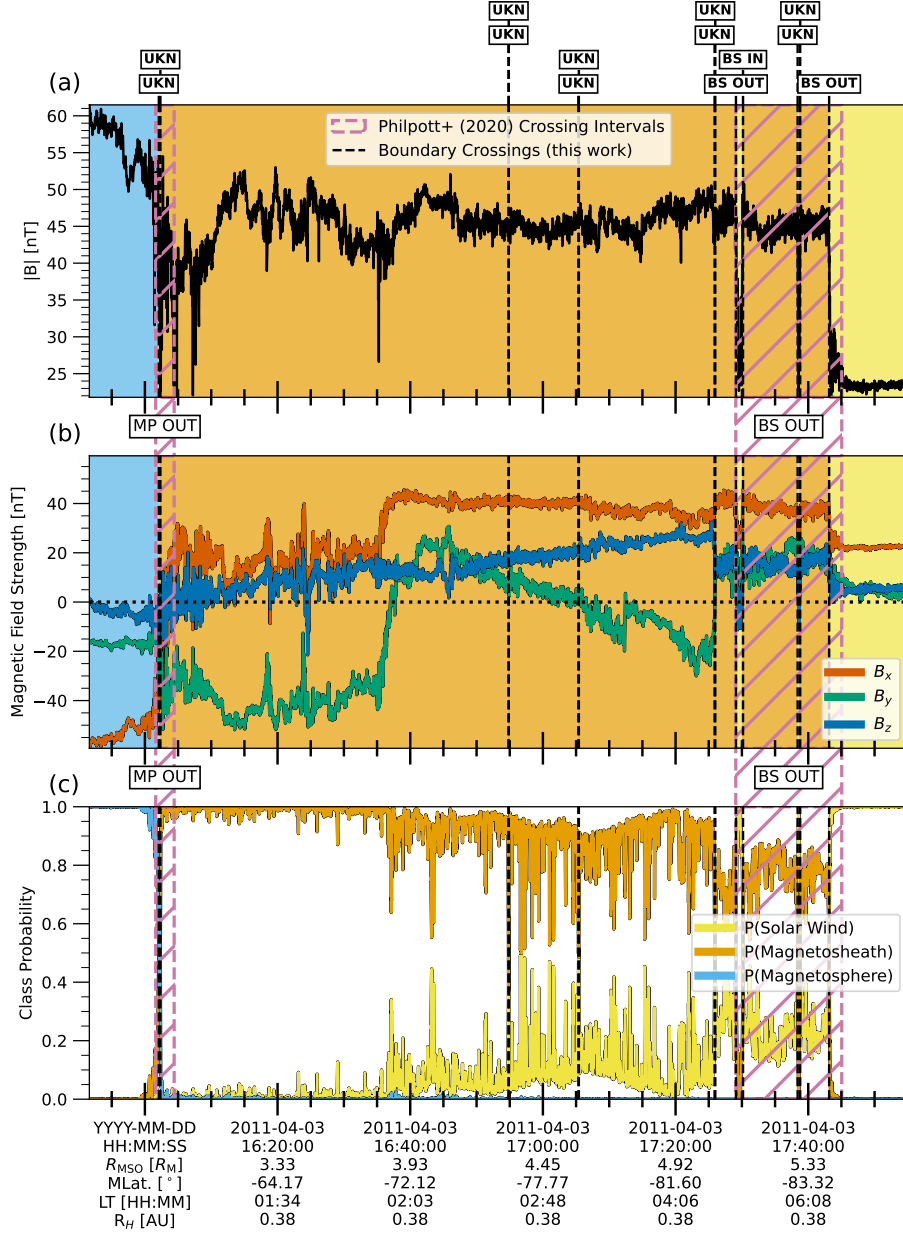
In addition to this, there were also applications with unexpected results. An example of this is shown in Figure 10, where no crossings were predicted within the crossing interval itself. The model predicts magnetosphere before, during, and after the magnetopause crossing interval, and only predicts magnetosheath  $\approx 5$  minutes after the end of the crossing interval. Furthermore, model probabilities (panel (c)) show that the change in prediction from magnetosphere to magnetosheath was gradual, indicating no distinct change in the data was present. We found approximately 5% of intervals didn't have a crossing predicted within 1 minute of the interval, and approximately 1% didn't have a crossing predicted within 5 minutes. Similarly, there were applications where the unexpected class was classified within a crossing interval. i.e. solar wind classifications within a magnetopause crossing interval. There were 13 of these cases ( $< 0.1\%$  of intervals), all of which occur during magnetopause crossing intervals. No further processing steps were taken with respect to these cases as they are rare, and indicative of events of physical interest that warrant further study. We list the intervals in which these cases occur in Appendix B1.

### 3.4 Post-Processing Steps

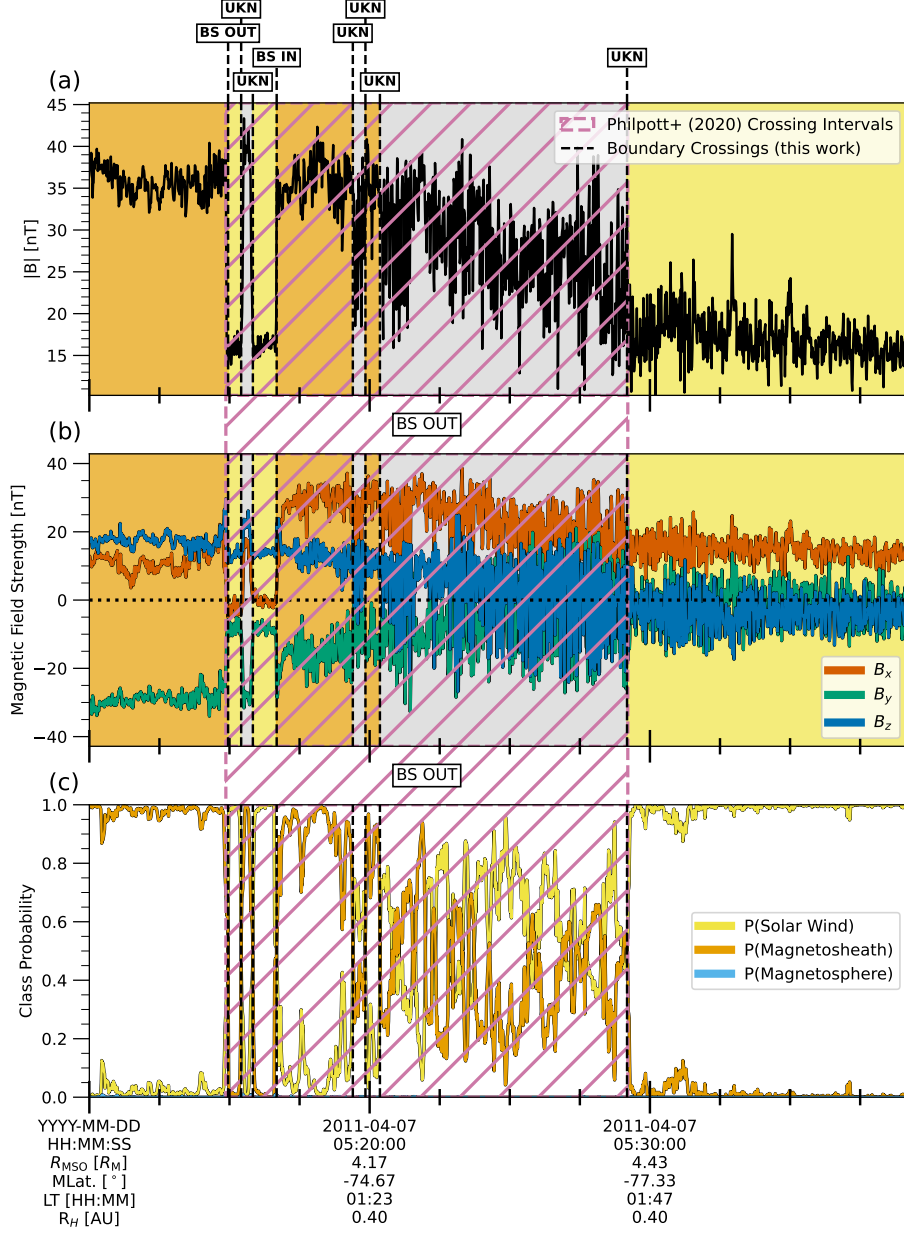
We introduced several post-processing steps to address issues in the application of the model. When applying the model, data was only provided between fixed points, and any window querying data outside of this (i.e. those at the beginning and end of the application,  $< 5$  seconds from the edges of the application) contained missing values. As a result, many of these were incorrectly classified. To amend this, we remove the first and last 5 seconds of each application from the model output prior to determining contiguous regions of like classifications, and prior to determining individual crossings.

Furthermore, as the applications were non-continuous in time, it was possible for the final region classification to conflict with the initial classification for consecutive applications. For example, an application to a pair of inbound crossing intervals: BS\_IN and MP\_IN, would typically end in magnetosphere classifications, matching the starting classifications for the following application: MP\_OUT and BS\_OUT. However, if instead the application to the inbound intervals ended with another classification, such as magnetosheath, this would not match the following application and the overall output would be discontinuous. Treating the crossing intervals as ground truth, we compared the starting and ending region predictions of each application - after determining contiguous region classifications - to what the expected classification was. If one or both of these book-ending regions were misclassified, we changed the region classification(s) to match what was expected. This step affected only 81 cases ( $\approx 0.01\%$ ).

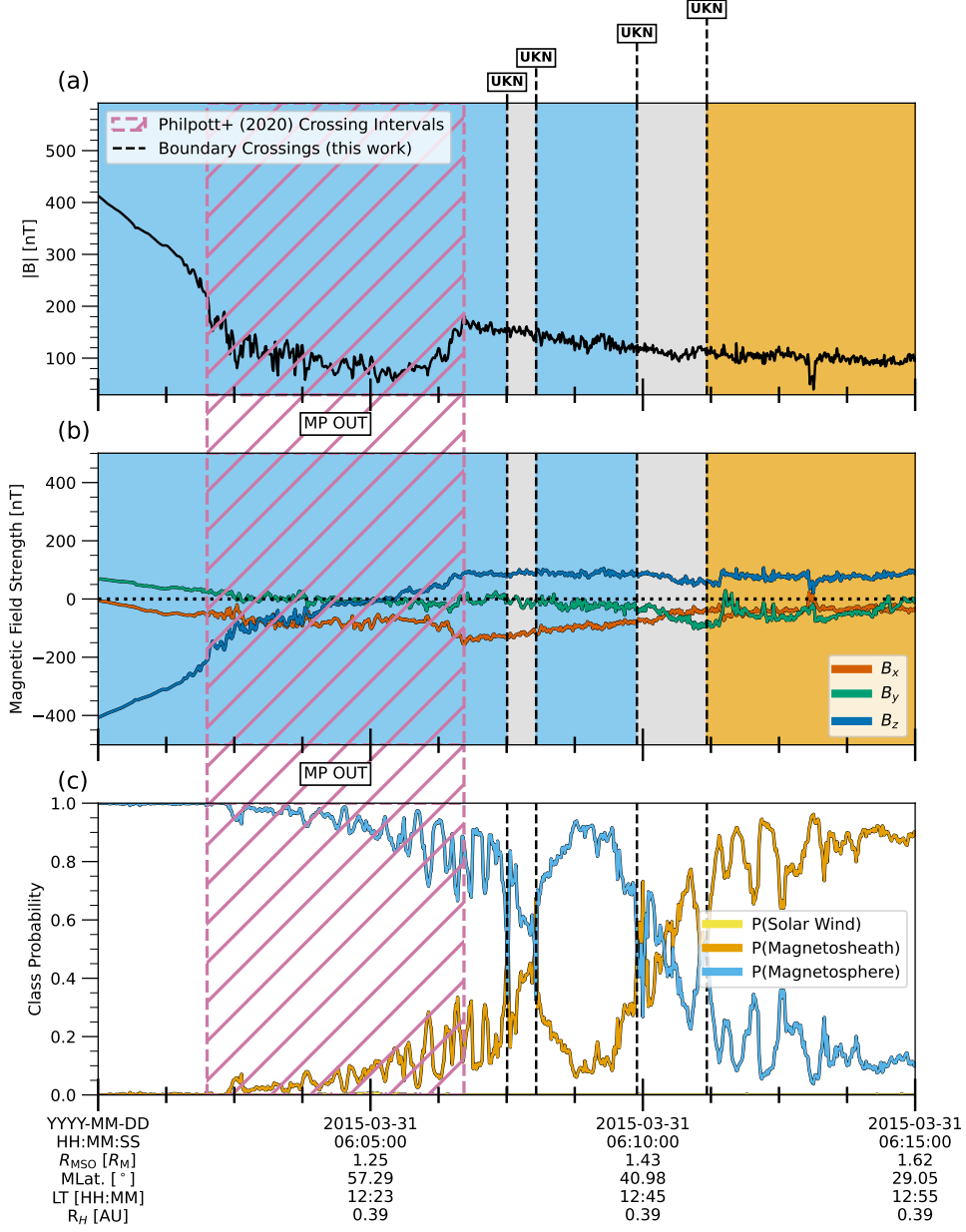
Additionally, due to the introduction of unknown regions, it was possible for region transitions between confident regions to occur through an unknown region - i.e. MSp  $\rightarrow$  UKN  $\rightarrow$  MSh. In these cases, we have only counted unknown crossings in Table 3, and not a direct crossing - such as MP\_OUT. This resulted in some crossing intervals having no direct crossings potentially skewing any statistics surrounding direct crossings only. Given that the scientific focus of our modelling approach is to produce an unambiguous and continuous list of crossings, we apply one further step to our model output: In cases where a known region transition occurs through an unknown region we placed a crossing that represents the region transition between the regions neighbouring the unknown region at the midpoint of the unknown region.



**Figure 8.** An example of the model application to an outbound section of MESSENGER MAG data from 3th April 2011, 16:02 to 17:55. Panels (a) and (b) show the magnitude and components respectively of the 1 second averaged magnetic field as measured by MAG. Panel (c) shows the model output probabilities for each class: solar wind (P(SW), yellow), magnetosheath (P(MSh), orange), and magnetosphere (P(MSp), blue). Unknown regions (UKN) are shaded grey. The colouring of the shaded background denotes classifications, with colour corresponding to the classes in panel (c). Pink diagonal hatching indicates boundary crossing intervals by Philpott et al. (2020). Individual boundary crossings are marked by black vertical dashed lines, and labelled at the top of the figure.



**Figure 9.** An example of the model application to an outbound section of MESSANGER MAG data from 7th April 2011, 05:00 to 05:39. Two unknown regions were placed at times where the model probabilities (panel (c)) were indistinct and changed prediction frequently. Figure formatting follows that described in Figure 8.



**Figure 10.** An example model application where model predictions diverge from expectations visually. Panel (a) shows MESSENGER MAG data from 31st March 2015, 06:00 to 06:15. Figure formatting follows that described in Figure 8. The model predicts magnetosphere (blue) before, during, and after this interval, and only predicts magnetosheath  $\approx 5$  minutes after the crossing interval ends. Figure formatting follows that described in Figure 8.

**Table 3.** Raw crossing counts as determined by region transitions in the initial model output, following the determination of unknown regions, along with updated crossing counts after the post-processing steps described in Section 3.4.

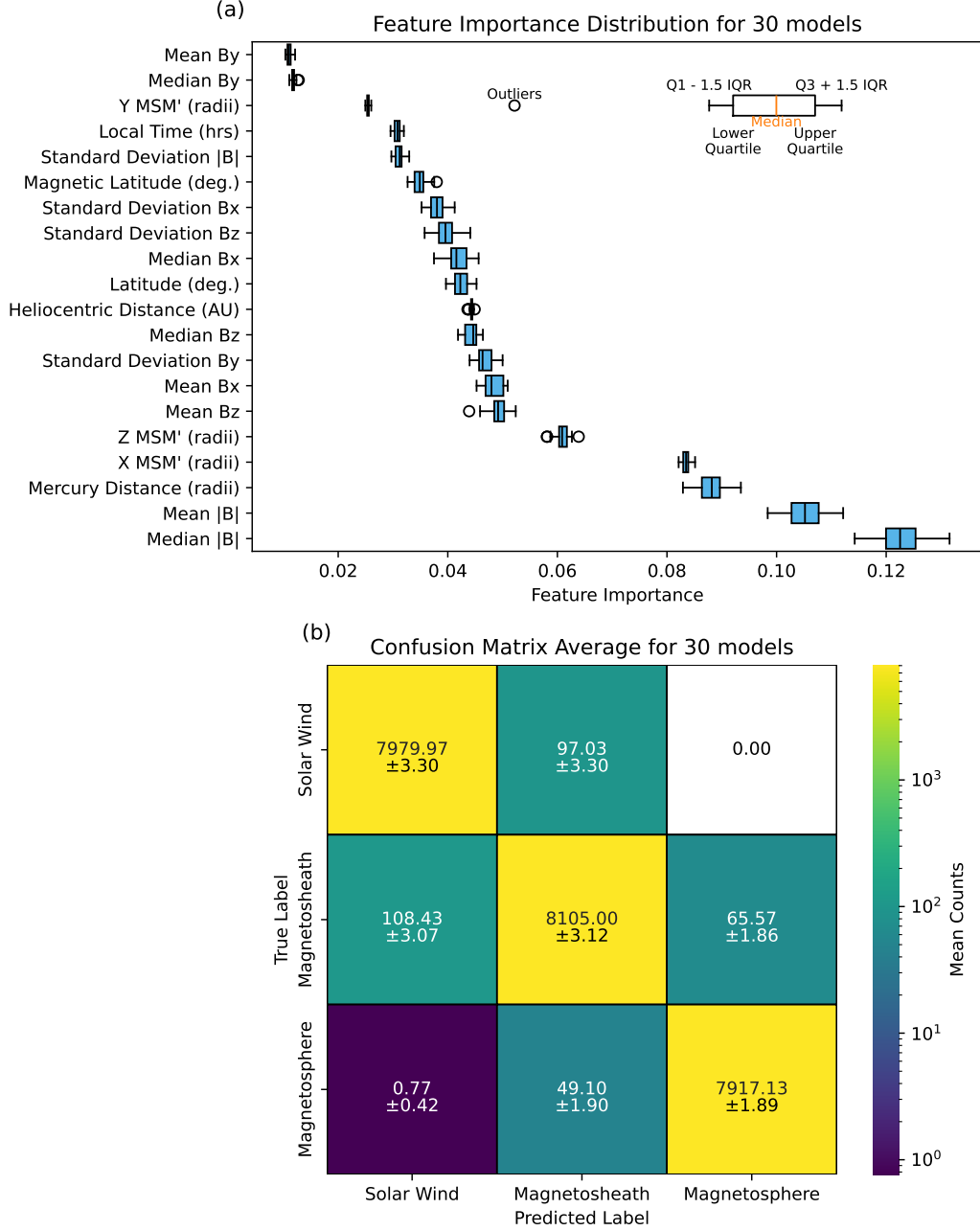
Crossing Type	Counts	Post-processed Counts
Total	102,153	31,326
BS_IN	6,975	10,074
BS_OUT	6,995	10,069
ALL BS	13,970	20,143
MP_IN	2,680	5,600
MP_OUT	2,688	5,580
ALL MP	5,368	11,180
UKN (SW $\rightarrow$ UKN)	11,486	
UKN (UKN $\rightarrow$ SW)	11,449	
UKN (MSh $\rightarrow$ UKN)	23,756	
UKN (UKN $\rightarrow$ MSh)	23,765	
UKN (MSp $\rightarrow$ UKN)	6,165	
UKN (UKN $\rightarrow$ MSp)	6,193	
UNPHYS (SW $\rightarrow$ MSp)	0	2
UNPHYS (MSp $\rightarrow$ SW)	1	1

In the remainder of this study we focus on the output of the model after the post-processing steps outlined above, however in Hollman et al. (2025c) we provide both the raw output and the post-processed outputs. Additionally, we updated Table 3 with the crossing counts after the post-processing described in this section.

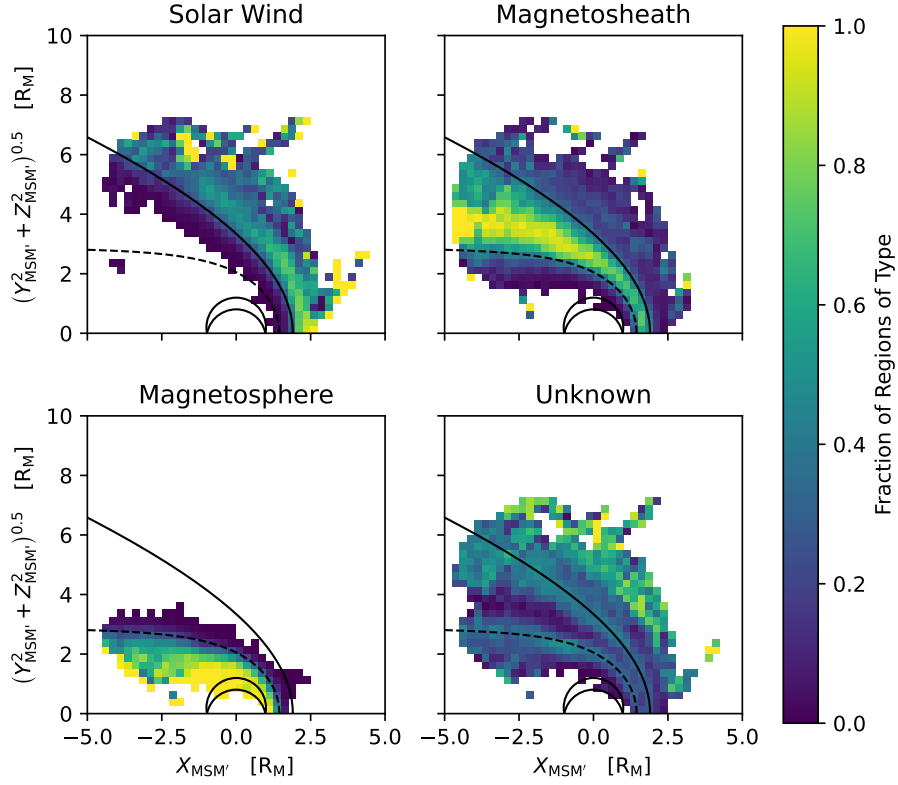
## 4 Results

### 4.1 Model Performance

In total we extracted 243,234 individual 10 second samples from the MESSENGER MAG data, with 218,911 (90%) of these dedicated to training the model, and 24,323 (10%) used for testing. Figure 11 shows feature importance (determined by mean decrease impurity (Ishwaran, 2015), implemented in the Scikit-learn classifier) along with a confusion matrix of classifications on the testing dataset. It is important to distinguish between the accuracy of the model and the accuracy of the application. Model accuracy was directly measurable as the fraction of correctly classified samples from the testing dataset. This model achieves an accuracy score of  $0.9868 \pm 0.0002$  on the testing dataset, however the accuracy score may not reflect the classification performance on data within the boundary crossing intervals, as we lack ground truth labels for these regions. Accuracy on the training dataset and OOB score were  $\sim 1 \pm \sim 0$  and  $0.9866 \pm 0.0001$  respectively. We note that while the training accuracy implies overfitting to the data, given the size of the training and testing datasets and the performance of the model on the unseen data, the model is clearly able to generalise well, and we don’t see this as an issue.



**Figure 11.** Average feature importance and confusion matrix for 30 random forest models with the optimised hyperparameters. Each model was initialised with a different, but fixed, random seed so that models could vary during training despite acting on the same dataset. We determined impurity-based feature importance each model and plotted their distributions in panel (a). Box and whisker plots show the distribution of importances across models as described in the key at the top right. The interquartile range (IQR) is marked with a blue box, with a black vertical line marking the median. Whiskers extend out to 1.5 times the IQR before and beyond the 1st and 3rd quarters respectively. Outliers to these are marked by circles. Panel (b) shows the mean and standard deviation of confusion matrices calculated on the testing dataset, coloured by number of classifications on a log scale. Correct classifications exist on the diagonal of this matrix. The 'True Label: Solar Wind, Predicted Label: Magnetosphere' cell is excluded as there were no misclassification of this type for any of the 30 models, and 0 is not defined on this log-space colour scale.



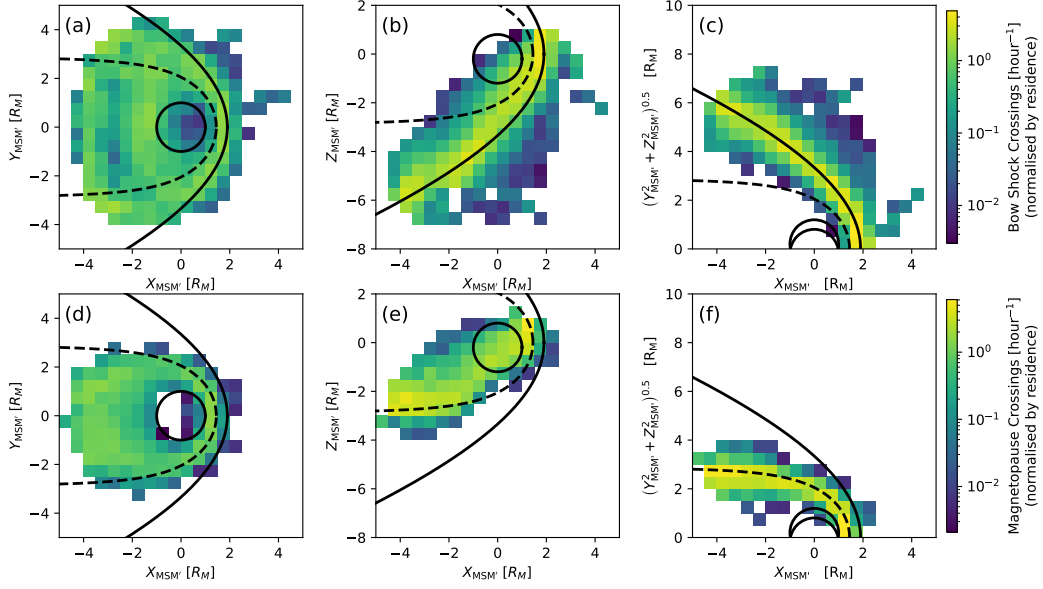
**Figure 12.** The spatial spread of solar wind, magnetosheath, magnetosphere, and unknown classifications as projected to the cylindrical plane in the aberrated Mercury-Solar-Magnetospheric (MSM') coordinate system. Two overlapping circles were drawn to represent the northern and southern extends of Mercury, folded into the plane. The average boundary shape by Winslow et al. (2013) for the bow shock and magnetopause are drawn as a solid and dashed curve respectively.

After the introduction of unknown regions as described in Section 3.3, we mapped the spatial occurrence of all region classifications in Figure 12. Each panel shows the fraction of regions of a given classification in spatial bins as projected to the cylindrical plane, with Mercury and the average bow shock and magnetopause boundary locations drawn in black. We observed an increase in unknown regions around these average boundary locations. This was expected, as the applications of the model occurred solely around boundary crossing intervals. We also see an increase in unknown classifications far from Mercury, where more extreme events were required for crossings to occur there.

## 4.2 Newly Identified Boundary Crossings

We identified a total of 102,153 new boundary crossings across the MESSENGER mission - an average of  $\approx 5$  individual boundary crossings per crossing interval. Table 3 shows counts for each crossing type. This method finds 19,338 direct boundary crossings and 82,815 unknown crossings. The imbalance between direct and unknown crossings exists as unknown regions were determined based on region duration, along with region con-





**Figure 13.** The spatial spread of individual bow shock (panels (a-c), above) and magnetopause (panels (d-f), below) crossings as determined based on changes in region probability are shown as projected to the  $XY$  and  $XZ$ , and cylindrical planes of the aberrated Mercury-Solar-Magnetospheric (MSM') coordinate system. A circle was drawn to represent the size and location of Mercury. In the case of the cylindrical planes two overlapping circles were drawn to represent the northern and southern sections of Mercury along the  $Z$  axis, folded into the plane. The average boundary shape by Winslow et al. (2013) for the bow shock and magnetopause are drawn as a solid and dashed curve respectively.

fidence. Shorter duration regions were more likely to be considered unknown (as shown in Figure 7) and hence the temporal density of unknown regions, and therefore also of unknown crossings, was larger than that for direct crossings. As discussed in section 3.2, direct crossings may, but don't necessarily, occur within unknown regions, though this method lacks the accuracy required to determine them. In addition to direct crossings, indirect (but definite) crossings also occur, where one region transitions to another through an unknown region (i.e. solar wind  $\rightarrow$  unknown  $\rightarrow$  magnetosheath). In these cases, we record two unknown crossings, though physically, at least one - but always an odd number - of direct crossings must have occurred. This issue is partially accounted for by post processing steps described in section 3.4, with updated counts for each crossing type recorded in Table 3.

After post-processing, we found a total of 31,326 crossings. We plotted the spatial distribution of these post-processed crossings in Figure 13 which visually compare well to those of the boundary crossing intervals in Figure 4, but leave further statistical analysis for future works. This list of post-processed crossings was made publicly available for use (Hollman et al., 2025c).

## 5 Discussion and Future Work

### 5.1 Evaluating Unknown Regions

Within the post-processing steps of this work, we defined some contiguous sections sharing the same classification label as unknown regions under the condition that they were shorter than a threshold on duration, and less confident than a threshold on region confidence. It is worthwhile discussing the interpretation of these unknown regions. They were constructed empirically, and hence we should be cautious when considering their meaning. We know the model has generalised well to the physics of this system. Its performance on the testing dataset has shown that it can separate solar wind, magnetosheath, and magnetosphere data with very high accuracy. By definition, these unknown regions occur where the model lacks 'confidence', at times where the ensemble is split in their classification, and is quite variable in subsequent time-steps. However, there are times when this can be informative. Consider Figure 9: before and after the outbound bow shock interval, we see distinct differences between magnetosheath and solar wind. The unknown region (highlighted in grey) contains data which differs significantly from both the magnetosheath before, and solar wind after. In this case, the model informs us that these times are dissimilar to its training dataset, and we can interpret this physically as processes at the boundary introducing turbulence and mixing of the plasma populations, resulting in data which was physically dissimilar from that which the model trained on. This is furthered by Figure 12, which plots the proportion of each region (solar wind, magnetosheath, and unknown) on a cylindrically symmetric spatial grid. We found unknown regions to be gathered near the average boundary determined by Winslow et al. (2013), where we would expect to see such near-boundary effects most commonly. Additionally, we see an increase in unknown regions with distance from the average bow shock location, where extreme conditions are required for application in these bins. We cannot however completely rely on the introduction of unknown regions to inform us of physical discrepancies as we also find cases where unknown regions occur without an obvious difference in data, for example, the brief unknown regions between the MP OUT and BS OUT crossing intervals in Figure 8.

It is important to highlight the quantity of unknown crossings determined compared to that of direct crossings. Prior to the removal of these unknown crossings in post-processing steps, the ratio of unknown crossings to direct crossings was roughly 3:1. This appears to directly conflict with the high accuracy of the model on the testing dataset, as if all classifications were highly confident we would solely have direct crossings. As discussed in Section 3.3, unknown regions were by definition shorter on average than any of the known classifications, and hence unknown crossings must have a larger temporal density than direct crossings, resulting in an increased count.

The post-processing steps introduced make an effort to replace these unknown crossings with direct crossings where applicable. In particular, we would like to highlight the first unknown region in Figure 8 (separating magnetosphere and magnetosheath), and the last (and largest) unknown region in Figure 9 (separating magnetosheath and solar wind). In both of these cases we require at least one physical crossing to have occurred. The post-processing steps replace the two unknown crossings bookending this region with a single direct crossing (labelled based on the region classifications before and after), placed in the middle of this region. For the example in Figure 8, this method works well as this region is short, and the placement is temporally accurate. However, for the example in Figure 9, the large length of this unknown region means that placement in the middle results in a large temporal inaccuracy. To eye, the true placement of the bow shock crossing within this unknown region is at the end, and hence, the placement of this crossing after the post-processing steps was offset by roughly 4 minutes. We can infer from the distribution of unknown regions there is an average temporal accuracy loss where unknown regions are present (Median:  $\pm 13$  s, Q1 =  $\pm 4$  s, Q3 =  $\pm 40$  s, 95th percentile:  $\pm 2$  minutes 36 seconds). For this reason, the post-processed version of this list is most

useful for large scale statistical studies, however for event studies, the raw model output is a better reference.

## 5.2 Future Work

This work opens numerous possibilities for further investigation. We present a random forest model as a method for magnetospheric region classification with magnetic field and ephemeris alone. As the model accuracy was high, the possible percentage gains in improving the model are low, however further tuning of hyperparameters is possible, and exploration of additional features could still prove useful. The set of features explored was not exhaustive. Additionally, further simplification of the model could be explored by removing low importance features. Both the tuning of hyper-parameters, and the further investigation of features could improve model performance in testing and application.

We have applied a new, generalisable method to detect Mercury boundary crossings. This manuscript, along with our open source code (Hollman et al., 2025b) will be a strong foundation for boundary detection for data from the future BepiColombo mission. This dual-spacecraft mission will be the first to offer multiple-viewpoint in-situ observations at Mercury. The Mercury Magnetospheric Orbiter (MMO) will enter into a highly eccentric orbit of Mercury, sampling from the solar wind through to the magnetosphere, while the Mercury Planetary Orbiter (MPO) will enter a relatively circular orbit sampling primarily the magnetosphere and magnetosheath (Benkhoff et al., 2021). Magnetospheric region classification and boundary detection will be vitally important for the coordination of science between both spacecraft. The two spacecraft will follow different orbits than MESSENGER, and while they may at times overlap with orbits by MESSENGER, they will sample a broader range of positions. As a result, ephemeris features extracted from BepiColombo when directly applying this model may fall outside the ranges observed during training, and affect model performance. To circumvent this, we performed a preliminary investigation into the performance of the model when excluding ephemeris features. This model performed similarly to the original model still achieving an average accuracy of  $0.929 \pm 0.001$ . A detailed description and comparison can be found in Appendix A. We hope to test both models with/without ephemeris features without any transfer learning on BepiColombo data in the future, however further labelling may be required to adapt this method to those spacecraft. Additionally, while the model used in this study is Mercury specific in terms of feature values, the method is entirely generalisable to any magnetised planet in the solar system (provided some labelling of magnetospheric regions has been carried out).

This work also paves the way for a wealth of new statistical studies. Previous statistical studies on the location of Mercury’s magnetospheric boundaries have been based off lists of crossing intervals, specifically taking the centre of the interval as the boundary location (Winslow et al., 2013; Philpott et al., 2020; He et al., 2022). However, with knowledge of the precise times (and locations) of each boundary crossing, as presented in this work, we can accurately connect perturbations of the boundary location to external and internal drivers (solar wind dynamic pressure, Alfvén Mach number, reconnection rates, core conduction, etc.). Additionally, we can unravel variability in the boundary shape, as crossing intervals contain a differing number of individual boundary crossings. Furthermore, the number of crossings for a given pass of each boundary gives an indication of boundary variability and could be explored further to determine differences in variability from the nose the flanks of the magnetosphere, along with differences in variability due to Mercury’s heliocentric distance.

## 6 Conclusions

We present a new list of crossings by the MESSENGER spacecraft of Mercury's bow shock and magnetopause based on automated region classification. A random forest was fit to MESSENGER magnetic field and ephemeris data to classify regions of solar wind, magnetosheath, and magnetosphere. The random forest is highly accurate on the testing dataset, with a score of  $0.9868 \pm 0.0002$ . Furthermore, our model and method are generalised and applicable to the future BepiColombo mission, and other magnetised planets in the solar system. We highlight how models trained on MESSENGER data have direct application to future data from BepiColombo.

**Table B1.**

Interval Label	Orbit Number	Interval Start Time [YYYY-MM-DD HH:MM:SS]	Interval End Time [YYYY-MM-DD HH:MM:SS]	Duration of Solar Wind Classifications [s] ([% of interval])
MP_OUT	28	2011-04-01 03:50:33	2011-04-01 04:11:11	291 (23.5)
MP_OUT	472	2011-11-07 22:33:40	2011-11-07 22:44:14	7 (1)
MP_IN	816	2012-04-22 13:48:22	2012-04-22 14:04:50	926 (93.7)
MP_OUT	817	2012-04-22 23:48:06	2012-04-23 00:13:15	565 (37.4)
MP_OUT	927	2012-05-29 15:48:26	2012-05-29 15:58:20	107 (18)
MP_OUT	1020	2012-06-29 16:26:30	2012-06-29 17:17:50	766 (24.9)
MP_IN	1241	2012-09-11 07:09:34	2012-09-11 07:11:08	20 (21.3)
MP_IN	1273	2012-09-21 22:47:23	2012-09-22 22:54:29	329 (77.1)
MP_OUT	1273	2012-09-22 01:26:12	2012-09-22 01:37:22	370 (55.2)
MP_OUT	1980	2013-05-15 17:06:32	2013-05-15 17:18:48	17 (2.3)
MP_IN	2536	2013-11-17 00:54:41	2013-11-17 01:02:52	64 (13)
MP_IN	2816	2014-02-18 09:38:01	2014-02-18 09:40:51	12 (7.1)

## Appendix A Evaluating a Model Excluding Ephemeris

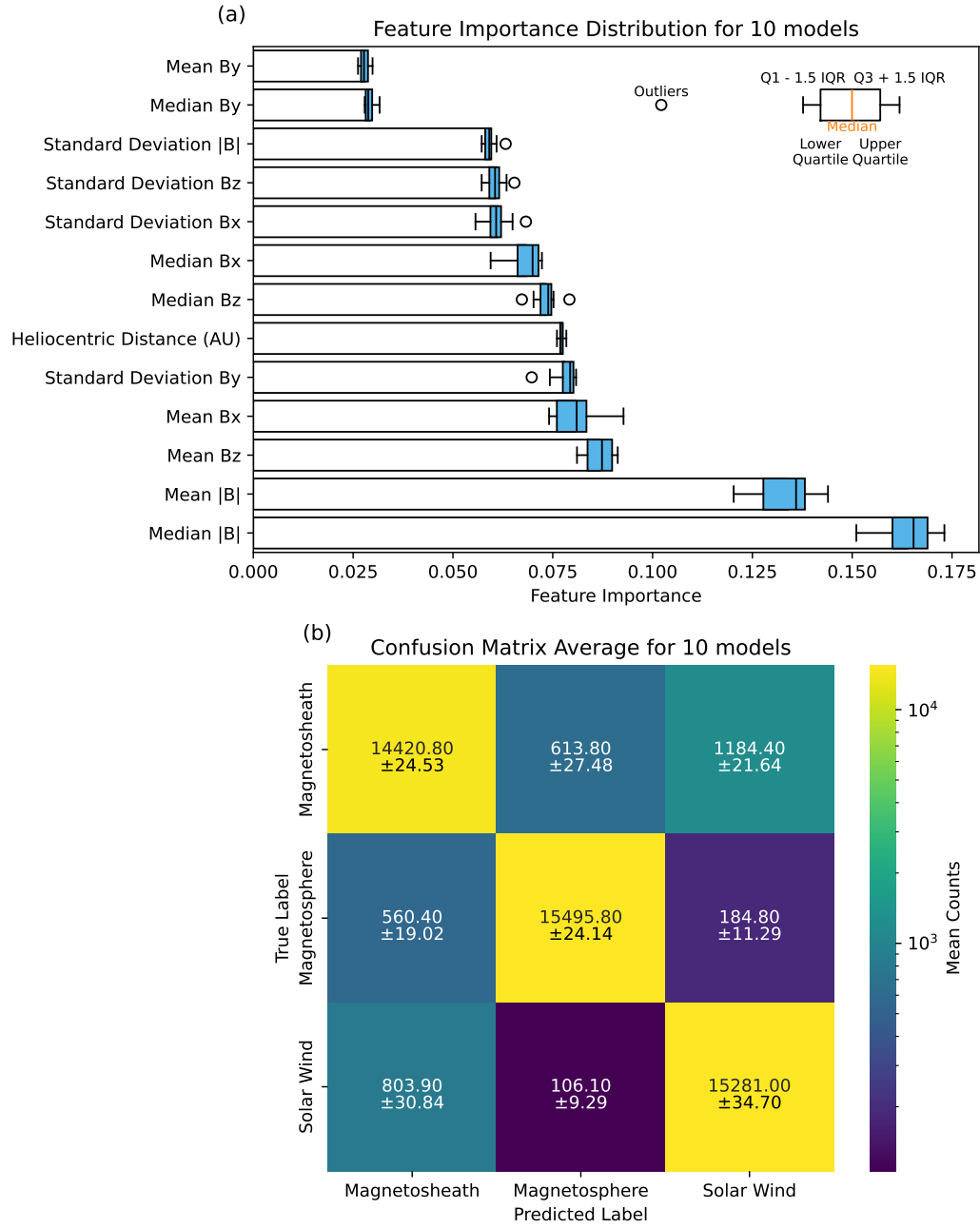
The random forest model used in this study makes predictions based on features, some of which are derived from MESSENGER’s trajectory (position, local time, latitude, etc.). These features are useful in achieving a high accuracy, with  $X_{\text{MSM}'}$  and  $Z_{\text{MSM}'}$  in the top four most important features (Figure 11). However, they restrict successful application of the model to the parameter-space which MESSENGER explored. Future spacecraft at Mercury, such as those from the BepiColombo mission, will orbit on different trajectories to MESSENGER’s and potentially take measurements in positions not sampled before. In these instances, a model which includes ephemeris features could produce inaccurate results. In order to showcase the generalisability of this method, we highlight how a version of this model retains a high accuracy when ephemerides are excluded. A random forest model was fit to the same training dataset as described in 3.1 with the exception that ephemeris features were removed. This model had reduced accuracy in comparison with the previous, as one would expect, but remains highly accurate with a score of  $0.929 \pm 0.001$ . Figures A1a, b show the feature importance and confusion matrix respectively. The feature importance has some differences in the ranking of features compared to Figure 11a for the five lowest importance features. In these cases the box and whisker plots demonstrate that different rankings were likely for each of the 10 models.

## Appendix B Crossing Intervals Containing Unexpected Classifications

As described in Section 3.3, our model made predictions of solar wind within the magnetopause crossing intervals previously labelled by Philpott et al. (2020). There were 16 of these intervals, which are detailed in Table B1.

## Data Availability Statement

MESSENGER data are available through the Planetary Plasma Interactions Node of the Planetary Data System (Korth & Anderson, 2016). The training dataset, raw model output, and post-processed crossing list created by this work are available for download (Hollman et al., 2025c). All code for dataset construction, modelling, and application for this work



**Figure A1.** Average feature importance and confusion matrix as described in Figure 11 for a model retrained without the inclusion of ephemeris features.

are publicly available (Hollman et al., 2025b). All scripts and resources to create the figures in this manuscript are also publicly available (Hollman et al., 2025a).

## Acknowledgments

Work at DIAS was supported by Taighde Éireann - Research Ireland: Postgraduate Scholarship GOIPG/2024/3906 (DMH); Research Ireland Laureate Consolidator award SOLMEX (CMJ, CFB, and ARF); Science Foundation Ireland award 18/FRL/6199 (CMJ, SJW, MJR).

## References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining*.
- Anderson, B. J., Acuña, M. H., Lohr, D. A., Scheifele, J., Raval, A., Korth, H., & Slavin, J. A. (2007). The Magnetometer Instrument on MESSENGER. *Space Science Reviews*, 131(1-4), 417-450. doi: 10.1007/s11214-007-9246-7
- Anderson, B. J., Acuña, M. H., Lohr, D. A., Scheifele, J., Raval, A., Korth, H., & Slavin, J. A. (2007). The magnetometer instrument on messenger. *Space Science Reviews*, 131(1-4), 417-450. doi: 10.1007/s11214-007-9246-7
- Anderson, B. J., Johnson, C. L., Korth, H., Purucker, M. E., Moldovan, R., Solomon, S. C., & McNutt Jr., R. L. (2011). The global-scale magnetic field of mercury. In *Epsc-dps joint meeting 2011* (Vol. 2011, p. 151).
- Andrews, G. B., Zurbuchen, T. H., Mauk, B. H., Malcom, H., Fisk, L. A., Gloeckler, G., ... Raines, J. M. (2007). The Energetic Particle and Plasma Spectrometer Instrument on the MESSENGER Spacecraft. *Space Science Reviews*, 131(1-4), 523-556. doi: 10.1007/s11214-007-9272-5
- Argall, M. R., Small, C. R., Piatt, S., Breen, L., Petrik, M., Kokkonen, K., ... Burch, J. L. (2020, September). Mms sitl ground loop: Automating the burst data selection process. *Frontiers in Astronomy and Space Sciences*, 7, 54. doi: 10.3389/fspas.2020.00054
- Baker, D. N., Odstrcil, D., Anderson, B. J., Arge, C. N., Benna, M., Gloeckler, G., ... Zurbuchen, T. H. (2011). The space environment of mercury at the times of the second and third messenger flybys. *Planetary Space Science*, 59(15), 2066-2074. doi: 10.1016/j.pss.2011.01.018
- Bakrania, M. R., Rae, I. J., Walsh, A. P., Verscharen, D., & Smith, A. W. (2020). Using dimensionality reduction and clustering techniques to classify space plasma regimes. *Frontiers in Astronomy and Space Sciences*, 7, 80. doi: 10.3389/fspas.2020.593516
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. doi: 10.1016/j.isprsjprs.2016.01.011
- Benkhoff, J., Murakami, G., Baumjohann, W., Besse, S., Bunce, E., Casale, M., ... Zender, J. (2021). Bepicolombo - mission overview and science goals. *Space Science Reviews*, 217. doi: 10.1007/s11214-021-00861-4
- Bowers, C. F., Jackman, C. M., Azari, A. R., Smith, A. W., Wright, P. J., Rutala, M. J., ... Healy, A. (2024). Estimating interplanetary magnetic field conditions at mercury's orbit from messenger magnetosheath observations using a feedforward neural network. *Journal of Geophysical Research: Machine Learning and Computation*, 1(4), e2024JH000239. (e2024JH000239 2024JH000239) doi: <https://doi.org/10.1029/2024JH000239>
- Bowers, C. F., Jackman, C. M., Jia, X., Slavin, J. A., Saur, J., Holmberg, M. K. G., ... Hollman, D. M. (2025, 3). Messenger observations of a possible alfvén



- wing at mercury driven by a low alfvénic mach number interplanetary coronal mass ejection. *Journal of Geophysical Research: Space Physics*, 130. doi: 10.1029/2024ja033619
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. doi: 10.1023/a:1010933404324
- Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020). Automatic classification of plasma regions in near-earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observations. *Frontiers in Astronomy and Space Sciences*, Volume 7 - 2020. Retrieved from <https://www.frontiersin.org/journals/astronomy-and-space-sciences/articles/10.3389/fspas.2020.00055> doi: 10.3389/fspas.2020.00055
- Case, N. A., & Wild, J. A. (2013). The location of the earth's magnetopause: A comparison of modeled position and in situ cluster data. *Journal of Geophysical Research: Space Physics*, 118(10), 6127–6135. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgra.50572> doi: <https://doi.org/10.1002/jgra.50572>
- Daigavane, A., Wagstaff, K. L., Doran, G., Cochrane, C. J., Jackman, C. M., & Rymer, A. (2022). Unsupervised detection of saturn magnetic field boundary crossings from plasma spectrometer data. *Computers Geosciences*, 161, 105040. doi: <https://doi.org/10.1016/j.cageo.2022.105040>
- Dimmock, A. P., Russell, C. T., Sagdeev, R. Z., Krasnoselskikh, V., Walker, S. N., Carr, C., ... Balikhin, M. A. (2019). Direct evidence of nonstationary collisionless shocks in space plasmas. *Science Advances*, 5(2), eaau9926. doi: 10.1126/sciadv.aau9926
- Edmond, J., Raeder, J., Ferdousi, B., Argall, M., & Innocenti, M. E. (2024). Clustering of global magnetospheric observations. *Journal of Geophysical Research: Machine Learning and Computation*, 1(4), e2024JH000221. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JH000221> (e2024JH000221 2024JH000221) doi: <https://doi.org/10.1029/2024JH000221>
- Grisson, B., Darrouzet, F., Maggiolo, R., Hajoš, M., Dvořák, M., Švanda, M., ... De Keyser, J. (2025, February). Localization of the cluster satellites in the geospace environment. *Scientific Data*, 12(1), 327. doi: 10.1038/s41597-025-04639-z
- Haaland, S., Reistad, J., Tenfjord, P., Gjerloev, J., Maes, L., DeKeyser, J., ... Dorville, N. (2014). Characteristics of the flank magnetopause: Cluster observations. *Journal of Geophysical Research: Space Physics*, 119(11), 9019–9037. doi: <https://doi.org/10.1002/2014JA020539>
- He, P., Xu, X., Yu, H., Wang, X., Wang, M., Chang, Q., ... Li, H. (2022). The mercury's bow-shock models near perihelion and aphelion. *The Astronomical Journal*, 164, 260. doi: 10.3847/1538-3881/ac9d89
- Hollman, D., Jackman, C., Domijan, K., Bowers, C., Walker, S., Rutala, M. J., & Fogg, A. R. (2025a, July). *daraghollman/messenger-random-forest-paper-figure*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.16357245> doi: 10.5281/zenodo.16357245
- Hollman, D., Jackman, C., Domijan, K., Bowers, C., Walker, S., Rutala, M. J., & Fogg, A. R. (2025b, July). *daraghollman/messenger-region-classification*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.16357230> doi: 10.5281/zenodo.16357230
- Hollman, D., Jackman, C., Domijan, K., Bowers, C., Walker, S., Rutala, M. J., & Fogg, A. R. (2025c, June). *Messenger magnetospheric boundary crossings - hollman et al. (2025)*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.15731194> doi: 10.5281/zenodo.15731194
- Ishwaran, H. (2015). The effect of splitting on random forests. *Machine Learning*,

- 99, 75–118. doi: 10.1007/s10994-014-5451-2
- Jackman, C. M., Thomsen, M. F., & Dougherty, M. K. (2019, November). Survey of saturn's magnetopause and bow shock positions over the entire cassini mission: Boundary statistical properties and exploration of associated upstream conditions. *Journal of Geophysical Research (Space Physics)*, 124(11), 8865–8883. doi: 10.1029/2019JA026628
- James, M. K., Imber, S. M., Raines, J. M., Yeoman, T. K., & Bunce, E. J. (2020). A Machine Learning Approach to Classifying MESSENGER FIPS Proton Spectra. *Journal of Geophysical Research (Space Physics)*, 125(6), e27352. doi: 10.1029/2019JA027352
- Jarvinen, R., Alho, M., Kallio, E., & Pulkkinen, T. I. (2020). Ultra-low-frequency waves in the ion foreshock of mercury: a global hybrid modelling study. *Monthly Notices of the Royal Astronomical Society*, 491, 4147–4161. doi: 10.1093/mnras/stz3257
- Johnson, C. L., Purucker, M. E., Korth, H., Anderson, B. J., Winslow, R. M., Al Asad, M. M. H., ... Solomon, S. C. (2012). Messenger observations of mercury's magnetic field structure. *Journal of Geophysical Research (Planets)*, 117, E00L14. doi: 10.1029/2012JE004217
- Joy, S. P., Kivelson, M. G., Walker, R. J., Khurana, K. K., Russell, C. T., & Ogino, T. (2002). Probabilistic models of the jovian magnetopause and bow shock locations. *Journal of Geophysical Research: Space Physics*, 107(A10), SMP 17-1-SMP 17-17. doi: <https://doi.org/10.1029/2001JA009146>
- Julka, S., Kirschstein, N., Granitzer, M., Lavrukhin, A., & Amerstorfer, U. (2023). Deep active learning for detection of mercury's bow shock and magnetopause crossings. In M.-R. Amini, S. Canu, A. Fischer, T. Guns, P. Kralj Novak, & G. Tsoumakas (Eds.), *Machine learning and knowledge discovery in databases* (pp. 452–467). Cham: Springer Nature Switzerland.
- Karlsson, T., Plaschke, F., Glass, A. N., & Raines, J. M. (2024, April). Short large-amplitude magnetic structures (slams) at mercury observed by messenger. *Annales Geophysicae*, 42(1), 117–130. doi: 10.5194/angeo-42-117-2024
- Kit Cheng, I., & Achilleos, N. (2022, July). Automated bow shock and magnetopause boundary detection with cassini using threshold and deep learning methods. In *44th cospar scientific assembly. held 16-24 july* (Vol. 44, p. 509).
- Korth, H., & Anderson, B. J. (2016). *MESSENGER MAG Calibrated Data Bundle*. NASA Planetary Data System. doi: 10.17189/1522383
- Kruparova, O., Krupar, V., Šafránková, J., Němeček, Z., Maksimovic, M., Santolik, O., ... Merka, J. (2019). Statistical survey of the terrestrial bow shock observed by the cluster spacecraft. *Journal of Geophysical Research: Space Physics*, 124(3), 1539–1547. doi: <https://doi.org/10.1029/2018JA026272>
- Leary, J. C., Conde, R. F., Dakermanji, G., Engelbrecht, C. S., Ercol, C. J., Fielhauer, K. B., ... Wiley, S. R. (2007). The messenger spacecraft. *Space Science Reviews*, 131, 187–217. doi: 10.1007/s11214-007-9269-0
- Milillo, A., Fujimoto, M., Murakami, G., Benkhoff, J., Zender, J., Aizawa, S., ... Wahlund, J.-E. (2020). Investigating mercury's environment with the two-spacecraft bepicolombo mission. *Space Science Reviews*, 216. doi: 10.1007/s11214-020-00712-8
- Ness, N. F., Behannon, K. W., Lepping, R. P., Whang, Y. C., & Schatten, K. H. (1974). Magnetic field observations near mercury: Preliminary results from mariner 10. *American Association for the Advancement of Science, Science*, 185(4146), 151–160.
- Nevskii, D. V., Lavrukhin, A. S., & Alexeev, I. I. (2023). Automatic detection of bow shock and magnetopause positions at mercury's magnetosphere using messenger magnetometer data. *Cosmic Research*, 61, 194–205. doi: 10.1134/s0010952522600081
- Nguyen, G., Aunai, N., Michotte de Welle, B., Jeandet, A., Lavraud, B., &

- Fontaine, D. (2022). Massive multi-mission statistical study and analytical modeling of the earth's magnetopause: 1. a gradient boosting based automatic detection of near-earth regions. *Journal of Geophysical Research: Space Physics*, 127(1), e2021JA029773. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029773> (e2021JA029773 2021JA029773) doi: <https://doi.org/10.1029/2021JA029773>
- Olshevsky, V., Khotyaintsev, Y. V., Lalti, A., Divin, A., Delzanno, G. L., Anderzén, S., ... Markidis, S. (2021). Automated classification of plasma regions using 3d particle energy distributions. *Journal of Geophysical Research: Space Physics*, 126. doi: 10.1029/2021ja029620
- 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.
- Philpott, L. C., Johnson, C. L., Anderson, B. J., & Winslow, R. M. (2020). The shape of mercury's magnetopause: The picture from messenger magnetometer observations and future prospects for bepicolombo. *Journal of Geophysical Research: Space Physics*, 125(5), e2019JA027544. (e2019JA027544 10.1029/2019JA027544) doi: <https://doi.org/10.1029/2019JA027544>
- Rutala, M. J., Jackman, C. M., Louis, C. K., Azari, A. R., Bagenal, F., Joy, S. P., ... Vogt, M. F. (2025). New models of jupiter's magnetopause and bow shock through the juno prime mission: Probabilistic location, shape, and internally-driven variation. *Journal of Geophysical Research: Space Physics*, 130. doi: 10.1029/2025ja033842
- Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops* (p. 166-171). doi: 10.1109/ICDCSW.2011.20
- Shue, J. H., Chao, J. K., Fu, H. C., Russell, C. T., Song, P., Khurana, K. K., & Singer, H. J. (1997). A new functional form to study the solar wind control of the magnetopause size and shape. *Journal of Geophysical Research: Space Physics*, 102(A5), 9497–9511. doi: 10.1029/97JA00196
- Shue, J.-H., & Song, P. (2002). The location and shape of the magnetopause. *Planetary and Space Science*, 50, 549–558. doi: 10.1016/s0032-0633(02)00034-x
- Slavin, J. A., Acuña, M. H., Anderson, B. J., Baker, D. N., Benna, M., Gloeckler, G., ... Zurbuchen, T. H. (2008). Mercury's magnetosphere after messenger's first flyby. *Science*, 321(5885), 85. doi: 10.1126/science.1159040
- Slavin, J. A., Middleton, H. R., Raines, J. M., Jia, X., Zhong, J., Sun, W.-J., ... Mays, M. L. (2019). Messenger observations of disappearing dayside magnetosphere events at mercury. *Journal of Geophysical Research: Space Physics*, 124(8), 6613–6635. doi: <https://doi.org/10.1029/2019JA026892>
- Sun, W. J. (2023). *Lists of magnetopause and bow shock crossings of mercury by messenger spacecraft*. Zenodo. doi: 10.5281/zenodo.8298647
- Sun, W. J., Slavin, J. A., Dewey, R. M., Chen, Y., DiBraccio, G. A., Raines, J. M., ... Akhavan-Tafti, M. (2020). Messenger observations of mercury's nightside magnetosphere under extreme solar wind conditions: Reconnection-generated structures and steady convection. *Journal of Geophysical Research: Space Physics*, 125. doi: 10.1029/2019ja027490
- Toy-Edens, V., Mo, W., Raptis, S., & Turner, D. L. (2024). Classifying 8 years of mms dayside plasma regions via unsupervised machine learning. *Journal of Geophysical Research: Space Physics*, 129(6), e2024JA032431. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JA032431> (e2024JA032431 2024JA032431) doi: <https://doi.org/10.1029/2024JA032431>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... Vázquez-Baeza, Y. (2020). Scipy 1.0: fundamental algo-

874 rithms for scientific computing in python. *Nature Methods*, 17, 261–272. doi:  
875 10.1038/s41592-019-0686-2

876 Wang, J., Liu, X., Dai, F., Zheng, R., Han, Y., Wang, Y., ... Baumjohann,  
877 W. (2025). Automated plasma region classification and boundary layer  
878 identification using machine learning. *Remote Sensing*, 17, 1565. doi:  
879 10.3390/rs17091565

880 Winslow, R. M., Anderson, B. J., Johnson, C. L., Slavin, J. A., Korth, H., Purucker,  
881 M. E., ... Solomon, S. C. (2013). Mercury's magnetopause and bow shock  
882 from messenger magnetometer observations. *Journal of Geophysical Research:*  
883 *Space Physics*, 118, 2213–2227. doi: 10.1002/jgra.50237

884 Zhong, J., Wan, W. X., Slavin, J. A., Wei, Y., Lin, R. L., Chai, L. H., ... Han,  
885 X. H. (2015). Mercury's three-dimensional asymmetric magnetopause.  
886 *Journal of Geophysical Research: Space Physics*, 120(9), 7658–7671. doi:  
887 <https://doi.org/10.1002/2015JA021425>